Proceedings of
International Conference on Emerging Technologies in Computer Science
(ICETCS)

Under the Auspices of

International Conference

On

Trends in Information, Management, Engineering and Sciences

(ICTIMES)

Malla Reddy College of Engineering(MRCE)

(Permanently Affiliated to JNTUH & Approved by AICTE,
New Delhi, An ISO 9001: 2008 Certified Institution)
Maisammaguda, Hyderabad-500014.
www.mrce.in

Editor in Chief

Dr. P. John Paul

Principal

Editor

Dr. Sunil Tekale

HOD - CSE Department

MALLA REDDY COLLEGE OF ENGINEERING (MRCE)

BEST WISHES

Sri.CH.MALLA REDDY Garu,

Founder chairman, MRGI

MP.MALKAJGIRI

I Congratulate CSE Department on conducting INTERNATIONAL CONFERENCE ON

" Emerging Technologies in Computer Science".

MALLA REDDY COLLEGE OF ENGINEERING (MRCE)

BEST WISHES FROM

Sri.CH.MAHENDER REDDY,
Secretary, MRGI

Dr.CH. BHADRA REDDY,
Treasurer, MRGI

Prof. R.MADAN MOHAN,
Director-Academics, MRGI

Col. G.RAM REDDY
Director/Administrations, MRGI

MALLA REDDY COLLEGE OF ENGINEERING (MRCE)

BEST WISHES

Mr.N.SUDHIR REDDY,

DIRECTOR.

My hearty wishes to department of CSE on conducting

INTERNATIONAL CONFERENCE ON

" Emerging Technologies in Computer Science".

MALLA REDDY COLLEGE OF ENGINEERING (MRCE)

BEST WISHES

Dr.P.JOHN PAUL,

Principal, MRCE

Editor in-Chief

Technology has to be invented or adopted. My wishes to Department of CSE on conducting

INTERNATIONAL CONFERENCE ON

" Emerging Technologies in Computer Science".

MALLA REDDY COLLEGE OF ENGINEERING (MRCE)

BEST WISHES

Dr.Seow Ta Wee,

University Tun Hussein Onn Malaysia,

MALAYSIA

My warmest congratulations to you, MRCE and all staff on conducting INTERNATIONAL CONFERENCE ON " Emerging Technologies in Computer Science".

MALLA REDDY COLLEGE OF ENGINEERING (MRCE)

BEST WISHES
Prof. Bharath Bhushan
WAIRCO Secretary,
INDIA

My congratulations to you, MRCE and all staff on conducting INTERNATIONAL CONFERENCE ON " Emerging Technologies in Computer Science".

MALLA REDDY COLLEGE OF ENGINEERING (MRCE)

Best Wishes
Dr.Sunil Tekale,
HOD CSE.
Convener

My Best wishes to department of CSE on conducting INTERNATIONAL CONFERENCE ON " Emerging Technologies in Computer Science".

INTERNATIONAL CONFERENCE ON EMERGING TECHNOLOGIES IN COMPUTER SCIENCE (ICETCS)

Conference by
Malla Reddy College of Engineering (MRCE)
Secunderabad/Hyderabad, India.

Chief Patron
Sri Ch. Malla Reddy, Founder Chairman, MRGI
(Honourable Member of Parliament, Govt. of INDIA)

Patrons
Mr. Ch. Mahender Reddy, Secretary MRGI
Dr. Ch. Bhadra Reddy, Treasurer MRGI

Co-Patrons
Col. G. Ram Reddy, Director (Admin), MRGI
Prof. R. Madan Mohan, Director (Academics), MRGI
Mr. N. Sudhir Reddy, Director, MRCE

International Advisory Committee
Dr. S. R. C. Murthy, University of Sydney, Australia
Dr. A.V. VidyaSagar, BELL, USA
Dr. K.V.S. S. Narayana Rao, NITIE, Bombay
Dr. Ch. A.V. Prasad, Senior Consultant, TCS
Dr. A. Govardhan, Principal, JNTUH
Dr. B. SudeerPrem Kumar, Chairman BOS, JNTUH
Dr. K. Venkateswar Rao, JNTUH
Dr. P. Dasharathan, JNTUH
Dr. B.N. Bhandari, Director DAP, JNTUH
Dr. M. Manzoor Hussain, Director Administrations, JNTUH
Dr. M. MadhaviLatha, Former Director, I-Tech, JNTUH
Mr. K. Praveen, Manager, DU PONT, USA
Mr. V.B Suresh, V.P., Virgin Mobile, USA
Mr. Koteshwar Rao, Manager, WIPRO

Chief Guest
Sri Ch. Malla Reddy, Founder Chairman, MRGI
(Honourable Member of Parliament, Govt. of INDIA)

Guest of Honour
Dr. Seow Ta Wee, University Tun Hussein Onn Malaysia, MALAYSIA

Keynote Speakers
Dr. Md. Zafar Ali Khan, IIT Hyderabad
Dr. N. Venkata Reddy, IIT Hyderabad
Dr. Paul Bharath Bushan, WAIRCO Secretary, INDIA

Conference General Chair
Dr. P John Paul, Principal, MRCE

Conference Committee MRCE
Dr. T. V. Reddy, Dean – H&S
Dr. Sunil Tekale, HOD - CSE
Dr. D.K. Nageswara Rao, HOD - MECH
Dr. V.V. PrathibhaBharathi, Head – R&D
Dr. Madhusudana Brahma, Head - English
Dr. T. Tharamani, Head – Maths

ICTIMES - INTERNATIONAL CONFERENCE ON EMERGING TECHNOLOGIES IN COMPUTER SCIENCE (ICETCS)

proceedings of International Conference on Emerging Technologies in Computer Science (ICETCS) ISBN 978-93-85101-62-5

AN EFFICIENT DISTRIBUTION VERIFICATION PROTOCOL (EDVP) FOR DATA STORAGE SECURITY IN CLOUD COMPUTING

J.Prashanth[1], K.Sagar[2], K..Sathvik Reddy[3], N.Ananth Ram Reddy[4]
Department of CSE,MRCE,JNTU, Hyderabad
e-mail[1]: prashanth23@gmail.com,e-mail[2] : sagar.kandula@gmail.com,e-mail[3] : sathvik54@gmail.com, e-mail[4] : ananth.narsireddy@gmail.com

Abstract -- *Cloud Computing (CC) is an emerging computing paradigm that provides large amount of computing and storage to the Clients provisioned as a service over the internet in a pay-as you-go pricing model, where the Clients pay only according to the usage of their services. In this thesis, we investigate this kind of security issues of cloud storage and propose New Probabilistic Efficient and Secure Protocols for data storage security. To avoid integrity availability & confidentiality for cloud storage. To provide better security to the consumers an efficient protocols and methodologies are to be used for cloud in order to store the data with third party members the main problem is security so in my thesis by using EDVP we can provide better security to the customers in cloud.*

***Keywords* -- CloudComputing,Storage,Security, Clients, Service, Protocols, Data.**

I. INTRODUCTION

This protocol implements the RSA-DPAP, ECC-DPAP and PVDSSP in a distributed manner which was discussed in chapters 5 and 6 respectively. Here, the *n* verifiers challenge the *n* servers uniformly and if *m* server‘s response is correct then, we can say that Integrity of data is ensured as to verify the Integrity of the data, the verification process uses multiple TPAs. Among the multiple TPAs, one TPA will act as main TPA and remaining are SUBTPAs. The main TPA uses all SUBTPAs to detect data corruptions efficiently, if main TPA fails, the one of the SUBTPA will act as main TPA. The SUBTPAs do not communicate witheach other and they would like to verify the Integrity of the stored data in cloud, and the 163 consistency of the provider‘s responses. The propose system guarantees *atomic* operations to all TPAS; this means that TPA which observe each SUBTPA operations are consistent, in the sense that their own operations, plus those operations whose effects they see, have occurred atomically in same sequence. The Centrally Controlled and Distributed Data paradigm, where all SUBTPAs areThe Centrally Controlled and Distributed Data paradigm, where all SUBTPAs are controlled by the TPA and SUBTPA‘s communicate to any Cloud Data Storage Server for verification. We consider a synchronous distributed system with multiple TPAs and Servers. Every SUBTPA is connected to Server through a synchronous reliable channel that delivers challenge to the server. The SUBTPA and the server together are called parties P. A protocol specifies the behaviors of all parties. An execution of P is a sequence of alternating states and state transitions, called events, which occur according to the specification of the system components. All SUBTPAs follow the protocol; in particular, they do not crash. Every SUBTPA has some small local trusted memory, which serves to store distribution keys and authentication values. The server might be faulty or malicious and deviate arbitrarily from the protocol such behavior is also called Byzantine failure. A party P that does not fail in an execution is correct.

II. APPROACH

Here, the Coordinator will randomly generates a bit string for each SUBTPA termed as *TaskDistribution Key* (TDK). Each SUBTPA will successively apply their TDK on the generated Sobol sequence as a mask upto the sequence will exhaust and take thecorresponding sequence number as block number for verification.

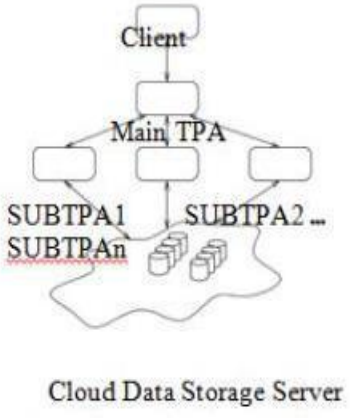


Fig. 1 Block Diagram of Distributed Audit System Architecture

For example, consider the TDK for the SUBTPA1 and SUBTPA2 are 10101 and 01010 respectively. Let, the generated Sobol random sequence is *{*1216, 5312, 3264, 7360, 704, 4800, 2752, 6848, 1728*}*, where file blocks are numbered from 0 to 8191. If we place the TDK for SUBTPA1 on the left end of the generated sequence and takes the block numbers corresponding to the 1, after that we slides the string to the right to the same length of the TDK and apply the same

procedure then it generates the subtask for SUBTPA1 in (1) and similarly for SUBTPA2 in (2).

{1216, 5312, 3264, 7360, 704, 4800, 2752, 6848, 1728}
1 0 1 0 1
{1216, 5312, 3264, 7360, 704, 4800, 2752, 6848,1728}
1 0 1 0 1

{1216, 3264, 704, 4800, 6848} (1)

{1216, 5312, 3264, 7360, 704, 4800, 2752, 6848, 1728}
0 1 0 1 0
{1216, 5312, 3264, 7360, 704, 4800, 2752, 6848, 1728}
0 1 0 1 0

{5312, 7360, 2752, 1728} (2)

In our protocols, we use two types of TDK for uniformly distribute the task among SUBTPAs and sometime, we adjust the TDK length to balance the subtask for each SUBTPA.

III. THEORETICAL BACKGROUND

A. SOBOL SEQUENCE

Sobol Sequence [3], [4] is a low discrepancy, quasi-random sequences that generates sequences between the interval [0, 1).

One salient features of this sequence is that the sequences are uniformly distributed over the interval [0, 1). Also, it maintains uniformity for any segment out of the sequence. That means SobolSequence is *segment wise uniform*. For generating this sequence, we need a primitivepolynomial, of degree over the finite field $\mathbb{Z}_2$, and direction numbers [3].

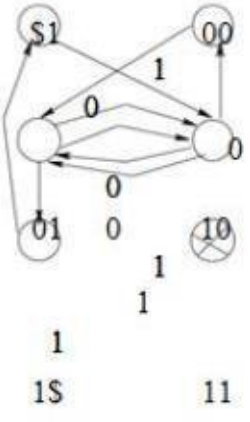


Fig. 2. Modified de Bruijn Graph Corresponding to the string 10010101

IV. EFFICIENT DISTRIBUTED VERIFICATION PROTOCOL

Our distributed verification protocols are based on the probabilistic verification scheme and we classify our protocols into two different types depending on the task distribution. First, we are describing our basic protocol based on the simple partition approach. In the second, we use TDK to partition the task. To enhance the performance of our protocols, we used (m, n) threshold scheme [15] with $<$, where Coordinator can stop the audit operation or detect the fault region after taking responses from any subset of out of SUBTPAs, because each subtask is uniformly distributed over the entire file blocks due to use of Sobol Sequence.

A. Protocol 1: Simple Partition with Threshold Seheme

In the first protocol, the Coordinator randomly chooses one Sobol random key , generate the Sobol Random Block Sequence by using $(\cdot)$, where consist one randomly chosen primitive polynomial, of order out of $(2-1)/$ primitive polynomials [4], randomly chosen initial values , where $\in \{1, 2, \ldots, \}$, and values respectively. Inthe next step, partition the generated sequence $\mathcal{L}$ by using partition function $(\cdot)$, with partition length and denotes each subsequence as , should maintain the equivalence relation property and also maintain the uniformity property. Algorithm 1 gives the detail of key generation and Distribution phase.

ALGORITHM 1: KEY GENERATION & DISTRIBUTION

1 Coordinator randomly chooses one Primitive Polynomial, of degree and initialization number
, $\in \{1,2, \ldots,\}$;
2 Coordinator decides Sobol Random Key as
= , , , , ,
3 Generated Sequence
$\mathcal{L} \leftarrow ()$
4 Multiply CONSTANT *powers of 2* with $\mathcal{L}$, to make each element as integer block number;
5 Coordinator Determines the Number of SUBTPAs, , and threshold value, ;
6 **for** $\leftarrow 1$*to* **do**
7 $\leftarrow (, \mathcal{L})$
8 **end**
9 whereis the length of each partition, and $\mathcal{L} \subseteq$ $\{1,\ldots,\}$
= $\in \{,1, \ldots,\}$;
10 Distributes Subtask, to;

ANALYSIS OF PROTOCOL1

Protocol 1 follows the *Centrally Controlled and DistributedData* paradigm, where all SUBTPAs are controlled by the Co-ordinator but communicate to any Cloud Data Storage Server for verification. Here, Coordinator will decide the partition length, , and divides the sequence to each . Due to the use of Sobol sequence each subsequence must be *uniform*. After partitioning the sequence, the Coordinator willsend the subsequence, , to each .

This protocol gives very good performance to detect er-rors in the file blocks. Nevertheless, for sending to , from the Coordinator takes extra network band-width. Although, it can not take any extra care about the critical data. To reduce the bandwidth usage and increase the efficiency, and also, taking extra care about critical data, we device the Task Distribution Key (TDK) to divide the sequence

ALGORITHM 2:DISRTIBUTEDCHALANDPROOFVERIFICATION( )

```
1        Calculates 10%
2          of; for each do
3          ← h( ); ←
4          ⌊(10/100)∗ ⌋;
5              End
6          for ←1to10 do
7            for  ←1toanddo
8        h , [ ] ←[ ];
9      end
10 Send h , as a challenge to the Cloud Server; 11 Wait
for the Proof, , from any Server;
12                    , ←();
13       if , equals to Stored Metadata then
14                    [ ] ←;
15                  Else
16                    [ ] ←;
17                Send Signal, , ,to the
                    Coordinator;
18     end
19   end
```

to subsequences. Our second scheme describes about TDK based techniques in more details.

B. Protocol 2: TASK DISTRIBUTION KEY BASED DISTRIBUTION SCHEME

In our second protocol, Coordinator and each will know the Sobol random key, , for generating the Sobol random sequence. In each new verification, Coordi-nator decides the parameters to generate the Sobol Random Key, and publicly send to all . In addition, Coordinator generates number of random TDKs, , and distributes among SUBTPAs by using *String ReconciliationProtocol* [1] *with* some modifications.

each SUBTPA will generate Sobol Random Sequence and interpret their subsequence by using their own TDK. We have given Sobol Random key, TDK generation and distribution in Algorithm 3. Algorithm 4, describes about subtask interpreta-tion, distributed challenge and verification for protocol 2.

In this protocol, we use two types of TDKs, one is *Non-Overlapping TDK* and another is *Overlapping TDK*. Overlap-ping TDK will apply when we want to verify *critical data*. We give the steps for generating Non-Overlapping TDK as follows:

ALGORITHM 3: KEY GENERATION & DISTRIBUTION

```
1  Coordinator randomly chooses one Primitive
   Polynomial, of degree d and initialization number
mi, i∈{1, 2, . . . , d} ;
2 Coordinator decides Sobol Random Key,
ser= (P,mi, SKIP, LEAP, CONSTANT, SeqLen)
3 Coordinator Determines the Number of SUBTPAs, n, and threshold value, m;
4 Coordinator send sets to all SUBTPAs;

5   Determine number of 1´ s, t, each TDKswill contain;
6   TDKLen← ser t;
7 Generates Random Permutation index from
1, 2, . . . , TDKLen;
8 for ← 1 to ndo
9 for ← 1 to ndo
10      {0}[ ← 0;
11 end
12 end
13      ← 1;
14      ← 1;
15      while <ndo
16      for ← 1 to ndo
17      ← RandS[++];
18      {0}[ ← 1;
19 end
20      ← + 1;
21 end
22      if (TDKLen, SeqLen) ← 1 then
23 TDKLength is acceptable;
24 else
25 TDK Length adjust to the next nearest Primes;
26 end
27 Generated TDKfor SUBTPAare represented as, os1, os2, os3, . . . , osnrespectively, distributes among SUBTPAs by
   using String Reconciliation Protocol.
```

ALGORITHM 4:
DISTRIBUTEDCHALANDPROOFVERIFICATION 2( ):

1 Each *SUBTPAi*generates Sequence
$\mathcal{L} \leftarrow for(SeqLen)$
2 Multiply CONSTANT *powers of 2* with $\mathcal{L}$, to make each element as integer block number.;
3 Interpret subsequence by using oS_ias $r_{i,}$ where
$\mathcal{L} \leftarrow \cup$

$i \in [1,\ldots,$

$n]$

$j \in [1,\ldots,$

$p] r_{i,j}$
4 *SUBTPAi*Calculates 10% of $r_{i,}$;
5 for *each SUBTPAi***do**
6 $k \leftarrow length(r_{i,j})$;
7 $Counter_i \leftarrow /(10/100) * l/$;
8 end
9 for $k \leftarrow 1$ *to 10* **do**
10 for $S \leftarrow 1$ *to Counteriand* $t \leqslant l$**do**
11 $Chal_i[s] \leftarrow r_{i,j}[t]$;
12 end
13 Send $(Chal_i,)$to *SUBTPAi*;
14 Wait for the proof, $PR_{i,}$ from any Cloud Server;
15 $PR_{i,} \leftarrow Rece()$;
16 if *PRi, equals to Stored Metadata* **then**
17 $Re[k] \leftarrow TRUE$;
18 else
19 $Re[k] \leftarrow FALSE$;
20 Send Signal, $(Packet_i,, FALSE)$to the Coordinator;
21 end
22 end

ANALYSIS OF PROTOCOL 2

In TDK generation phase, we take the mask length asco-prime to sequence length or prime length, because afterapplying TDK on $\mathcal{L}$, subsequence, $r_{i,}$becomes nonuniform,and to make it uniform, we use these adjustment. In algorithm3, Coordinator generates Sobol Random Key and send to theSUBTPAs. In addition, send different TDK, , for each*SUBTPAi*. In Algorithm 4, *SUBTPAi*generates the Sobol Random sequence by using key, $o_r$and stored in $\mathcal{L}$. Then, each *SUBTPAi*interpret their task by using corresponding TDK, , and we denoted subtask for *SUBTPAi*as $r_{i,j}$and. defined as
$\mathcal{L} = \cup$

$i \in [1,\ldots,$

$n]$

$j \in [1,\ldots,$

$p] r_{i,j}$

where
$p =$
$SequenceLength$
$TDKLength$
$+ \pounds$
$\pounds = Numberof1$
$'$
$sinfirst$
$(SequenceLength\%\ TDKLength)\ lengthinTDK$
Then, *SUBTPAi*will calculate 10% of $r_{i,}$ and createschallenge, $Chal_{i,}$ and send to the server and waits for theproof, $PR_{i,}$. After receiving the proof SUBTPA will verifywith the stored matadata, and if the proof is correct then storeTRUE in its table and if not match then store FALSE andsend a signal to the Coordinator for corrupt file blocks. TheCoordinator will receive *signals* from any subset of *m*out of*n*SUBTPAs and ensures the fault location or stop the Auditoperation. In the final step, Main TPA will give the Auditresult to the Client.

Here, we generalize the integrity verification protocol in adistributed manner. Therefore, we can use our protocols onexisting RSA based [11] [13] or ECC [10] based protocol tomake distributed RSA or ECC protocols. In the next section,we discuss about the performance of our protocols.

V. IMPLEMENTATION AND EXPERIMENTAL RESULTS

It is very natural that audit activities would increase thecommunication and computational overhead of audit services.To enhance the performance, we used the String Reconciliation Protocol to distribute the TDK,that reduces the communication overhead between Main TPA and SUBTPAs. For each new verificationCoordinator can change the TDK for any SUBTPA and send only the difference part of the multyset element to the SUBTPA. In addition, we used probabilistic verification scheme based on Sobol Sequence that provides not only uniformity for whole sequence but also for each subsequence, so each SUBTPA will independently verify over the whole file blocks. Thus, there is a high probability to detect fault location very efficiently and quickly. Therefore, Sobol sequence provides strong integrity proof for the remotely stored data. Table I shows comparison between two protocols.

TABLE I
PERFORMANCE COMPARISON BETWEEN TWO PROPOSED PROTOCOLS

	Protocol 1	**Protocol 2**
Public Verifiability	Yes	Yes
Coordinator Controlled	Yes	Yes
Probabilistic	Yes	Yes
Privacy Preserving	Yes	Yes
Task Distribution	uniform	uniform
Fault Detection	fast	very first
Coordinator Computation	more	more
Communication Complexity	morc	less

detection probability for Sobol Random Sequence and Pseudo Random Sequence.

We have shown our experimental results in Table II.

TABLE-II

DETECTION PROBABILITY FOR 1% CORRUPTION OUT OF 300000 BLOCKS

Number of samples as Percentage of total samples		Detection Probability
SobolSequence	*PseudorandomSequence*	
10%	20%	0.6
14%	27%	0.7
16%	30%	0.8
21%	37%	0.85
23%	41%	0.9
26%	49%	0.95
30%	54%	0.9999

VI. CONCLUSIO

In this paper, we addressed the efficient Distributed Verification protocol based on the Sobol Random Sequence. We have shown that our protocols uniformly distribute the task among SUBTPAs. Most importantly, our protocols can handle failures of SUBTPAs due to its uniform nature and also gives better performance in case of unreliable communication link. Here, we mainly focussed on the uniform task distribution among SUBTPAs to detect the erroneous blocks as soon as possible. We used String Reconciliation Protocol to minimize the communication bandwidth between Coordinator and SUBTPA side. In addition, we reduce the workload at the Server side and also reduce the chance of network congestion at the Server side as well as Coordinator side by distributing the task. Thus,our Distributed Verification Protocol increases the efficiency and robustness of data integrity in Cloud Computing. generaterandom block numbers by using Sobol Random generatorfor a given length, then it must be uniform. In addition,if we simply partition the sequence into subsequence anddistributes among various SUBTPAs, then each subsequencemust be maintain the uniformity. But, when we use TDK thensubtask may or may not be uniform. We saw that when theTDK length is powers of 2, then generated subtask does notmaintain the uniformity property. Because, Sobol sequencemaintain some pattern, if we take 4 consecutive number thenwe can see that these numbers are from four region over theSequence, if we divide the full sequence into four region, and

for 8, 16, 32,. . . it also hold. When we placed the TDK overthe generated Sequence then Subtask contain those numberswhose corresponding TDK bit is 1 and successively applyingthis TDK to generate the subsequence. Thus, if the TDK length power of two then for each successive TDK shifting, the chosen block numbers must be very close to each other andform cluster. If, we take TDK length as prime then in each successive shifting the chosen block numbers are spreadingover the segment. Therefore, maintains the uniformity for each subtask or subsequence. Now, if the TDK length is Co-primemean$gcd(TDKLength, SequenceLength) = 1$Then there is no factor equals to the power of 2, that means for each successive TDK shifting block numbers are spreadingover the whole sequence and maintain the uniformity propertyfor each subtask. Therefore, generated subtask must be uniform if the TDK length relatively prime or prime to the sequence length

VII. REFERENCE

[1]. Aaram, Y., Chunhui, S., and Yongdae, K., ―On Protecting Integrity and Confidentiality ofCryptographic File System for Outsourced Storage‖, In proc.of CCSW‘09, Chicago, Illinois,USA November 13, 2009.

[2]. Alexander, H., Bernardo, P., Charalampos, P., and Roberto, T., ―Efficient IntegrityChecking of Untrusted Network Storage‖, In Proceedings Of StorageSS‘08, Fairfax, Virginia, October 31, 2008.

[3]. Alexander, S., Christian, C., Asaf, C., Idit, K., Yan, M., and Dani, S., ―Venus: Verificationfor Untrusted Cloud Storage‖, In Procedings Of CCSW‘10, Chicago, Illinois, USA October8, 2010.

[4]. Amazon.com, Amazon Web Services (AWS), *Online at http://aws.amazon.com(2008).*

[5]. Anjie P., and Lei W., ―One Publicly Verifiable Secret Sharing Scheme based on LinearCode‖, In Proc. Of 2010 2nd Conference on Environmental Science and InformationApplication Technology, Jul-2010, pp.260-262.

[6].Apple―ICloud‖Onlineat http://www.apple.com/icloud/what-is.html 2010.

[7]. Armbrust, M., Fox, A., Rean, G., Anthony, D. J., Randy, H. K., Andrew K., Gunho L,David, A. P., Ariel, R., Ion, S., and Matei, Z., ―A view of cloud computing‖, Commun.ACM 53, 2010, pp.50–58.

[8]. Armbrust, M., Fox, A., Rean, G., Anthony, D., J.,, Randy., H. K., Andrew K., Gunho L,David, A., P., Ariel, R., Ion, S., and Matei, Z., ―Above the Clouds: A Berkeley View ofCloud Comput-ng,‖ Tech. Rep. UCBEECS-2009, Univ. California, Berkeley, February 28,2009.

DATA SECURITY IN MOBILE AD HOC NETWORKS

G.Pravallika[1],J.Pradeep Reddy[2], K.HarshaVardhanReddy [3], G.Anupama[4]

Department of CSE, MRCE, JNTU Hyderabad
E-mail[1]: pravalika_cse@gmail.com, E-mail[2] : pradeep_cse@gmail.com,
E-mail[3]: harsha_cse@gmail.com, E-mail[4] : anu.guthireddy@gmail.com

ABSTRACT

Today organizing innovation has risen quickly and now it has reached out to the development of remote systems. Diverse sorts of systems can be shaped to share the assets in light of necessities. As the development advances, the issue of security to the information in the system gets to be lasting significance. The information exchanged from one framework to other framework in versatile specially appointed system, has more powerlessness. Giving security at various levels is an absolute necessity in light of the fact that the assailants assault the systems at various levels, to get entrance the data. In spite of the fact that there are a few strategies rehearsed for giving security to the information, each time a gatecrasher finds distinctive approaches to access to the system on the grounds that the system continues developing. For giving security to the information, all security administrations to be considered are secrecy, trustworthiness, verification, non-revocation. Aside from this, security level is likewise considered, and the get to benefits for this level are resolved by client.

The exploration is centered around examining the standard security administrations accessible in portable specially appointed system environment. The objective is to keep the different assaults and to distinguish a superior course to move the date in the portable impromptu systems. The proactive approach expands remediation effectiveness, evaluates the genuine effect of potential assaults and doles out security assets wisely. Thus, an approach based shared plan is proposed for giving better security to the information that is moved in versatile specially appointed system. Securing information is done through privacy confirmation and honesty.

At first, Trust based parcel sending plan is proposed to ascertain the trust record of the hub and the courses are chosen by trust esteem with a view to enhance the trustworthiness. Keeping in mind the end goal to give confirmation dispersed declaration power strategy is given to build an authentication. A novel encryption and unscrambling instrument, which is a blend of both symmetric and topsy-turvy key cryptographic strategies is proposed to give classification. The three plans are consolidated to shape a common plan to give security to the information in light of the necessities of the client. The assurance conspire gives the sought level of security, in view of the arrangement by commonly coordinating the plan, as indicated by the prerequisite of the client. The proposed systems are joined to frame a strategy based shared plan for information security that can give finish insurance to the information in MANET correspondence.

Keywords—data security, networks, mobile networks

I. INTRODUCTION

In recent years mobile ad hoc networks (MANETs) have received tremendous attention because of their self-configuration and self-maintenance capabilities. While early research effort assumed a friendly and cooperative environment and focused on problems such as wireless channel access and multihop routing, security has become a primary concern in order to provide protected communication between nodes in a potentially hostile environment. Although security has long been an active research topic in wireless networks, the unique characteristics of MANETs present a new set of nontrivial challenges to security design. These challenges include open network architecture, shared wireless medium, stringent resource constraints, and highly dynamic network topology. Consequently, the existing security solutions for wired networks do not directly apply to the MANET domain.

The ultimate goal of the security solutions for MANETs is to provide security services, such as authentication, confidentiality, integrity, anonymity, and availability, to mobile users. In order to achieve this goal, the security solution should

provide complete protection spanning the entire protocol stack. In this article,we consider a fundamental security problem in MANET: the protection of its basic functionality to deliver data bits from one node to another. In other words, we seek to protect the network connectivity between mobile nodes over potentially multihop wireless channels, which is the basis to support any network security services.

Multihop connectivity is provided in MANETs through two steps: (1) ensuring one-hop connectivity through link-layer protocols (e.g., wireless medium access control, MAC); and (2) extending connectivity to multiple hops through networklayer routing and data forwarding protocols (e.g., ad hoc routing). Accordingly, we focus on the link- and network- layer security issues, challenges, and solutions in MANETs in this article. One distinguishing characteristic of MANETs from the security design perspective is the lack of a clear line of defense. Unlike wired networks that have dedicated routers, each mobile node in an ad hoc network may function as a router and forward packets for other peer nodes.The wireless channel is accessible to both legitimate network users and malicious attackers. There is no well defined place where traffic monitoring or access control mechanisms can be deployed. As a result, the boundary that separates the inside network from the outside world becomes blurred. On the other hand, the existing ad hoc routing protocols, such as Ad Hoc On Demand Distance Vector (AODV) [1] and Dynamic Source Routing (DSR) [2], and wireless MAC protocols, such as 802.11 [3], typically assume a trusted and cooperative environment. As a result, a malicious attacker can readily become a router and disrupt network operations by intentionally disobeying the protocol specifications.There are basically two approaches to protecting MANETs: proactive and reactive. The proactive approach attempts to prevent an attacker from launching attacks in the first place, typically through various cryptographic techniques.

In contrast, the receptive approach tries to identify security dangers a posteriori and respond likewise.Because of the nonattendance of an unmistakable line of barrier, a finish security answer for MANETs ought to coordinate both methodologies and envelop all three segments: aversion, discovery, and response. For instance, the proactive approach can be utilized to guarantee the rightness of directing states, while the receptive approach can be utilized to secure bundle sending operations. security is a chain, and it is just as secure as the weakest connection. Missing a solitary part may essentially debase the quality of the general security arrangement. Security never wants free. Whenever more security elements are brought into the system, in parallel with the improved security quality is the continually expanding calculation, correspondence, and administration overhead. In conclusion, we talk about open difficulties and conceivable future headings in this range.

II. PROPOSED SYSTEM

Trust based packet forwarding scheme

Trust is an important factor in the design and deployment of security systems. In MANET, trust evaluation can be applied to node authentication, access control and trust routing. By evaluating the trustworthiness of a node, it not only enhances the security but also improves the routing performance in MANETs. To define a suitable trust evaluation model for MANETs there are several issues that need to be taken into consideration. Here the trust index value is calculated and it is used to forward the packets.

Trust Index Calculation

The trust model uses a structured approach. A group of nodes or mobile devices is considered in a network. Trust value is calculated by every individual node in the network. A node is not trusted until it presents a trust value. The trust values are incremented or decremented according to the behavior of the node. The route for sending the packet is selected according to the trust index value of the node. The trust index value is compared with the threshold value. If the trust index value is less than the threshold value, the packet may be dropped else the packet is forwarded.

III. PROBLEM IDENTIFICATION AND PROPOSED SOLUTION

The future system situations don't comprise essentially of one get to innovation however have various get to advancements. Clients like to get associated at all times, with the best get to organize accessible. The versatile systems administration may require arrangement of the best security benefits that will end up being a fundamental element in portable terminals. Security administrations are regularly offered by organizations and data innovation experts.

Some security administrations are produced for individual buyers or little scale organizations. These framework frequently require less asset to oversee. There are distinctive routes in which security administrations can be given, contingent upon the necessities of individual organizations. Huge organizations may have interior groups and security specialists to manage arrange security and dangers to information. Dangers to PC framework and systems develop exponentially with innovative progression. Each time another arrangement of hazard takes after and a PC organize security pro requires to re-design his level of learning. The most hazardous hazard to security frequently originates from outside source. Subsequently versatile specially appointed system security has been liable to broad late study. While much consideration has been spent taking a gander at security of directing conventions in specially appointed systems, it is similarly essential to secure interchanges in versatile systems.

The topology of the system continues evolving powerfully. The hubs have constrained physical assurance. There is absence of brought together observing. An appealing thought is to have a mix of security administrations that can give best results. To accomplish this, the cell phones should be more canny to offer best security administrations among themselves. In the proposed work, outline of security arrangement depends on the accompanying necessities:

1. Prerequisites for Integrity and drop

The information transmitted between the hubs in MANET ought to be gotten to the expected elements without change or unapproved alteration. There the nonattendance of sufficient information honesty insurance, arrangements are to be planed for ensuring each sort of information paying little heed to its sort, where it is put away, whether it is in stationary or in travel.

The accompanying prerequisites are distinguished:

a. Make arrangements and systems for information quality and information honesty.

b. Make arrangements and systems to recognize the degree of the issue.

c. Embrace risk appraisal of esteemed information.

Trust based packets sending plan is proposed for relieving the information drop assaults. The trust record is to be figured for every one of the hubs in the system. Trust values support bundle sending by keeping up motivating forces and punishments for every hub. Every transitional hub denote the bundles by including its hash esteem and advances the parcel towards the goal hub. The goal hub checks the motivators and punishments and confirms the hash esteem for hubs with low impetus and high punishment.

2. Necessities for validation

Verification is basic to check the personality of every hub in MANET and its qualification to get to the systems. The utilization of advanced testament issued and checked by an endorsement power as a major aspect of open key framework is viewed as liable to end up a standard approach to perform verification on the versatile specially appointed systems. Clients or hubs need to have admittance to the receptive testament dissemination component utilizing Certificate Authority (CA) hubs, The hubs trusted and being trusted by more than one CA need to apply for an authentication and private-key- offers from every CA. A hub without endorsement or expecting to restore its testament must approach different hubs in the MANET for
a declaration.

3. Necessities for classification

Contingent upon various application necessities, the payload part might be alternatively encoded with the common key between the source and the goal. Information or data is not made accessible or revealed to unapproved people or procedures. Contingent upon the way of information and client prerequisites, strategies with the accompanying decisions are held by any client:

a. Respectability and classification.

b. Privacy and confirmation.

c. Validation and honesty.

d. Any of privacy, honesty, validation.

e. All privacy, honesty, validation.

Policy based mutual scheme

Depending upon the nature of data and user requirements, user policies can be specified which can take the following values:

1. I - Only Integrity
2. A - Only Authentication

3. C - Only Confidentiality.
4. IA - Both Integrity and Authentication.
5. IC - Both Integrity and Confidentiality.
6. AC – Both Authentication and Confidentiality.
7. IAC - Integrity, Authentication and confidentiality

Based on the policy of the user, the corresponding security module(s) can be executed, as per the following algorithm.

Algorithm : Policy based Mutual Scheme for data Security. If Policy = "I", then

Calculate the trust index of all the nodes according to algorithm.

Else if Policy = "A", then

DCA private key is applied to d eliver security. Share updation is done among the cluster heads.

Else if Policy = "C," then Encryption and Decryption are done according to the algorithm

Else if Policy ="I" and Policy = "A", then
Calculate the trust index of all the nodes according to algorithm .

The DCA private key is applied to deliver security services. Share updating is done among the cluster heads.

Else if Policy = "I" and Policy ="C", then

Calculate the trust index of all the nodes according to algorithm.
Encryption and Decryption are done according to the algorithm 3.
Else if Policy ="A" and Policy ="C", then

The DCA private key is applied to deliver security services. Share updating is done among the cluster heads.

Encryption and Decryption are done according tothe algorithm 3.

Else if Policy = "I" and Policy ="A" and Policy="C", then

Calculate the trust index of all the nodes according to algorithm.

End if

The above scheme is simple and robust in the sense that there is no need to synchronize, as the combined scheme work based on the users requirements. The above said policy based mutual scheme algorithm focus on security requirement services based on minimum recourses available in the mobile ad hoc networks.

IV. CONCLUSION

It is getting to be apparent that future system situations are probably not going to comprise of basically one get to innovation yet will incorporate numerous get to advancements, adding complexities to the portability and security of the frameworks. Arrange clients will incline toward get to organize accessible with the best security and availability. The general target of this examination is to give security administrations to versatile specially appointed system by keeping up arranged security and directing in capricious environment. In this examination work, outlined and actualized security conspire in view of the client require has been proposed. Moreover, composed and executed trust based bundle sending plan for hub verification, get to control and trust steering hav been proposed. The proposed conveyance of testament power plan is to give verification utilizing private key.

V. SCOPE FOR FUTURE W ORK

The work proposed in this article can be utilized as a beginning stage for different lines of research identified with proactive approach based security in heterogeneous systems. Some of them can be identified with the upgrade of the proposition made in this postulation. New techniques may be investigated to characterize approaches to accomplish the underlying goals.

The execution of the proposed security plan can be upgraded to accomplish an ideal usage as

far as execution. The security of the framework can be upgraded by encryption of the approaches while they are put away in the arrangement storehouse

For expansive appropriated systems, between operation requires countless to be characterized, put away in the vault, and actualized and when required premise. The arrangement depends on the client prerequisite. As the systems extend the proposed framework can be improved to have dynamic strategy for particular clients and their particular prerequisites in view of security and directing.

VI. LIST OF REFERENCES

1. Douglas E. Corner and Narayanan M.S. (2004),"Computer Networks and Internets with Internet Applications" Pearson Education, Fourth Edition.
2. Andrew Tanenbaum S. (2003), "Computer Networks", Prentice Hall India, Fourth Edition.
3. Marwa, Altayeb and Imad Mohgoub "A survey on vehicular ad hoc network routing protocols" International journal of Innovation and applied studies, ISSN:2028-9324,Vol. 3, No. 3 , pp. 829-846.Stefan Ruhrup. (2009), "Network Protocol Design and Evaluation", University of Freiburg.
4. Martin P. and Clark (2003), "Data Networks, Tthe Internet Protocols Design and Operation", ISBN: 0-470-84856-13, John Wiley & Sons Ltd.
5. Pahlavan K. and Prashant Krishnamurty (2002), "Principles of Wireless Networks: A unified Approach" , ISBN-978-81-203-2380-3, Prentice Hall India, Second Edition.
6. Praveen Kumar B.,Chandra Sekhar P., Papanna N and Bharath Bhushan B. (2013), "A Survey on MANET Security Challenges and Routing Protocols", International Journal of Computer Technology & Applications, Vol. 4, No. 2, pp. 248-256.
7. Rahman A., Islam S. and Talevski A (2009), "Performance Measurement of Various Routing Protocols in ad-hoc Network", Proceedings of the International Multi • Conference of Engineers and Computer Scientists, IMECS'09, Vol. 1, pp.18- 20.
8. Papadimitraos Panagiotis. and Haas J. (2003), "Secure Message Transmission in Mobile ad hoc Networks", www.elsewier.com/ locate/adhocl, pp. 193- 209.
9. Capkun S., Buttyan L. and Hubaux J.P. (2003), " Self-Organized Public-Key Management for Mobile Ad Hoc Networks", IEEE Transactions on Mobile Computing, Vol. 2, No. 1, pp.52-64.
10. Pravin Ghosekar and Pradip Ghorade. (2010), "Mobile ad hoc Networking: Imperative and Challenges", International Journal of Computer Applications of MANET, Vol. 5, pp.153-158.
11. Priyanka Goyal, Vinti Parmar and Rahul Rishi(2011), "MANET: Vulnerabilities, Challenges, Attack and Application", International Journal of Computational Engineering and Management, ISSN:2230-7893, Vol. 11,pp.32-40.
12. Frodigh M., Johansson P. and Larsson P. (2000), "Wireless ad hoc Networking: The Art of Networking without a Network", Ericsson Review, No. 4, pp. 248-263.
13. Shahram Gilanina., Seyed Mousavian J. and Orang Taheri (2012), "Information Security Management on Performance of Information System Management", Journal of Basic and Applied Scientific Research, ISSN:2090-4304, pp.2582-2587.
14. Guru Baskar T. and Girija Manimegalai M. (2011), "Performance comparisons of routing protocols in mobile ad hoc networks", International Journal of Wireless and Mobile ad hoc Networks, Vol. 3, pp.133-140.

CRYPTOGRAPHY BASED PRIVACY PRESERVING DATACOMMUNICATION IN HYBRID WIRELESS NETWORKS

G APOORVA[1],K SRILATHA[2],G VENKATESH[3],M AHARONU[4]
Department Of CSE, MRCE, Hyderabad, Telangana, India.
Apoorva.g@gmail.com[1],Sri.k@gmail.com[2],Venky.g@gmail.com[3],aharonu_cse@mrce.in[4]

ABSTRACT

Distributed Three-hop Routing protocol.DTR is used for data transmission in Hybrid wireless network. DTR divide a data into segments and transmits the segment in a distributed way. It uses at most two hops in ad-hoc transmission mode and one hop in cellular transmission mode. However, the selection of trust nodes for data transmission is difficult in DTR which in turn creates security issues. This paper proposes a TEEN APTEEN SPEED (TAS) protocol for conviction node selection. TAS protocol allocate a threshold value to each node in a network. Based on the threshold value, a trust node is selected for efficient data transmission in Hybrid Wireless Network. The threshold value is also to preserve security in the network in order that unauthorized spoofing nodes can't enter the network. Furthermore, this paper implements overhearing technique in which the sending node share the content with one or more other nodes before data transmission with the purpose that failure node can be exposed and replaced.

Index Terms – Hybrid wireless networks, Cryptography, Trust node, Overhearing

1. INTRODUCTION

Hybrid wireless network merge mobile ad-hoc network and infrastructure wireless network. It is to be an [3]improved network arrangement for the next generation network. According to the environment situation, it can select base station transmission mode or mobile ad-hoc transmission mode. The mobile ad-hoc network is an infrastructure-less network. The devices in a mobile ad-hoc network can shift in any path and the link between the devices can altered regularly. In this network, the data is transmitted from starting place to target in a multi-hop way through in-between nodes. In an infrastructure wireless network (e.g. Cellular network), each device communicates with other device through base stations. Each cell in a cellular network has a base station. These base stations are linked via cable or fiber or wirelessly through switching centers.

If the region has no communication infrastructure or the existing infrastructure, communication between nodes are complex or not suitable to use. In this location [2] hybrid wireless network may still be able to communicate through the construction of an ad-hoc network. In such a network, every mobile node operates as a host and also as a router. Forwarding packets to new mobile nodes in the network may not be within straight wireless transmission range. Each node participates in an ad-hoc routing and infrastructure routing, for this [1] Distributed three hop routing protocol is used. It allows to discovering a "Three-hop" path to any other node during the network is introduced in this effort The first two hops in ad-hoc networking is sometimes called infrastructure-less networking, since the mobile nodes in the network animatedly make routing between themselves to form their personal network. The third hop is created in infrastructure networking. Most Wi-Fi networks task in an infrastructure approach. Devices in this network communicate through a single access point, which is generally the wireless router. For example, consider the two laptops are placed next to each other, each connected to the same wireless network. still the two laptops are sited next to each other, they're not communicating in a straight line in infrastructure network. Some possible uses of hybrid wireless network consist of students using laptop, computers to participate in an interactive instruct, trade associates and sharing information during a gathering soldiers communicate information about the condition attentiveness on the emergency failure release and personnel coordinating efforts after a storm or shaking

Spread Code is normally used for safe data transmission in wireless communication as a way to measure the excellence of wireless connections. In wired networks, the existence of a wired path between the sender and receiver are determining the correct reception of a message. But in wireless networks, path defeat is a main trouble. The wireless communication network has to obtain a lot of environmental parameters to report background noise and interfere power of other simultaneous transmission. SINR attempts to produce a demonstration of this aspect. So the TAS protocol is implemented to keep the details about the dispatcher and receiver and the communication media in the network. This is implemented through overhearing concept. This TAS implements grouping of nodes depending on the threshold value so that the communication will be simple. In overhearing, the data is transferred to many nearby nodes in a cluster . The cluster is a grouping of nodes, which enclose cluster head and gateway. So the fundamental idea is to individually learn unknown and possibly random mobility parameters and to group the mobile node with related mobility prototype to the same cluster. The nodes in a cluster can then interchangeably distribute their resources for load balancing and overhead reduction, aiming to achieve scalable and proficient routing.

In TAS protocol, a secured code called threshold value is used. The nodal contact[7] probability are updating with the help of threshold value, it established to join the true contacts probabilities. Subsequently, a set of functions are devised to form clusters and choose entrance nodes based on nodal contact probabilities. lastly gateway nodes switch the network information and make routing. The result demonstrate that it is get higher delivery ratio and considerably lower overhead and end-to-end wait when compared to non-clustering matching part.

2. EXISTING WORK

The Base stations are coupled by means of a wired backbone, so that there are no power constraints and bandwidth during transmission among BS. The in-between nodes are utilized to indicate convey nodes that task as gateways connecting an infrastructure wireless network and mobile ad hoc network. DTR aims to move the routing load from the ad hoc network to the infrastructure network by taking advantage of extensive base stations in a hybrid wireless network. Rather than using one multi-hop path to forward a message to one BS, DTR uses at most[3] two hops to relay the segments of a message to different BS in a distributed way, and relies on BS to merge the segments. When a source node needs to propose a message stream to a destination node, it partition the message flow into a number of partial streams called segments and spread each segment to a neighbor node. Upon receiving a segment from the source node, a neighbor node decides between direct transmission and relay transmission based on the QoS requirement of the application. The neighbor nodes encourage these segments in a distributed way to nearby BS. Relying on the infrastructure network routing, the BS further transmit the segment to the BS where the destination node resides.

The ending BS reorganizes the segments into the original order and forwards the segments to the destination. It uses the cellular IP transmission method to begin segments to the destination if the destination moves to another BS through segment transmission.DTR works on the Internet layer. It receives packets from the TCP layer and routes it to 1the destination node, where DTR forwards the packet to the TCP layer. The data routing process in DTR can be separated into two processes: uplink from a source node to the first BS and downlink from the last BS to the data's destination. In uplink process, one hop to forward the segments of a message in a distributed way and uses another hop to find high-capacity forwarder for high show routing. As a result, DTR restrictions the path length of [8]uplink routing to two hops in order to keep away from the problems of long-path multi-hop routing in the ad-hoc networks. particularly, in the uplink routing, a source node divides its message flow into a number of segments, then transmits the segments to its neighbor nodes. The neighbor nodes promote segments to BS, which will forward the segments to the BS where the destination resides. In this work, throughput and routing speed are taken as a QoS requirement. The bandwidth/queue metric is to reflect node capacity in throughput and fast data forwarding. A larger

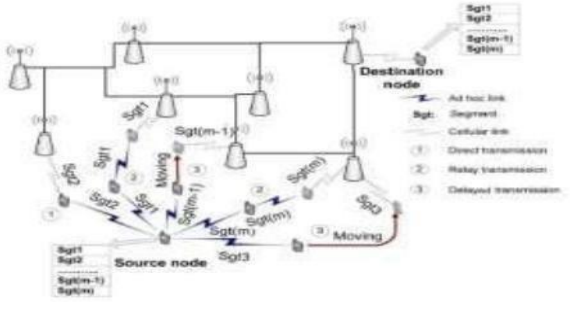
bandwidth/queue value means higher throughput and message forwarding speed, and vice versa. When selecting neighbors for data forwarding, a node needs the capacity information of its neighbors. Also, a chosen neighbor should have enough storage space for a segment. To find the capacity and storage space of its neighbors, each node periodically interactions its current information with its neighbors. If a node's capacity and storage space are altered, it again sends its present information to the segment forwarder. After that, the segment forwarder will select the maximum capacity nodes in its neighbors based on the updated information. That is, after a neighbor node receives a segment from the source, it uses either direct transmission or convey transmission. If the capacity of each of its neighbors is no greater than itself, relay node make use of direct transmission. If not, it uses convey transmission. In direct transmission, the relay nodes pass on the segment to a BS if it is in a BS's region. Or else, it stores the segment while moving until it goes into a BS's region. In relay transmission, relay node chooses its highest-capacity neighbor as the second relay node based on the QoS requirement. The second relay node will use through transmission to forward the segment directly to a BS. As a result, the number of transmission hops in the ad-hoc network component is confined to no more than two. The small number of hops helps to increase the capacity of the network and reduce channel conflict in ad-hoc transmission. The intention of the second hop choice is to find a higher capacity node as the message forwarder in order to pick up the performance of the QoS requirement.

If a source node has the maximum capability in its region, the segments will be forwarded rear to the source node according to the DTR protocol. The source node then forwards the segments to the BS straight due to the three-hop limit. This case occurs only when the source nodes is the maximum capacity node within its[9] two-hop neighborhood. Since the data transmission rate of the ad hoc interface is more than 10 times earlier than the cellular interface example 3G and GSM. Thus, the transmission wait for sending the data back and forth in the ad-hoc transmission is negligible in the total routing latency. After a BS receives a segment, it needs to forward the segment to the BS, where the destination node resides (i.e., the destination BS)..However, the destination BS recorded in the home BS may not be the most up-to-date destination BS since destination mobile nodes switch between the coverage regions of different BS during data transmission to them. For instance, data is transmitted to BS Bi that has the data's destination, but the destination has moved to the range of BS Bj before the data arrives at BS Bi. To deal with this problem, the[4] Cellular IP protocol is used for tracking node locations. With this protocol, a BS has a home agent and a foreign agent. The foreign agent keeps track of movable nodes moving into the ranges of other BS. The home agent intercepts in-coming segments, reconstructs the original data, and re-routes it to the foreign agent, which then forwards the data to the destination mobile node. After the destination BS receives the segments of a message, it rearranges the segments into the original message and then sends it to the destination mobile node. DTR specify the segment structure format for reschedule message. Each segment contains eight fields, including: (1) source node IP address; (2) destination node IP address; (3) message sequence number; (4) segment sequence number;(5) QoS indication number; (6) data; (7)length of the data; and (8) checksum.

3. PROPOSED WORK

Establishing the Network

The first step of network establishment is forming the cluster. The cluster is the group of related nodes formed in order to make the data transmission easier. every cluster will have Cluster top, Gateway and other nodes. The first criterion in wireless medium was to discover the available routes and establish them earlier than transmitting. The network consists of n nodes in which two nodes must be source and destination others will be used for data transmission. The path selection for data transmission is based on the availability of the nodes in the area using the ad-hoc on demand distance vector routing algorithm. Using the Ad-hoc on Demand Distance Vector routing protocol, the routes are created on demand as needed.



Threshold allocation

Threshold value distribution is done using TEEN, APTEEN and SPEED protocol. Based on the threshold value, trust node can be chosen also malicious node can be unobserved.

3.2.1 Threshold-sensitive Energy Efficient sensor Network protocol (TEEN)

It is a immediate protocol proposed for time-risky applications. The major idea of this technique is to produce the threshold value to every node in the network. After create the threshold value, the node is set in a hierarchical[6] clustering scheme in which some nodes act as a 1st level and 2nd level cluster heads. After forming the cluster head, the nodes get the data for transmission. Once the data is received the cluster head broadcasts the data to this cluster member.

Adaptive Threshold-sensitive Energy Efficient sensor Network protocol (ATEEN)

APTEEN is a hybrid [10]routing protocol planned for both time cyclic data collection and critical events. The main idea is to keep the statistical information. In this APTEEN method, the threshold value of each node in the cluster will be communicated with other cluster. Each cluster will have an APTEEN values.

SPEED Protocol

SPEED is a stateless protocol which provides real time communication by maintaining preferred release speed across the network. SPEED protocol is to discover geographic location. In this protocol whenever source nodes are[5] transmits a packet, the

next hop neighbor is acknowledge using Stateless Non deterministic Geographic Forwarding (SNGF). The SNGF identifies a node as next hop neighbor, if it belongs to neighboring set of nodes, lies within the range of destination area and having speed larger than confident desired speed.

Overhearing Technique

The path selection, preservation and data transmission is repeated process which happens in split seconds in real time transmission. Hence the path allocated priory is used for data transmission. The first path allocated previously is used for data transmission. The data is transferred through the tinted path. But the transmission lane may be unsuccessful some times. At that moment second path is selected for data transmission. It takes additional time to find the second path. In order to deal with these overhearing is used. The overhearing is the idea in which the sending nodes allocate data to more than one node in a network. If the node collapse occurs in a network, that can be substituted by other active node.

Three hop Routing

Three hops are used for data transmission in a network. Two hops at mobile ad-hoc network and one hop at infrastructure network. The usage of this amalgamation will pick up the reliability. In this technique, the network is silent until a connection is needed. The new nodes forwarded this message, and documentation the node that they heard it from, creating an blast of temporary routes is back to the wanted node. while a node receives such a message, it will send the message backwards through a fleeting route to the requesting node. The deprived node then begins using the route that is the least number of hops through other nodes. Idle entries in the routing table

4. CONCLUSION

Distributed Three-hop Routing protocol integrate the features of infrastructure and ad-hoc network in the data transmission process. In Distributed Three-hop Routing, source node divides a message flow into segments and broadcast them to its mobile neighbors and it further advance the segments to their target via an infrastructure network. Distributed Three-hop Routing restrictions the routing path length to three, and always arranges for high ability nodes to forward data. Distributed Three-hop Routing produces appreciably lower overhead by eliminating route find and maintenance.

TAS protocol is implemented in this work which distributes a threshold value to each and every node in a network for the collection of trust nodes. In addition, Overhearing technique is applied to find out and change the failure node in the network. . It has the characteristics of short path length, short-distance transmission, and balanced load distribution provides high routing reliability with high efficiency and also include congestion control algorithm which can avoid load congestion in Bs in the case of unbalanced traffic distributions in networks. Besides the data transmission in hybrid wireless network is highly secure and more efficient.

REFERENCES

[1]. Bengfort, W. Zhang, and X. Du "Efficient resource allocation in hybrid wireless networks," In Proc. of WCNC, 2011.

[2] L. M. Feeney, B. Cetin, D. Hollos, M. Kubisch, S. Mengesha, and H. Karl, "Multi-rate relaying for performance improvement in IEEE 802.11 wlans," In Proc. of WWIC, 2007.

[3] X. J. Li, B. C. Seet, and P. H. J. Chong, "Multi-hop cellular networks: Technology and economics," Computer Networks, 2008.

[4] K. Akkarajitsakul, E. Hossain, and D. Niyato, "Cooperative packet delivery in hybrid wireless mobile networks: A coalitional game approach," IEEE Trans. Mobile Computing 2013.

[5] P. Thulasiraman and X. Shen, "Interference aware resource allocation for hybrid hierarchical wireless networks," Computer Networks, 2010.

[6] L. B. Koralov and Y. G. Sinai, "Theory of probability and random processes," Berlin New York Springer, 2007.

[7] D. M. Shila, Y. Cheng, and T. Anjali, "Throughput and delay analysis of hybrid wireless networks with multi-hop uplinks," In Proc. of INFOCOM, 2011.

[9] T. Liu, M. Rong, H. Shi, D. Yu, Y. Xue, and E. Schulz, "Reuse partitioning in fixed two-hop cellular relaying network," In Proc. of WCNC, 2006.

[10] C. Wang, X. Li, C. Jiang, S. Tang, and Y. Liu, "Multicast throughput for hybrid wireless networks under Gaussian channels model," TMC, 2011.

proceedings of International Conference on Emerging Technologies in Computer Science (ICETCS) ISBN 978-93-85101-62-5

TESTING THE PERFORMANCE OF EAACK IN AUTHENTIC NETWORK ENVIRONMENT

A. Bala Chandana[1], D. Lakshminarayana reddy[2], A. Abhilash Reddy[3], A. Nandini [4]
Department of Computer Science & Engineering, MRCE, JNT University, Hyderabad, India,
e-mail[1]: balachandanacse501@gmail.com, email[2]: dlnreddy_cse@mrce.in
e-mail[3]: abhilashcse508@gmail.com, email[4]: nandinicse511@gmail.com

Abstract— The development to remote framework from wired framework has been a general example inside the late decades. The quality and quantifiability brought by remote framework make its potential in a couple of utilizations. Among all the best in class remote net-satisfies desires, Mobile Adhoc Network (MANET) is one in everything about overwhelming essential and diverse applications. On the notwithstanding matured assurance, MANET needn't trouble with a relentless framework base; every one single center point works as each a transmitter and a recipient. Center points talk particularly with each other once they are in degree between times steady correspondence shifts. Else, they place confide in their neighbors to exchange messages. The engineering toward oneself limit of center points in MANET made it in vogue among crucial mission applications like military usage or emergency recovery. In any case, the open medium and wide dispersal of center points make MANET subject to malicious aggressors. In the midst of this case, its crucial to make moderate intrusion acknowledgment parts to shield MANET from ambushes. With the upgrades of the designing and cut in fittings costs, we tend to range unit seeing a present example of extending MANET into mechanical applications. To figure out how to such example, we tend to persuasively acknowledge that its fundamental to handle its potential security issues. In the midst of this paper, we tend to propose and realize a fresh out of the box new interference area and revolution system named EAACK based Intrusion Detection and shirking structure using ECC approach phenomenally expected for MANET. Appeared differently in relation to extraordinary approaches, our strategy indicates higher threatening behavior revelation rates in without question conditions while doesn't unfathomably affect the framework presentations.

Keywords— Digital signature, Enhanced Adaptive ACKnowledgment (AACK) (EAACK), Mobile Adhoc NETwork (MANET), Elliptic Curve Cryptography (ECC)

I. INTRODUCTION

Because of their trademark quality and quantifiability, remote frameworks go unit unendingly most pervasive since the basic day of their creation. As a consequence of the upgraded designing and reduced costs, remote frameworks have grabbed rather more slant over wired frameworks inside the late decades. By definition, Mobile Ad hoc Network (MANET) is an arranged of flexible center points outfitted with each a remote transmitter and a recipient that talk with each other through bidirectional remote joins either direct or by suggestion. Advanced remote get to and organization by method for remote frameworks are getting additional and additional in style starting now [35]. One in all the key blessings of remote frameworks is its ability to permit electronic correspondence between absolutely particular get-togethers and still keep up their quality. Regardless, this correspondence is restricted to the move of transmitters. This underwear that 2 centers can't talk with each other once the space between the 2 centers is on the far side the correspondence changes of their own. MANET comprehends this inconvenience by permitting midway get-togethers to

hand-off information transmissions. This is consistently refined by segregating MANET into 2 blends of frameworks, to be particular, single hop and multihop. In the midst of a single bob sort out, all center points among a practically identical radio vary relate clearly with each other. On the reverse hand, in the midst of a multihop framework, centers surrender unmistakable center points to transmit if the end of the line center point is out of their radio vary. In instead of the customary remote framework, MANET joins a suburbanized framework establishment. MANET needn't trouble with a steadfast establishment; thusly, all centers locale unit unengaged to move self-emphatically [10], [27], [29]. MANET is prepared for making a sorting out toward oneself and keeping toward oneself up framework while not the support of a bound together establishment, that is for the most part unfeasible in huge mission applications like military conflict or emergency recovery. Tokenize setup Associate in fast preparation make MANET prepared to be utilized as a part of emergency conditions wherever a base is out of reach or impracticable to put in circumstances like trademark or human-influenced disasters, military conflicts, and restorative emergency things [19], [30].

Owing to these different attributes, MANET will be getting to be extra and extra wide authorized inside the exchange [14], [28]. Notwithstanding, considering the very truth that MANET is in style among urgent mission applications, system security will be of essential imperativeness. Unfortunately, the open medium and remote appropriation of MANET make it inclined to shift mixed bags of assaults. For example, as a consequence of the hubs' need of physical security, vindictive aggressors will basically catch and bargain hubs to accomplish assaults. uniquely, considering the extremely certainty that the lion's share steering conventions in MANETs accept that every hub inside the system acts hand and glove with distinctive hubs and possibly not malignant [5], assailants will just trade off MANETs by embeddings vindictive or no agreeable hubs into the system. Furthermore, subsequently of MANET's disseminated plan and dynamical topology, a traditional incorporated recognition strategy isn't any longer conceivable in MANETs. In such case, it's pivotal to create Associate in Nursing interruption identification framework (IDS).

II. Related Work

A. Interruption Detection in Manets:

As said some time recently, as an aftereffect of the limitations of most MANET steering conventions, hubs in Manets accept that diverse hubs constantly work with each other to transfer data. This supposition leaves the assailants with the chances to accomplish imperative effect on the system with just one or 2 traded off hubs. to handle this downside, partner IDS should to be extra to fortify the assurance level of Manets. On the off chance that MANET will watch the aggressors as a little while later as they enter the system, we'll be capable to completely dispose of the potential harms brought on by bargained hubs at the essential time. Idss now and again act in light of the fact that the second layer in Manets, and that they zone unit an fantastic supplement to existing proactive approaches [27]. Anantvalee and Wu tongue [4] given a dreadfully exhaustive overview on up to date Idss in Manets. In this segment, we tend to fundamentally depict 3 exist ing approaches, to be specific, Watchdog [17], TWOACK [15], and adjustive Acknowledgment (AACK) [25].1)

1) Watchdog: [17] anticipated a topic named Watchdog that expects to help the yield of system with the vicinity of malignant hubs. In actuality, the Watchdog subject is comprised of 2 components, to be specific, Watchdog and Path rater. Guard dog is partner IDS for Manets. It's subject for police examination malevolent hub mischievous activities in the system. Guard dog recognizes pernicious mischievous activities by wantonly being mindful to its next bounce's transmission. In the event that a Watchdog hub catches that its next hub comes up short to forward the parcel among a specific sum of your time, it will build its disappointment counter. At the point when ever a hub's disappointment counter surpasses a predefined edge, the Watchdog hub reports it as acting up. Amid this case, the Path rater coordinates with the steering conventions to dodge the reported hubs in future transmission. Numerous after

investigation studies and executions have demonstrated that the Watchdog topic is conservative. Besides, contrasted with an alternate plans, Watchdog will be proficient of police examination malignant hubs rather of connections. These profits have made the Watchdog topic an overall enjoyed option inside the field. A few MANET Idss zone unit either upheld or created as partner change to the Watchdog subject [15], [20], [21], [25]. All the same, as seen by Marti et al. [17], the Watchdog subject comes up short to watch noxious mischievous activities with the vicinity of the accompanying: 1) uncertain impacts; 2) collector crashes; 3) Restricted transmission power; 4) false offense report;5) arrangement; and 6) fractional dropping.

2) Twoack: with respect to the six shortcomings of the Watchdog subject, a few scientists anticipated new methodologies to unwind these issues. TWOACK anticipated by Liu et al. [16] will be one in all the first vital approaches among them. On the as opposed to a few distinctive plans, TWOACK is not partner sweetening or a Watchdog-based topic. Getting to purpose the collector impact and limited transmission power issues Of Watchdog, TWOACK distinguishes making trouble Joins by recognizing every data bundle transmitted over every 3 successive hubs on the trail from the supply to the goal. Upon recovery of a parcel, each hub on the course is expected to test partner affirmation bundle to the hub that is 2 jumps remote from it down the course. TWOACK is expected to figure on steering conventions like Dynamic supply Routing (DSR) [11]. The working technique for TWOACK is demonstrated in Fig. one: Node an essential advances Packet 1 to hub B, and then, hub B advances Packet one to hub C. when hub C gets Packet one, on the grounds that it will be 2 bounces detached from hub A, hub C will be obligation -bound to come up with a TWOACK bundle, that contains converse course from hub A to hub C, and sends it over to hub A. The recovery of this TWOACK bundle at hub A demonstrates that the transmission of Packet one from hub A to hub C is blessed. Something else, if this TWOACK parcel isn't gotten in an exceedingly predefined period, every hubs B and C range unit reported pernicious. Indistinguishable system applies to every 3 successive hubs on the rest of the course.

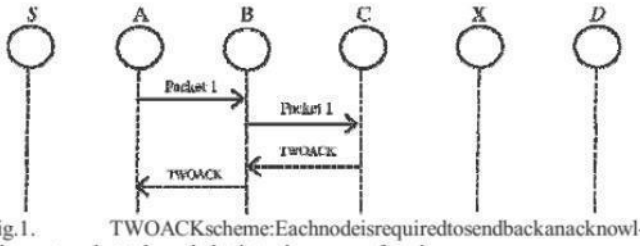


Fig.1. TWOACKscheme:Eachnodeisrequiredtosendbackanacknowl-edgment packettothenodethatistwohops awayfromit.

The TWOACK topic with achievement understands the beneficiary impact and limited transmission force issues uncover by Watchdog. Be that as it may, the affirmation technique required in every parcel transmission strategy extra an enormous amount of undesirable system overhead. as a consequence of the limited battery power nature of Manets, such repetitive transmission technique will essentially corrupt the lifetime of the entire system. Notwithstanding, a few examination studies zone unit working in vitality social event to handle this disadvantage [25], [28], [29].

3) AACK: backed TWOACK, [25] ace uncover another subject alluded to as AACK. practically like TWOACK, AACK will be partner affirmation based system layer subject which may be thought-about as a mixture of a topic alluded to as TACK (indistinguishable to TWOACK) related an end-to-end affirmation topic alluded to as Acknowledge (ACK). Contrasted with TWOACK, AACK significantly decreased system overhead though still fit of keeping up or maybe surpassing indistinguishable system yield. The end-to-end affirmation topic in ACK is demonstrated in Fig. 2.

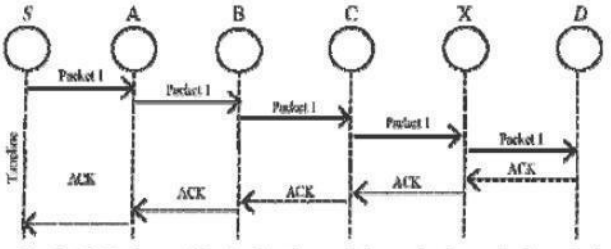


Fig.2.ACKscheme:Thedestinationnodeisrequiredtosendacknowledg-ment packetstothesourcenode.

In the ACK topic indicated in Fig. 2, the supply hub S sends out Packet one with none overhead aside from two b of banner showing the parcel sort. All the halfway hubs only forward this parcel. Once the end hub D gets Packet one, its required to test partner ACK affirmation bundle to the supply hub S on the reverse request of indistinguishable course. Among a predefined period, if the supply hub S gets this ACK affirmation bundle, then the parcel transmission from hub S to hub D will be lucky. Something else, the supply hub S can change to TACK topic by causation out a TACK bundle. The origination of embracing a half breed topic in AACK incredibly diminishes the system overhead, however every TWOACK and AACK still experience the ill effects of the matter that they fizzle to watch vindictive hubs with the vicinity of false offense report and cast affirmation bundles.

B. Computerized Signature:

Computerized marks have consistently been partner fundamental a part of cryptography in history. Cryptography will be that the study of numerical systems connected with parts of information security like secrecy, learning respectability, substance validation, and learning starting point confirmation [18].

The occasion of cryptography strategy offers a long and fascinating history. The quest for secure correspondence has been directed by individual since 4000 years agone in Egypt, in keeping with Kahn's book [30] in 1963. Such advancement drastically quickened since the globe War II, that some accept is essentially on account of the financial {process} process. The security in Manets is plot as a mixof methodologies, methodology, and frameworks won't to ensure secrecy, verification, respectability, accessibility, and non-revocation [18]. Computerized mark may be a wide embraced methodology to affirm the confirmation, honesty, and non-denial of Manets. To guarantee the legitimacy of the computerized signature, the sender Alice will be committed to constantly keep her individual key Pr Alice as a mystery while not uncovering to anybody else.

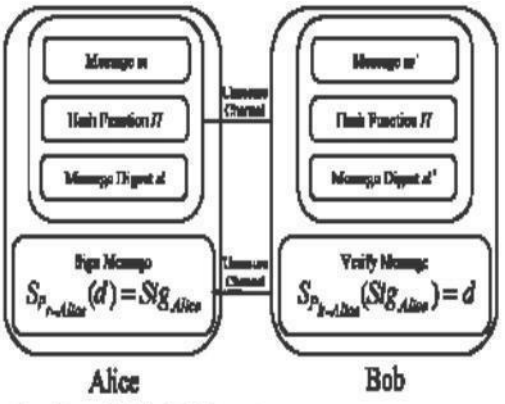


Fig.3.Communicationwithdigitalsignature.

Something else, if the aggressor Eve gets this mystery individual key, she will be capable to capture the message and basically produce vindictive messages with Alice's signature and send them to Bob. As these malevolent messages will be digitally marked by Alice,

Bob sees them as genuine and true messages from Alice. Along these lines, Eve will immediately achieve pernicious assaults to Bob or maybe the complete system. Next, Alice will send a message m in conjunction with the signature Sigalice to Bob by means of partner unsecured channel.

Weave then processes the got message m against the preagreed hash work H to urge the message digest d. This system are frequently summed up as H (m) = d . (3) Bob will confirm the signature by applying Alice'sopen keypk−alice on Sigalice , by utilizing can be summed up as an data string, that partners a message (in computerized structure) with some starting element, or partner electronic Spk Alice (Sigalice) = d. (4) Digital signature plans will be regularly in the fundamental separated into the consequent 2 classes.

1) Digital signature with index: the starting message is required inside the signature confirmation recipe. Samples exemplify an advanced mark recipe (DSA) [33].

2) Digital signature with message recuperation: this sort of topic doesn't need the other information other than the signature itself inside the check strategy.

Illustrations epitomize RSA [23]. On the off chance hat d == d, then its safe to say that the message M transmitted through partner unsecured channel will be So sent from Alice furthermore the message itself is unbroken.

III. Problem Definition

Our proposed methodology EAACK with ECC is intended to handle three of the six shortcomings of Watchdog plan, specifically, false bad conduct, constrained transmission force, and beneficiary crash and to give Security in bundle conveyance. In this area, we talk about these three shortcomings in point of interest.

In a normal sample of collector crashes, indicated in Fig. 4, once hub A sends Packet one to hub B, it tries to take in if hub B sent this bundle to hub C; meanwhile, hub X is sending Packet a couple of to hub C. In such case, hub A catches that hub B has with achievement sent Packet one to hub C however didn't watch that hub C neglected to get this parcel as a result of a impact between Packet one and Packet a couple of at hub C.

On account of limited transmission control, in order to safeguard its own particular battery assets, hub B intentionally restricts its transmission control in place that its sufficiently strong to be caught by hub A however not sufficiently hearty to be gotten by hub C, as demonstrated in Fig. 5.

For false wrongful transmit report, however hub A with achievement caught that hub B sent Packet one to hub C, hub A still reputed hub B as acting mischievously, as demonstrated in Fig. 6. As a result of the open medium and remote circulation of average Manets, aggressors will just catch and bargain one or 2 hubs to achieve this false wrongful behavior report assault.

As said in past areas, TWOACK and AACK comprehend 2 of those 3 shortcomings, In particular, recipient crash and limited transmission power. Notwithstanding, every of them range unit at hazard of the false wrongful convey assault. amid this investigation work, our objective will be to propose a brand new IDS extraordinarily planned for Manets, that unravels not exclusively beneficiary crash and limited transmission power however moreover the false wrongful behavior drawback. Besides, we have a inclination to stretch out our investigation to receive an advanced sig nature subject all through the parcel transmission strategy. As all told affirmation based Idss, its critical to verify the uprightness and validity of all affirmation bundles. In this segment, we have a tendency to portray our anticipated EAACK topic altogether. The approach depicted amid this examination paper depends on our past work [12], wherever the spine of EAACK was anticipated and assessed through execution. Amid this work, we tend to amplify it with the presentation of advanced signature to hinder the assaulter from arrangement affirmation bundles.

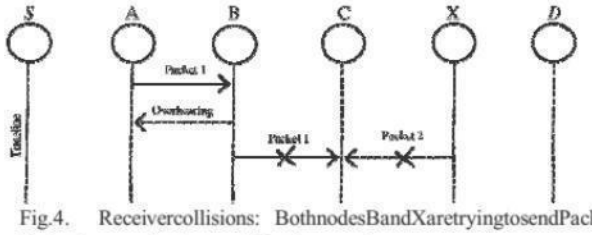


Fig.4. Receivercollisions: BothnodesBandXaretryingtosendPacket1and Packet2,respectively,tonodeCatthesame time.

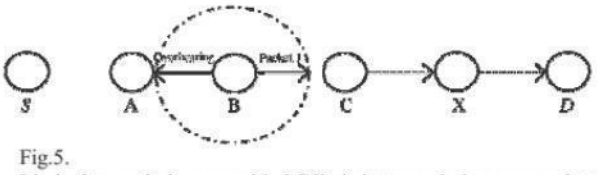


Fig.5. Limitedtransmissionpower:NodeBlimitsitstransmissionpowersothatthepack ettransmissioncanbeoverheardbynodeA buttooweaktoreachnode C.

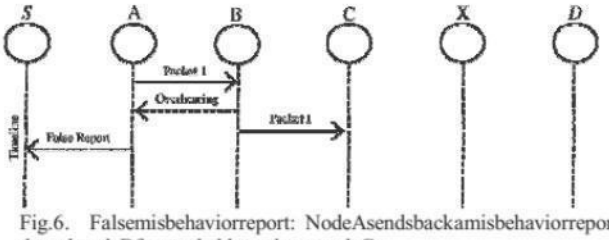


Fig.6. Falsemisbehaviorreport: NodeAsendsbackamisbehaviorreporteven thoughnodeBforwardedthepackettо nodeC.

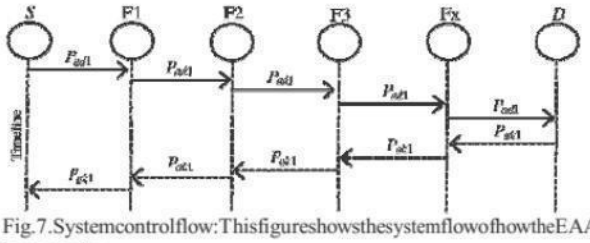


Fig.7.Systemcontrolflow:Thisfigureshowsthesystemflowofhowthe EAACKsc hemeworks.

IV. System Design

EAACK will be comprised of 3 major segments, to be specific, ACK, secureACK (S-ACK),and wrongdoing report confirmation (MRA). so as completely

separate to tell apart} distinctive parcel mixtures in diverse plans, we tend to encased a 2-b bundle header in EAACK. Concurring to the web draft of DSR [11], there's about six b held inside the DSR header. In EAACK, we tend to use two b of the about six b to banner contrasting sorts of bundles. Fig. 7 (indicated later) presents a stream graph depicting the EAACK subject. If its not too much trouble note that, in our anticipated subject, we Tend to accept that the join between each hub inside the system will be bifacial. What will be more, for Each correspondence strategy, every the supply hub furthermore the goal hub don't appear to be malevolent. Unless nominative, all affirmation parcels depicted amid this examination square measure expected to be digitally marked by its sender and checked by its collector.

A. ACK

As specified in the recent past, ACK is basically partner end-to end affirmation subject. It acts as a locale of the cross breed topic in EAACK, going to scale back system overhead once no system unfortunate behavior is discovered. In Fig. 8, in ACK mode, hub S starting conveys partner ACK data bundle Pad1 to the end of the line hub D. In the event that all the middle of the road hubs on the course between hubs S and D square measure agreeable and hub D with achievement gets Pad1 , hub D will be required to remand partner ACK affirmation bundle Pak1 on a comparative course however in a extremely reverse request. inside a predefined principal amount, if hub S gets Pak1 , then the parcel transmission from hub S to hub D is winning. Something else, hub S can change to S-ACK mode by creating out partner S-ACK data parcel to sight The acting mischievously hubs inside the course.

B. S-ACK

The S-ACK topic will be partner Enhanced variant of the TWOACK subject anticipated by Liu et al. [16]. The standard is to let every 3 back to back hubs add a gaggle to sight getting rowdy hubs. for every 3 successive hubs inside the course, the third hub will be required to send partner S-ACK affirmation bundle to the essential hub. The proposition of presenting S-ACK mode will be to sight acting up hubs inside the vicinity of beneficiary impact or confined transmission power.

As demonstrated in Fig. 9, in S-ACK mode, the 3 successive hubs (i.e., F1, F2, and F3) include a gaggle to sight getting into mischief hubs inside the system. Hub F1 introductory sends out S-ACK data bundle Psad1 to hub F2. At that point, hub F2 advances this bundle to hub F3. When hub F3 gets Psad1 , on the grounds that it is that the third hub amid this three-hub group, hub F3 will be required to remand partner S-ACK affirmation parcel Psak1 to hub F2. Hub F2 advances Psak1 once again to hub F1. On the off chance that hub F1 doesn't get this affirmation bundle inside a predefined major amount, every hubs F2 and F3 square measure supposed as noxious. In addition, an unfortunate behavior report are created by hub F1 and sent to the supply hub S. In any case, not like the TWOACK subject, wherever the supply hub like a shot trusts the wrongdoing report, EAACK needs the supply hub to adjust to MRA mode and verify this unfortunate behavior report. This can be an essential step to sight false wrongdoing report in our anticipated subject.

C. MRA

The MRA topic will be implied to resolve the shortcoming of Watchdog once it comes up short to sight getting out of hand hubs with the vicinity of false wrongdoing report. The false offense report might be created by malevolent assailants to inaccurately report honest hubs as malignant. This assault might be lethal to the complete system once the aggressors break down enough hubs thus cause a system division. The center of MRA subject is to bear witness to whether the goal hub has gotten the supposed missing parcel through an extraordinary course. To launch the MRA mode, the supply hub introductory ventures its local mental object and looks for for an interchange course to the objective hub. In the event that there's no option that exists, the supply hub begins a DSR directing appeal to search out an alternate course. Attributable to the character of Manets, its regular to search out different courses between 2 hubs. By receiving a substitute course to the objective hub, we tend to bypass the unfortunate behavior newsperson hub. Once the end hub gets partner MRA parcel, it hunts its local information base and analyzes if the supposed bundle was gotten. On the off chance that its now gotten, then its safe to close that this can be a

false offense report and whoever created this report will be checked as vindictive. Something else, the offense report is trusty and acknowledged. By the reception of MRA subject, EAACK is equipped for sleuthing pernicious hubs notwithstanding the presence of false misconduct report. D. Advanced Signature As specified some time recently, EAACK will be partner affirmation based IDS. All 3 segments of

EAACK, in particular, ACK, S-ACK, and MRA, square measure affirmation based discovery plans. Every one of them accept on affirmation parcels to sight mischievous activities inside the system. Hence, its phenomenally important to verify that each one affirmation bundles in EAACK square measure legitimate and untainted. Something else, if the assailants square measure great enough to fashion affirmation parcels, all of the3 schemes can be powerless. With reference to this basic concern, we tend to consolidated computerized signature in our anticipated topic. so as to verify the trustworthiness of the IDS, EAACK needs all affirmation bundles to be digitally marked before they're sent out and checked till they're acknowledged. Nonetheless, we tend to completely see the extra assets that square measure required with the presentation of advanced signature in Manets. To bargain with this worry, we have a tendency to authorized every DSA [33] and RSA [23] computerized signature plans in our anticipated methodology. The objective will be to search out the principal best determination for exploitation computerized signature in Manets.

V. Conclusion

Bundle dropping assault has constantly been a genuine danger to the security in MANETs. In this examination paper, we tend to have arranged totally remarkable IDS named EAACK convention exceptionally de marked for MANETs and looked at it against diverse standard components in a few circumstances through recreations. The results will be positive exhibitions against Watchdog, TWOACK, and AACK inside the cases of recipient impact, confined transmission control, and false wrongful behavior report.

References

[1] K. Al Agha, M.-H.Bertin, T. Dang, A. Guitton, P. Minet, T. Val, and J.-B. Viollet, "Which wireless technology for industrial wireless sensor networks? The development of OCARI technol,"IEEE Trans. Ind.Electron., vol. 56, no. 10, pp. 4266–4278, Oct. 2009.

[2] R. Akbani, T. Korkmaz, and G. V. S. Raju, "Mobile Ad hoc Network Security," inLecture Notes in ElectricalEngineering, vol. 127. New York: Springer-Verlag, 2012, pp. 659–666.

[3] R. H. Akbani, S. Patel, and D. C. Jinwala, "DoS attacks in mobile ad hoc networks: A survey," inProc. 2nd Int.Meeting ACCT, Rohtak, Haryana, India, 2012, pp. 535–541.

[4] T. Anantvalee and J. Wu, "A Survey on Intrusion Detection in Mobile Ad Hoc Networks," inWireless/MobileSecurity. New York: SpringerVerlag, 2008.

[5] L. Buttyan and J. P. Hubaux, Security and Cooperation in Wireless Networks. Cambridge, U.K.: CambridgeUniv. Press, Aug. 2007.

[6] D. Dondi, A. Bertacchini, D. Brunelli, L. Larcher, and L. Benini, "Modeling and optimization of a solar energyharvester system for self-powered wireless sensor networks,"IEEE Trans. Ind. Electron., vol. 55, no. 7, pp.2759–2766, Jul. 2008.

[7] V. C. Gungor and G. P. Hancke, "Industrial wireless sensor networks: Challenges, design principles, andtechnical approach,"IEEE Trans. Ind. Electron., vol. 56, no. 10, pp. 4258–4265, Oct. 2009.

[8] Y. Hu, D. Johnson, and A. Perrig, "SEAD: Secure efficient distance vector routing for mobile wireless ad hocnetworks," inProc. 4th IEEE Workshop Mobile Comput. Syst. Appl., 2002, pp. 3–13.

[9] Y. Hu, A. Perrig, and D. Johnson, "ARIADNE: A secure on-demand routing protocol for ad hoc networks," in Proc. 8th ACM Int. Conf. MobiCom, Atlanta, GA, 2002, pp. 12–23.

[10] G. Jayakumar and G. Gopinath, "Ad hocmobile wireless networks routing protocol—A review," J. Comput.Sci., vol. 3, no. 8, pp. 574–582, 2007.

[11] D. Johnson and D. Maltz, "Dynamic Source Routing inadhocwireless networks," inMobile Computing.Norwell, MA: Kluwer, 1996, ch. 5, pp. 153–181.

[12] N. Kang, E. Shakshuki, and T. Sheltami, "Detecting misbehaving nodes in MANETs," in Proc. 12th Int. Conf.iiWAS, Paris, France, Nov. 8– 10, 2010, pp. 216–222.

[13] N. Kang, E. Shakshuki, and T. Sheltami, "Detecting forged acknowledgements in MANETs," inProc. IEEE25th Int. Conf. AINA, Biopolis, Singapore, Mar. 22–25, 2011, pp. 488–494.

[14] K. Kuladinith, A. S. Timm-Giel, and C. Görg, "Mobilead-hoccommunications in AEC industry,"J. Inf.Technol. Const., vol. 9, pp. 313–323, 2004.

[15] J.-S. Lee, "A Petri net design of command filters for semiautonomous mobile sensor networks,"IEEE Trans.Ind. Electron., vol. 55, no. 4, pp. 1835–1841, Apr. 2008.

[16] K. Liu, J. Deng, P. K. Varshney, and K. Balakrishnan, "An acknowledgment-based approach for the detectionof routing misbehaviour in MANETs," IEEE Trans. Mobile Comput., vol. 6, no. 5, pp. 536–550, May 2007.

[17] S. Marti, T. J. Giuli, K. Lai, and M. Baker, "Mitigating routing misbehaviour in mobile ad hoc networks,"inProc.6th Annu.Int. Conf. Mobile Comput.Netw., Boston, MA, 2000, pp. 255–265.

[18] A. Menezes, P. van Oorschot, and S. Vanstone, Handbook of Applied Cryptography. Boca Raton, FL: CRC,1996, T-37.

[19] N. Nasser and Y. Chen, "Enhanced intrusion detection systems for discovering malicious nodes in mobile ad hoc network," inProc. IEEE Int. Conf. Commun., Glasgow, Scotland, Jun. 24–28, 2007, pp. 1154–1159.

[20] J. Parker, J. Undercoffer, J. Pinkston, and A. Joshi, "On intrusion detection and response for mobile adhocnetworks," in Proc. IEEE Int. Conf. Perform., Comput.,Commun., 2004, pp. 747–752.

[21] A. Patcha and A. Mishra, "Collaborative security architecture for black hole attack prevention in mobile ad hoc networks," inProc. Radio Wireless Conf., 2003, pp. 75–78.

[22] A. Patwardhan, J. Parker, A. Joshi, M. Iorga, and T. Karygiannis, "Secure routing and intrusion detection in ad hoc networks," inProc.3rd Int. Conf. Pervasive Comput.Commun., 2005, pp. 191–199.

[23] R. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and publickeycryptosystems,"Commun. ACM, vol. 21, no. 2, pp. 120–126, Feb. 1983.

CHALLENGES AND OPPORTUNITIES WITH BIG DATA

[1]Mr.A.Saideep [2]Dr.Sunil Tekale,[3]Mr.P.Amarnath, [4]Ms.N.vanaja
[2]Prof-CSE ,[3]Asso Prof ,[4]Asso Prof
[2]Sunil.tekale2010@gmail.com [3]amar.sap16@gmail.com,[4]vanajacse@gmail.com
Department of Computer Science & Engineering

ABSTRACT

We are awash in a flood of data today. In a broa d range of application areas, data is being collected at unprecedented scale. Decisions that previously were based on guesswork, or on painstakingly constructed models of reality, can now be made based on the data itself. Such Big Data analysis now drives nearly every aspect of our modern society, including mobile services, retail, manufacturing, financial services, life sciences, and physical sciences.

Heterogeneity, scale, timeliness, complexity, and privacy problems with Big Data impede progress at all phases of the pipeline that can create value from data. The problems start right away during data acquisition, when the data tsunami requires us to make decisions, currently in an ad hoc manner, about what data to keep and what to discard, and how to store what we keep reliably with the right metadata. Much data today is not natively in structured format; for example, tweets and blogs are weakly structured pieces of text, while images and video are structured for storage and display, but not for semantic content and search: transforming such content into a structured format for later analysis is a major challenge.

Review:

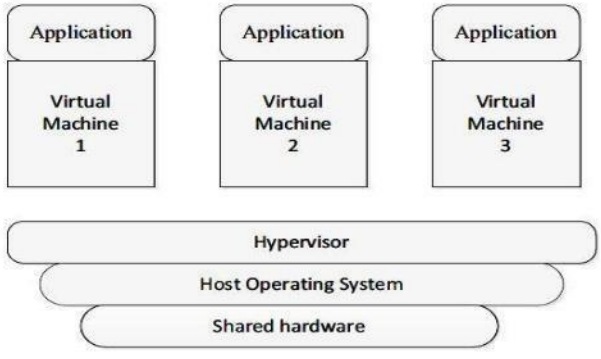
Data analysis is a clear bottleneck in many applications, both due to lack of scalability of the underlying algorithms and due to the complexity of the data that needs to be analyzed[3]. Finally, presentation of the results and its interpretation by non-technical domain experts is crucial to extracting actionable knowledge. A problem

with current Big Data analysis is the lack of coordination between database systems, which host the data and provide SQL querying, with analytics packages that perform various forms of non-SQL processing, such as data mining and statistical analyses. Today's analysts are impeded by a tedious process of exporting data from the database, performing a non-SQL process and bringing the data back[5]. This is an obstacle to carrying over the interactive elegance of the first generation of SQLdriven OLAP systems into the data mining type of analysis that is in increasing demand. A tight coupling between declarative query languages and the functions of such packages will benefit both expressiveness and performance of the analysis.

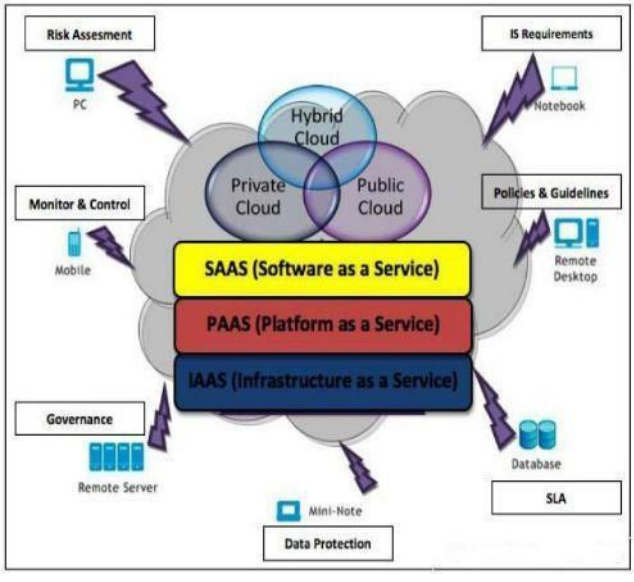
Keywords: Semantic, data base, data mining, data acquisition, Data extraction, cleaning

Introduction

Big Data has the potential to revolutionize not just research, but also education [1]. A recent detailed quantitative comparison of different approaches taken by 35 charter schools in NYC has found that one of the top five policies correlated with measurable academic effectiveness was the use of data to guide instruction [2]. Imagine a world in which we have access to a huge database where we collect every detailed measure of every student's academic performance. This data could be used to design the most effective approaches to education, starting from reading, writing, and math, to advanced, college-level, courses. We are far from having access to such data, but there are powerful trends in this direction.



In particular, there is a strong trend for massive Web deployment of educational activities, and this will generate an increasingly large amount of detailed data about students' performance. It is widely believed that the use of information technology can reduce the cost of healthcare while improving its quality [3], by making care more preventive and personalized and basing it on more extensive (home-based) continuous monitoring. McKinsey estimates a savings of 300 billion dollars every year in the US alone. In a similar vein, there have been persuasive cases made for the value of Big Data for urban planning (through fusion of high-fidelity geographical data), intelligent transportation (through analysis and visualization of live and detailed road network data), environmental modeling (through sensor networks ubiquitously collecting data) [4], energy saving (through unveiling patterns of use), smart materials (through the new materials genome initiative , computational social sciences 2 (a new methodology fast growing in popularity because of the dramatically lowered cost of obtaining data) , financial systemic risk analysis (through integrated analysis of a web of contracts to find dependencies between financial entities), homeland security (through analysis of social networks and financial transactions of possible terrorists), computer security (through analysis of logged information and other events, known as Security Information and Event Management (SIEM)), and so on.



Challenges in Big Data Analysis

There are various challenges that needs to be addressed in handling Big Data. As the name suggest big data, its really very big to mange because of

1. Heterogeneous Formats
2. Scale
3. Timeliness
4. Incompleteness
5. Privacy
6. Human Collaboration

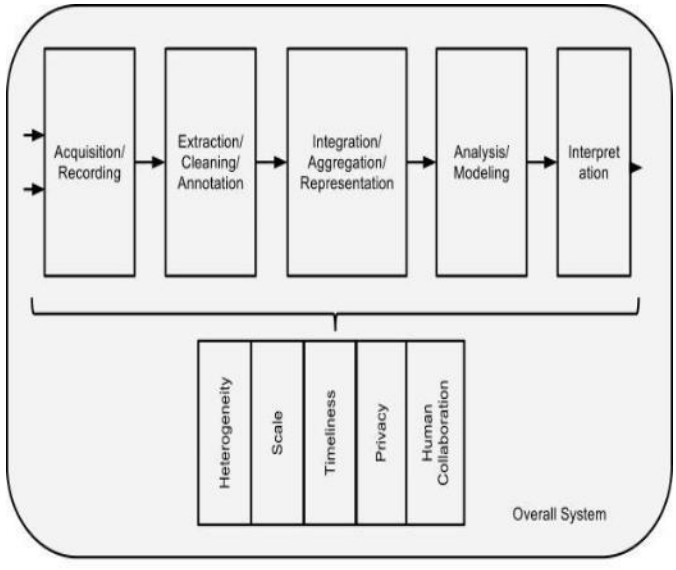


Figure 1: The Big Data Analysis Pipeline. Major steps in analysis of big data are shown in the flow at top. Below it are big data needs that make these tasks challenging.

Data Acquisition and Recording

Big Data does not arise out of a vacuum: it is recorded from some data generating source. For example, consider our ability to sense and observe thworld around us, from the heart rate of an elderly citizen, and presence of toxins in the air we breathe, to the planned square kilometer array telescope, which will produce up to 1 million terabytes of raw data per day. Similarly, scientific experiments and simulations can easily produce petabytes of data today.

Information Extraction and Cleaning

Frequently, the information collected will not be in a format ready for analysis. For example, consider the collection of electronic health records in a hospital, comprising transcribed dictations from several physicians, structured data from sensors and measurements (possibly with some associated uncertainty), and image data such as x-•rays. We cannot leave the data in this form and still effectively analyze it. Rather we require an information extraction process that pulls out the required information from the underlying sources and expresses it in a structured form suitable for analysis. Doing this correctly and completely is a continuing technical challenge. Note that this data also includes images and will in the future include video; such extraction is often highly application dependent (e.g., what you want to pull out of an MRI is very different from what you would pull out of a picture of the stars, or a surveillance photo). In addition, due to the ubiquity of surveillance cameras and popularity of GPS enabled mobile phones, cameras, and other portable devices, rich and high fidelity location and trajectory (i.e., movement in space) data can also be extracted.

Data Integration, Aggregation, and Representation

Given the heterogeneity of the flood of data, it is not enough merely to record it and throw it into a repository. Consider, for example, data from a range of scientific experiments. If we just have a bunch of data sets in a repository, it is unlikely anyone will ever be able to find, let alone reuse, any of this data. With adequate metadata, there is some hope, but even so, challenges will remain due to differences in experimental details and in data record structure.

Query Processing, Data Modeling, and Analysis

Methods for querying and mining Big Data are fundamentally different from traditional statistical analysis on small samples. Big Data is often noisy, dynamic, heterogeneous, inter-•- related and untrustworthy. Nevertheless, even noisy Big Data could be more valuable than tiny samples because general statistics obtained from frequent patterns and correlation analysis usually overpower individual fluctuations and often disclose more reliable hidden patterns and knowledge. Further, interconnected Big Data forms large heterogeneous information networks, with which information redundancy can be explored to compensate for missing data, to crosscheck conflicting cases, to validate trustworthy relationships, to disclose inherent clusters, and to uncover hidden relationships and models.

Interpretation

Having the ability to analyze Big Data is of limited value if users cannot understand the analysis. Ultimately, a decision-maker, provided with the result of analysis, has to interpret these results. interpretation cannot happen in a vacuum. Usually, it involves examining all the assumptions made and retracing the analysis. Furthermore, as we saw above, there are many possible sources of error: computer systems can have bugs, models almost always have assumptions, and results can be based on erroneous data. For all of these reasons, no responsible user will cede authority to the computer system. Rather she will try to understand, and verify, the results produced by the computer. The computer system must make it easy for her to do so.

Opportunities in Big Data

1. Soaring Demand for Analytics Professionals

2. Huge Job Opportunities & Meeting the Skill Gap

3. Big Data Analytics: A Top Priority in a lot of Organizations

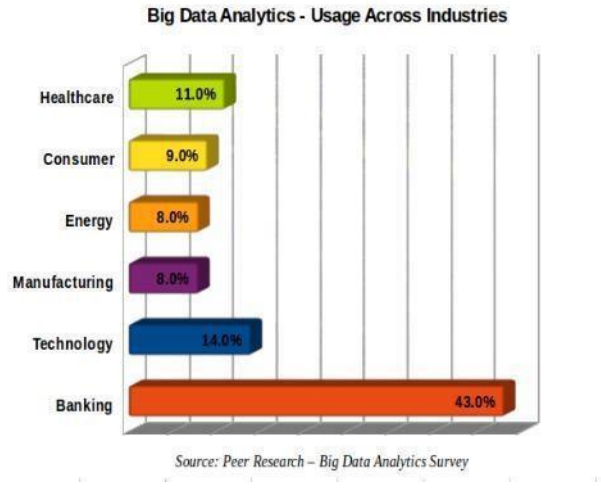
4. Adoption of Big Data Analytics is Growing:

5. Analytics: A Key Factor in Decision Making

6. The Rise of Unstructured and Semistructured Data Analytics:

7. Surpassing Market Forecast / Predictions for Big Data Analytics:

8. Numerous Choices in Job Titles and Type of Analytics **:**



Proposed System:

Roughly there are two types of approaches for big data analytics

1. Parallelize existing (single-machine) algorithms.

2. Design new algorithms particularly for distributed settings

The problem now is we take many things for granted on one computer On one computer, have you ever worried about calculating the average of some numbers? Probably not. You can use Excel, statistical software (e.g., R and SAS), and many things else We seldom care internally how these tools work Can we go back to see the early development on one computer and learn some lessons/experiences.

Consider the example of matrix-matrix products C = A × B, A ∈ R n×d , B ∈ R d×m where Cij = X d k=1 AikBkj This is a simple operation. You can easily write your own code

A segment of C code (assume n = m here) for (i=0;i<n;i++)

```
for (j=0;j <N;J++)
{
C[I][J]=0;
For(k=0;k<n;k++)
C[i][j]+=a[i][k]*b[k][j];
}
```

But on Matlab (single-thread mode) $ matlab -singleCompThread >> tic; c = a*b; toc Elapsed time is 4.095059 seconds.

CPU ↓ Registers ↓ Cache ↓ Main Memory ↓ Secondary storage (Disk) ↑: increasing in speed ↓: increasing in capacity Optimized BLAS: try to make data

For big-data analytics, we are in a similar situation We want to run mathematical algorithms (classification and clustering) in a complicated architecture (distributed system) But we are like at the time point before optimized BLAS was developed.

Conclusion We have entered an era of Big Data. Through better analysis of the large volumes of data that are becoming available, there is the potential for making faster advances in many scientific disciplines and improving the profitability and success of many enterprises. However, many technical challenges described in this paper must be addressed before this potential can be realized fully. The challenges include not just the obvious issues of scale, but also heterogeneity, lack of structure, error-handling, privacy, timeliness, provenance, and visualization, at all stages of the analysis pipeline from data acquisition to result interpretation. These technical challenges are common across a large variety of application domains, and therefore not cost-effective to address in the context of one domain alone. Furthermore, these challenges will require transformative solutions, and will not be addressed naturally by the next generation of industrial products. We must support and encourage fundamental research towards addressing these technical challenges if we are to achieve the promised benefits of Big Data.

REFERENCES

[1] Big Data. Nature (http://www.nature.com/news/specials/bigdata/index.html), Sep 2008.

[2] Data, data everywhere. The Economist (http://www.economist.com/node/15557443), Feb 2010.

[3] Drowning in numbers – Digital data will flood the planet—and help us understand it better. The Economist (http://www.economist.com/blogs/dailychart/2011/11/bigdata-0), Nov 2011.

[4] D. Agrawal, P. Bernstein, E. Bertino, S. Davidson, U. Dayal, M. Franklin, J. Gehrke, L. Haas, A. Halevy, J. Han, H. V. Jagadish, A. Labrinidis, S. Madden, Y. Papakonstantinou, J. M. Patel, R. Ramakrishnan, K. Ross, C. Shahabi, D. Suciu, S. Vaithyanathan, and J. Widom. Challenges and Opportunities with Big Data – A community white paper developed by leading researchers across the United States. http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf, Mar 2012.

[5] S. Lohr. The age of big data. New York Times (http://www.nytimes.com/2012/02/12/sunday-review/bigdatas-impact-in-the-world.html), Feb 2012.

[6] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, May 2011.

[7] Y. Noguchi. Following Digital Breadcrumbs to Big Data Gold. National Public Radio (http://www.npr.org/2011/11/29/142521910/the-digitalbreadcrumbs-that-lead-to-big-data), Nov 2011.

[8] Y. Noguchi. The Search for Analysts to Make Sense of Big Data. National Public Radio (http:// www.npr.org/ 2011/11/30/ 142893065

CLASSIFIER BASED INFORMATION MINING APPROACHES

Mr.G.Siva[1], Ms. G. Bhavani[2], Mr. G. Rajinikanth[3], Ms. Udaya Deepti P.[4]
Department of CSE, MRCE, JNT University, Hyderabad, *e-mail*[1]: *siva_561@gmail.com*, *e*-mail[2] : *bhavani.g@gmail.com*, *e-mail*[3]:*rajiniknath.g@gmail.com* & e-*mail*[4]:*deeptipolisetti@gmail.com.*

Abstract*— *Content mining is a procedure of separating the data from an unstructured content. This examination work manages a few classifiers including k-Nearest Neighbor (k-NN), Radial Basis Function (RBF), Multilayer Perception (MLP), and Support Vector Machine (SVM) which are utilized as prepared classifiers for performing order of information into pertinent and non-significant information. This study means to look at the productivity of the different existing grouping calculations with the proposed arrangement calculations on the premise of runtime, blunder rate and exactness.

Keywords: **k-NN, RBF, MLP, SVM**

I. INTRODUCTION

Information mining can diminish data over-burden and enhance basic leadership. This is accomplished by separating and refining valuable learning through a procedure of hunting down connections and examples from the broad information gathered by associations. The separated data is utilized to anticipate, order, display, and outline the information being mined. A content mining methodology will include order of content, content bunching, and extraction of ideas, granular scientific classifications creation, estimation examination, record outline and demonstrating. It includes a two phase handling of content. In the initial step a portrayal of archive and its substance is finished. This procedure is called arrangement prepare. In the second step called as arrangement, the record is isolated into expressive classifications and an entomb archive relationship is set up. Content mining has been helpful in numerous zones, i.e. security applications, programming applications, scholarly applications and so forth.

k-closest neighbor is a directed learning calculation where the aftereffect of new occasion question is characterized in light of dominant part of k-closest neighbor classification. The motivation behind this calculation is to characterize another question in view of traits and preparing tests.

A spiral capacity or an outspread premise work (RBF) is a class of capacity whose esteem reductions (or increments) with the separation from an essential issue. A RBF has a Gaussian shape, and a RBF system is regularly a Neural Network with three layers. The info layer is utilized to just information the information. The Gaussian enactment capacity is utilized at the shrouded layer, while a direct actuation capacity is utilized at the yield layer. The goal is to have the shrouded hubs figure out how to react just to a subset of the info, to be specific, that where the Gaussian capacity is entered. This is normally refined by means of administered learning.

The bolster vector machine (SVM) is a preparation calculation for taking in order and relapse rules from information. It can be connected for arrangement and relapse issues. It utilizes a non straight mapping to change the first preparing information into a higher measurement. Order calculations are progressively being utilized for critical thinking. The proficiency of calculations has been thought about on the premise of runtime, blunder rate, precision utilizing Weka machine learning device.

II. REVIEW OF LITERATURE

Numerous scientists have examined the procedure of consolidating the expectations of different classifiers to create a solitary classifiers (Breiman 1996c; Clemen, 1989; Perrone, 1993; Wolpert, 1992). The subsequent classifier (in the future alluded to as a troupe) is for the most part more exact than any of the individual classifiers making up the outfit. Both hypothetical (Hansen and Salamon, 1990; Krogh and Vedelsby, 1995) and experimental (Hashem, 1997; Opitz and Shavlik,1996a, 1996b) inquire about has exhibited that a decent outfit is one where the individual classifiers in the group are both exact and make their mistakes on various parts of the information space. Two mainstream strategies for making precise troupes are packing (Breiman, 1996c) and Boosting (Freund and Schapire, 1996; Schapire,1990). These strategies depend on "resampling" systems to get distinctive preparing sets for each of the classifiers. This work exhibits an extensive assessment of sacking on information mining issues utilizing four premise arrangement strategies: k-Nearest Neighbor (k-NN), Radial Basis Function (RBF), Multilayer Perceptron (MLP), and Support Vector Machine (SVM). Rachid Beghdad (2008) introduce a basic learn about the utilization of some neural systems (NNs) to identify and group interruptions. The point of research is to figure out which NN groups well the assaults and prompts to the higher location rate of every assault. This study concentrated on two order sorts of records: a solitary class (ordinary, or assault), and a multiclass, where the classification of assault is additionally recognized by the NN. Five distinct sorts of NNs were tried: multilayer perceptron (MLP), summed up bolster forward (GFF), spiral premise work (RBF), self-arranging highlight delineate), (and primary part examination (PCA) NN. In the single class case, the PCA NN plays out the higher recognition rate

III. DATABASE

Information gathering assumes a vital part in the information mining issues. In this paper, the dataset utilized for the second worldwide information disclosure and information mining devices rivalry, which was held in conjunction with KDD-98 the fourth universal meeting on learning revelation and information mining.

IV. PROPOSED PROCEDURES

The issue is to recognize purchasers utilizing information gathered from past battles, where the item is to be advanced is typically settled and the best figure is about who are probably going to purchase. Reaction demonstrating has turned into a key variable to direct promoting. By and large, there are two phases accordingly displaying. The main stage is to distinguish respondents from a client database while the second stage is to gauge buy measures of the respondents. (Dongil Kim, Hyoung-joo Lee, Sungzoon Cho, 2008) concentrated on the second stage where a relapse, not a characterization, issue is comprehended. As of late, a few non-direct models in light of machine adapting, for example, bolster vector machines (SVM) have been connected to reaction demonstrating.

Organizations worldwide are starting to understand that surviving a seriously focused and worldwide commercial center requires nearer associations with clients. Thusly, upgraded clients connections can help benefit three ways: 1) lessening costs by drawing in more reasonable clients; 2) creating benefits through cross-offering and up-offering exercises; and 3) amplifying benefits through client maintenance.

k-closest neighbor (Margaret H.Dunham, 2003) is a regulated learning calculation where the aftereffect of new occasion inquiry is ordered in light of larger part of k-closest neighbor class. The motivation behind this calculation is to characterize another protest in light of properties and preparing tests. The classifiers don't utilize any model to fit and just in view of memory. Given a question point,k number of articles (k=1) are discovered nearest to the inquiry point. The order is utilizing lion's share vote among the characterization of the k objects. Any ties can be broken aimlessly. k-Nearest neighbor calculation utilized neighborhood characterization as the expectation estimation of the new inquiry occurrence. The Euclidean separation between two focuses or tuples, say X1 = (x11, x12, … ,x1n) and X2 = (x21, x22, … ,x2n) is

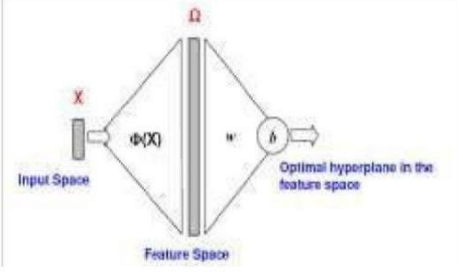
The easiest neural system is known as a perceptron. A perceptron is a solitary neuron with various data sources and one yield. The first perceptron proposed the utilization of a stage actuation work, yet it is more regular to see another sort of capacity, for example, a sigmoidal capacity. A basic perceptron can be utilized to characterize into two classes. Utilizing a unipolar actuation work, a yield of 1 would be utilized to order into one class, while a yield of 0 would be utilized to go in alternate class. Multilayer perceptrons with L layers of synaptic associations and L + 1 layers of neurons are considered. This is now and again called a L-layer organize, and some of the time a L + 1-layer arrange. A system with a solitary layer can inexact any capacity, if the concealed layer is sufficiently extensive. This has been demonstrated by various individuals, for the most part utilizing the Stone-Weierstrass hypothesis. In this way, multilayer perceptrons are representational capable.

We should outline the system as $x^{0}\ w1b1\ x^{1}\ w2b2\ \ldots.\ w^{LbL}\ x^{L}$, where $x^{l} \in R^{n}$ for all l 0....., L and W^{l} is a n xn network for all l 1,....., L. There are L+1 layers of neurons, and L layers of synaptic weights. It should change the weights W and predispositions b so that the genuine yield x^{L} turns out to be nearer to the fancied yield d.

The back proliferation calculation comprises of the accompanying strides.

1. Forward pass. The info vector x^{0} is changed into the yield vector x^{L}, by assessing the condition $X_i^{l}\ f(u_i^{l})\ f(i^{nl}\ W^{l}x\ b)^{l}$ for l= 1 to L
2. Mistake calculation. The distinction between the craved yield d and real yield xL is processed $i^{L}\ f(u_i^{L})(d_i x^{L})_i$
3. In reverse pass. The blunder motion at the yield units is Propagated in reverse through the whole system, by assessing $^{l1}\ I\ f(u_j^{l1})^{n1}\ i1 W_i^{l}\ ij^{l}$ from l = L to 1

SVM were initially proposed by Vapnik in the 1960s for grouping and have as of late turned into a range of extraordinary research inferable from advancements in the methods and hypothesis combined with expansions to relapse and thickness estimation. SVM convey the condition of workmanship execution in genuine applications, for example, content order, manually written character acknowledgment, picture grouping, money related estimating et cetera (Bao, 2003). The bolster vector machine (SVM) is a preparation calculation for taking in characterization and relapse rules from information. It is another machine-learning worldview that works by finding an ideal hyper plane as to take care of the learning issues.



$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{m} (x_{1i} - x_{2i})^2}$$

Figure 4.3: Support Vector Machine

The blunder rate is figured utilizing mean square mistake (MSE) assessed by relative cross approval is

$$MSE\ i1^{n(ai\ pi)2}/n$$

V. EXPERIMENTAL RESULTS

WEKA is an open source information mining programming that contains java executions of numerous well known machine learning-calculations including some prominent arrangement calculations. It has usage of different characterization calculations. The calculations require the information to be in particular organizations. The information incorporates 87 characteristics, for example,

State, postal district, age, date of birth, pay, sexual orientation, riches data and so forth.

Fig 5.1: Running Time in Existing and Proposed Classifiers

Fig 5.2: Error rate in Existing and Proposed Classifiers

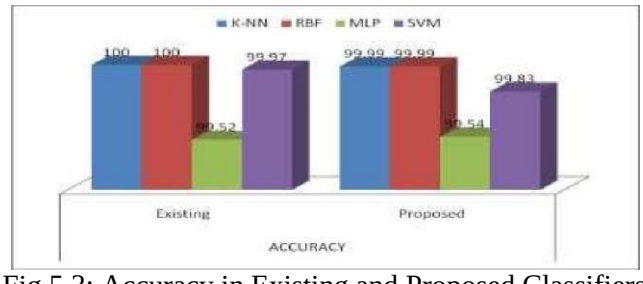


Fig 5.3: Accuracy in Existing and Proposed Classifiers

VI. CONCLUSION

The study has endeavored to build up another procedure called similar cross approval for information mining issues. The strategy assesses the mistake rate, precision and run time for base classifiers. This examination paper presents exhaustive exact assessment of four diverse methodologies to be specific k-Nearest Neighbor, outspread premise work, Multilayer perceptron, Support vector machine with direct advertising. Weka information mining programming is utilized to look at the different calculations and the outcomes have been accounted for.

REFERENCES

[1] Berk. R. A. (2004) "Data Mining within a Regression Framework", in Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, oded Maimon and Lior Rokach (eds.), Kluwer Academic Publishers.

[2] Fayyad, U., Piatetsy-Shapiro, G., Smyth, P., and Uthurusamy, R. (1996). From data mining to knowledge discovery. In Advances in Knowledge Discovery and Data Mining.

[3] Freund, Y., and Schapire, R. (1996).Experiments with a new boosting algorithm. In Proceedings of the Thirteenth International Conference on Machine Learning, 148-156, Bari, Italy.

[4] Friedman, J. H. (1997). On bias, variance,0/1 loss and the curse of dimensionality. Data Mining and Knowledge Discovery, 1:55–77.

[5] Govindarajan, RM.Chandrasekaran, (2009) "Performance optimization of data mining application using radial basis function classfier", International Scholarly and Scientific Research and Innovation, 3 (2),Pages 405 – 410

[6] Hansen, L., and Salamon, P. (1990). Neural Network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12: 993-1001.

[7] Ian H.Witten and Eibe Frank, (2005). "Data Mining-Practical Machine Learning Tools and Techniques", Elsevier, 177-178.

[8] Jiawei Han , Micheline Kamber, (2003) " Data Mining – Concepts and Techniques" Elsevier, pp. 359-366.

[9] Margaret H.Dunham, (2003), "Data Mining-Introductory and Advanced Topics", Pearson Education, pp. 90-113

[10] Oliver Buchtala, Manual klimek and Bernhard Sick, Member, IEEE, "Evolutionary Optimization of Radial Basis Function Classifier for Data Mining Applications", IEEE transactions on systems, man, and cybernetics—part B: cybernetics vol.35, No.5, pp. 928 – 947

[11] Rachid Beghdad. (2008) "C r i t i c a l study of neural networks in detecting intrusions", Computers & security, 27(5-6): 168-175.

[12] Schapire, R. (1990). The strength of weak learnability. Machine Learning, 5(2):197 – 227.

EFFECTIVE AND SECURE KAC SCHEME FOR DISTRIBUTED CLOUD STORAGE

[1]KVM Raghavendra, [2] SH Mehar Tabassum , [3] I Akhil Reddy, [4]P Pavani 1&4:Department of Computer Sceience & Engg,MRCE,JNT University, Hyderabad,India.
email:[1]Raghu_cse@gmail.com email: [2] Tabassum_cse@gmail.com email:
[3]Akhil_cse@gmail.com email :[4] pavani20891@gmail.com

***Abstract*—*Information sharing is a critical usefulness in cloud stor-age. In this past work, we demonstrate to safely, proficiently, and adaptable impart information to others in distributed storage. The current work shows the Key-Aggregate Cryptosystem (KAC) utilized for helpfully sent to others or be put away in a brilliant card with extremely constrained secure storage. An impediment of existing work is the predefined bound of the quantity of most extreme figure content classes and key is provoke to spillage. Our proposed work for the most part concentrates on over two issues. Our first work powerfully holds number of greatest figure content classes in distributed storage. If there should arise an occurrence of Stream figure the quantity of classes chose powerfully, in light of the fact that the figure content size is excessively bigger than piece figure. We propose an impeccable decentralized get to control conspire with total key encryption for information put away in cloud. This plan gives secure information stockpiling and recovery. Alongside the security the get to approach is additionally covered up for concealing the client's character. This plan is so effective since we utilize total encryption and string coordinating calculations in a solitary plan. The plan distinguishes any change made to the first document and if discovered clear the error's. The calculation utilized here are extremely basic so vast number of information can be put away in cloud with no issues. The security, confirmation, confidentiality are equivalent to the incorporated methodologies*.**

Keywords: Cloud Storage, Data Sharing, Asymmetric Encryption, String matching algorithms, Key- Aggregate Cryptonyms-tem

I. INTRODUCTION

Distributed storage is picking up fame as of late. In big business settings, we see the ascent sought after for information out sourcing, which helps with the vital administration of corporate information. It is likewise utilized as a center innovation behind numerous online administrations for individual applications. Presently a days,

it is anything but difficult to apply with the expectation of complimentary records for email, photograph collection, document sharing and additionally remote access, with capacity estimate more than 25 GB (or a couple of dollars for more than 1 TB). Together with the present remote innovation, clients can get to al-most the majority of their documents and messages by a cell phone in any side of the world. Considering information security, a traditional approach to guarantee it is to depend on the server to uphold the get to control after verification, which implies any startling benefit acceleration will uncover all information. In a common ten a cycloid processing environment, things turn out to be much more dreadful. Information from various customers can be facilitated on discrete virtual machines (VMs) yet live on a solitary physical machine.

Information in an objective VM could be stolen by instantiating an-other VM inhabitant with the objective one. As to capacity of documents, there are progressions of cryptographic plans which go similarly as permitting an outsider reviewer to check the accessibility of records for the information proprietor without spilling anything about the information, or without Compromising the information proprietor's secrecy. In like manner, cloud users presumably won't hold the solid conviction that the cloud server is benefiting work as far as secrecy. These clients are roused to encode their information with their own keys before transferring them to the server clouds can give a few sorts of administrations like applications (e.g., Google Apps, Microsoft on the web), foundations Security is required in light of the fact that information put away in mists is exceptionally touchy, for instance, therapeutic records and interpersonal organizations.

So encryption must be done in a flawless way. A few late encryption calculation bombs in seeking process. Be that as it may, the best encryption calculation which likewise improves hunt is total sort encryption [1].Thus this encryption strategy is utilized generally. Giving security just is extremely straightforward yet furnishing security with privacy[2] is especially troublesome. Keeping up the

security is particularly important on the grounds that it is simple for gatecrashers to get to the classified information. Since exceptionally secret information's are put away in cloud it is particularly expected to keep up the security and protection. Utilizing homomorphic encryption, the cloud gets figure content of the information and performs computations on the cipher ext and give back's the encoded esteem. Presently the client changes over the esteem, yet the cloud does not realize what information it has worked on. These are the regular issues in cloud. So this region must be concentrated.

Exchanges done in the cloud ought to likewise be noted periodically. The client ought to be confirmed and ought to give fitting authorization for them. Authorization criteria are painstakingly taken care of on the grounds that clients may change the information un-essentially. So this region ought to be focused excessively. Including this sort of highlight may consequently lessen the effectiveness of the calculation, so the calculation composed must be extremely proficient. It must consider all the extra components and the framework ought to be looked after in like manner. Consider the accompanying circumstance: An understudy from a school discovered a few acts of neglect done by a few representatives in school. At that point the understudy finds a way to inform the insights regarding the negligence done in the school.

Presently he will report the negligence done by the workers of the school to the college which controls the school. While reporting there are a few conditions to be checked genuinely. To begin with the understudy ought to demonstrate the personality be-cause the college ought to trust that the message originated from an approved individual. Second there ought not be any obstruction. Additionally if any change is accomplished for the first message then it ought to be discovered and the record is recovered. Subsequently in this paper the above issues are depicted and amended. A territory where get to control is generally being utilized is wellbeing care[14]. Mists are being utilized to store touchy data about patients to empower access to medicinal experts, healing facility staff, scientists, and arrangement producers. It is imperative to control the entrance of information so that lone approved clients can get to the information. Utilizing Aggregate key encryption [1], the records are encoded under some get to arrangement and put away in the cloud. Clients are given arrangements of keys. Only when the clients have coordinating arrangement of keys, would they be able to decode the data put away in the cloud. Get to control is likewise picking up significance in online long range interpersonal communication.

II. RELATED WORK

Attribute based encryption [7] [8] [12] [13] (ABE) was pro-posed by Sahai and Waters [26]. In ABE, a user has a set of attributes based on the user in addition to its unique ID. In Key-policy ABE or KP-ABE (Goyal et al.[27]), the sender has an access policy to encrypt data. A writer whose attributes and keys have been revoked cannot write back stale information. The receiver receives attributes and secret keys from the attribute authority and is able to decrypt information if it has matching attributes. In Cipher text-policy, CP-ABE ([28],[29]), the receiver has the access policy in the form of a tree, with attributes as leaves and monotonic access structure with AND, OR and other threshold gates.

All the approaches take a centralized approach and allow only one KDC, which is a single point of failure. Chase proposed a multi-authority ABE, in which there are several KDC authorities (coordinated by a trusted authority) which distribute attributes and secret keys to users. Multi-authority ABE protocol was studied in [7], [8], which required no trusted authority which requires every user to have attributes from at all the KDCs. Recently, Lewko and Waters [9] proposed a fully decentralized ABE where users could have zero or more attributes from each authority and did not require a trusted server.

In all these cases, decryption at user's end is computation intensive. So, this technique might be inefficient when users access using their mobile devices. However, as mentioned earlier in the previous section it is prone to replay attack. To reduce or block replay attack we use string matching algorithms [3][5] which is more efficient and perfect in security. It works more efficient than all other matching algorithms.

II.I EXISTING SYSTEM

Encryption keys additionally accompany two flavors—symmetric key or deviated (open) key. Utilizing symmetric encryption, when Alice needs the information to be begun from an outsider, she needs to give the encrypted her mystery key; clearly, this is not generally alluring. By differentiation, the encryption key and decoding key are diverse in public key encryption. The utilization of open key encryption gives more adaptability for our applications. For instance, in big business settings, each worker can transfer encoded information on the distributed storage server without the learning of the organization's lord mystery key. Presenting an exceptional kind of open key encryption which we call key-total cryptosystem (KAC). In KAC, clients encode a message under an open key, as well as under an identifier of figure content called class.

That implies the figure writings are further arranged into various classes. The key proprietor holds an ace mystery called ace mystery key, which can be utilized to concentrate mystery keys for various classes. All the more vitally, the removed key have can be a total key which is as smaller as a mystery key for a solitary class, however totals the force of numerous such keys, i.e., the decoding power for any subset of figure content classes. The sizes of figure content, open key, ace mystery key, and total key in KAC plans are all of steady size. General society sys-tem

parameter has measure straight in the quantity of figure content classes, yet just a little piece of it is required every time and it can be gotten on request from expansive (yet non confidential) cloud storage. Issues

•This work is the predefined bound of the number of maximum cipher text classes.

•When one carries the delegated keys around in a mobile device without using special trusted hardware, the key is prompt to leakage.

III. AUDIT SYSTEM ARCHITECTURE

The audit system architecture for outsourced data in clouds in which can work in an audit service outsourcing approach. In this architecture, we reflect on a data storage service containing four entities:

1) Data owner (DO): who has data files to be Stored in the cloud and relies on the cloud for data maintenance, can be an individual customer or an organization.
2) Cloud Storage Service Provider (CSP): who provides data storage service and has enough storage space to maintain client's data.
3) Third Party Auditor (TPA): a trusted person who manage or monitor outsourced data under request of the data owner.
4) Authorized Application (AA): who have the right to access and manipulate stored data.

The information which the information proprietor needs to store in cloud first achieves the approved application which will make advanced mark and sends the information to the distributed storage. On the off chance that the client needs to check information implies the confirmation demand ought to be send to outsider examiner (TPA), the TPA will recover the advanced mark from the database and will send the confirmation demand to the administration server. The administration server thus will produce the computerized signature for the information put away in the cloud and it will send just that advanced mark rather than the entire information to the TPA. The TPA will unscramble the computerized mark and thinks about the message process for confirming rightness of information.

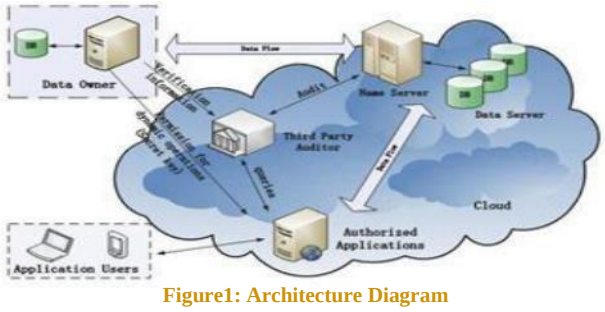


Figure1: Architecture Diagram

This building is known as the survey advantage outsourcing on account of data trustworthiness confirmation. Plan contains the data proprietor and permitted update their data for various application purposes. How-ever, we neither acknowledge that cloud advantage provider is trust to guarantee the security of set away data, or expect that the data proprietor has the ability to assemble the affirmations of cloud organization providers fault after slip-ups happen. Subsequently, pariah inspector, as a trust untouchable (TTP), is used to ensure the limit security of their outsourced data. We expect the outcast controller is tried and true and independent, and therefore has no support to unite with either the cloud advantage provider or the clients in the midst of the inspecting system:

- TPA must have the ability to make reliable be careful with the uprightness and availability of these named data at appropriate intervals;
- TPA must have the ability to take the affirmations for the level headed discussion about the anomaly of data to the extent genuine records for all data operations. To support insurance defending open looking at for cloud data stockpiling underneath the building, the tradition setup should accomplish following security and execution guarantees:

1) Audit-without-downloading: to allow TPA (or other clients with the help of TPA) to authenticate the correctness of cloud data on demand without recovering a copy of whole data or bring in additional on-line burden to the cloud users;
2) Verification-correctness: to make sure there exists no unethical CSP that can pass the audit from TPA without indeed storing users‟ data intact;
3) Privacy-preserving: to make sure that there exists no way for TPA to derive users‟ data from the in sequence collected during the auditing process;
4) High-performance: to allow TPA to perform auditing with minimum overheads in storage, communication and computation, and to maintain statistical audit sampling and optimized audit schedule with a long enough period of time.

IV. PROPOSED METHOD

A. Framework:

The premise or diagram of the key-total encryption conspires comprises of five polynomial-time calculations, which are clarified underneath: Setup guarantees that the proprietor of the information can build the general population framework structure or dad parameter. KeyGen, as the name proposes creates a bar lic/ace mystery (not to be mistaken for the assigned key clarified later) key combine. By utilizing this open and ace mystery key figure content class list he can change over plain content into figure content by means of utilization of Encrypt. Using Extract, the ace mystery can be used to generate a total decoding key for an arrangement of figure content classes. These produced keys can be securely transported to the representatives by utilization of secure components with appropriate efforts to establish safety clung to. In the event that and just if the figure content's class record is encased in the single key, then every client with a total key can decode the given figure content gave using Decrypt.

B. Algortihm:

1. Setup(Security level parameter, number of figure content classes): Setup guarantees that the proprietor of the information can construct general society framework stricture or parameter he make account on cloud. Subsequent to entering the info, the aggregate of cipher content classes n and a security level parameter 1, the general population framework parameter is given as yield, which as a rule skipped from the contribution of different calculations with the end goal of brevity.
2. KeyGen: it is for era of open or ace key mystery combine.
3. Encrypt(public key,index,message):run any individual who need to change over plaintext into figure content utilizing open and ace mystery key
4. Extract(master key, Set): Give contribution as ace mystery key and S records of various cipher text class it create yield total key. This is finished by executing extricate by the information proprietor himself. The yield is shown as the total key spoke to by Ks, when the info is entered in the frame the set S of records identifying with the different classes and master secret key msk.
5. Decrypt (Ks,S,i,C): When a nominee gets an aggregate key Ks as displayed by the past stride, it can execute Decrypt. The unscrambled unique message m is shown on entering Ks, S, i, and C, if and just in the event that I be-yearns to the set S.

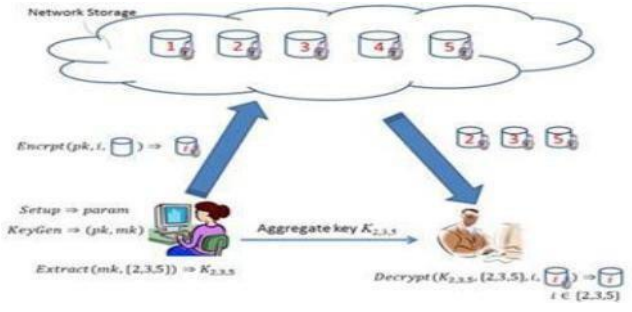


Fig2. Proposed KAC for data sharing in cloud storage system

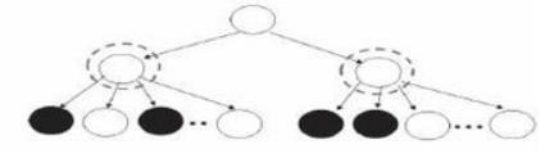


Fig.3.Key Assignment

V. CRYPTOSYSTEM ME6

Plaintext is discernable information (for instance, a spreadsheet document), and cipher text is the consequence of scrambling plaintext. A cryptosystem is an arrangement of methodology and traditions for covering up and uncovering data in a controlled way. A cryptosystem for the most part has two particular segments:

(a) the forms used to encipher and decode information and

(b) the set of keys used to impact the operation of these procedures so that the cipher text is subject to the key utilized for encryption. The security of a cryptosystem lies not in the mystery of the strategies used to encipher and translate the information yet rather in the trouble of decoding cipher text without learning of the key used to create it. Cryptosystem ME6 scrambles information in documents put away on plate. A record might be considered as an arrangement of no less than one byte and maybe a large number of bytes. ME6 peruses in plaintext from a record in obstructs whose size is between 6 KB and 10 KB (the correct size of every piece relies on upon the encryption key), encodes every square and composes the subsequent cipher text to circle. This is accomplished for each of the squares making up the document. Every square is initially compacted, if conceivable, before being encoded, so normally the cipher text pieces are littler than the plaintext obstructs, with the outcome that the record containing the scrambled information is generally littler than the information document.

VI. RESULT AND DISCUSSION

Our methodologies change the pressure issue (F =n in our plans) to be a tunable parameter, at the cost of O(n)-measured framework parameter. cryptography is drained constant time, while coding is drained O(|S|) bunch multiplications (or reason expansion on elliptic bends) with 2 matching operations, where S is that the arrangement of cipher text classes decrypt able by the allowed blend key and $|S| \leq n$. obviously, key extraction needs O(|S|) group multiplications moreover, that a substitution progress on the stratified key task (an old approach) that pre-serves ranges giving the sums of the key-holders have comparable edges is our approach of "compacting" secret enters openly key cryptosystems.

These open key cryptosystems produce figure writings of steady size ostensible sparing assignment of mystery composing rights for any arrangement of figure writings is conceivable. This not only upgrades client protection and privacy of information in distributed storage, however it'll this by supporting the distribution or designating of mystery keys fluctuated for diverse} figure content classes and creating keys by various derivation of figure content class properties of the data and its related keys. This totals up the extent of our paper. As there is a point of confinement assault determination the amount the

quantity} of figure content classes in advance and notwithstanding the exponential development inside the amount of figure messages in distributed storage, there is an interest for reservation of figure content classes for future utilize. With respect to potential modifications and upgrades to our present cause, in future, the parameter measure territory unit generally adjusted ostensible it's independent the very pinnacle of style of figure content classes. to boot, an exceptionally outlined cryptosystem, with the utilize of an exact security equation, as partner degree case, the Diffie-Hellman Key-Exchange methodology, which can then be imperviable, or at the premier confirmation against overflowing at the part of conservative key naming, will affirm that one can transport same keys on cell phones without dread of overflowing.

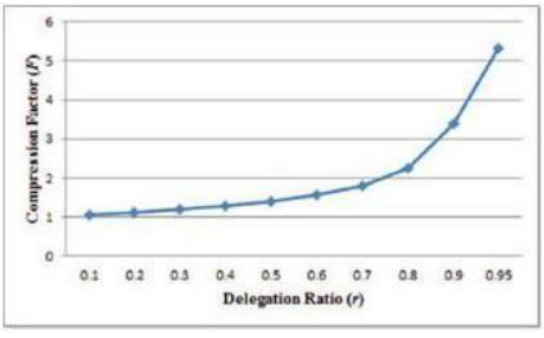


Fig 4. (A) Compression achieved by the tree-based ap-proach for delegating different ratio of the classe

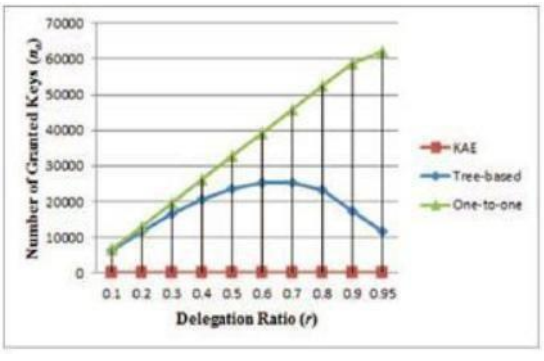


Fig 4. (B) Number of granted keys (na) required for different approaches in the case of 65536 classes of data.

VII. CONCLUSION

We consider how to ―compress‖ mystery enters openly key cryptosystems which bolster designation of mystery keys for various figure content classes in distributed storage. Regardless of which one among the power set of classes, the delegate can simply get a total key of steady size. Our approach is more adaptable than various leveled key dole out which can just spare spaces if every single key-holder share a comparable arrangement of benefits. The work is giving an effective security saving stockpiling contrasted with different works.

Despite the fact that there are many methodologies in the writing for alleviating the worries in protection, no approach is fully refined to give a security safeguarding capacity that beats the various security concerns. Along these lines to manage the worries of protection, we have to create privacy– saving system that defeats the stresses in privacy security and urge clients to receive distributed storage benefits all the more unhesitatingly. Our approach is more adaptable than various leveled key task which can just spare spaces if every single key-holder share a comparative arrangement of benefits. An impediment in our work is the predefined bound of the quantity of most extreme cipher text classes. In distributed storage, the quantity of cipher texts more often than not develops quickly. So we need to hold enough cipher text classes for the future augmentation.

REFERENCES

[1] Yan Zhua,b, Hongxin Huc, Gail-Joon Ahnc, Stephen S.Yauc. “Efficient audit service outsourcing for data in-tegrity in clouds”. In “The Journal of Systems and Soft-ware 85 (2012) 1083– 1095”.

[2] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A.Konwinski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica, and M.Zaharia, “A View of Cloud Computing,” Comm. ACM, vol. 53, no. 4, pp. 50-58, 2010.

[3] T. Velte, A. Velte, and R. Elsenpeter, Cloud Comput-ing: A Practical Approach, first ed., ch. 7. McGraw-Hill, 2010.

[4] A. Juels and B.S. Kaliski Jr., “PORs: Proofs of Re-trievability for Large Files,” Proc. 14th ACM Conf. Com-puter and Comm. Security (CCS ‟07), pp. 584-597, Oct. 2007.

[5] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z.Peterson, and D. Song, “Provable Data Posses-sion at Untrusted Stores,” Proc. 14th ACM Conf. Com-puter and Comm. Security (CCS‟07), pp. 598-609, Oct. 2007.

[6] M.A. Shah, M. Baker, J.C. Mogul, and R. Swamina-than, “Auditing to Keep Online Storage Services Honest,” Proc. 11th USENIX Workshop Hot Topics in Operating Systems (HotOS ‟07), pp. 1-6,2007.

[7] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z.Peterson, and D. Song, “Provable Data Posses-sion at Untrusted Stores,” Proc. 14th ACM Conf. Com-puter and Comm. Security (CCS‟07), pp. 598-609, 2007.

[8] M.A. Shah, R. Swaminathan, and M. Baker, “Priva-cy-Preserving Audit and Extraction of Digital Contents,” Cryptology ePrint Archive, Report 2008/186, 2008.

[9] A. Juels and J. Burton, S. Kaliski, “PORs: Proofs of Retrievability for Large Files,” Proc. ACM Conf. Com-puter and Comm. Security (CCS ‟07), pp. 584-597, Oct. 2007.

[10] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "En-abling Public Auditability and Data Dynamics for Storage Security in Cloud Computing," IEEE Trans. Parallel Dis-tributed Systems, vol. 22, no. 5,pp. 847-859, May 2011.

[11] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing," Proc. IEEE INFOCOM, pp. 525-533, 2010.

[12] C. Wang, K. Ren, W. Lou, and J. Li, "Toward Public-ly Auditable Secure Cloud Data Storage Services," IEEE Network, vol. 24, no. 4, pp. 19-24, July/Aug. 2010.

[13] Q. Wang et al., "Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Comput-ing," Proc. ESORICS „09, Sept. 2009, pp. 355–70.

[14] K. Yang and X. Jia, "Data Storage Auditing Service in Cloud Computing: Challenges, Methods and Oppor-tunities," World Wide Web, vol. 15, no. 4, pp. 409-428, 2012.

Particle Swarm Optimization Based K-mean Clustering - A Survey

B Mounika, Shiva Teja, Shubham Srivastav, Madhurima Rana
B.Tech Student, Assistant Professor
Malla Reddy College of Engineering
Hyderabad, India

***Abstract*— In Data Mining, Clustering is an important research topic and wide range of unsupervised classification application. Clustering is technique which divides a data into meaningful groups. K-mean is one of the popular clustering algorithms. K-mean clustering is widely used to minimize squared distance between features values of two points reside in the same cluster. Particle swarm optimization is an evolutionary computation technique which finds optimum solution in many applications. Using the PSO optimized clustering results in the components, in order to get a more precise clustering efficiency. In this paper, we present the comparison of K-mean clustering and the Particle swarm optimization.**

Keywords*— *Clustering, K-mean Clustering, Particle Swarm Optimization

I. INTRODUCTION

Clustering is a technique which divides data objects into groups based on the information found in data that describes the objects and relationships among them, their feature values which can be used in many applications, such as knowledge discovery, vector quantization, pattern recognition, data mining, data dredging and etc. [1] There are mainly two techniques for clustering: hierarchical clustering and partitioned clustering. Data are not partitioned into a particular cluster in a single step, but a series of partitions takes place in hierarchical clustering, which may run from a single cluster containing all objects to n clusters each containing a single object. And each cluster can have sub clusters, so it can be viewed as a tree, a node in the tree is a cluster, the root of the tree is the cluster containing all the objects, and each node, except the leaf nodes, is the union of its children. But in partitioned clustering, the algorithms typically determine all clusters at once, it divides the set of data objects into non-overlapping clusters, and each data object is in exactly one cluster. Particle swarm optimization (PSO) has gained much attention, and it has been applied in many fields [2]. PSO is a useful stochastic optimization algorithm based on population. The birds in a flock are represented as particles, and particles are considered as simple agents flying through a problem area. And in the multi-dimensional problem space, the particle's location can represent the solution for the problem. But the PSO may lack global search ability at the end of a run due to the utilization of a linearly decreasing inertia weight and PSO may fail to find the required optima when the problem to be solved is too complicated and complex. K-means is the most widely used and studied clustering algorithm. Given a set of n data points in real d-dimensional space (Rd), and an integer k, the clustering problem is to determine a set of k points in Rd, the set of points is called cluster centres, the set of n data points are divided into k groups based on the distance between them and cluster centres. K means algorithm is flexible and simple. But it has some limitation, the cluster result mainly depends on the selection of initial cluster centroids and it may converge to the local optima [3]. However, the same initial cluster centre in a data space can always generate the same cluster results, if a good cluster centre can always be obtained, the K-means will work well.

II. K-MEAN CLUSTERING

James MacQueen, the one who proposed the term "k-means"[4] in 1967. But the standard algorithm was firstly introduced by Stuart Lloyd in 1957 as a technique pulse-code modulation. The K-Means clustering algorithm is a partition-based cluster analysis method [5]. According to the algorithm we firstly select k objects as initial cluster centres, then calculate the distance between each cluster centre and each object and assign it to the nearest cluster, update the averages of all clusters, repeat this process until the criterion function converged. Square error criterion for clustering.

$$E = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \|x_{ij} - m_i\|$$

x_{ij} is the sample j of i-class, m_i is the center of i-class, n_i is the number of samples i-class, Algorithm step are shown in the fig(1).

K- means clustering algorithm is simply described as follows:

Input: N objects to be cluster {x1, x2.....xn}, the number of clusters k;

Output: k clusters and the sum of dissimilarity between each object and its nearest cluster center is the smallest;

• Arbitrarily select k objects as initial cluster centers (m1, m2... mk);

• Calculate the distance between each object xi and each cluster center, then assign each object to the nearest cluster, formula for calculating distance as:

$$d(x_i, m_i) = \sqrt{\sum_{j-1}^{d} (x_{i1} - m_{f1})}$$

i = 1,2.....N

j = 1,2.....k d (xi, mj) is the distance between data i and cluster j;

• Calculate the mean of objects in each cluster as the new cluster centers,

$$m_i = \frac{1}{N_i}\sum_{j=1}^{N_i} x_{ij}$$

=1 i=l, 2 . . . k; Ni is the number of samples of current cluster i;

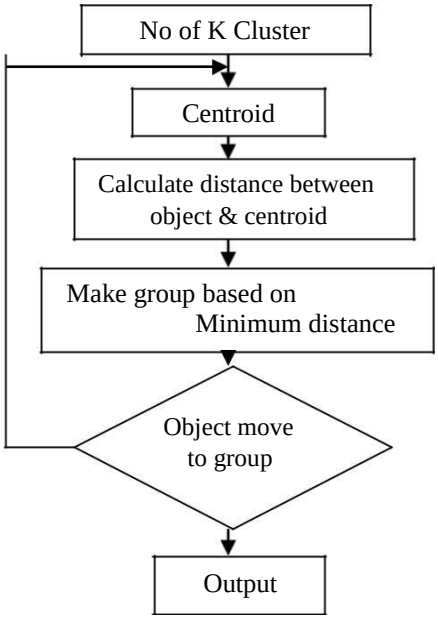


Fig-1 K-means algorithm.

III. PARTICLE SWARM OPTIMIZATION

PSO was introduced by Kennedy and Eberhart[6], it was based on the swarming behaviour of animals and human social behaviour. A particle swarm is a population of particles, in which each particle is a moving object which can move through the search space and can be attracted to the better positions. PSO must have a fitness evaluation function to decide the better and best positions, the function can take the particle's position and assigns it a fitness value. Then the objective is to optimize the fitness function. In general, the fitness function is pre-defined and is depend on the problem. Each particle has own coordinate and velocity to change the flying direction in the search space. And all particles move through the search space by following the current optimum particles.

Each particle consists of a position vector z, which can represent the candidate solution to the problem, a velocity vector v, and a memory vector pid, which is the better candidate solution encountered by a particle. Suppose the search space is n-dimensional, then the ith individual can be represented as:

Zi = {Zi1, Zi2,…….Zin}

Vi = {Vi1, Vi2,…….Vin}

i=1,2,3……..n.

Where n is the size of swarm. The best previous experience of the ith particle is represented as:

pid_i = {pid_i1, pid_i2……..pid_in}

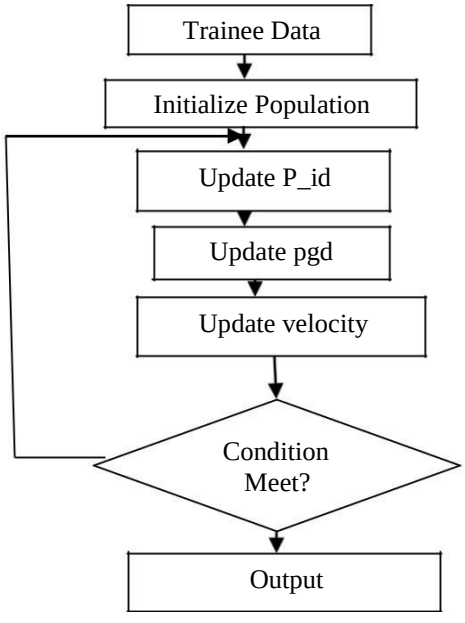


Fig.2 flowchart of PSO

Another memory vector pgd is used, which is the best candidate solutions encountered by all particles. The particles are then manipulated according to the following equations:

V_id(t+1)=wv_id(t)+η1rand(pid_i_Z_id(t))+η2rand(pgdZ_id(t)),

Z_id (t+1)=Z_id (t)+V_id (t+1),

d= 1,2….n

Where w is an inertia weight, which is used for controlling the effect of previous history of velocities on current velocity, and controls the tradeoff between the local and global exploration abilities of the swarm. A small inertia weight facilitates local exploration, while a big one tends to facilitate global exploration. In order to get a better global exploration, w can be gradually decreased to get a better solution.η1 and η2 are two positive constants, rand is a uniformly generated random number. The equation shows that in calculating the next velocity for a particle, the previous velocity of the particle, the best location in the neighborhood about the particle, the global best location all contribute some influence to the next velocity. Particle's velocities in each dimension can arrive to a maximum velocity vmax, which is defined to the range of the search space in each dimension. [3]

The process of the PSO can be described as follows: First, It will initialize a population of particles with velocities and random positions in search space. Secondly, for each particle i, update the position and velocity according to ,compute the fitness value according to the fitness function, update pid_i and pgd if necessary, repeat this process until termination conditions are met. Flowchart shown in Fig-2.

IV. ADVANTAGE AND DISADVANTAGE OF K-MEAN AND PSO

Advantages of K-mean clustering

• K-mean clustering is simple and flexible.

• K-mean clustering algorithm is easy to understand and implements.

Disadvantages of K-mean clustering

• In K-mean clustering user need to specify the number of cluster in advanced [7].

• K-mean clustering algorithm performance depends on a initial centroids that why the algorithm doesn't have guarantee for optimal solution [7].

Advantages of PSO

• PSO based on the intelligence and it is applied on both scientific research and engineering.

• PSO have no mutation and overlapping calculation. The search can be take place by the speed of the particle. Most optimist particle can able to transmit the information onto the other particles during the development of several generations, and the speed of researching is faster.[8]

• PSO accepts the real number code, and that is decided directly by the solution. Calculation in PSO is simpler and efficient in global search [8]

Disadvantages of PSO

• It is slow convergence in refined search stage and weak local search ability.

• The method cannot work on the problems of non-coordinate systems like the solution of energy field and the moving rules for the particles in the energy field.

V. CONCLUSION

Study of the k-mean clustering and Particle swam optimization we say that the k-mean which is depend on initial condition, which cause the algorithm may converge to suboptimal solution. On the other side Particle swarm optimization is less sensitive for initial condition due to its population based nature. So Particle swarm optimization is more likely to find near optimal solution.

REFERENCES

[1] A. Jain, M. Murty and P. Flynn, "Data Clustering: A Review", ACM Computing Surveys, Vol.31, No. 3, Sep 1999, pp. 264– 323.

[2] H. M. Feng, C.Y. chen and F. Ye, "Evolutionay fuzzy particle swarm optimization vector quantization learning scheme in image compression", Expert Systems with Applications. Vol. 32, No. 1, 2007, pp. 213-222.

[3] Jinxin D. And Minyong Q., "A new Algorithm for clustering based on particle swarm optimization and k-Means", International Conference Intelligence,2009,pp 264-268.

[4] Shalove Agarwal, Shashank Yadav and Kanchan Singh, "K-mean versus k-mean++ clustering Techniques", in IEEE 2012.

[5] Juntao Wang and Xiaolong Su, "An improved k-mean clustering algorithm", in IEEE, 2011, pp 44-46.

[6] R. Eberhart and J. Kennedy, " Particle swarm optimization ", Proc. of the IEEE Int. Conf. on Neurad l Networks, Piscataway, NJ., 1995, pp. 1942–1948.

[7] Garbriela derban and Grigoreta sofia moldovan, "A comparison of clustering techniques in aspect mining", Studia University, Vol LI, Number1, 2006, pp 69-78.

[8] Qinghai B., "The Analysis of Particle Swarm Optimization Algorithm", in CCSE, February 2010, vol.3.

SECURE BYOD ENVIRONMENTS ON REMOTE MOBILE SCREEN (RMS)

Bhandaram Manogna[1], Benchi Raja Reddy[2], Deekonda Sai Priya[3], CH. Mahender Reddy[4]
Department of CSE, Malla Reddy College of Engineering, JNTU Hyderabad, Telangana, INDIA
E-mail1:manogna_cse@gmail.com, E-mail[2]:rajaredy_cse@gmail.com
E-mail[3]:saipriya_cse@gmail.com, E-mail[4]:mahender.chilukala@gmail.com,

Abstract -- ***The presentation of bring your own particular gadget (BYOD) technique in the common world makes benefits for organizations and also work fulfillment for the representative. In any case, it additionally delivers tests as far as security as new liabilities emerge. Specifically, these difficulties incorporate space detachment, information security, and strategy consistence and also taking care of the asset requirements of cell phones and the lack of care made by introduced applications trying to perform BYOD capacities. We show Remote Mobile Screen (RMS), an approach for secure BYOD situations that reports every one of these analyses. So as to accomplish this, the endeavor furnishes the representative with a trusted virtual machine running a versatile working framework, which is situated in the undertaking system and to which the worker interfaces utilizing the portable BYOD gadget.***

Key words -- Privacy, Remote Mobile System, Remote Mobile Screen

1. INTRODUCTION

Cell phones have gotten to be vital components in our everyday life, and they have ended up pervasive. For instance, in 2013, the selection of cell phones and associations developed to 7 billion units, as indicated by a report from Cisco [1]. To put this consider along with point of view, as per the United Nations there are 7.2 billion occupants on the planet [2].

Cell phones have tremendously affected organizations, since they increment the profitability of the representatives, and in addition give adaptability as far as time and space. Thusly, organizations have been furnishing their workers with cell phones to empower them to play out their occupation related assignments. Be that as it may, the broad utilization of these gadgets has made burdens for undertakings since they should handle the expenses connected with acquiring and keeping up such gadgets. Furthermore, the extraordinary speeds in which new advances are presented make the present models of these gadgets less engaging the representatives after a brief timeframe.

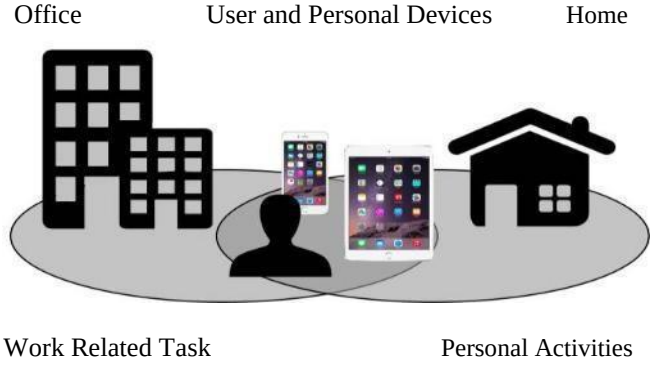


Figure 1.1: Representation of a BYOD environment

These outcomes in circumstances where representatives need to change their gadgets quicker than the undertakings can furnish them with new ones. As an aftereffect of this decision and customization in cell phones, representatives frequently ask for their organizations to permit them to utilize their own cell phones for business related assignments while likewise holding them for individual utilize [3]. Due to this converging of utilization, these gadgets are known as double utilize gadgets [4].

Description of BYOD

In these situations, organizations have embraced arrangements as new strategies. This arrangement of strategies is known as Bring Your Own Device (BYOD), which permits a worker to utilize the cell phones they want to perform business related undertakings. In a late study did by Cisco, it was found that 89% of IT offices empower BYOD in some frame [5]. A common BYOD environment is portrayed in Figure 1.1. Where a representative uses an individual cell phone and an individual tablet for individual exercises as well as for business related errands.

BYOD gives a progression of favorable circumstances to both representatives and the undertaking, which are depicted underneath:

Job Satisfaction

The utilization of BYOD arrangements delivers an expansion in occupation fulfillment in the workers. As specified some time recently, representatives can choose the gadget they feel good with and supplant it at the season of their picking. They likewise abstain from conveying extra gadgets by utilizing a solitary gadget for both individual and

work employments. A Cisco report [5] notices that the principle purposes behind representatives to utilize individual gadgets in a BYOD situation are the "any gadget work style", the likelihood of consolidating work and individual exercises, the evasion of utilization confinements, the "shopper encounter" at work, and the entrance to individual portable applications.

Productivity

In the wake of applying BYOD strategies, organizations have seen an expansion in profitability regarding yield and an augmentation in cooperation among workers. As indicated by Cisco [5], this is the most essential finding in their examination, since it demonstrates that the utilization of individual gadgets in the BYOD environment does not bring about diversions, but rather it has the inverse impact. Promote, Assign et al [6]. said that staff profitability is expanded by the way that versatility permits the workers to be gainful from anyplace.

Recruitment

Free et al [7]. Demonstrated that there is an exceedingly huge relationship between's the means by which appealing an undertaking is to a future representative and the selection of BYOD by that venture. Hayes [8] bolsters this idea by demonstrating that an organization that applies BYOD strategies is all the more speaking to potential representatives, as the venture is viewed as an adaptable workplace.

Cost Saving

Allocate et al [8]. Express that BYOD has its inceptions in the way that organizations needed to decrease expenses of gear accessible to representatives by giving them a chance to buy the cell phones they want. Furthermore, undertakings don't need to bring about in costs identified with giving specialized support to such gadgets. In addition, specialized bolster assets can be centered around a superior administration, distributed to other range of the venture, or essentially diminished.

2. SECURITY CHALLENGES RELATED TO BYOD ENVIRONMENT

In view of the danger show displayed in the past area, we can present an arrangement of difficulties that a protected BYOD environment introduce. While the initial three difficulties were distinguished by Wang et al., the last one was examined by Miller et al. We display these four difficulties as takes after.

Data spillage

We sort the corporate information as open or private. In the main classification, the data is openly appropriated by the venture. Notwithstanding, for the second classification,

The endeavor spends assets to counteract information spillage.

In a BYOD domain, representatives have entry to the secret data through their cell phones. Considering every one of the dangers that influence cell phones, there are difficulties identified with how to secure the corporate data once it achieved such gadgets.

Unauthorized sharing of spaces

In BYOD situations we can characterize two spaces: an individual space and a corporate space. From one perspective, the individual space incorporates all applications and archives possessed by the worker, for example, family photographs, individual contacts, or relaxation applications like amusements. Then again, the corporate space incorporates every one of the applications and records identified with the undertaking, for example, corporate messages and contact records and in addition profitability applications gave by the venture, similar to a spreadsheet supervisor. The test in securing BYOD situations is the manner by which to keep these two spaces detached from each other. At the end of the day, to give systems to keep the entrance of individual information from big business related undertakings, and the other way around.

Lack of security consistence

In numerous BYOD situations the endeavor thinks that its hard to implement its security arrangements because of the way that the workers are the proprietors of the cell phones. For instance, if the approaches express those cell phones must run antiviruses to forestall malware, the undertaking must watch that all gadgets consent to this mandate. Encourage, testing gadget by-gadget is impossible since it is tedious and does not scale well.

Employee protection

There is concern identified with worker's protection, as organizations may screen the representative's close to home exercises and also break down his or her own data. This is exceptionally a worry when he or she is associated with the corporate system, as the later could conceivably track every one of the information in the system. Therefore, the representative won't not feel great utilizing his or her cell phone, which contrarily influences efficiency and occupation fulfillment, and annihilations the motivation behind BYOD approaches.

Goals for a Secure BYOD Environment

Considering the difficulties portrayed in area 2.2, an arrangement of objectives can be characterized for secure BYOD situations. Wang et al., have portrayed the initial three objectives in the accompanying rundown, while the staying three have been distinguished by Gimenez Ocano et al [9]. The last creators expressed that the primary arrangement of objectives are fundamental however not adequate to accomplish the objective of a protected BYOD environment, since they don't consider the asset limitations of the cell phones, the security intrusion that a worker may involvement, and different circumstances that the main arrangement of objectives does not address.

Space Isolation

This objective addresses the test of unapproved sharing of spaces. This is accomplished by disconnecting the individual space from the corporate space in a manner that no information can be sent starting with one space then onto the next. Be that as it may, the usage of space disconnection does not anticipate circumstances where the venture plays out reclamation to manufacturing plant default design, with the reason for erasing all classified data situated in the gadget. This is not attractive in light of the fact that, under these conditions, the representative would lose every one of the information spared in the individual space.

Corporate Data Protection

The private data of the undertaking must stay mystery even after a cell phone is sold, lost or stolen. Along these lines, just approved workers can get to the data. Cryptographic calculations can be utilized to figure the corporate information with the end goal that lone the representative that has the key can get to such information.

Security Policy Enforcement

Since strategy requirement is difficult to accomplish for cell phones, one of the objectives is to make this implementation programmed using programming. Also, in light of the fact that it comes about unfeasible to check every cell phone at the time, arrangement authorization must be performed through programmed checkups in view of programming. Furthermore, an answer ought to incorporate instruments to make an interpretation of approaches into programming setup that agree to them.

3. REMOTE MOBILE SCREEN (RMS): DESCRIPTION AND EXPERIMENTS

Before we display our answer, Remote Mobile Screen (RMS), we begin by depicting BSF a structure arrangement, which is identified with our work. At that point, we give a dialog on how RMS accomplishes every one of the objectives for a safe BYOD environment. We give the means expected to a session start and end. We examine the components and difficulties that RMS presents. At that point, we depict an execution of our system that utilizations normally accessible programming. We play out a security investigation and recognize security dangers identified with our answer. Encourage, we give a security examination of the design. At long last, we give trial brings about request to address an arrangement of these difficulties.

BYOD Security Framework (BSF)

BSF is a structure exhibited by Wang et al. This system has been intended to accomplish three objectives. To begin with, space disengagement is required so that the individual space and the corporate space get to be isolated from each other, and permit strategies to be executed for each of them separately. Second, corporate information insurance is required so that unapproved access to this information gets to be unfeasible, which is accomplished by encoding all the corporate information put away on the BYOD gadget. In conclusion, security approach authorization must be actualized so that the gadgets conform to the venture's prerequisites.

With a specific end goal to meet these three prerequisites, BSF characterizes two elements: the endeavor side and the gadget side. The previous is created by all the corporate assets, for example, venture's servers, portals to the Internet and the corporate information. In this side a Network Access Control (NAC) component is responsible for giving access control when the BYOD gadgets attempt to get to these assets. This get to is either approved or dismisses in view of the corporate strategies. Furthermore, the NAC needs to separate between the solicitations from the individual space and the solicitations from the corporate space, which is accomplished by actualizing authentications for each of them. So as to deal with the corporate arrangements a security approach database is sent. These arrangements incorporate data on the most proficient method to handle the get to demand when it originates from a client space on a BYOD gadget, which gadgets are permitted to get to the system, and the parameters of the association. At last, cell phones are overseen by coordinating a MDM arrangement, which depends on the strategy database and authorizes these approaches on the BYOD gadgets. Figure 3.1 demonstrates a representation of the considerable number of segments found in BSF.

At the gadget side, we can discover space segregation between the individual space and the corporate space. Therefore, the individual space contains all the versatile

applications and information claimed by the representative, while the corporate space has the portable applications and data required by the endeavor. Since the corporate space must consent to the security arrangements of the endeavor, a MDM operator is introduced in this space, which gives the heads administration capacities on the cell phone. Encourage, a security strategy authorization element is likewise some portion of this space.

These approaches are put away in this space through the execution of a security strategy database. At last, corporate information security is accomplished by actualizing cryptographic calculations and additionally gets to instruments so as to keep the information to be replicated without appropriate approval.

Architecture introduced by RMS

RMS alters the BSF's design by moving the corporate space situated in the cell phone to the venture organize. Also, RMS includes another part that we signified Corporate Space Manager, which is utilized to deal with the entrance to portable virtual machines situated in the endeavor arrange. At last, RMS utilizes the Virtual Network Computing (VNC) convention (which is thus in light of the Remote Frame cushion (RFB) convention) to permit the client to get to his or her legitimate corporate space. Similarly as in BSF, RMS presents a BYOD side and a venture side, which are depicted as takes after.

BYOD side

Contrasted with the BSF, this part of the engineering is less complex. The cell phone just contains the representative's close to home space. This implies the gadget ca exclude any segment except for the individual information and utilizations of the representative. Thus, there is no compelling reason to introduce either a MVM or a MDM operator on the cell phone. The main prerequisite of our system is that VNC customer application must be introduced in the BYOD side. This customer is utilized by the representative to get to the endeavor space situated at the corporate system. Section (An) of Figure 3.1 demonstrates the BYOD agree with all the specified segments.

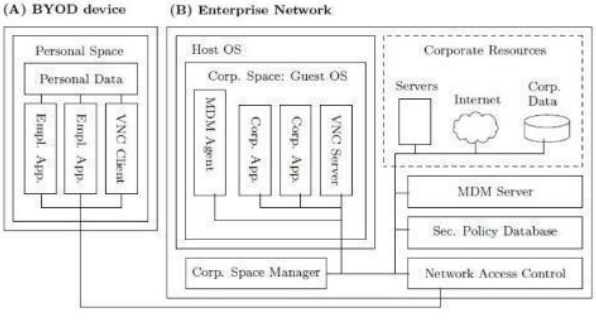


Figure 3.1: RMS architecture.

The oddity of our engineering depends in transit the workers get to the corporate assets. So as to get to these assets, the representative must introduce and utilize a VNC customer from an application store, for example, Apple's App Store, Google's Play Store, or an application store gave by the venture. At that point, the client does not get to a desktop OS (i.e. Windows, Linux, Mac OS X), however he or she is given a versatile OS (i.e. Android, iOS). This necessity addresses the poor level of ease of use that desktop OSes have when they are gotten to from a cell phone.

Desktop OSes are not intended to be scaled to the little screens that cell phones introduce, since the graphical interface and desktop applications are produced for greater screens. Advance, desktop OSes don't actualize signals (e.g. squeeze to zoom, swipe, and so forth.) nor the kind of information (i.e. finger or stylus) standard to a portable OSes. Thusly, when the worker gets to the undertaking side, he or she is given an interface intended for a cell phone. To our comprehension, the idea of a cell phone that gets to a versatile OS over a system has not been utilized as a part of BYOD situations, or for some other sort of reason.

Enterprise Side

This side is made out of the majority of the components required in the RMS engineering. This is on account of the undertaking side does not experience the ill effects of the constraints regarding assets that a cell phone has. As a result, in the venture side we can locate the Corporate Resources, a Network Access Control, a Security Policy Database, a MDM server, Corporate Spaces and the Corporate Space Manager. The undertaking side, and in addition its segments, is portrayed in Part (B) of Figure 3.1.

The corporate assets are made by gadgets and administrations, for example, email servers, web servers, doors to the Internet, or any restrictive application that the venture has. The Network Access Control is accountable for verifying substantial workers and approving them to get to the undertaking assets. The Network Access Control not just investigates demands from outside of the undertaking system, additionally assesses the solicitations that originate from within the venture arrange. The Security Policy Database stores all the strategy definitions that the venture has. With a specific end goal to give or reject get to, the Network Access Control depends on the security arrangements that the Security Policy Database contains. The MDM server is accountable for implementing all the security arrangements on the Corporate Spaces.

At the venture side we locate the corporate space, which contains all the corporate information and applications that the representatives requirements for working. We can characterize this space as a VM gave a versatile OS. Thusly, the venture side contains a server running VM programming

that conveys a few corporate spaces as visitor OSes. Each of these corporate spaces is appointed to one representative as it were. Further, each of the versatile OSes has introduced a VNC server, which is designed in a manner that the representative can get to his or her corporate space utilizing the VNC customer found as a part of his or her BYOD gadget. What's more, each corporate space is given a MDM specialist, which is accountable for actualizing the security strategies authorized by the MDM server.

4. CONCLUSION

The presentation of BYOD arrangements are valuable for both the undertaking and its workers, as it builds work fulfillment, the representatives turn out to be more gainful, the endeavor can utilize it to select potential workers, while it diminishes costs identified with resources and operation.

In any case, BYOD strategies and the versatility way of BYOD gadgets posture security dangers to the endeavor data, and also the representative protection. This makes the difficulties of information spillage, unapproved sharing of spaces, absence of security consistence, and worker protection.

With a specific end goal to address these difficulties, a safe BYOD environment must meet the objectives of space detachment, corporate information assurance, and security approach implementation, genuine space disconnection, non-meddling, and low asset utilization.

An order of the present answers for BYOD has been exhibited. We can arrange the arrangements into Mobile Virtual Machine, Agent-based, Cloud-based, Virtual Private Network, Trusted Environments, and Framework. We abridge which objectives they meet, and we demonstrate that right now there is no arrangement that accomplishes these objectives.

In this theory we proposed another structure for a protected BYOD environment, Remote Mobile Screen (RMS), which meets all the fundamental objectives for a safe BYOD environment. Our system predominantly comprises on sending an individual space on the cell phone, and conveying a corporate space at the corporate system. At that point, the representative gets to his or her corporate space using a VNC customer.

At last, in our inactivity explore we demonstrated that the application dormancy experienced in a specific customer increments as an element of the quantity of simultaneous customers getting to a server, and the ping delay from the customer to the server. Facilitate we demonstrated how the application deferral can be diminished beneath a satisfactory estimation of 150 milliseconds by sending numerous servers near the customers and by utilizing high-pressure encoding designs as a part of the VNC convention.

5. BIBLIOGRAPHY

1. "Cisco visual networking index: Global mobile data traffic forecast update, 2013-2018," White Paper, Cisco, Feb. 2014. [Online].
   a. Available: http://www.cisco.com
2. "The world population situation in 2014," Department of Economic and Social Affairs Population Division, 2014. [Online].
   a. Available: http://www.un.org
3. "BYOD: From company-issued to employee-owned devices," McKinsey & Company, June 2012. [Online]. Available: http://www.mckinsey.com/
4. M. Silic and A. Back, "Factors impacting information governance in the mobile device dual-use context," Records Management Journal, vol. 23, no. 2, pp. 73– 89, 2013.
5. J. Bradley, J. Loucks, J. Macaulay, R. Medcalf, and L. Buckalew, "BYOD: A global perspective. Harnessing employee-led innovation," 2012. [Online].Available: http://www.cisco.com/web/about/ac79/docs/re/BYOD_Horizons- Global.pdf
6. D. Assing and S. Cal'e, Mobile access safety: Beyond BYOD. Hoboken,NJ: John Wiley & Sons, 2013.
7. M. Loose, A. Weeger, and H. Gewald, "BYOD - The next big thing in recruiting? Examining the determinants of BYOD service adoption behavior from the perspective of future employees," in Proc. 19th Americas Conf. Information Systems, Chicago, IL, Aug. 2013.
8. J. Hayes, "The device divide," Engineering & Technology, vol. 7, no. 9, pp. 76–78, Oct. 2012. 130
9. S. Gimenez Ocano, B. Ramamurthy, and Y. Wang, "Remote Mobile Screen (RMS): an approach for secure BYOD environments," in Computing, Networking and Communications (ICNC), Int. Conf., Anaheim, CA, Feb. 2015.
10. ——, "Security challenges in and solutions for the Bring Your Own Device (BYOD) environment: a survey," 2015, under review.
11. ——, "Implementation challenges of Remote Mobile Screen (RMS) for secure BYOD environments," 2015, under review.

12. Y.Wang, K. Streff, and S. Raman, "Smartphone security challenges," Computer, vol. 45, no. 12, pp. 52–58, Dec. 2012.
13. "Mobile threat report, July - September 2013," F-Secure Corporation, 2013. [Online]. Available: http://www.f-secure.com/documents/996508/1030743/Mobile_Threat_Report_Q3_2013.pdf
14. "Mobile threat report, Q 2014," F-Secure Corporation, 2014. [Online]. Available: http://www.f-secure.com/documents/996508/1030743/Mobile_Threat_Report_Q1_2014.pdf
15. Y. Zhou, Z. Wang, W. Zhou, and X. Jiang, "Hey, you, get off of my market: Detecting malicious apps in official and alternative Android markets," in Proc. 19th Annual Network & Distributed System Security Symp., San Diego, CA, Feb. 2012.
New York State Attorney General, June 2014. [Online]. Available: http://www.ag.ny.gov/pdfs/SOS%201%20YEAR%20REPORT.pdf

AUTHORSHIP IDENTIFICATION BASED ON STYLOMETRY FEATURES

D.Mounica[1], B.Aravind Reddy[2], B.Panhindra Reddy[3], Mekala sreenivas[4] Department of CSE, Malla Reddy College of Engineering, JNTU Hyderabad, Telangana, India

email1[1]: mounica_monu111@gmail.com, email[2]:aravind_cse@gmail.com,

email [3]: phanindra_cse@gmail.com, email [4] :mekala_sreenivase@mrce.in

Abstract-- ***Electronic communication is one of the popular ways of communication in this era. E-mail communication is the most popular way of electronic communication. Internet works as the backbone for these communications. In digital forensics, questions is arises that the authors of documents and the author identity, demographic background is linked to other documents or not. So identification of the author(s) of the message(s) and non repudiation are some of the major challenges. Author identification is a critical point to be ensured, because many people are used to copy the content of others. Stylometry can be used for the author identification for text documents. As the non–repudiation and integrity of the message are the major concerns, Stylometry is not only identifying a writing pattern but we can also identify the gender of the human. So this document discussed about identification of author, authentication through stylometry technique. In this paper different stylometric techniques are discussed.***

Key words-- **Stylometry, author identification, email, gender.**

I. INTRODUCTION

In 1851 the utilization of tools i.e. statistical tools to test inquiries of origin was done when mathematician Augustus de Morgan proposed utilizing normal word length to numerically describe initiation style [1]. After that Thomas Mendenhall who was a physicist and recommended that a author has a "trademark bend of arrangement" controlled by how a author utilizes expressions of various lengths every now and again, in year 1887 [2]. In 1888 a mathematician (William Benjamin Smith) distributed two papers depicting a "bend of style" to recognize authorial styles in view of normal sentence lengths, this strategy was connected to the Pauline Epistles [3]. A book "Principes de stylometrie" 1890 was given by the Polish logician Wincenty Lutosławski to depict the nuts and bolts of stylometry. Order of Plato's Dialogs was given by Lutosławski by utilizing this strategy. At that point Lucius Sherman, an educator of English in 1893, found that composition style after some time changes with normal sentence length [4]. Because of the expanding figuring power, accessibility of the Internet, development of ultrahigh dimensional factual devices the stylometric methods are developing quickly step by step. In this paper, fundamentally we concentrated on the different sorts of stylometry procedures. This paper is composed as takes after: in area 2; we have the portrayal of stylometry and in segment 3. This study talked about the writing Review. In the segment 4, is giving the near investigation of research in light of logical articles, email writer recognizable proof utilizing stylometry and as a part of the end in area 4, is giving the conclusion over the examination given in the paper.

II. RELATED WORK

The fields of stylistics, computational linguistics, and non-conventional origin attribution to build up a conceivable structure for the ID of email content initiation. The fields like text classification, machine learning, software forensics, and forensic linguistics also affect on the present study. Written falsification discovery [8] can be viewed as integral to stylometric initiation attribution: it endeavors to recognize normal substance between records, regardless of the possibility that the style may have been changed. Origin attribution and initiation portrayal are very particular issues from unoriginality recognition. Initiation investigation has been utilized as a part of various application territories, for example, recognizing authors in writing, in program code, and so on. In the initiation attribution writing there are three sorts of proof that can be utilized to set up origin i.e. outside, interpretive and phonetic.

•External confirms incorporates the penmanship or a marked original copy of author.

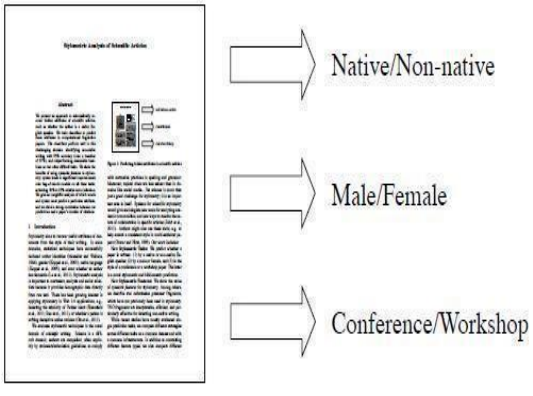
•Interpretive confirmation is the investigation of archive i.e. when it was composed, what the author implied by it and how that can be contrasted with different works by a similar author.

•Linguistic confirmation is concentrating on the examples of words and the real words that are utilized as a part of an archive.

In a few spaces, factual procedures have effectively found author character. Stylometric examination is critical to social researchers, advertisers and experts since it gives demographic information specifically from crude content or

information [9]. Stylometric study is utilized to recognize and verify the origin of email instant messages [14]. The intrigue has been developing in applying stylometry to the substance era where the substance is checked whether it is unique or duplicated from others style. Shane Bergsma, Matt Post, David Yarowsky are assessing stylometric strategies in the novel space of logical written work. Authors may likewise utilize these devices, e.g., to guarantee a predictable style in multi-wrote papers , or to decide areas of a paper requiring update. The commitments of paper incorporate, new Stylometric Tasks. They are foreseeing whether a paper is composed.

(1) By a local or non-local speaker
(2) By a male or female, and
(3) In the style of a gathering or workshop paper.



Legal etymology has a sub-field that is scientific stylistics and the author recognizable proof should be possible by applying stylistics. The elaborate depends on two premises. Two authors don't write in a similar example (having same first language).

• The essayist itself does not write in a similar example constantly. The elaborate can be classifications into two diverse methodologies:
• Qualitative
• Quantitative

In the subjective approach mistakes and individual conduct of the writers are evaluated though in the quantitative approach concentrate on promptly calculable and countable dialect highlights, e.g. length of word, length of sentence, expression length, recurrence of vocabulary, appropriation of expressions of various lengths [10]. Men and ladies actually talk a similar dialect. There are loads of studies have been done to think about the relationship between dialect utilize and sexual orientation. Sexual orientation recognizable proof issue can be dealt with as a paired grouping issue in (an), i.e., given two classes {male; female}, relegate an unknown email to one of them as indicated by the sex of the relating author:

e Є { Class1 if the author of e is male
e Є { Class2 if the author of e is female (a)

To test the twofold speculation (an), arrangement of components must be chosen that remain generally consistent for same sexual orientation composed substantial number of messages. At the point when the list of capabilities has been chosen, a n-dimensional vector speak to a given email, where n is the aggregate number of elements. An arrangement of known pre-ordered messages, a classifier (or model) can be worked by characterization procedures and the classification of another email can be resolved [11].

A. Gender: Male vs. Female

The authors [9] have taken information of Bergsma and Lin (2006).This information has been generally utilized as a part of gathering determination however never in stylometry. Every line in the information records how regularly a thing co-happens with male, female, impartial and plural pronouns. On the off chance that the name has a total check >30 and female likelihood >0.85, mark as female; generally if the total tally is >30 and male likelihood >0.85, name male [9]. For the sexual orientation ID of the author of an email is unique in relation to alternate sorts of initiation distinguishing proof issues. The email length is generally not long when contrasted with different sorts of writings like books and books. The messages style may fluctuate as indicated by the sort or economic wellbeing of beneficiaries, for instance, in business messages we take after formal style and in individual messages we take after casual style. Some extraordinary etymological components, for example, outward appearances regularly show up in messages. The organization or the messages structure may change among various clients. In this way, particular email-based sex separating highlight sets must be considered alongside conventional stylometric highlights [11]. Brennan and Greenstadt [12] clarified that present initiation attribution calculations are exceptionally precise in the non-ill-disposed case, however neglect to trait amend creation when a writer intentionally veils his written work style. The two types of ill-disposed assaults were characterized and tried: impersonation and confusion.

In the impersonation assault, writers conceal their written work style by copying another writer. In the jumbling assault, writers shroud their composition style in a way that won't be perceived. Conventional creation acknowledgment techniques perform not as much as arbitrary possibility in ascribing origin in both cases. These outcomes demonstrate that

successful stylometry procedures need to perceive and adjust to tricky written work. The writer contended that some etymological components change when individuals conceal their written work style and by distinguishing those elements, beguiling archives can be perceived. Misleading requires extra intellectual push to shroud data, which regularly presents unobtrusive changes in human conduct. These behavioral changes influence verbal and composed correspondence [13].

The task of distinguishing the author of a given content is author recognizable proof, in this way, it can be figured as a run of the mill grouping issue, which relies on upon discriminant components to speak to the style of a author [6]. Literary theft location [8] can be viewed as reciprocal to stylometric initiation attribution: it endeavors to distinguish regular substance between records, regardless of the possibility that the style may have been changed. Written falsification is not generally purposeful or taking a few things from another person; it can be unexpected or inadvertent and may involve self taking.

III. COMPARATIVE STUDY

In this section we have looked at the exploration in view of logical articles, email writer recognizable proof utilizing stylometry on the premise of their outcome examination. Shane Bergsma, Matt Post, David Yarowsky [9] have given the perforances investigation that they have considered NativeL (i.e Native versus Non-Native English Speaker) and Venue (i.e Top-Tier Vs. workshop). For NativeL, they had made Strict control (i.e. English name/nation) and just plot papers set apart as local. The papers which get the most minimal Native L-scores acquire less reference, yet they soon level off (Figure 2(a)). They have investigated that numerous lesser scientists at English colleges are non-local speakers; early-profession non-locals may get less references than understood companions. The connection amongst's references and Venue-scores is considerably more grounded (Figure 2(b)).The writer [9] effectively ascertained critical new errands and strategies in the stylometric examination of logical article, incorporated the novel determination of production scene in view of paper style, and novel syntactic elements in light of tree substitution linguistic use sections. In all above cases, there syntactic and elaborate components essentially enhance over a sack of-words (BOW) gauge, accomplishing 10%to 25% relative blunder decrease in each of the three noteworthy errands. The writer [14] made a program which was composed in the C# programming dialect, and it is having a Graphical User Interface (GUI) to improve the undertakings of deciding creation via computerizing the ID procedure. For the assurance of the creation they included accumulation of information, extraction of highlight, and arrangement. Clients helps the program to perceive authors by at first selecting an arrangement of test messages marked with known authors (counting author demographics) and thusly selecting an arrangement of test messages by obscure authors for correlation. They consider fifty-five complex components. There were 12 members and every member made ten messages, which found the middle value of one hundred and fifty (150) words, each on a particular subject. In [14] they have utilized different methods as a part of example arrangement, for example, Bayesian Theory, Decision Trees, Neural Networks or k-closest neighbor (KNN). Their program utilizes the KNN calculation which used to arrange objects in light of the premise of their similarity or separation metric. KNN classifiers depend on learning by similarity. On the premise of stylistics elements they discover the outcome in the expository way. The polarity information for the Stylometry confirmation tests contained 1770 records for every subset of six subjects. Every subset was keep running against the other yielding 76.72% and 66.72% exactness. The author [14] confronts a few challenges and their future work is to extend the validation assignment to distinguish designs in every now and again utilized incorrectly spelled and abused words.

IV. CONCLUSION AND FUTURE WORK

Through the overall discussion the paper we discussed first the basic behind the stylometry then later in the discussion move to the literature review where we have discussed about stylometry that can be used for the identification and authentication of the author in different fields like Author identification; detection of hoaxes, frauds, and deception in writing styles; gender identification from emails, plagiarism detection etc.. We have also analyzed the result on the basis of the stylometric features for the scientific articles and email author identification. So in this manner, we just see that the stylometry can be used in many broad areas. A still lot of research has to be done in field of author identification but we have chosen to implement it for the security of email by identifying the author and with this the security of the email system will be improved.

V.REFERENCES

[1] David I. Holmes, "*The Evolution of Stylometry in Humanities Authorship*," Literary and Linguistic Computing 13/3: Pages: 111-117, 1998.

[2] T. C. Mendenhall, "*The Characteristic Curves of Composition*," Science 214, Pages : 237– 246, 1887.

[3] C.Mascol "*Curves of Pauline and Psudo-Pauline Styles I*" Unitarian Review 30 pages: 452-60 1888:

[4] L. A. Sherman, "*Analytics of Literature: A Manual for the Objective Study of English Prose and Poetry"*, Boston: Ginn, 1893.

[5] Stefan Gruber, and Stuart Noven, "*Tool support for plagiarism detection in text documents*", Symposium on Applied Computing archive Proceedings of the 2005 ACM , Pages: 776 – 781, 2005.

[6] D. Pavelec, L. S. Oliveira, E.Justino, F. D. Nobre Neto, and L. V. Bastista, "*Compression and Stylometry for Author Identification*",

Proceedings of International Joint Conference on Neural Networks, 2009.

[7] Burrows, J. F., "Computers and the Study of Literature",. In: C. S. Butler (ed.): Computers and Written Texts. Oxford: Blackwell, Pages: 167–204, 1992.

[8] H. Maurer, F. K., "*Plagiarism - a survey*", Journal of Universal Computer Science, vol. 12, no. 8 , Pages: 1050-1084, 2006.

[9] Daniel Pavelec, Edson Justino, and Luis S.Oliveira, "*Author Identification using Stylometric Features*", Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. Vol 11, No 36, Pages: 59-65, 2007.

[10] Na Cheng, Xiaoling Chen, R. Chandramouli,

K. P. Subbalakshmi, "*Gender Identification from E-mails*", Computational Intelligence and Data Mining , Pages: 154-158, 2009.

[11] Michael Brennan, Rachel Greenstadt, "*Practical Attacks Against Authorship Recognition Techniques",* Innovative Applications of Artificial Intelligence (IAAI), 2009.

[12] Sadia Afroz, Michal Brennan and Rachel Greenstadt, "*Detecting Hoaxes, Frauds, and Deception in Writing Style Online*", IEEE Symposium on Security and Privacy, 2012 .

[13] K. Calix, M. Connors, D. Levy, H. Manzar, G. McCabe, and S. Westcott, "*Stylometry for E-mail Author Identification and Authentication*" Proceedings of CSIS Research Day, Pace University, May 2008.

CRYPTOGRAPHY BASED PRIVACY PRESERVING DATACOMMUNICATION IN HYBRID WIRELESS NETWORKS

G APOORVA[1],K SRILATHA[2],G VENKATESH[3],M AHARONU[4]
Department Of CSE, MRCE, Hyderabad, Telangana, India.
Apoorva.g@gmail.com[1],Sri.k@gmail.com[2],Venky.g@gmail.com[3],aharonu_cse@mrce.in[4]

ABSTRACT

Distributed Three-hop Routing protocol.DTR is used for data transmission in Hybrid wireless network. DTR divide a data into segments and transmits the segment in a distributed way. It uses at most two hops in ad-hoc transmission mode and one hop in cellular transmission mode. However, the selection of trust nodes for data transmission is difficult in DTR which in turn creates security issues. This paper proposes a TEEN APTEEN SPEED (TAS) protocol for conviction node selection. TAS protocol allocate a threshold value to each node in a network. Based on the threshold value, a trust node is selected for efficient data transmission in Hybrid Wireless Network. The threshold value is also to preserve security in the network in order that unauthorized spoofing nodes can't enter the network. Furthermore, this paper implements overhearing technique in which the sending node share the content with one or more other nodes before data transmission with the purpose that failure node can be exposed and replaced.

Index Terms – Hybrid wireless networks, Cryptography, Trust node, Overhearing

1. INTRODUCTION

Hybrid wireless network merge mobile ad-hoc network and infrastructure wireless network. It is to be an [3]improved network arrangement for the next generation network. According to the environment situation, it can select base station transmission mode or mobile ad-hoc transmission mode. The mobile ad-hoc network is an infrastructure-less network. The devices in a mobile ad-hoc network can shift in any path and the link between the devices can altered regularly. In this network, the data is transmitted from starting place to target in a multi-hop way through in-between nodes. In an infrastructure wireless network (e.g. Cellular network), each device communicates with other device through base stations. Each cell in a cellular network has a base station. These base stations are linked via cable or fiber or wirelessly through switching centers.

If the region has no communication infrastructure or the existing infrastructure, communication between nodes are complex or not suitable to use. In this location [2] hybrid wireless network may still be able to communicate through the construction of an ad-hoc network. In such a network, every mobile node operates as a host and also as a router. Forwarding packets to new mobile nodes in the network may not be within straight wireless transmission range. Each node participates in an ad-hoc routing and infrastructure routing, for this [1] Distributed three hop routing protocol is used. It allows to discovering a "Three-hop" path to any other node during the network is introduced in this effort The first two hops in ad-hoc networking is sometimes called infrastructure-less networking, since the mobile nodes in the network animatedly make routing between themselves to form their personal network. The third hop is created in infrastructure networking. Most Wi-Fi networks task in an infrastructure approach. Devices in this network communicate through a single access point, which is generally the wireless router. For example, consider the two laptops are placed next to each other, each connected to the same wireless network. still the two laptops are sited next to each other, they're not communicating in a straight line in infrastructure network. Some possible uses of hybrid wireless network consist of students using laptop, computers to participate in an interactive instruct, trade associates and sharing information during a gathering soldiers communicate information about the condition attentiveness on the emergency failure release and personnel coordinating efforts after a storm or shaking

Spread Code is normally used for safe data transmission in wireless communication as a way to measure the excellence of wireless connections. In wired networks, the existence of a wired path between the sender and receiver are determining the correct reception of a message. But in wireless networks, path defeat is a main trouble. The wireless communication network has to obtain a lot of environmental parameters to report background noise and interfere power of other simultaneous transmission. SINR attempts to produce a demonstration of this aspect. So the TAS protocol is implemented to keep the details about the dispatcher and receiver and the communication media in the network. This is implemented through overhearing concept. This TAS implements grouping of nodes depending on the threshold value so that the communication will be simple. In overhearing, the data is transferred to many nearby nodes in a cluster . The cluster is a grouping of nodes, which enclose cluster head and gateway. So the fundamental idea is to individually learn unknown and possibly random mobility parameters and to group the mobile node with related mobility prototype to the same cluster. The nodes in a cluster can then interchangeably distribute their resources for load balancing and overhead reduction, aiming to achieve scalable and proficient routing.

In TAS protocol, a secured code called threshold value is used. The nodal contact[7] probability are updating with the help of threshold value, it established to join the true contacts probabilities. Subsequently, a set of functions are devised to form clusters and choose entrance nodes based on nodal contact probabilities. lastly gateway nodes switch the network information and make routing. The result demonstrate that it is get higher delivery ratio and considerably lower overhead and end-to-end wait when compared to non-clustering matching part.

2. EXISTING WORK

The Base stations are coupled by means of a wired backbone, so that there are no power constraints and bandwidth during transmission among BS. The in-between nodes are utilized to indicate convey nodes that task as gateways connecting an infrastructure wireless network and mobile ad hoc network. DTR aims to move the routing load from the ad hoc network to the infrastructure network by taking advantage of extensive base stations in a hybrid wireless network. Rather than using one multi-hop path to forward a message to one BS, DTR uses at most[3] two hops to relay the segments of a message to different BS in a distributed way, and relies on BS to merge the segments. When a source node needs to propose a message stream to a destination node, it partition the message flow into a number of partial streams called segments and spread each segment to a neighbor node. Upon receiving a segment from the source node, a neighbor node decides between direct transmission and relay transmission based on the QoS requirement of the application. The neighbor nodes encourage these segments in a distributed way to nearby BS. Relying on the infrastructure network routing, the BS further transmit the segment to the BS where the destination node resides.

The ending BS reorganizes the segments into the original order and forwards the segments to the destination. It uses the cellular IP transmission method to begin segments to the destination if the destination moves to another BS through segment transmission.DTR works on the Internet layer. It receives packets from the TCP layer and routes it to 1the destination node, where DTR forwards the packet to the TCP layer. The data routing process in DTR can be separated into two processes: uplink from a source node to the first BS and downlink from the last BS to the data's destination. In uplink process, one hop to forward the segments of a message in a distributed way and uses another hop to find high-capacity forwarder for high show routing. As a result, DTR restrictions the path length of [8]uplink routing to two hops in order to keep away from the problems of long-path multi-hop routing in the ad-hoc networks. particularly, in the uplink routing, a source node divides its message flow into a number of segments, then transmits the segments to its neighbor nodes. The neighbor nodes promote segments to BS, which will forward the segments to the BS where the destination resides. In this work, throughput and routing speed are taken as a QoS requirement. The bandwidth/queue metric is to reflect node capacity in throughput and fast data forwarding. A larger

bandwidth/queue value means higher throughput and message forwarding speed, and vice versa. When selecting neighbors for data forwarding, a node needs the capacity information of its neighbors. Also, a chosen neighbor should have enough storage space for a segment. To find the capacity and storage space of its neighbors, each node periodically interactions its current information with its neighbors. If a node's capacity and storage space are altered, it again sends its present information to the segment forwarder. After that, the segment forwarder will select the maximum capacity nodes in its neighbors based on the updated information. That is, after a neighbor node receives a segment from the source, it uses either direct transmission or convey transmission. If the capacity of each of its neighbors is no greater than itself, relay node make use of direct transmission. If not, it uses convey transmission. In direct transmission, the relay nodes pass on the segment to a BS if it is in a BS's region. Or else, it stores the segment while moving until it goes into a BS's region. In relay transmission, relay node chooses its highest-capacity neighbor as the second relay node based on the QoS requirement. The second relay node will use through transmission to forward the segment directly to a BS. As a result, the number of transmission hops in the ad-hoc network component is confined to no more than two. The small number of hops helps to increase the capacity of the network and reduce channel conflict in ad-hoc transmission. The intention of the second hop choice is to find a higher capacity node as the message forwarder in order to pick up the performance of the QoS requirement.

If a source node has the maximum capability in its region, the segments will be forwarded rear to the source node according to the DTR protocol. The source node then forwards the segments to the BS straight due to the three-hop limit. This case occurs only when the source nodes is the maximum capacity node within its[9] two-hop neighborhood. Since the data transmission rate of the ad hoc interface is more than 10 times earlier than the cellular interface example 3G and GSM. Thus, the transmission wait for sending the data back and forth in the ad-hoc transmission is negligible in the total routing latency. After a BS receives a segment, it needs to forward the segment to the BS, where the destination node resides (i.e., the destination BS)..However, the destination BS recorded in the home BS may not be the most up-to-date destination BS since destination mobile nodes switch between the coverage regions of different BS during data transmission to them. For instance, data is transmitted to BS Bi that has the data's destination, but the destination has moved to the range of BS Bj before the data arrives at BS Bi. To deal with this problem, the[4] Cellular IP protocol is used for tracking node locations. With this protocol, a BS has a home agent and a foreign agent. The foreign agent keeps track of movable nodes moving into the ranges of other BS. The home agent intercepts in-coming segments, reconstructs the original data, and re-routes it to the foreign agent, which then forwards the data to the destination mobile node. After the destination BS receives the segments of a message, it rearranges the segments into the original message and then sends it to the destination mobile node. DTR specify the segment structure format for reschedule message. Each segment contains eight fields, including: (1) source node IP address; (2) destination node IP address; (3) message sequence number; (4) segment sequence number;(5) QoS indication number; (6) data; (7)length of the data; and (8) checksum.

3. PROPOSED WORK

Establishing the Network

The first step of network establishment is forming the cluster. The cluster is the group of related nodes formed in order to make the data transmission easier. every cluster will have Cluster top, Gateway and other nodes. The first criterion in wireless medium was to discover the available routes and establish them earlier than transmitting. The network consists of n nodes in which two nodes must be source and destination others will be used for data transmission. The path selection for data transmission is based on the availability of the nodes in the area using the ad-hoc on demand distance vector routing algorithm. Using the Ad-hoc on Demand Distance Vector routing protocol, the routes are created on demand as needed.

Threshold allocation

Threshold value distribution is done using TEEN, APTEEN and SPEED protocol. Based on the threshold value, trust node can be chosen also malicious node can be unobserved.

3.2.1 Threshold-sensitive Energy Efficient sensor Network protocol (TEEN)

It is a immediate protocol proposed for time-risky applications. The major idea of this technique is to produce the threshold value to every node in the network. After create the threshold value, the node is set in a hierarchical[6] clustering scheme in which some nodes act as a 1st level and 2nd level cluster heads. After forming the cluster head, the nodes get the data for transmission. Once the data is received the cluster head broadcasts the data to this cluster member.

Adaptive Threshold-sensitive Energy Efficient sensor Network protocol (ATEEN)

APTEEN is a hybrid [10]routing protocol planned for both time cyclic data collection and critical events. The main idea is to keep the statistical information. In this APTEEN method, the threshold value of each node in the cluster will be communicated with other cluster. Each cluster will have an APTEEN values.

SPEED Protocol

SPEED is a stateless protocol which provides real time communication by maintaining preferred release speed across the network. SPEED protocol is to discover geographic location. In this protocol whenever source nodes are[5] transmits a packet, the next hop neighbor is acknowledge using Stateless Non deterministic Geographic Forwarding (SNGF). The SNGF identifies a node as next hop neighbor, if it belongs to neighboring set of nodes, lies within the range of destination area and having speed larger than confident desired speed.

Overhearing Technique

The path selection, preservation and data transmission is repeated process which happens in split seconds in real time transmission. Hence the path allocated priory is used for data transmission. The first path allocated previously is used for data transmission. The data is transferred through the tinted path. But the transmission lane may be unsuccessful some times. At that moment second path is selected for data transmission. It takes additional time to find the second path. In order to deal with these overhearing is used. The overhearing is the idea in which the sending nodes allocate data to more than one node in a network. If the node collapse occurs in a network, that can be substituted by other active node.

Three hop Routing

Three hops are used for data transmission in a network. Two hops at mobile ad-hoc network and one hop at infrastructure network. The usage of this amalgamation will pick up the reliability. In this technique, the network is silent until a connection is needed. The new nodes forwarded this message, and documentation the node that they heard it from, creating an blast of temporary routes is back to the wanted node. while a node receives such a message, it will send the message backwards through a fleeting route to the requesting node. The deprived node then begins using the route that is the least number of hops through other nodes. Idle entries in the routing tables are recycled after a time.

4. CONCLUSION

Distributed Three-hop Routing protocol integrate the features of infrastructure and ad-hoc network in the data transmission process. In Distributed Three-hop Routing, source node divides a message flow into segments and broadcast them to its mobile neighbors and it further advance the segments to their target via an infrastructure network. Distributed Three-hop Routing restrictions the routing path length to three, and always arranges for high ability nodes to forward data. Distributed Three-hop Routing produces appreciably lower overhead by eliminating route find and maintenance. TAS protocol is implemented in this work which distributes a threshold value to each and every node in a network for the collection of trust nodes. In addition, Overhearing technique is applied to find out and change the failure node in the network. . It has the characteristics of short path length, short-distance transmission, and balanced load distribution provides high routing reliability with high efficiency and also include congestion control algorithm which can avoid load congestion in Bs in the case of unbalanced traffic distributions in networks. Besides the data transmission in hybrid wireless network is highly secure and more efficient.

REFERENCES

[1] Haiying shen, Ze Li, and Chenxi Qiu, "A Distributed Three-
Hop routing protocol to increase thehybri capacity of d wireless networks," IEEE Transactions Mobi on le computing, 2015.
sss

[2] B. Bengfort, W. Zhang, and X. Du "Efficient resource allocation in hybrid wireless networks," In Proc. of WCNC, 2011.

[3] L. M. Feeney, B. Cetin, D. Hollos, M. Kubisch, S. Mengesha, and H. Karl, "Multi-rate relaying for performance improvement in IEEE 802.11 wlans," In Proc. of WWIC, 2007.

[4] X. J. Li, B. C. Seet, and P. H. J. Chong, "Multi-hop cellular networks: Technology and economics," Computer Networks, 2008.

[5] K. Akkarajitsakul, E. Hossain, and D. Niyato, "Cooperative packet delivery in hybrid wireless mobile networks: A coalitional game approach," IEEE Trans. Mobile Computing 2013.

[6] P. Thulasiraman and X. Shen, "Interference aware resource allocation for hybrid hierarchical wireless networks," Computer Networks, 2010.

[7] L. B. Koralov and Y. G. Sinai, "Theory of probability and random processes," Berlin New York Springer, 2007.

[8] D. M. Shila, Y. Cheng, and T. Anjali, "Throughput and delay analysis of hybrid wireless networks with multi-hop uplinks," In Proc. of INFOCOM, 2011.

[9] T. Liu, M. Rong, H. Shi, D. Yu, Y. Xue, and E. Schulz, "Reuse partitioning in fixed two-hop cellular relaying network," In Proc. of WCNC, 2006.

[10] C. Wang, X. Li, C. Jiang, S. Tang, and Y. Liu, "Multicast throughput for hybrid wireless networks under Gaussian channels model," TMC, 2011.

REVIEW ON PARAMETERIZED ALGORITHMS AND KERNELIZATION

T.Aishwarya[1], S.Rajkumar[2], S.Santhoshi[3], Vijayakumari Ch[4]
Department of CSE, Malla Reddy College of Engineering, JNTU Hyderabad, Telangana, INDIA
E-mail[1]: aishwarya_cse@gmail.com, E-mail[2]: rajkumar_cse@gmail.com,
E-mail[3]: santoshi_cse@gmail.com, E-mail[4]: nagvijji@gmail.com

Abstract— *Huge numbers of the fascinating computational issues are NP-Hard. Down to earth uses of these issues made to handle these issues in numerous bearings. Correct answer for these NP-Hard issues is out of degree for sensibly greater occurrences of the issue. Heuristics, surmised arrangements are one approach to handle the issue. Parameterized many-sided quality comprehends these issues as for various different parameters. With the assistance of parameterized calculations a portion of the NP-Hard issues can be unraveled effectively for the little estimations of the information parameters. On the off chance that n is size of the information and k is the span of the parameter, an issue is Fixed Parameter Tractable (FPT) if the issue can be reasonable in time O (f(k) nc), where f(k) is a capacity just subject to k and c is a consistent. That is running time of the calculation is just polynomial ward of n.*

Kernelization is an intriguing idea to decrease the issue estimate. In this paper we audit this area of parameterized multifaceted nature, particularly parameterized calculations and kernelization procedures for a NP-Hard issue of registering Vertex Cover of a chart.

Keywords—Parameterized Complexity, Parameterized Algorithm, Kernelization, Vertex Cover

I. INTRODUCTION

To comprehend the NP-difficult issues in more point by point, Downey and Fellows (1999) presented the idea of parameterized many-sided quality. The traditional computational intricacy measures the running time of a calculation as a component of information size (say n). An issue is accepted to have productive arrangement if the issue can be understood in time corresponding to nc (that is O(nc)), where c is a consistent. Under the supposition that P ≠ NP there are numerous computational issues which might not have polynomial time calculations. These issues have exponential time calculations (That is O(cf(n))) for some steady c > 1. As n develops, the issue can't be settled by a PC. With the development of parameterized calculations, an issue in handled in numerous measurements. That is, aside from information measure, some different parameters are additionally given. Understood parameters incorporate, most extreme level of a chart, yield arrangement estimate, tree width et cetera. On the off chance that we can propose a calculation with running time O(f(k) nc), where c is a consistent and f(k) is a capacity exclusively ward of k (can be an exponential capacity on k), for little estimations of k the issue is resolvable and the issue is called settled parameter tractable (FPT). FPT is presently regarded as computational class and contains every one of the issues which has FPT calculations. There are issues turned out to be in FPT and there are an issue ended up being to be not has a place with the FPT class. There are issues which are yet demonstrate their enrollment to the FPT class.

There are numerous approaches to demonstrate an issue is in FPT. In writing numerous procedures are utilized to demonstrate the presence of a FPT calculation. The strategies incorporate limited pursuit tree, iterative pressure and kernelization. There are different systems additionally yet in this paper we concentrate just on these three methods. For more points of interest on parameterized many-sided quality you can allude to the book [1] by Downey and Fellows and a late book [2] by Cyganet. al. on parameterized calculations.

Let G = (V, E), with the end goal that |V| = n and |E| = m, be a basic undirected chart. Level of a vertex v is the quantity of edges occurrence on the vertex v. The open neighborhood of a vertex v is the arrangement of all the vertices which are adjoining v and signified by N(v). Shut neighborhood of a vertex v is the arrangement of all the vertices adjoining the vertex v including the vertex v itself, indicated by N[v] = N(v) ∪ {v}.

The vertex cover issue is characterized as takes after: A vertex cover is a subset S of the vertex set V (S ⊆ V) with the end goal that, for each edge (u, v) ∈ E either u ∈ S or v ∈ S. There might be numerous vertex covers for a chart, for instance the whole arrangement of vertices V is a vertex front of the diagram. Be that as it may, the set S with least cardinality among all the vertex spreads is called least vertex front of the chart. Finding the base vertex front of a diagram is NP-Complete [3]. The parameterized variation of the vertex cover issue (k-vertex cover issue) is characterized as takes after:

Input Instance: Input diagram G = (V, E) and a positive number parameter k

Yield: Vertex front of size at generally k

Proportionate choice issues: (Answer to the choice issue is either "YES" or 'NO')

Input Instance: Input diagram G = (V, E) and a positive number parameter k

Yield: Does the diagram has vertex front of size at generally k

II. BOUNDED SEARCH TREE TECHNIQUE

We hunt down an answer by taking after tree like pursuit, where the tree has limited profundity and each hub has consistent number of branches. For instance in tackling the k-vertex cover issue, in the event that we take any edge (x, y), either x is a piece of the vertex cover or y is a piece of the vertex cover. Thus, we can have seek tree with two branches. When we achieve level k of the pursuit tree, we will check vertices included so far structures a vertex cover, if not we will backtrack to alternate branches of the inquiry tree. This basic calculation requires some investment O(2k nc). The procedure is delineated in the Fig. 1.

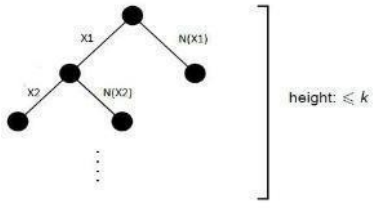


Fig. 1. First Bounded Search Tree Technique

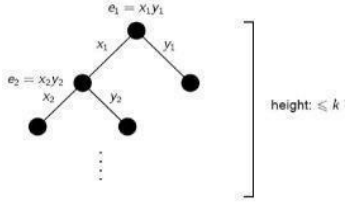


Fig. 2. Second Bounded Search Tree Technique

Presently we give an enhanced calculation. A little perception is, for a chart with each vertex has degree at most 1, vertex cover is registered effortlessly in direct time. Since, the chart is union of disjoint edges and we can incorporate one vertex of every edge in least vertex cover. In the event that the diagram has a vertex of degree ≥ 2, the inquiry tree can be altered as takes after: Take a vertex v with degree ≥ 2 then either v is a piece of the vertex cover or all its open neighbors N(v) are in the vertex cover. Subsequently we have a hunt tree with two branches. In one branch one vertex is added to the vertex cover and in other branch no less than two vertices (N(v)) are added to the vertex cover. The procedure is delineated in the Fig. 2.

The running time of the calculation can be communicated as T(k) ≤ T(k-1) + T(k-2), which is like Fibonacci arrangement. Henceforth T (n) = O(1.618k m).

A. Better Bounded Search Tree Techniques for Vertex Cover Problem

Balasubramanianet. al. [4] gave a calculation time O(kn + (1.324718)kk2). The best known calculation for vertex cover issue utilizing limited inquiry tree system is by Chen et. al. [5]. Their calculation requires some serious energy O(1.2852k + kn). Chen et. al. [6] likewise gave a calculation time O(1.2738k + kn).

III. KERNELIZATION

Utilizing certain diminishment rules iteratively, the issue example size is decreased. At certain point none of the diminishment tenets would apply, by then we would demonstrate the upper bound on the info example measure. These decrease principles would change an issue example P1 to an issue case P2 of lesser size. That is |P2| ≤ |P1| and P1 is a "YES" occurrence if and just if P2 is a 'YES example'. On the off chance that the information parameters for the issues P1 and P2 are individually k and k' and the lessening tenets would ensure that k' ≤ k. The decrease guidelines ought to require significant investment polynomial in information measure n. On the off chance that we can demonstrate the diminished issue P2 size is an element of k (not reliant on n) then the issue is dealt with as in FPT. Formally, kernelization is a polynomial change which maps an issue case (P1, k) to an issue occurrence (P2, k') with the end goal that the accompanying holds:

1) (P1, k) is a "YES" case if and just if (P2, k') is a "YES" occasion

2) k' ≤ k and

3) |P2| ≤ f(k), for some capacity f(k)

Any lessening principle taking after the over three conditions are called safe.

A. Kernelization for Vertex Cover Problem

The accompanying diminishment principles are connected to decrease the vertex cover issue example: ((G, k) be the info issue occurrence)

Rule1: Remove disengaged vertex. In the event that G has a confined vertex v, then the issue example can be diminished to (G\v, k)

Rule2: If G has a vertex v with degree > k, then the vertex v must be in the vertex cover, else we won't have a vertex front of size at generally k. Along these lines, the issue example can be diminished to (G\v, k-1)

Rule3: If G has a vertex of degree 1, then we can expect its neighbor is a piece of the vertex cover. Thus, the issue occasion can be decreased to (G\{u,v}, k-1)

The over three tenets are protected. In the wake of applying the lessening rules iteratively until none of the tenets are relevant, then if the diagram has |E| >k2 then obviously the issue has no vertex front of size at generally k. Subsequently the "YES" case will have at most k2 edges. Thus, the vertex cover has piece of size k2.

B. Polynomial Kernels

An issue has polynomial piece if the extent of the part is O(kc) for a consistent c. An issue has a direct piece if the span of the part is O(k). Generally individuals search for polynomial parts (direct portions) for parameterized issues. Kernelization infers the issue is in FPT. Not all issues has polynomial parts. There are procedures to demonstrate the nonexistence of polynomial portions. Indeed, even there are methods to demonstrate the part bring down limits.

C. Better Kernels for Vertex Cover Problem

Chen et. al. [5] has demonstrated the presence of 2k bit for the vertex cover issue. They have utilized the idea called Crown Reduction to demonstrate the 2k part.

IV. ITERATIVE COMPRESSION

The pressure routine is essential for the iterative pressure system. The pressure routine address the issue of separating a littler arrangement of the issue gave a greater arrangement of the issue. That is it takes an answer of size k + 1 and returns an answer of size littler than k + 1 on the off chance that it exists else it gives back no arrangement showing that the issue does not have arrangement littler than k + 1. Thus, the iterative pressure procedure takes a greater answer for the issue which is inconsequential to remove and iteratively searches for littler and littler arrangement and returns the arrangement.

Iterative pressure is utilized to demonstrate the presence of FPT calculations for issues like input vertex set issue.

V. CONCLUSIONS

In this paper we have audited fundamental computational instruments of parameterized calculations and kernelization. We took the vertex cover issue and delineated the strategies like limited hunt tree method and kernelization. We have additionally depicted about iterative pressure. Despite the fact that it is not connected to vertex cover issue it is very much utilized as a part of parameterized calculations. We have additionally recorded the best known calculations for the vertex cover issue. Both iterative pressure and kernelization best known calculations are appeared.

VI. REFERENCES

[1] R. G. Downey and M. R. Fellows, Parameterized Complexity. Springer Publishing Company, Incorporated, 2012.

[2] M. Cygan, F. V. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, and S. Saurabh, Parameterized Algorithms. Springer Publishing Company, Incorporated, 2015.

[3] R. M. Karp, "Reducibility Among Combinatorial Problems," in Com-plexity of Computer Computations. Plenum Press, 1972, pp. 85–103.

[4] R. Balasubramanian, M. R. Fellows, and V. Raman, "An improved fixed-parameter algorithm for vertex cover," Inf. Process. Lett., vol. 65, no. 3, pp. 163–168, 1998.

[5] J. Chen, I. Kanj, and W. Jia, "Vertex cover: Further observations and further improvements," in Proceedings of the 25th International Workshop on Graph-Theoretic Concepts in Computer Science, 1999, pp. 313–324.

[6] J. Chen, I. A. Kanj, and G. Xia, "Improved upper bounds for vertex cover," Theoretical Computer Science, vol. 411, no. 4042, pp. 3736 – 3756, 2010.

proceedings of International Conference on Emerging Technologies in Computer Science (ICETCS) ISBN 978-93-85101-62-5

COOPERATIVE PROVABLE DATA POSSESSION FOR INTEGRITY VERIFICATION IN MULTI-CLOUD STORAGE

M.Priyanka[1], Mohammed Abdul[2], P.Mukesh[3], N.Keerthi[4]
Department of CSE, MRCE, JNTU , Hyderabad.
e-mail[1] : priyanka.m@gmail.com , e-mail[2] : abdul.m23@gmail.com , e-mail[3] : mukesh.p21@gmail.com , e-mail[4] : keerthi.narvaneni@gmail.com

Abstract -- ***Provable data possession (PDP) is a procedure for guaranteeing the trustworthiness of information away outsourcing. In this paper, we address the development of a proficient PDP conspire for disseminated distributed storage to bolster the versatility of administration and information movement, in which we consider the presence of various cloud benefit suppliers to agreeably store and keep up the customers' information. We exhibit a cooperative PDP (CPDP) conspire based on homomorphic verifiable reaction and hash record chain of command. We demonstrate the security of our plan in view of multi-proven zero-learning confirmation framework, which can fulfill culmination, information soundness, and zero-learning properties. Also, we explain execution streamlining systems for our plot, and specifically show a productive strategy for selecting ideal parameter qualities to minimize the calculation expenses of customers and capacity benefit suppliers. Our investigations demonstrate that our answer presents bring down calculation and correspondence overheads in correlation with non-agreeable methodologies.***

Index Terms-- Provable Data Possession, Zero-Knowledge, Storage Security, POR, Multiple-Cloud

I. INTRODUCTION

Lately, distributed storage benefit has turned into a quicker benefit development point by giving an equivalently ease, versatile, position-autonomous stage for customers' information. Since distributed computing environment is built in view of open models and interfaces, it has the ability to consolidate numerous inner as well as outer cloud benefits together to give high interoperability. We call such a disseminated cloud environment as a multi-Cloud (or cross breed cloud).Frequently, by utilizing virtual infrastructure management (VIM) [1], a multi-cloud permits customers to effortlessly get to his/her assets remotely through interfaces, for example, Web administrations gave by Amazon EC2. Provable data possession (PDP) [2] (or proofs of retrievability (POR) [3]) is such a probabilistic confirmation procedure for a capacity supplier to demonstrate the honesty what's more, responsibility for information without downloading information. The verification checking without downloading makes it particularly imperative for vast size records and envelopes (regularly including many customers' records) to check whether these information have been altered then again erased without downloading the most recent adaptation of information. Along these lines, it can supplant conventional hash and signature works away outsourcing. Different PDP plans have been as of late proposed, for example, Versatile PDP [4] and Dynamic PDP [5]. Nonetheless, these plans predominantly concentrate on PDP issues at untrusted servers in a solitary distributed storage supplier and are not appropriate for a multi-cloud environment.

There exist different devices and innovations for multi cloud, for example, Platform VM Orchestrator, VMware VSphere, and Ovirt. These devices cloud suppliers develop a disseminated distributed storage stage for dealing with customers' information. Notwithstanding, if such a critical stage is powerless against security assaults, it would convey hopeless misfortunes to the customers. For instance, the classified information in an undertaking might be illicitly gotten to through a remote interface gave by a multi-cloud, or important information and chronicles might be lost or messed with when they are put away into an· indeterminate capacity pool outside the endeavor. Along these lines, it is fundamental for cloud benefit suppliers to give security systems to dealing with their capacity administrations.

II. RELATED WORK

To check the accessibility and uprightness of outsourced information in cloud stockpiles, scientists have proposed two essential methodologies called Provable Data Possession (PDP) [2] and Proofs of

Retrievability (POR) [3]. Ateniese et al. [2] initially proposed the PDP demonstrate for guaranteeing ownership of documents on untrusted stockpiles and gave a RSA-based plan to a static case that accomplishes the (1) correspondence taken a toll. They additionally proposed a freely unquestionable variant, which permits anybody, not only the proprietor, to challenge the server for information ownership. This property significantly expanded application territories of PDP convention due to the partition of information proprietors and the clients. In any case, these plans are shaky against replay assaults in dynamic situations due to the conditions on the record of pieces. Additionally, they don't fit for multi-distributed storage because of the loss of homomorphism property in the check procedure. Keeping in mind the end goal to bolster dynamic information operations, Ateniese et al. built up an element PDP arrangement called Adaptable PDP [4].

They proposed a lightweight PDP plot in light of cryptographic hash work and symmetric key encryption, however the servers can hoodwink the proprietors by utilizing past metadata or reactions because of the absence of arbitrariness in the difficulties. The quantities of upgrades and difficulties are constrained and settled ahead of time and clients can't perform piece inclusions anyplace.

In synopsis, a confirmation conspire for information honesty in dispersed stockpiling situations ought to have the accompanying elements

- **Usability aspect:** A customer ought to use the honesty check in the method for cooperation administrations. The plan ought to cover the points of interest of the capacity to diminish the weight on customers.
- **Security aspect:** The plan ought to give sufficient security components to oppose some current assaults, for example, information spillage assault and label fraud assault.
- **Performance aspect:** The plan ought to have the lower correspondence and calculation overheads than non-helpful arrangement.

ALGORITHMS USED

Algorithm	Description	Evaluation
PDP	Ensuring possession of files on untrusted storages and provided an RSA-based scheme for communication.	Insecure again replay attacks dynamic scenarios .
Compact	Uses homomorphic a proof in authenticator value with O(1) and t challenge blocks O(t).	Supports only for static data and could not prevent the leakage of data blocks in the verification.
Scalable PDP	Suitable for the limited dynamic nature and require pre-computed answers as metadata which allows limited and fixed a prior no of updates and challenges.	Requires lot off pre computations to improve the performance and supporting only append type insertions.
DPDP	Based on PDP model for dynamic files which can be updated online.	Complexity of the order of O(log n).
Improved DPDP	Improved the model based on DPDP model, and reduces the computational and communication complexity to constant.	----
Cooperative PDP	Provable data possession in distributed cloud environments from the aspects : high security , transparent verification , and high performance.	Model is evaluated on simulator by using hadoop file system.

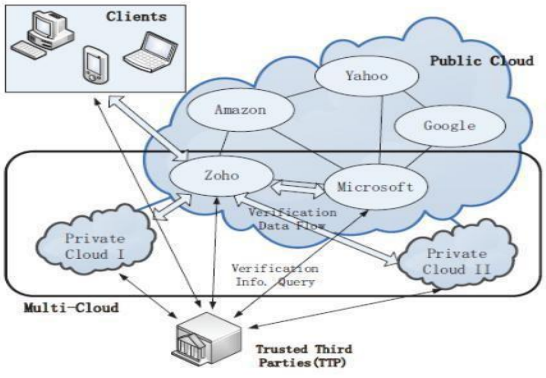
III PROPOSED WORK

In this paper, we address the issue of provable information ownership in dispersed cloud more, homomorphic Verifiable response (HVR). We then exhibit that the likelihood of developing an agreeable PDP (CPDP) conspire without trading off information security in view of cutting edge cryptographic methods, for example, interactive proof systems (IPS). We promote present a compelling development of CPDP plan utilizing previously mentioned structure. Besides, we give a security examination of our CPDP conspire from the IPS demonstrate. We demonstrate that this development is a multi-proven zero-knowledge provable system (MP-ZKPS) [11], which has culmination, information soundness, and zero-learning properties. These properties guarantee that CPDP plan can actualize the security against information spillage assault and label imitation assault. To enhance the framework execution as for our plan, we examine the execution of probabilistic questions for recognizing irregular circumstances. This probabilistic technique additionally has a natural advantage in diminishing calculation and correspondence overheads. At that point, we display a productive strategy for the choice of ideal parameter qualities to minimize the calculation overheads of CSPs and the customers' operations. What's more, we examine that our plan is reasonable for existing disseminated distributed storage frameworks. At last, our tests demonstrate that our answer presents extremely constrained calculation and correspondence overheads.

IV. STRUCTURE AND TECHNIQUES

In this area, we exhibit our confirmation system for multi-distributed storage and a formal meaning of CPDP. We present two central strategies for developing our CPDP plot: hash index hierarchy (HIH) on which the reactions of the customers' difficulties processed from different CSPs can be com- This article has been acknowledged for production in a future issue of this diary, however has not been completely altered. Substance may change preceding last distribution. bined into a solitary reaction as the last result; and homomorphic verifiable response (HVR) which bolsters disseminated distributed storage in a multi-distributed storage furthermore, executes a proficient development of collision resistant hash work, which can be seen as a arbitrary prophet demonstrate in the check convention.

situations from the accompanying viewpoints: high security, straightforward check, and superior. To accomplish these objectives, we first propose a check system for multi-distributed storage alongside two principal systems: hash index hierarchy (HIH) what's



1. Multi distributed storage:

Conveyed registering is utilized to allude to any extensive coordinated effort in which numerous individual PC proprietors permit some of their PC's handling time to be put at the administration of an expansive issue. In our framework the every cloud administrator comprise of information pieces. The cloud client transfers the information into multi cloud. Distributed computing environment is developed in light of open models and interfaces; it has the ability to fuse numerous inner as well as outer cloud benefits together to give high interoperability. We call such a conveyed cloud environment as a multi-Cloud .A multi-cloud permits customers to effortlessly get to his/her assets remotely through interfaces.

2. Agreeable PDP

Agreeable PDP (CPDP) plans embracing zero-information property and three-layered record order, separately. Specifically productive strategy for selecting the ideal number of parts in every square to minimize the calculation expenses of customers and capacity benefit suppliers. Helpful PDP (CPDP) plot without trading off information security in light of present day cryptographic procedures

3. Information Integrity

Information Integrity is vital in database operations specifically and Data warehousing and Business insight by and large. Since Data Integrity guaranteed that information is of high caliber, right, steady and open.

4. Outsider Auditor

Trusted Third Party (TTP) who is trusted to store confirmation parameters and offer open inquiry administrations for these parameters. In our framework the Trusted Third Party, see the client information squares and transferred to the circulated cloud. In dispersed cloud environment every cloud has client information squares. In the event that any alteration attempted by cloud proprietor an alarm is send to the Trusted Third Party.

5. Cloud User

The Cloud User who has a lot of information to be put away in numerous mists and have the consents to get to and control put away information. The User's Data is changed over into information pieces. The information squares is transferred to the cloud. The TPA sees the information squares and Uploaded in multi cloud. The client can upgrade the transferred information. On the off chance that the client needs to download their records, the information's in multi cloud is coordinated and downloaded.

V. SECURITY ANALYSIS

We give a brief security examination of our CPDP development. This development is specifically inferred from multi-proven zero-information evidence framework (MPZKPS), which fulfills taking after properties for guaranteed declaration L:

1. **Completeness:** at whatever point x ∈ L, there exists a procedure for the provers that persuades the verifier that this is the situation.
2. **Soundness**: at whatever point x ∉ L, whatever procedure the provers utilize, they won't persuade the verifier that x ∈ L.
3. **Zero-information**: no tricking verifier can learn something besides the veracity of the announcement. As per existing IPS explore [15], these properties can shield our development from different assaults, for example, information spillage assault (security spillage), label falsification assault (possession conning), and so forth.

VI. CONCLUSION AND FUTURE SCOPE

We displayed the development of a proficient PDP plot for dispersed distributed storage.In view of homomorphic certain reaction and hash List chain of importance, we have proposed a helpful PDP plan to bolster dynamic adaptability on various stockpiling servers. We likewise demonstrated that our plan Given all security properties required by zero information intelligent evidence framework, with the goal that it can oppose different assaults regardless of the possibility that it is sent as an open Audit benefit in mists. Moreover, we enhanced the probabilistic inquiry and intermittent check to enhance the review execution. Our examinations unmistakably showed that our methodologies just present a little measure of calculation and correspondence overheads. Thusly, our answer can be dealt with as another possibility for information respectability confirmation in outsourcing information stockpiling frameworks. As a feature of future work, we would extend our work to investigate more powerful CPDP developments. At long last, it is still a testing issue for the era of labels with the length insignificant to the measure of information squares. We would investigate such an issue to give the support of variable-length piece confirmation.

VII. REFERENCES

1. A.Juels and B. S. K. Jr., "Pors: proofs of retrievability for large files," in *ACMConference on Computer and Communications Security*, P. Ning, S. D. C. di Vimercati, and P. F. Syverson, Eds. ACM, 2007, pp. 584–597.

2. G. Ateniese, R. C. Burns, R. Curtmola, J.Herring, L. Kissner, Z. N. J. Peterson, and D. X. Song, "Provable data possession at untrusted stores," in *ACM Conference on Computer and Communications Security*, P. Ning, S. D. C. di Vimercati, and P. F. Syverson, Eds. ACM, 2007, pp. 598–609.

3. Y. Zhu, H. Wang, Z. Hu, G.-J. Ahn, H. Hu, and S. S. Yau, "Dynamic audit services for integrity verification of outsourced storages in clouds," in *SAC*, W. C. Chu, W. E. Wong, M. J. Palakal, and C.-C. Hung, Eds. ACM, 2011, pp. 1550–1557.

4. G. Ateniese, R. D. Pietro, L. V.Mancini, and G. Tsudik, "Scalable and efficient provable data possession," in Proceedings of the 4th international conference on Security and privacy in communication netowrks,2008, pp. 1–10.
5. R. Curtmola, O. Khan, R. Burns, and G. Ateniese. Mr. pdp: Multiple-replica provable data possession. In Proc. of The 28th IEEE International Conference on Distributed Computing Systems (ICDCS'08), 2008, to appear.
6. A. Juels and B. Kaliski. PORs: Proofs of retrievability for large files. In ACM CCS'07, Full paper available on e-print (2007/243), 2007

WEB USAGE MINING THROUGH LESS COST

P.V.Vara Yeswanth[1] , P. Akhil Chandra[2] ,P. Surabh[3], R. Bangari[4]
1,2,3&4 Department of Computer Science and Engineering
Malla Reddy College of Engineering , JNTUniversity
Hyderabad , Telengana, India
vara@gmail.com[1] , akhil@gmail.com[2] , surabhi@gmail.com[3] , ramarapun@gmail.com[4]

Abstract: In this paper we present the web mining using Cloud Computing Technology. Web mining includes how to extract the useful information from the web and gain knowledge using data mining techniques. Here so many resources and techniques are available i.e. web content mining, web structure mining, web usage mining and access through the web servers. Web mining techniques (specially web usage mining techniques) and applications are much needed in cloud computing. The implementation of these techniques through cloud computing will allow users to retrieve relevant and meaningful data from virtually integrated data warehouse which reduces cost and infrastructure.

Keywords— Data mining, Web Mining, Data Warehouse, Knowledge Discovery, Cloud Mining, Web Content Mining, Web Structure Mining, Web Usage Mining.

INTRODUCTION

Web mining - is the application of data mining techniques to discover patterns from the Web. According to analysis targets, web mining can be divided into three different types, which are Web usage mining, Web content mining and Web structure mining. Web usage mining is the process of extracting useful information from server logs e.g. use Web usage mining is the process of finding out what users are looking for on the Internet using cloud computing . Some users might be looking at only textual data, whereas some others might be interested in multimedia data. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. Web usage mining itself can be classified further depending on the kind of usage data considered. Several data mining methods are used to discover the hidden information in the Web. However, Web mining does not only mean applying data mining techniques to the data stored in the Web. The algorithms have to be modified such that they better suit the demands of the Web [1].

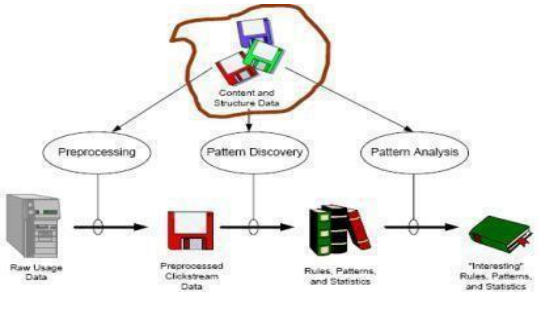
The web usage mining generally includes the following several steps: data collection, data retreatment, and knowledge discovery and pattern analysis

A] Data collection:

Web usage mining is the process of extracting useful information from server logs e.g. use Web usage mining is the process of finding out what users are looking for on the Internet.

Approach of Web usage mining

2.1. Concept of web usage mining



Some users might be looking at only textual data, whereas some others might be interested in multimedia data. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. Web usage mining itself can be classified further depending on the kind of usage data considered:

Web Server Data: The user logs are collected by the Web server. Typical data includes IP address, page reference and access time. Application Server Data: Commercial application servers have significant features to enable e-commerce applications to be built on top of them with little effort. A key feature is the ability to track various kinds of business events and log them in application server logs[3].

Application Level Data: New kinds of events can be defined in an application, and logging can be turned on for them thus generating histories of these specially defined events. It must be noted, however, that many end applications require a combination of one or more of the techniques applied in the categories above in the figure() [3].

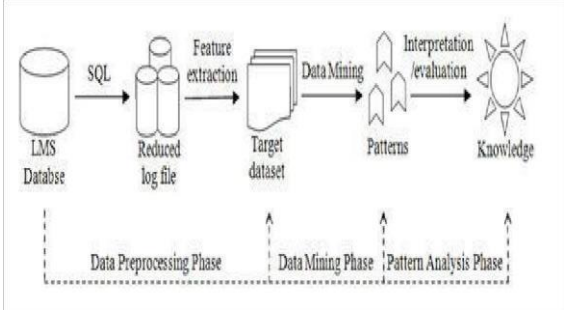
B] Data preprocessing:

Web Usage Mining in cloud computing is one of the categories of data mining technique that identifies usage patterns of the web data, so as to perceive and better serve the requirements of the web applications. The working of WUM involves three steps - preprocessing, pattern discovery and analysis. The first step in WUM - Preprocessing of data is an essential activity which will help to improve the quality of the data and successively the mining results. This research paper studies and presents several data preparation techniques of access stream even before the mining process can be started and these are used to improve the performance of the data preprocessing to identify the unique sessions and unique users in cloud computing . The methods proposed will help to discover meaningful pattern and relationships from the access stream of the user and these are proved to be valid and useful by various research tests.

The paper is concluded by proposing the future research directions in this space [2].

In the data pretreatment work, mainly include data cleaning, user identification, session identification and path completion.

1. Data Cleaning*:* The most important task of the Web Usage Mining in cloud computing process is data preparation. This process is diagrammatically represented in Fig(2) . The success of the project is highly correlated to how well the data preparation task is executed. It is of utmost importance to ensure, every nuance of this task is taken care of. This process deals with logging of the data; performing accuracy check; putting the data together from disparate sources; transforming the data into a session file; and finally structuring the data as per the input requirements. The data used for this project is from the RIT Apache server logs, which is in the Common Log File format. This access log includes the agent and the referrer in the data as one of the attributes[4].



2. Path completion*:* An implementation of data preprocessing system for Web usage mining and the details of algorithm for path completion are presented. After user session identification, the missing pages in user access paths are appended by using the referrer-based method which is an effective solution to the problems introduced by using proxy servers and local caching. The reference length of pages in complete path is modified by considering the average reference length of auxiliary pages which is estimated in advance through the maximal forward references and the reference length algorithms. As verified by practical Web access log, the proposed path completion algorithm efficiently appends the lost information and improves the reliability of access data for further Web usage mining calculations[5].

C] Knowledge Discovery: In general, knowledge discovery can be defined as the process of identifying interesting new patterns in data. These patterns can be, e.g., relations, events or trends, and they can reveal both regularities and exceptions[3].

D] Pattern analysis: Challenges of Pattern Analysis are to filter uninteresting information

and to visualize and interpret the interesting patterns to the user. First dlete the less significance rules or models from the interested model storehouse; Next use technology of OLAP and so on to carry on the comprehensive mining and analysis; Once more, let discovered data or knowledge be visible; Finally, provide the characteristic service to the electronic commerce website[3].

LITERATURE SURVEY

I read many paper related to web usage mining these are following :

A Framework for Personal Web Usage Mining: In this paper, I got to mine Web usage data on client side, or personal Web usage mining, as a complement to the server side Web usage mining. By mining client side Web usage data, more complete knowledge about Web usage can be obtained.

A Research Area in Web Mining: This paper also discusses an application of WUM, an online Recommender system that dynamically generates links to pages that have not yet been visited by a user and might be of his potential interest. Differently from the recommender systems proposed so far, it does not make use of any off-line component, and is able to manage Web sites made up of pages dynamically generated.

Cloud Mining: This paper also discussed about Web usage mining and user behavior analysis using fuzzy C-means clustering: In this paper I got methodologies used for classifying the user using Web Usage data. This model analysis the users behaviors and depend on the interests of similar patterns provides appropriate recommendations for active user.

Discovery and Applications of Usage Patterns from Web Data: This paper provides an up-to-date survey of the rapidly growing area of Web Usage mining. With the growth of Web-based applications, specifically electronic commerce, there is significant interest in analyzing Web usage data to better understand Web usage, and apply the knowledge to better serve users.

Web Mining Using Cloud Computing: This paper provide present the technology of cloud computing using web mining .Web mining include how to extract the useful information from the web and gain knowledge using data mining techniques. Here so many online resources are available i.e. web content mining and access through the web servers.

RELATED WORK

Many researchers have looked for way of represent the web mining and future of web mining in Cloud Computing. Some of these are said that cloud mining is the future of web mining. This paper describes the web usage mining in Cloud Computing technology. Web usage mining model is a kind of mining to server logs. Web Usage Mining plays an important role in realizing enhancing the usability of the website design, the improvement of customers' relations and improving the requirement of system performance and so on. Web usage mining provides the support for the web site design, providing personalization server and other business making decision.

Web usage mining in Cloud Computing is clearly one of today's most seductive technology area in research field due to its cost efficiency and flexibility. However, despite increased activity and interest, there are significant, persistent concerns about cloud computing that are impeding momentum and will eventually compromise the vision of cloud computing as a new IT procurement model. The term cloud is a symbol for the Internet, an abstraction of the Internet‘s underlying infrastructure, used to mark the point at which responsibility moves from the user to an external provider. Basically Cloud Mining is new approach to faced search interface for your data. SaaS (Software-as-a-Service) is u sed for reducing the cost of web mining and try to prov ide security that become

with cloud mining technique. Now a day we are read y to modify the framework of web mining for demand c loud computing. In terms of ―mining clouds, the Hadoop and Map Reduce communities who have developed a powerful framework for doing predictive analytics against complex distributed information sources.

Currently, Web usage mining finds patterns in Web server logs. The logs are preprocessed to group requests from the same user into sessions. A session contains the requests from a single visit of a user to the Web site. During the preprocessing, irrelevant information for Web usage mining such as background imag es and unsuccessful requests is ignored. The users are id entified by the IP addresses in the log and all requests from the same IP address within a certain time-window are put into a session.

ONLINE WEB USAGE MINING IN CLOUD SYSTEM:

Web based recommender systems are very hel pful in directing the users to the target pages in particular web sites. Moreover, Web usage mining cloud model systems have been proposed to predict user's intention and their navigation b ehaviors. In the following, we review some of the most significant WUM syst

terms and architecture that can be compared with our syste m. Cloud system proposed a Cloud model for navigation pattern mining thrrough Web usage mining to predict user future moveme nts. The approach is based on the graph partitioning clustering algorithm to m odel user navigation patterns for the navigation patterns m ining phase.

CONCLUSION

Cloud Computing is a broad term that describ es a broad range of services. As with other significant developments in technology, many vendors have seized the term "Cloud" and are using it for products that sit outside of the common definition. In order to truly understand how the Cloud can be of value to an organization, it is first important to understand wha t the Cloud really is and its different components. Since the Cloud is a bro ad collection of services, organizations can choose wh ere, when, and how they use Cloud Computing. In this paper I explained th e different types of Cloud Computing services commonl y referred to as Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) with the help o f web using mining in cloud

REFERENCE

[1] Web Mining , www.wikipedia.com,10-02-2015.
[2] Data preprocessing,http://ieeexplore.ieee.org,10-02-2015.
[3] M.Rathamani et al,"Web usage mining and user behavior analysis using fuzzy c-means clustering"IOSR journal
Of Computer Engineering ISSN:2278-0661,ISBN:2278-8727 Volume 7,Issue2(Nov-Dec.2012),PP09-15.
[4] Yongjian Fu et al, "A Frame work for Personal Web Usage Mining"
[5] Rajni Pamnani,Pramila Chawan Department of computer technology,VJTI University,Mumbai "Web Usage
Mining A Research Area in Web Mining"

Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud

N.Praveen kumar Reddy[1] , N.Shashank Reddy[2], D. Abhishek[3], K.Navya[4]
Department of CSE, MRCE, JNTU Hyderabad
e-mail[1]: praveenkumarreddy@gmail.com, e- mail[2] : shashankreddy@gmail.com, e-mail[3] : abhishek@gmail.com, e-mail[4] : navya.koochana@gmail.com

ABSTRACT

In recent years ad-hoc parallel data processing has emerged to be one of the killer applications for Infrastructure-as-a-Service (IaaS) clouds. Major Cloud computing companies have started to integrate frameworks for parallel data processing in their product portfolio, making it easy for customers to access these services and to deploy their programs. However, the processing frameworks which are currently used have been designed for static, homogeneous cluster setups and disregard the particular nature of a cloud. Consequently, the allocated compute resources may be inadequate for big parts of the submitted job and unnecessarily increase processing time and cost. In this paper we discuss the opportunities and challenges for efficient parallel data processing in clouds and present our research project. It is the first data processing framework to explicitly exploit the dynamic resource allocation offered by today's IaaS clouds for both, task scheduling and execution. Particular tasks of a processing job can be assigned to different types of virtual machines which are automatically instantiated and terminated during the job execution.

Keywords: Task computing, query processing, dynamic resource allocation,Task Computing

I. INTRODUCTION

For organizations that exclusive need to process vast sums Today a developing number of organizations need to prepare gigantic measures of information in a cost-effective way. Great agents forthese organizations are administrators of Internet web indexes, similar to Google, Yahoo, or Microsoft. The boundless measure of information they need to manage each day has made customary database arrangements restrictively costly [5]. Rather, these organizations have advanced a design worldview in light of countless servers. Issues like preparing slithered archives or recovering a web file are part into a few autonomous subtasks, appropriated among the accessible hubs, and computedin parallel.Cloud processing has developed as a promising way to deal with lease a huge IT in-frastructure on a fleeting pay-per-use premise. Administrators of alleged Infrastructure-as-a-Service (IaaS) mists, similar to Amazon EC2 [1], let their clients apportion, get to, and control an arrangement of Virtual Machines (VMs) which keep running inside their server farms and just charge them for the timeframe the machines are designated. The VMs are regularly offered in various sorts, every sort with its own qualities (number of CPU centers, measure of primary memory, and so forth.) and cost.

This paper is an extended It includes further detailson scheduling strategies and extended experimental results. The paper is structured as follows: Section II, starts with analyzing the above men-tioned opportunities and challenges and derives some important design principles for our new framework.
In Section III, we present Nephele's basic architecture and outline how jobs can be described and executed in the cloud. Section IV, provides some first figures on Nephele's performance and the impact of the optimizations we propose. Finally, our work is concluded by related work (Section V) and ideas for future work

II. CHALLENGES AND OPPORTUNITIE

Current data processing frameworks like Google's MapReduce or Microsoft's Dryad engine have been designed for cluster environments. This is reflected in a number of assumptions they make which are not necessarily valid in cloud environments. In this section we discuss how abandoning these assumptions raises new opportunities but also challenges for efficient parallel data processing in clouds.

A. OPPORTUNITIES

Today's processing frameworks typically assume the re-sources they manage consist of a static set of homogeneous compute nodes. Although designed to deal with individ- ual nodes failures, they consider the number of available machines to be constant, especially when scheduling the processing

job's execution. While IaaS clouds can certainly be used to create such cluster-like setups, much of their flexibility remains unused. One of an IaaS cloud's key features is the provisioning of compute resources on demand. New VMs can be allocated at any time through a well-defined interface and become available in a matter of seconds. Machines which are no longer used can be terminated instantly and the cloud customer will be charged for them no more. Moreover, cloud operators like Amazon let their customers rent VMs of different types, i.e. with different computational power, different sizes of main memory, and storage. Hence, the compute resources available in a cloud are highly dynamic and possibly heterogeneous.

B. CHALLENGES

The cloud's virtualized nature helps to enable promising new use cases for efficient parallel data processing. However, it also imposes new challenges compared to classic cluster setups. The major challenge we see is the cloud's opaqueness with prospect to exploiting data locality: In a cluster the compute nodes are typically interconnected through a physical high performance network. The topology of the network, i.e. the way the compute nodes are physically wired to each other, is usually wellknown and, what is more important, does not change over time. Current data processing frameworks offer to leverage this knowledge about the network hierarchy and attempt to schedule tasks on compute nodes so that data sent from one node to the other has to traverse as few network switches as possible [9]. That way network bottlenecks can be avoided and the overall throughput of the cluster can be improved. In a cloud this topology information is typically not exposed to the customer [29]. Since the nodes involved in processing a data intensive job often have to transfer tremendous amounts of data through the network, this drawback is particularly severe; parts of the network may become congested while others are essentially unutilized. Although there has been research on inferring likely network topologies solely from end- to-end measurements (e.g. [7]), it is unclear if these techniques are applicable to IaaS clouds. For security reasons clouds often incorporate network virtualization techniques (e.g. [8]) which can hamper the inference process, in particular when based on latency measurements.

III. DESIGN

Based on the challenges and opportunities outlined in the previous section we have designed Nephele, a new data processing framework for cloud environments. Nephele takes up many ideas of previous processing frameworks but refines them to better match the dy- namic and opaque nature of a cloud.

A. ARCHITECTURE

Nephele's architecture follows a classic master-worker pattern as illustrated in fig. 1.Fig.

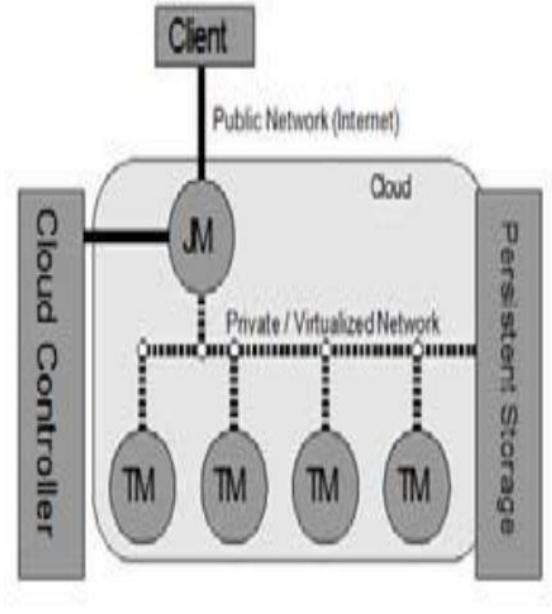


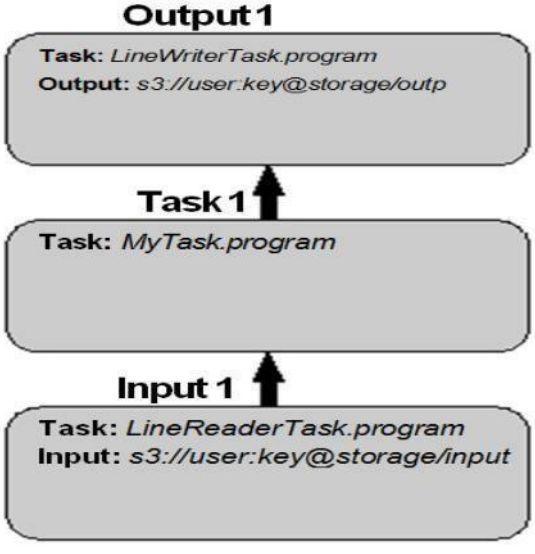
Fig. 1: Structural Overview of Nephele Running in an Infrastructure-as-a-Service (IaaS) Cloud

Before submitting a Nephele compute job, a user must start a VM in the cloud which runs the so called Job Manager (JM). The Job Manager receives the client's jobs, is responsible for scheduling them, and coordinates their execution. It is capable of communicating with the interface the cloud operator provides to control the instantiation of VMs. We call this interface the Cloud Controller. By means of the Cloud Controller the Job Manager can allocate or deallocate VMs according to the current job execution phase. We will comply with common Cloud computing terminology and refer to these VMs as instances for the remainder of this paper. The term instance type will be used to differentiate between VMs with different hardware characteristics. E.g., the instance type "m1.small" could denote VMs with one CPU core, one GB of RAM, and a 128 GB disk while the instance type "c1. xlarge" could refer to machines with 8 CPU cores, 18 GB RAM,and a 512 GB disk.

The actual execution of tasks which a Nephele job consists of is carried out by a set of instances. Each instance runs a so-called Task Manager (TM). A

Task Manager receives one or more tasks from the JobMan- ager at a time, executes them, and after thatinforms the Job Manager about their completion or possible errors. Unless a job is submitted to the Job Manager, we expect the set of instances (and hence the set of Task Managers) to be empty. Upon job reception the Job Manager then decides, depending on the job's particular tasks, how many and what type of instances the job should be executed on, and when the respective instances must be allocated/deallocated to ensure a continuous but costefficient processing. Our current strategies for these decisions are highlighted at the end of this section. The newly allocated instances boot up with a previously compiled VM image. The image is configured to automatically start a Task Manager and register it with the Job Manager. Once all the necessary Task
Managers have successfully contacted the Job Manager, it triggers the execution of the scheduled job.

After having specified the code for the particular tasks of the job, the user must define the DAG to connect these tasks. We call this DAG the Job Graph. The Job Graph maps each task to a vertex and determines the communication paths between them. The number of a vertex's incoming and outgoing edges must thereby comply with the number of input and output gates defined inside the tasks. In addition to the task to execute, input and output vertices (i.e. vertices with either no incoming or outgoing edge) can be associated with a URL pointing to external storage facilities to read or write input or output data, respectively. Fig. 2, illustrates the simplest possible Job Graph. It only consists of one input, one task, and one output vertex. Fig



1. Number of Subtasks

A developer can declare his task to be suitable for parallelization. Users that include such tasks in their Job Graph can specify how many parallel subtasks Nephele should split the respective task into at runtime. Subtasks execute the same task code, however, they typically process different fragments of the data.

2. Number of Subtasks Per Instance

By default each subtask is assigned to a separate instance. In case several subtasks are supposed to share the same instance, the user can provide a corresponding an- notation with the respective task.

3. Sharing Instances Between Tasks

Subtasks of different tasks are usually assigned to different (sets of) instances unless prevented by another scheduling restriction. If a set of instances should be shared between different tasks the
user can attach a cor- responding annotation to the Job Graph.

4. Channel Types

For each edge connecting two vertices the user can determine a channel type. Before executing a job, Nephele requires all edges of the original Job Graph to be replaced by at least one channel of a specific type. The channel type dictates how records are transported from one subtask to another at runtime. Currently, Nephele supports network, file, and in-memory channels. The choice of the channel type can have several implications on the entire job schedule. A more detailled disussion on this is provided
in the next subsection.

IV. EVALUATION

In this section we want to present first performance results of Nephele and compare them to the data processing framework Hadoop. We have chosen Hadoop as our competitor, because it is an open source software and currently enjoys high popularity in the data processing community. We are aware that Hadoop has been designed to run on a very large number of nodes (i.e. several thousand nodes). However, according to our observations, the software is typically used with significantly fewer instances in current IaaS clouds. In fact, Amazon itself limits the number of available instances for their MapReduce service to 20 unless the respective customer passes an extended registration process [2]. The challenge for both frameworks consists of two

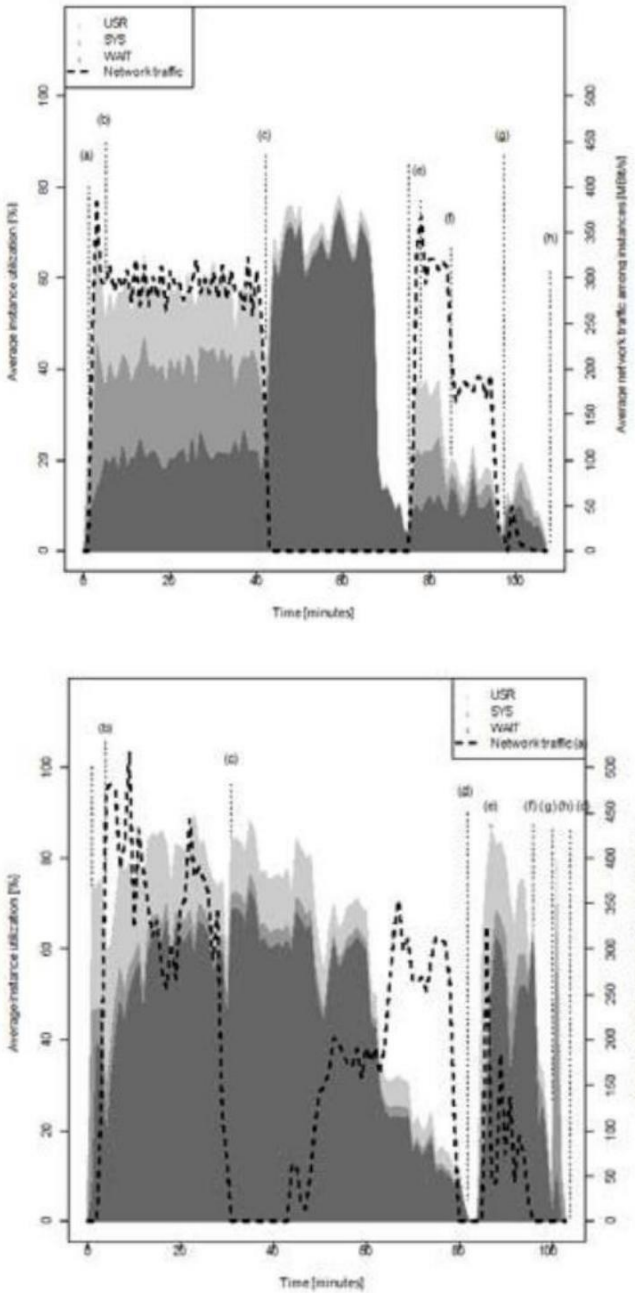
abstract tasks: Given a set of random integer numbers, the first task is to determine the k smallest of those numbers. The second task subsequently is to calculate the average of these k smallest numbers. The job is a classic representative for a variety of data

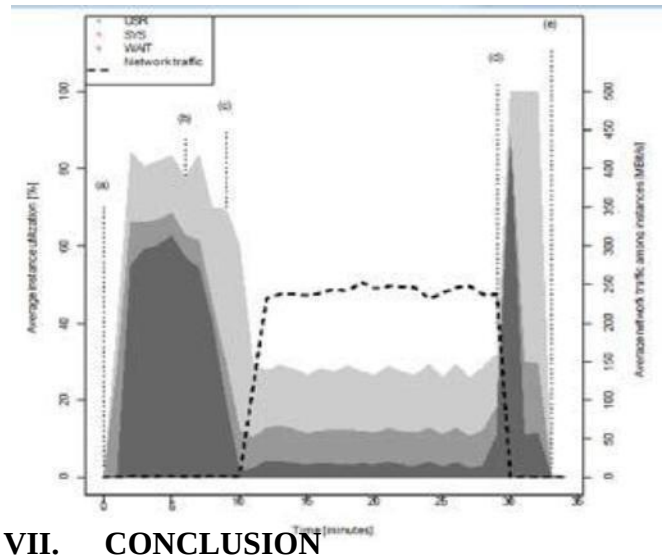
V. RELATED WORK

In recent years a variety of systems to facilitate MTC has been developed. Although these systems typically share common goals (e.g. to hide issues of parallelism or fault tolerance), they aim at different fields of application. MapReduce [9] (or the open source version Hadoop [25]) is designed to run data analysis jobs on a large amount of data,which is expected to be stored across a large set of share-nothing commodity servers. MapReduce is highlighted by its simplicity:Once a user has fit his program into the required map and reducepattern, the execution framework takes care of splitting the job into subtasks, distributing and executing them. A single MapReduce job always consists of a distinct map and reduce program. However, several systems have been introduced to coordinate the execution of a sequence of MapReduce jobs [17,19]. MapReduce has been clearly designed for large static clusters. Although it can deal with sporadic node failures, the available compute resources are essentially considered to be a fixed set of homogeneous machines. The Pegasus framework by Deelman et al. [10] has

VI. RESULTS

show the performance results of our three experiment, respectively. All three plots illustrate the average instance utilization over time, i.e. the average utilization of all CPU cores in all instances allocated for the job at the given pointin time. The utilization of each instance has been monitored with the Unix command "top" and is broken down into the amount of time the CPU cores spent running the respective data processing framework (USR), the kernel and its processes (SYS), and the time waiting for I/O to complete (WAIT). In order to illustrate the impact of network communication, the plots additionally show the average amount of IP traffic flowing between the instances over time. We begin with discussing Experiment 1 (MapReduce and Hadoop): For the first MapReduce job, TeraSort, fig. 7, shows a fair resource utilization. During the map (point (a) to (c)) and reduce phase (point (b) to (d)) the overall system utilization ranges from 60 to 80%. This is reasonable since we configured Hadoop's MapReduce engine to perform best for this kind of task. For the following two MapReduce jobs, however, the allocated instances are oversized: The second job, whose map and reduce phases range from point (d) to (f) and point (e) to (g), respectively, can only utilize about one third of the available CPU capacity. The third job (running between point (g) and (h)) can only consume
about 10 % of the overall resources.





VII. CONCLUSION

In this paper we have discussed the challenges and opportunities for efficient parallel data processing in cloud environments and presented Nephele, the first data processing framework to exploit the dynamicresource provisioning offered by today's IaaS clouds. We have described Nephele's basic architecture and presented a performance comparison to the well-established data processing framework Hadoop. The performance evaluation gives a first impression on how the ability to assign specific virtual machine types to specific tasks of a processing job, as well as the possibility to automatically allocate/deallocate virtual machines in the course.

VIII. REFERENCES:

[1] Amazon Web Services LLC,"Amazon Elastic Compute Cloud", (Amazon EC2). [Online] Available: http://www. aws.amazon.com/ec2/, 2009.

[2] Amazon Web Services LLC,"Amazon Elastic Map Reduce", [Online] Available: http://www.aws.amazon.com/ elasticmapreduce/, 2009.

[3] Amazon Web Services LLC,"Amazon Simple Storage Service", [Online] Available: http: //www.aws.amazon.com/ s3/, 2009.

[4] D. Battr´ S. Ewen, F. Hueske, O. Kao, V. Markl, D. Warneke. e,"Nephele/PACTs: A Programming Model and Execution Frame- work for Web-Scale Analytical Processing", In SoCC '10: Proceed- ings of the ACM Efficient Parallel Processing of Massive Data Sets", Proc. VLDB Endow., 1(2):1265-1276, 2008.

[5] H. chih Yang, A. Dasdan, R.-L. Hsiao, D. S. Parker. "MapReduce-Merge: Simplified Relational Data Processing on Large Clusters", In SIGMOD '07: Proceedings of the 2007 ACM SIGMOD international conference on Management of data, pp. 1029-1040, New York, NY, USA, 2007. ACM.

proceedings of International Conference on Emerging Technologies in Computer Science (ICETCS) ISBN 978-93-85101-62-5

Strongly Providing Security in Multi-cloud Computing Environments Framework

Dr.Sunil Tekale, Malla Reddy College of Engineering ,Secunderabad-100.
Mr.CH.Vengaiah, Malla Reddy College of Engineering ,Secunderabad-100.

ABSTRACT

A proposed proxy-based multicloud computing framework allows dynamic, on the fly collaborations and resource sharing among cloud-based services, addressing trust, policy, and privacy issues without reestablished collaboration agreements or standardized interfaces. The recent surge in cloud computing arises from its ability to provide software, infrastructure, and platform services without requiring large investments or expenses to manage and operate them. Clouds typically involve service providers, infrastructure/resource providers, and service users (or clients). They include applications delivered as services, as well as the hardware and software systems providing these services. Cloud computing characteristics include a ubiquitous (network-based) access channel; resource pooling; multitenancy;automatic and elastic provisioning and release of computing capabilities; and metering of resource usage (typically on a pay-per-use basis). Virtualization of resources such as processors, network, memory, and storage ensures scalability and high availability of computing capabilities. Clouds can dynamically provision these virtual resources to hosted applications or to clients that use them to develop their own applications or to store data. Rapid provisioning and dynamic reconfiguration of resources help cope with variable demand and ensure optimum resource utilization. A proposed proxy-based multicloud computing framework allows dynamic, on- the-fly collaborations and resource sharing among cloud-based services, addressing trust, policy, and privacy issues without pre-established collaboration agreements or standardized interfaces.

INTRODUCTION

Multicloud applications, including the following: Increase in the attack surface due to system complexity. Loss of client's control over resources and data due to asset migration. Threats that target exposed interfaces due to data storage in public domains, and Data privacy concerns due to multitenancy. The research community is beginning to develop architectures, technologies, and standards to support collaboration among multiple cloud systems. Some specific security issues associated with collaboration among heterogeneous clouds include establishing The recent surge in cloud computing arises from its ability to provide software, infrastructure, and platform services without requiring large investments or expenses to manage and operate As more organizations adopt cloud computing, cloud service providers (CSPs) are developing new technologies to enhance the cloud's capabilities. Cloud mashups are a recent trend; mashups combine services from multiple ate them. Clouds typically involve service providers, Infrastructure/resource providers, and service users (or clients). They include applications delivered as services, as well as the hardware and software systems providing these services. Cloud computing characteristics include a ubiquitous (network-based) access channel; resource pooling; Multi- tenancy; automatic and elastic provisioning and release of computing capabilities; and metering of resource usage (typically on a pay-per-use basis).1 Virtualization of resources such as processors, network, memory, and storage ensures scalability and high availability of computing capabilities. Clouds can dynamically provision these virtual resources to hosted applications or to clients that use them to develop their own applications or to store data. Rapid provisioning and dynamic reconfiguration of resources help cope with variable demand and ensure optimum resource utilization. Clouds into a single service or application, possibly with on-premises (client-side) data and services. This service composition lets CSPs offer new functionalities to clients at lower development Costs. Examples of cloud mashups and technologies to support them include the following:

- IBM's Mash up Center, a platform for rapid creation, sharing, and discovery of reusable application building blocks (like widgets and feeds) with tools to easily assemble them into new Web applications.
- Appirio Cloud Storage, a cloud-based storage service that lets Salesforce.com cloud customers store information about accounts, opportunities, and so on in the Amazon S3 cloud.
- Force.com for the Google App Engine, a set of libraries that enable development of Web and business applications using resources in the Salesforce.com and Google clouds.

Collaboration among multiple cloud-based services, like cloud mashups, opens up opportunities for CSPs to offer more-sophisticated services that will benefit the next generation of clients.

For example, cloud-based electronic medical record (EMR) management systems like Practice Fusion, Verizon Health Information Exchange, Med scribbler, and GE Healthcare Centricity Advance are emerging. In addition, government agencies are working toward building interoperable healthcare information systems that promote electronic exchange of data across multiple organizations. These developments will influence healthcare providers to interact with multiple cloud-based EMR systems in the future. Today, cloud mashups require preestablished agreements among providers as well as the use of custom-built, proprietary tools that combine services through low-level, tightly controlled and constraining integration techniques. This approach to building new collaborative services does not support agility, flexibility, and openness. Realizing multi- cloud collaboration's full potential will require implicit, transparent, universal, and on-the-fly interaction involving different services spread across multiple clouds

that lack pre- established agreements and proprietary collaboration tools. As the name suggests, provider-centric approaches require CSPs to adopt and implement the changes that facilitate collaboration—changes such as standardized interfaces, protocols, formats, and other specifications, as well as new architectural and infrastructure components. Without these provider-centric changes, current proposals do not provide facilities for client-centric, on-the-fly, and opportunistic combinations of heterogeneous cloud-based services. While cloud standardization will promote collaboration, there are several hurdles to its adoption.6, 7 from a market perspective; it is unlikely that multiple CSPs will agree on an easy and standardized way to access services, as this would give clients total freedom in changing providers, leading to increased open and direct competition with other providers. CSPs often offer differentiated services with specialized proprietary products and services. Standardization also reduces the efficacy of CSPs that use such differentiated service offerings to attract and maintain more clients. For cloud collaboration to be viable in the current environment, researchers need to develop mechanisms that allow opportunistic collaboration among services without requiring standards and extensive changes to the cloud service delivery model. This approach will allow incremental provisioning of collaborative services to clients, which will continue to improve as more cloud services become interoperable in the future Cloud-based computing also introduces new security concerns that affect collaboration across trust among different cloud providers to encourage collaboration; addressing policy heterogeneity among multiple clouds so that composite services will include effective monitoring of policy anomalies to minimize security breaches; and maintaining privacy of data and identity during collaboration. Mechanisms for collaboration across multiple clouds must undergo a rigorous, in-depth security analysis to identify new threats and concerns resulting from collaboration. They must have the support of innovative, systematic, and usable mechanisms that provide effective security for data and applications. Such security mechanisms are essential for gaining the trust of the general public and organizations in adopting this new paradigm.

EXISTING SYSTEM:

Many existing cloud data services provide similar access control models, in which individual and organizational privacy, a key requirement for digital identity management, is unprotected. Also, with cloud computing initiatives, the scope of insider threats, a major source of data theft and privacy breaches, is no longer limited to the organizational perimeter. Multicloud environments exacerbate these issues because proxies can access data (which the environment might dynamically move or partition across different clouds) on behalf of clients. Revealing sensitive information in identity attributes to proxies that grant them authorization to access the data on behalf of clients is not an attractive solution. Thus, assuring the private and consistent management of information relevant to ABAC becomes more complex in multicloud systems.

PROPOSED SYSTEM:

Our proposed framework for generic cloud collaboration allows clients and cloud applications to simultaneously use services from and route data among multiple clouds. This framework supports universal and dynamic collaboration in a multicloud system. It lets clients simultaneously use services from multiple clouds without prior business agreements among cloud providers, and without adopting common standards and specifications. As more organizations adopt cloud computing, cloud service providers (CSPs) are developing

new technologies to enhance the cloud's are a recent trend; mashups combine services from multiple clouds into a single service or application, possibly with on-premises (client-side) data and services. This service composition lets CSPs offer new functionalities to clients at lower development costs.

MODULE DESCRIPTION:

Number of Modules

After careful analysis the system has been identified to have the following modules:

1. Collaboration Framework for Multicloud System Module.
2. Client/Users Module.
3. Cloud Service Provider Module.
4. Proxy Service Provider Module.

1. Collaboration Framework for Multicloud System Module:

Cloud collaboration allows clients and cloud applications to simultaneously use services from and route data among multiple clouds. This framework supports universal and dynamic collaboration in a multicloud system. It lets clients simultaneously use services from multiple clouds without prior business agreements among cloud providers, and without adopting common standards and specifications.

2. Client/Users Module:

Client sends a request to cloud C1, which dynamically discovers the need to use

Services from clouds C2 and C3. C1 employs proxies to manage these interactions. A client that wishes to simultaneously use services from multiple clouds must individually interact with each cloud service, gather intermediate results, process the collective data, and generate final results. Proxies can facilitate collaboration without requiring prior agreements between the cloud service providers. First, the requesting entity chooses proxies to act on its behalf and to interact with cloud applications. A client or a CSP might employ multiple proxies to interact with multiple CSPs. It can select proxies based on, for example, latencies between proxies and capabilities. Cloud mashups clouds or workload conditions at various proxies.

3. Cloud Service Provider Module:

Cloud service providers (CSPs) deploy proxies as an autonomous cloud system and offer it as a service to clients. A client employs two proxies to interact with CSPs C1 and C2. Alternatively, a client initiates a service request with C1, which then discovers the need for a service from C2. PSP: proxy service provider. Clients deploy proxies within the infrastructure of their organization. A client employs two proxies to interact with CSPs C1 and C2. A client initiates a service request with C1, which then discovers the need for a service from C2.

4. Proxy Service Provider Module:

It involves deploying proxies as an autonomous cloud that offers collaborative services to clients and CSPs. A group of CSPs that are willing to collaborate can manage this proxy-as-a-service cloud, or a third-party entity, a proxy service provider (PSP), can provide management. Clients directly subscribe to the proxy cloud service and employ them for inter cloud collaboration. To protect data at rest and data in transit, proxies must provide a trusted computing platform that prevents malicious software from taking control and compromising sensitive client and cloud application data.

COLLABORATION FRAMEWORK FOR MULTICLOUD SYSTEMS

Our proposed framework for generic cloud collaboration allows clients and cloud applications to simultaneously use services from and route data among multiple clouds. This framework supports universal and dynamic collaboration in a multicloud system. It lets clients simultaneously use services from multiple clouds without prior business agreements among cloud providers, and without adopting common standards and specifications.

RESEARCH FEATURES:

Use of proxies for collaboration

In the current environment, a client that wishes to simultaneously use services from multiple clouds must individually interact with each cloud service, gather inter- mediate results, process the collective data, and generate final results. The following restrictions in the current cloud computing model prevent direct collaboration among applications hosted by different clouds: **Heterogeneity and tight coupling.** Clouds implement proprietary interfaces for service access, configuration, and management as well as for interaction with other cloud components. Each service layer of a cloud tightly integrates with lower service layers or is highly dependent on the value-added proprietary solutions that the cloud offers. This heterogeneity and tight coupling prohibit interoperation between services from different clouds. Proxies can facilitate collaboration without requiring prior agreements between the cloud service providers. Preestablished business agreements. The current business model requires preestablished agreements between CSPs before collaboration can occur. These agreements are necessary for clouds to establish their willingness to collaborate and establish trust with one another. The lack of such agreements prohibits multicloud collaborative efforts due to incompatible intentions, business rules, and policies. Moreover, collaborations resulting from preestablished agreements typically exhibit tight integration between the participants and cannot be extended to provide universal and dynamic collaboration. Service delivery model. Clouds use a service delivery model that provides service access to legitimate sub- scribing clients and denies all other requests because of security and privacy concerns. This prevents direct interaction between services from different clouds. Also, CSPs typically package their service offerings with other resources and services. This results in a tight dependency of a service on the hosting CSP. Such a service delivery model limits a client's ability to customize a service and use it in combination with service offerings from different CSPs.

A technique that could overcome these restrictions uses a network of proxies. A proxy is an edge-node-hosted soft- ware instance that a client or a CSP can delegate to carry out operations on its behalf. Depending on the context, the system can regard a network of proxies as a collection of virtual software instances connected via a virtual network or a set of physical nodes connected via an underlying net- work infrastructure. The basic idea is to enable proxies that act on behalf of a subscribing client or a cloud to provide a diverse set of functionalities: cloud service interaction on behalf of a client, data processing using a rich set of operations, caching of intermediate results, and routing, among others. With these additional functionalities, proxies can act as mediators for collaboration among services on different clouds. Proxy deployment can be strategic—in close geographical proximity to the clouds, for example—to improve performance and facilitate execution of long - lived applications without additional user intervention. As an example of proxy-facilitated collaboration between clouds, consider a case in which a client or CSP wishes to simultaneously use a collection of services that multiple clouds offer. First, the requesting entity chooses proxies to act on its behalf and to interact with cloud applications. A client or a CSP might employ multiple proxies to interact with multiple CSPs. It can select proxies based on, for example, latencies between proxies and clouds or workload conditions at various proxies. Once it chooses proxies, the client or CSP delegates the necessary service-specific privileges to the proxies to carry out the service request using the necessary security precautions. These proxies can further delegate to other proxies if necessary and initiate the service request. In some in- stances, clients or CSPs can assign special roles to one or more proxies in the network to coordinate the operations in a service request among the multiple delegate proxies. Following delegation, the requesting

entity need not further interact with the proxy network until the proxies complete the service request.

During execution of a service request, proxies would interact with cloud-based applications, collaboration without requiring prior agreements between the CSPs. Proxies can also perform operations to help overcome incompatibilities among services to allow data exchange between them.

Architectural overview

Clouds consist of multiple network-connected resource clusters such as server farms, data warehouses, and so on that host geographically distributed virtual machines and storage components that ensure scalability, reliability, and high availability. A multicloud system that employs proxies for collaboration consists of three architectural components: multiple cloud computing systems, networks of proxies, and clients (or service users). Such systems can use several possible strategies for placing proxies in the proxy network.

Cloud-hosted proxy.

As Figure 1 shows, each CSP can host proxies within its cloud infrastructure, manage all proxies within its administrative domain, and handle service requests from clients that wish to use those proxies for collaboration. The proxy instances might need to be CSP- specific. For example, in Figure 1, both C1 and C2 might mutually and dynamically provision sharing and collaboration logic as proxy virtual instances within their respective administrative domains.

Proxy as a service:

As Figure 2 shows, this scenario involves s deploying proxies as an autonomous cloud that offers collaborative services to clients and CSPs. A group of CSPs that are willing to collaborate can manage this proxy-as-a-service cloud, or a third-party entity, a proxy ser- vice provider (PSP), can provide management. playing the role of the service subscriber(s). By independently requesting services from the clouds, and by routing data between each other in a manner transparent to cloud applications, proxies can facilitate Clients directly sub- scribe to the proxy cloud service and employ them for inter cloud collaboration

Peer-to-peer proxy

Proxies can also interact in a peer-to-peer network managed by either a PSP or a group of CSPs that wish to collaborate. Another possibility is for proxies to have no collective management: each proxy in the peer-to-peer network is an independent entity that manages itself. In this case, the proxy itself must handle requests to use its services

On-premise proxy

In the scenario shown in Figure 3, a client can host proxies within its organization infrastructure and manage all proxies within its administrative domain. A client that wishes to use proxies for collaboration will employ its on-premises proxies, whereas CSPs that wish to collaborate with other CSPs must employ proxies that are within the domain of the service-requestingclient

Figure 1. Client sends a request to cloud C1, which dynamically discovers the need to use Services from clouds C2 and C3. C1 employs proxies to manage these interactions.

Client

.

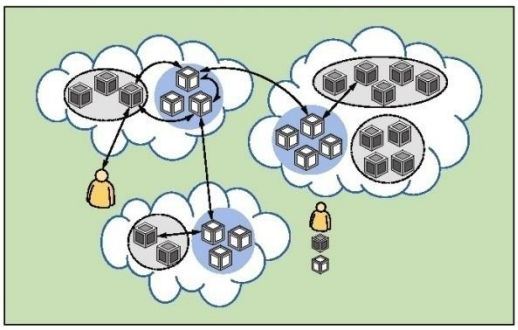


Figure 2. Proxy as a service. In this scenario, cloud service providers (CSPs) deploy proxies as an autonomous cloud system and offer it as a service to clients. (a) A client employs two proxies to interact with CSPs C1 and C2. (b) Alternatively, a client initiates a service request with C1, which then discovers the need for a service from C2. PSP: proxy service provider.Client

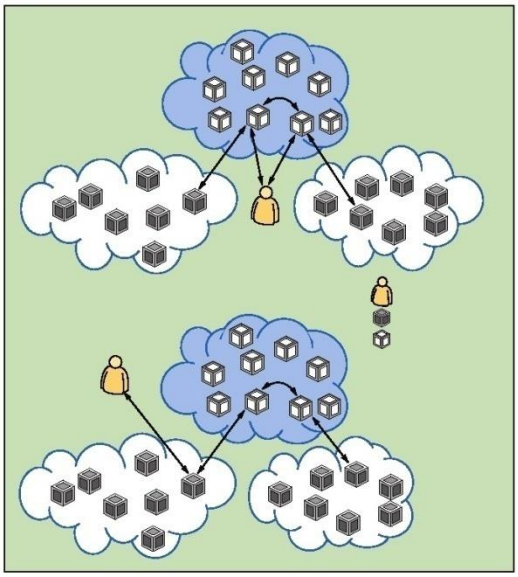
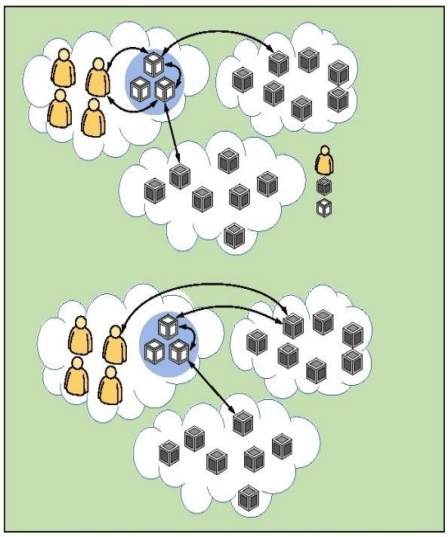


Figure 3. On-premises proxy. Clients deploy proxies within the infrastructure of their Organization. (a) A client employs two proxies to interact with CSPs C1 and C2. (b) A client initiates a service request with C1, which then discovers the need for a service from C2. Client Service VM instance Proxy VM instance C1, C2 Cloud service providers Service VM instance Proxy VM instance C1, C2 Cloud service providers PSP Proxy service provider



Hybrid proxy infrastructure.

A hybrid infrastructure can include on-premises, CSP- and PSP-maintained, and peer-to-peer proxies. Selecting proxies for collaboration will depend on the type of service being requested and the entity that initiates collaboration, among other factors. For example, clients that must initiate a service request with two CSPs can employ on-premises proxies for collaboration. On the other hand, a cloud-based application that discovers it needs a service from another CSP to fulfill a client's request can employ a CSP-maintained proxy.

The proposed architectures illustrate the various options that are available for deploying proxies to support collaboration. Developing these architectures serves as the first step in building a proxy-based, collaborative, multicloud computing environment. A complete solution will entail several additional tasks. For example, an important task is a comprehensive study and evaluation of the proposed proxy-based architectures. Such an evaluation must cover each architecture's possible variations under diverse practical use cases and scenarios for multicloud collaboration. Based on this study, researchers can refine the proposed architectures, develop new variations to support different scenarios and use cases, and, if possible, merge the architectures into a universal proxy-based architecture for multicloud collaboration.

Another important task is developing a full suite of protocols and mechanisms that proxies must implement to support all the functionalities necessary for acting as mediators among services from multiple clouds. For example, supporting collaboration scenarios that migrate a client-subscribed virtual machine from one cloud to another re- quires techniques for translation between various virtual machine packages and distribution formats.

SECURITY ISSUES IN MULTICLOUD COLLABORATION

Researchers and industry specialist s have highlighted several security issues in cloud computing, including isolation management, data exposure and confidentiality, virtual OS security, trust compliance, and mission assurance.8 Specific security issues emerge during dynamic sharing and collaboration across multiple clouds. In particular, issues pertaining to trust, policy, and privacy are a concern in multicloud computing environments

Establishing trust and secure delegation

As in other IT systems, security in clouds relies heavily on establishing trust relationships among the involved entities. The need for trust arises because a client relinquishes direct control of its assets' security and privacy to a CSP. Doing so exposes a client's assets to new risks that are otherwise lessened or avoidable in an internal organization. These risks include insider security threats, weakening of data ownership rights, transitive trust issues with third-party providers in composite cloud services, and diminished oversight of system security.8 A client must confer a high level of trust to a CSP with regard to its ability to implement effective controls and processes to secure assets. Thus, a client must be able to accept the higher levels of risk in using cloud-based services. Using proxies moves the trust boundary one step further: clients and CSPs now must establish trust relationships with proxies, which includes accepting a proxy's security, reliability, availability, and business continuity guarantees. Moreover, CSPs responding to service requests that a proxy makes on behalf of a client or another CSP must trust the proxy to legitimately act on behalf of the requesting entity. Establishing a trust relationship with proxies depends on the strategy used to establish, manage, and administer the proxy network. The entity managing the proxies must provide guarantees of its own trustworthy operation; additionally, it must provide assurances of the proxies' security, reliability, and availability. From the client's point of view, employing on-premises proxies that are within the client's administrative domain can exacerbate trust issues. By using on-premises proxies, a client maintains control over its assets while proxies process them during a collaborative service request. Similarly, using proxies within the CSP's administrative domain lets the CSP exercise control over the proxies' operations, and thus it can trust the proxies to enable collaboration. Proxy networks are a potential platform for developing proxy-based security architectures and solutions for multicloud systems. At a minimum, the proxy network must implement security and privacy mechanisms that mirror, extend, or complement similar mechanisms offered by clouds8 to maintain asset protection outside the domain of clouds and client organizations. For example, to protect data at rest and data in transit, proxies must pro- vide a trusted computing platform that prevents malicious software from taking control and compromising sensitive client and cloud application data? They must also

guarantee data confidentiality and integrity during transmission through the proxy network, possibly by using standards such as the Transport Layer Security protocol. In addition, clients, clouds, and proxies must implement mechanisms that ensure secure delegation, which entails the following:

• **On-the-fly agreements.** Delegating to a proxy must establish, on the fly, an explicit agreement between the delegator and proxy that lets the proxy act on the delegator's behalf. Techniques for delegation to a proxy must include mechanisms that restrict the proxy's behavior, including data and resource access, to comply with delegator-specified constraints.

• **Expected behavior.** After delegation, a proxy must not deviate from the expected behavior. It must act only on behalf of the delegator (a client or a CSP). After the proxy fulfills the service request, it can no longer act on the delegator's behalf. The proxy cannot modify the intended service request or misuse client assets, and it must not transitively delegate its capabilities to other proxies without the delegator's explicit consent.

The technologies for secure delegation to proxies include the use of warrant-based proxy signatures for delegation of signing rights to provide authentication of proxies.9 Simple

Public-key infrastructure authorization certificates10 or the OAuth protocol can facilitate the secure delegation of access rights and permissions. Researchers must thoroughly evaluate existing secure-delegation techniques and develop a comprehensive protocol suite that implements mechanisms that support secure delegation and proxy operation.

Policy heterogeneity and conflicts:

When proxies enable dynamic collaboration between multiple CSPs, heterogeneous security policies can be the source of policy conflicts that result in security breaches. Proxies must monitor for and defend against such breaches. Even though existing policy evaluation mechanisms can verify individual domain policies, security violations can easily occur during integration.12 In multicloud collaborations using proxies, service requirements can drive dynamic, transient, and intensive interactions among different administrative domains. Thus, a proxy's policy integration tasks must address challenges such as semantic To protect data at rest and data in transit, proxies must provide a trusted computing platform that prevents malicious software from taking control and compromising sensitive client and cloud application data. Heterogeneity, secure interoperability, and policy evolution management. The design of access control policies for multicloud collaboration must permit careful management by proxies while ensuring that policy integration does not lead to security breaches. Policy analysis generally includes property verification and conflict detection, as well as an analysis of the differences between policy versions. For a collaborative service, a proxy must deal with several registered services from multiple clouds as well as proxies. This requires various proxies to locally conduct policy integration and decomposition.

Policy integration aims to generate agreement on access rights for each party involved in a collaborative project. A policy integration process for inter cloud collaboration must systematically handle potential conflicts and resolution problems. Proxies must analyze relationships between policies to detect and resolve policy anomalies using mechanisms that easily adapt to handle composite policies evaluated as a whole. Possible policy anomalies include policy inconsistencies and inefficiencies.

Policy inconsistency. Access control policies reflect security requirements, which should be consistent within and across multiple participating parties to accommodate the dynamic and complex nature of multicloud environments. Policy inconsistencies can result in security and availability problems; they include the following: *Contradiction*. Two policies are contradictory if they Have different effects on the same subjects, targets, and conditions. Contradictions are the most common form of policy conflicts • Exception. A policy is an exception of another policy if they have different effects, but one policy is a subset of the other. The exception might not be a policy conflict, but access policy evaluation mechanisms commonly use exceptions to exclude a specific access request from general access permission.

• Correlation. Two policies are correlated if they have different effects but intersect each other. In this case, one policy permits the intersection, but the other does not. This is a partial policy conflict.

Policy inefficiency: The composition of policies from multiple origins can result in a large collection of policies controlling the access to federated applications in multi- clouds. Since an access request's response time largely depends on the number of policies that proxies must parse, inefficiencies in composite policies can adversely affect performance. Inefficiencies in composite policies include •

Redundancy—a policy is redundant if every access request that matches the policy also matches another policy with the same effect; and *verbosity*—similar to data element merging in data integration, policy composition can merge similar

policies from different origins; resolving the policy verbosity during composition affects the policy size.

With cloud computing initiatives, the scope of insider threats, a major source of data theft and privacy breaches, is no longer limited to the organizational perimeter.

Once proxies identify conflicts, they must use conflict resolution strategies to resolve them. However, current conflict resolution mechanisms have limitations. For ex- ample, existing conflict resolution mechanisms—including Extensible Access Control Markup Language (XACML) policies—are too restrictive because they only allow the selection of one resolution algorithm to resolve all the identified conflicts. Multicloud environments require adaptively applying different algorithms to resolve different conflicts. It is therefore necessary to develop a flexible and extensible conflict resolution approach to achieve fine-grained conflict resolution. Such an approach must let a proxy automatically map different conflict resolution strategies that resolve different conflicts. Situations in which a policy component becomes involved in multiple conflicts also require a correlation mechanism that identifies dependent relationships among conflicting segments. Such a mechanism ensures that conflict resolution does not introduce new policy violations during the resolution process. Earlier research applied an approach for detecting and resolving policy anomalies to healthcare domains.13, 14 specifically, our preliminary study demonstrates how to achieve compliance and conflict analysis in EMR management systems, as applied to data sharing and Health Insurance Portability and Accountability Act (HIPAA) policies. Use cases, and, if possible, merge the architectures into a universal proxy-based architecture for multicloud collaboration. Another important task is developing a full suite of protocols and mechanisms that proxies must implement to support all the functionalities necessary for acting as mediators among ser- vices from multiple clouds. For example, supporting collaboration scenarios that migrate a client-subscribed virtual machine from one cloud to another requires techniques for translation between various virtual machine packages and distribution formats.

Identity attributes and data privacy

In shared computing environments like clouds, protecting the privacy of client assets is critical.8 the privacy issues pertaining to both data and identity.

Identity attributes privacy. Data as a service (DaaS), such as Amazon S3 and Microsoft Azure, is an emerging cloud service in which organizations can seamlessly store data in the cloud and retrieve it based on access control policies that cover legal requirements and organizational policies. An expressive access control model, such as XACML, can specify access control policies on protected objects in terms of a subject's properties, called identity attributes. These can include a subject's email address, organizational role, age, and location of access. Such an attribute-based access control (ABAC) model provides fine-grained data access and expresses policies closer to organizational policies.

A crucial issue in this context is that identity attributes required by subjects to access protected objects often encode sensitive information. Many existing cloud data services provide similar access control models in which individual and organizational privacy, a key requirement for digital identity management, is unprotected.

Also, with cloud computing initiatives, the scope of insider threats, a major source of data theft and privacy breaches, is no longer limited to the organizational perimeter. Multicloud environments exacerbate these issues because proxies can access data on behalf of clients. Revealing sensitive information in identity attributes to proxies that grant them authorization to access the data on behalf of clients is not an attractive solution. Thus, assuring the private and consistent management of information relevant to ABAC becomes more complex in multicloud systems. In multicloud environments, where proxies use ABAC to retrieve client data from the clouds, clients need to hide their identity attributes from both proxies and CSPs to preserve the privacy of sensitive client information. However, clients must still give proxies the information that grants them access to requested data. This requirement calls for the use of identity attribute and data encoding techniques that, used together, permit oblivious data transfer between CSPs, proxies, and clients while providing privacy-preserving ABAC.

The techniques for encoding client identity attributes must permit clients to transfer the encoded attributes to proxies; the proxies, in turn, must convince CSPs of the ownership and validity of the encoding, without having the client reveal its identity attributes to either entity. Data and identity attribute encoding techniques must ensure that decoding the data is possible when the identity attributes match the ABAC policies, without revealing the attribute to the proxy or the CSP.

Client data privacy. Often, clients must protect data privacy before sharing the data.

Considering example in which multiple medical insurance companies, each of which has a designated CSP, would like to share customer data to have a much larger customer database from which to obtain

useful statistical query results. On e CSP might have an application that requires information on the percentage of male construction workers in the US who are younger than 40 and have respiratory diseases. This would require collecting data from multiple CSPs for the analytical results to be meaningful, since the data from one CSP might be inadequate (after filtering for multiple selective predicates) or atypical (say, one CSP only has data for customers in a particular region of the US).

In this example, the disease attribute of records is sensitive and requires protection when shared among multiple CSPs. Using encryption is not a viable option because maintaining the data's utility is a key requirement for many applications. Most applications require a well-balanced tradeoff between formal privacy and practical utility. Privacy-protection methods (other than encryption) fall broadly into two categories: 15 work. Compression methods such as dictionary encoding can reduce both communication and query processing costs—for example, CSPs and proxies can perform much of the query processing over the encoded format.17 Favorable solutions to ensure data privacy must employ flexible data perturbation methods that provide control over the tradeoff between the privacy guarantee and the utility of the query results. Perturbation methods must pro- vide high accuracy for queries that involve a large number of records. At the same time, they must introduce large amounts of noise in the results for queries over a few re- cords, which is desirable for privacy. In the multiple-CSP context, a CSP can use local data perturbation techniques to perturb its sensitive data and then ship it to another CSP for collaborative query processing. Local techniques permit query processing at one site to avoid on-the-fly data communication costs. Moreover, Favorable solutions to ensure data privacy must employ flexible data perturbation methods that provide control over the tradeoff between the privacy guarantee and the utility of the query results. When a query itself must be private, a CSP can limit query processing to its own site by using local techniques. In some applications, the receiving CSP need not perturb its own sensitive data. These situations present opportunities to further optimize the accuracy and efficiency of query processing that researchers can explore by judiciously determining which CSP should answer a particular query (when queries are not private and shareable). Finally, multicloud scenarios require new privacy definitions that will allow formal proofs of privacy guarantees for protection schemes. Which adds some form of noise to the data itself, and facilitate dynamic collaboration between clouds, we proposed a framework that uses proxies to act as mediators between applications in multiple clouds that output perturbation, which adds noise to the otherwise accurate query answers. Earlier research studied data privacy in outsourcing data aggregation services.16 Regardless of the methods used to maintain data privacy, the resulting solution must scale to use for large amounts of data and many CSPs. Perturbation methods often produce data with high redundancies, which can lead to scalability issues in multi- cloud environments; a client's request for data can result in a large communication overhead in the proxy net- must share data. Our proposed framework has the potential to overcome several restrictions in the current cloud computing model that can prevent dynamic collaboration among applications hosted by different cloud systems.

Future research directions for the proposed framework include refining the proxy deployment scenarios and development of infrastructural and operational components of a multicloud system. This must be accompanied by implementation of an experimental plat for musing open source tools and libraries that work in combination with real-world cloud services to evaluate the system's functionality and limitations, and make further refinements. Currently, our research team is working toward a single viable proxy deployment strategy based on use cases, trust, and security requirements. We are also developing specifications to instantiate, deploy, maintain, and release proxy virtual machines reliably and securely, along with a suite of proxy services to support various collaboration use cases. Our incremental approach to the development of proxy services for collaboration initially provides support for simple use cases, later progressing to more complex use cases.

References:

1. P.Mell and T. Grance, The NIST Definition of Cloud Comput- ing, special publication 800-145, Nat'l Inst. Standards and Technology, 2011, p. iii + 3.
2. D. Bernstein and D. Vij, "Intercloud Security Consider- ations," Proc. 2nd Int'l Conf. Cloud Computing (CloudCom 10), IEEE Press, 2010, pp. 537-544.
3. R. Buyya et al., "Market-Oriented Cloud Computing: Vision, Hype, and Reality of Delivering Computing as the 5th Util- ity," Proc. 9th IEEE/ACM Int'l Symp. Cluster Computing and the Grid (CCGRID 09), IEEE CS, 2009, pp. 599-616.
4. B. Rochwerger et al., "Reservoir—When One Cloud Is Not Enough," Computer, Mar. 2011, pp. 44-51.
5. M.P. Papazoglou and W. van den Heuvel, "Blueprinting the Cloud," IEEE Internet Computing, Nov./Dec 2011, pp. 74-79.
6. S. Ortiz Jr., "The Problem with Cloud Computing Standard- ization," Computer, July 2011, pp. 13-16.
7. P. Mell and T. Grance, "Perspectives on Cloud Computing and Standards, NIST Information Technology Laboratory," Nat'l Inst. Standards and Technology, 2008; http://csrc.nist.gov/groups/SMA/ispab/documents/minutes/2008-12/cloud- computing-standards_ISPAB-Dec2008_P-Mell.pdf.
8. W. Jansen and T. Grance, Guidelines on
9. S. Chandrasekhar et al., "Efficient Proxy Signatures Based on Trapdoor Hash Functions," IET Information Security, Dec. 2010, pp. 322-332.
10. C.M. Ellison et al., SPKI Certificate Theory, IETF RFC 2693, Sept. 1999; www.ietf.org/rfc/rfc2693.txt.
11. E. Hammer-Lahav, ed., The OAuth 1.0 Protocol, IETF RFC 5849, Apr. 2010; http://tools.ietf.org/html/rfc5849.
12. Y. Zhang and J.B.D. Joshi, "Access Control and Trust Management for Emerging Multidomain Environments," Ann. Emerging Research in Information Assurance, Security and Privacy Services, Emerald Group Publishing, 2009, pp. 421-452.
13. J. Jin et al., "Patient-Centric Authorization Framework for Electronic Healthcare Services," Computers & Security, Mar.-May 2011, pp. 116-127.
14. R. Wu, G.J. Ahn, and H. Hu, "Towards HIPAA-Compliant Healthcare Systems," Proc. 2nd ACM Int'l Symp. Health In- formatics (IHI 12), ACM, 2012, pp. 593-602.
15. N.R. Adam and J.C. Wortmann, "Security-Control Methods for Statistical Databases: A Comparative Study," ACM Com- puting Surveys, Mar. 1989, pp. 515-556.

Hypothetical Analysis on The Transitory Characteristics Of EDFA In Optical Fiber Communication

1.Mr.Amarnath .P Asso Professor CSE Malla Reddy College of Engineering
2.Mr.Ch.Mahende Reddyr Assistant Professor CSE Malla Reddy College of Engineering

Abstract:

Erbium laser amplifier has become one of the important components indispensable in optical fiber communication for its high gain, high pumping efficiency, polarization-independent and small crosstalk between signals, etc. The transient characteristic of the EDFA is an inevitable phenomenon based on the mechanism of EDFA amplification by stimulated emission of radiation. This paper focuses on the EDFA transient effects caused by the signal power, pump power , and gain saturation recovery time from the aspects of EDFA transient rate equation、the relation between the signal power and the output power、The relations between pump power and the output power and the transient effect caused by gain recovery time.

Keywords: EDFA, transient characteristics, optical fiber communication

Introduction

In the future, optical fiber communication will occupy the leading position in the communication's industry inevitably as its large capacity, long distance, security and good performance of adaptability. While as a representative, Erbium laser amplifier has become one of the important components indispensable in optical fiber communication for its high gain, high pumping efficiency, polarization-independent and small crosstalk between signals, etc. However when a bunch of light pulses pass through an optical amplifier, the former pulse will have some impact on the amplification behavior of the latter one, even when it is a single light pulse, the leading edge will also affect the amplification behavior of the Trailing edge, which is an inevitable phenomenon based on the mechanism of EDFA amplification by stimulated emission of radiation, known as the transient characteristics of the EDFA.

The main reason causing the transient gain effect is the initial state of stimulated emission and the temporal correlation , the leading edge of signal pulse makes a large number of particles on the upper level transit by absorbing energy ,and then supply the lost particles for stimulated radiative transition on the upper level by the two means of transverse relaxation of the particles on the upper level and non-radiative transitions of a large number of particles in the pump energy level, in a sufficiently short period of time, if the number of particles in the upper level can’t be replenished , there will be a gain difference between the former and the latter one of a series of pulses, or even the former and the latter edge of a pulse, causing changes in the output pulse waveform [1].

The main parameters of transient effects of EDFA are almost the same as the one of steady stat , both are associated with the pump power , signal power and noise - related gain of EDFA , the only difference is that the number of particles in the upper and lower levels is not certain value at transient changes, which changes with time, when the signal light passes through the EDFA , the

transient power changes will cause the gain saturation phenomenon of EDFA, which is described by the EDFA gain recovery time. So when we analyze the transient changes of EDFA, we also need to analyze the recovery time of EDFA gain [2]. This paper focuses on the EDFA transient effects caused by the signal power, pump power, and gain saturation recovery time.

EDFA transient rate equation

Assume that the number of particles on the upper level is constant when the amplifier works at the steady state. However, the time-varying response of the number of particles on the upper level must be considered when it comes to the transient state, and the transient rate equation [3] should be adopted to analyze the characteristics of the amplifier. The level population change equation is as follows: (ignoring pump state absorption

2 The relation between the signal power and the output power

The power transmission equation of the signal can be acquired by the rate equation and the power transmission equation of the level system. Replace the optical power by the number of photons, the output light power of the k- beam can be achieved of the fiber. The relationship between the gain with the input power and the length of the fiber is shown in figure 1[4].

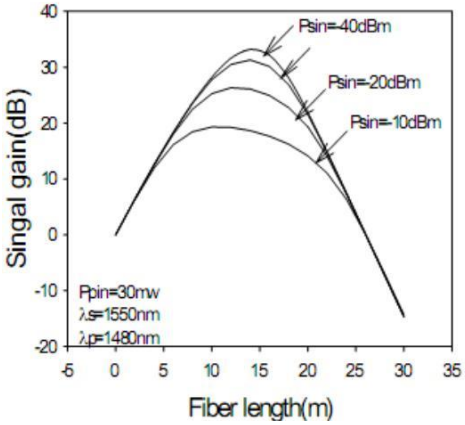


Fig 1 The relationship between the gain with the input power and the length of the fiber

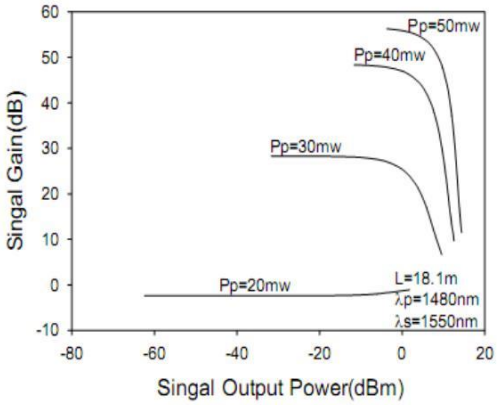


Fig 2. The relationship between the pump power and the signal output power

It can be seen from the figure that the gain is affected by the input power and the length of the fiber. For a given signal power, there is always an optimal fiber length which makes the maximum gain. While for a fixed fiber length, the gain increases as the input signal power [5].

The relations between pump power and the output power

$P_p(L)$ represent the input and output normalized pump power respectively.

According to (14) and the relationship between gain G and the normalized input pump power $P_p(0)$ as well as the length L of the amplifier, it can be known that for a given input signal power ,when changing the pump power, the gain changes as the pump power. However the optical amplification of EDFA is due to the principle of stimulated emission, so the impact on the inversion population from the pump power will also affects EDAF gain [6] . When the input signal power is small , the pump increases, the gain will also increase. Until the pump power reaches a certain level, the gain no longer increases, which is due to the complete saturation of EDFA gain, the number of particles in he lower level are completely reversed at this time , there is little effect on the population inversion when increasing the pump power continuously, so the output power of the

signal light hardly changes with pump light. However when the power of input signal light is larger, the front gain of the signal is larger, while the rear gain may decrease, which is also decided by the inversion population of EDFA, in which case the pump power is increased, the output power increases , which can slow down the gain reduction of the rear end . The relationship between the pump power and the signal output power is shown in Figure 2.

It can be seen from the figure that the amplifier is in the state of small signal amplification when the pump power is small, in which case the signal output power is almost constant. As the pump power increases, the signal output power increases, but the speed for EDFA to enter into saturation also increases, until the pump power increases to a certain value, it can be seen that the output power of the signal is into saturation, which is a fixed value that does not change with the increase of the pump power.

The transient effect caused by gain recovery time

Analysis on the change of EDFA inversion population should be done before the analysis on gain recovery time. Assuming that the input pump is uniform, and the signal light is pulse signal,

the pump rate R_{13} , the stimulated emission rate W_{21} and the spontaneous emission rate taken as constants, it can be obtained by the transient rate equation: so the recovery time is always larger than the saturation time [9] . Therefore the frontier edge of the signal light pulse consume the inversion population, which may be insufficient because there is not enough time to replenish when the trailing edge comes, leading to the reduction of the trailing edge gain ,and a great distortion of the output waveform. Figure 4 shows the distortion waveform for a low frequency signal passing through the EDFA.

Fig 4 The distortion waveform for a low frequency signal

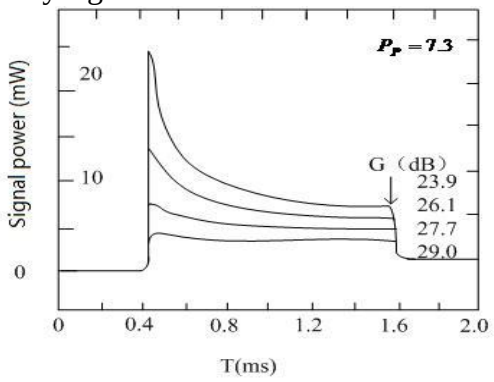


figure above shows that there will be an amplification distortion when a square wave signal pulse of lower frequency passes through the EDFA system. It can be seen that the waveform of the signal light pulse will be deformed in the process of transmission through the EDFA system and amplification. This is because that in the transmission process, the signal has been consuming the particles in the upper levels , leading to that the inversion population is a variable that changes with

time , finally resulting in that the frontier edge gain of the square wave pulse is higher than the trailing edge gain [10] . The size of the signal modulation frequency has some impact on the

inversion population changes, and therefore it can be drawn that the modulation frequency of the signal light is related with the deformation of the pulse. In the case of the recovery time of EDFA gain, it can be analyzed by the change of EDFA output power in Figure 5.

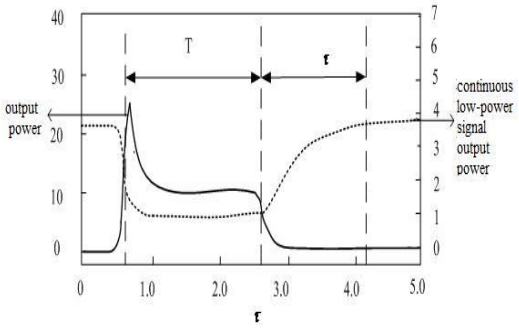


Fig 5 the output power changes of EDFA

The dotted line (continuous low-power signal light) indicates that when the input signal is not coupled with the optical pulse , EDFA is in the steady state , the continuous signal light output power is constant, when coupled with the input signal pulse of larger power, EDFA starts to be in the state of gain saturation , continuous signal output power decreases along. When the input pulse is over, the complementary carriers increase, the gain starts to slowly return to the initial value, which is the small signal gain of no saturation, and the output power is gradually return to stability. The solid line indicates that at the beginning of the entrance of the signal light, the gain is not saturated , the signal output power is high , the gain factor is large, then with the consumption of the inversion population by the input pulse, although by pump supplement, but the inversion population is not enough to compensate for the lost particles used for amplification before, leading to that the number of particles on the upper level decreases, the gain saturation occurs , the output power

decreases until the pulse width $\tau = T$, the input pulse transmission of number is over, the output

power is reduced to zero. Gain recovery time is the duration from the minimum gain value of the pulse end ($\tau = T_s$) to the value when the gain has been restored to the steady state.

We conclude that if the time constant t_{rec} is much smaller than the signal modulation cycle

(modulation $T >> t_{rec}$), that is the signal symbol rate is low , the signal is low-speed long-pulse

signal , the EDFA inversion population change is small, resulting in a weak graphic effects .

When the time constant t_{rec} is much larger than the signal modulation cycle (modulating

T<< t_{rec}), that is the signal symbol rate is higher, for this narrow pulse signal , EDFA do not appear
graphical effects basically.

When the time constant t_{rec} is several times of the period of the modulation signal （t_{rec} ~T）,

the frontier pulse waveform will be distorted and along comes the latter pulse , then the gain is not stabilized , the particles have not been pumped transitions , the gain of the latter pulse is 0 , as time increases , the gain slowly return . A larger effect of graphics generated at this time. The figure of waveform distortion should be as follows [2] Figure 6.

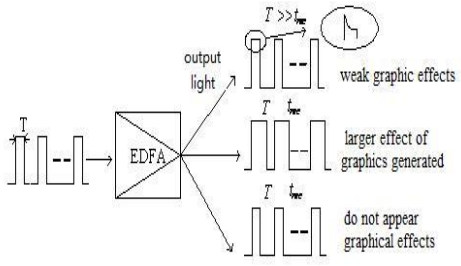


Fig 6 The figure of waveform distortion of EDFA signal for different pulse width

Conclusion

Due to the analysis above, the article determined the relationship between the signal output power, pump power and signal power. Based on the theoretical study and simulation, the article draws the following conclusions: When the pump power is fixed, as the input signal power increases, the gain recovery time increases, the saturation time slightly decreases. When the input signal power is fixed, with the increase of pump power, the gain recovery time and saturation time decreases. Modulation signal pulse width has an effect on the EDFA transient effect, only when the pulse

width T and the gain recovery time is a particular value, the distortion is obvious. When

results hardly any graphic effect; when t_{rec} <<T, it results a weak graphic effects.

References:

[1]Song-Yingxiong, Li-Yingchun. All-Optical Gain-Clamped EDFA with High Gain and Low Noise Factor. JOURNAL OF SHANGHAI UNIVERSITY(NATURAL SCIENCE).2006,12(2)：120-124

[2] Jiang-Nan. Measure of Recovery Time and Research of lain-flattening Technology of Erbium-doped Fiber Amplifier. Beijing Jiaotong University.2006.12

[3]Zhao-Lanlan,Zhu-Yiqing. Pump Power's Impact on the Gain Tilt of C-Band EDFA [J]. Journal of Jiangnan University(Natural Science Edition),2010,9(3): 289-291

[4] Yu-Qiaoyan.Research on Application of EDFA in High Speed Fiber Communication. Wuyi Universtiy.2010

[5]Awaji.Y , Furukawa.H ,Wada.N ,Chan.P ,Man.R et al .Burst-mode EDFA .

[6]Shi-Fengqi . Study of High Concentration Erbium-doped Fiber. Beijing Jiaotong University.2008

[7]Toru Shiozaki, Masaru Fuse, Susumu Morikura. A Study of Gain Dynamics of Erbium-Doped Fiber Amplifiers for Burst Optical Signals . Networkings and switching.2003

[8]P.E.Barnsley,H.J.Wickes,G.E.Wickens et al . All-Optical Clock Recovery from 5 Gb/s RZ Data Using a Self-Pulsating 1.56um Laser Diode.IEEE.1991

[9] G.Talli ,M.J.Adams .Gain recovery acceleration in semiconductor optical amplifiers employing a holding beam .Optics Communications.2005,245:363–370

[10] E.Desurvire,M.Zirngibl et al .Dynamic Gain Compensation in Saturated Erbium-Doped Fiber Amplifiers.IEEE.1991:453-455

Improving Security and Quality of Service (QOS) Desktop Grids

Mr.P.Amarnath Associate Professor MRCE Secunderabad-100.
Mr CH.Vengaiah Assistant Professor MRCE Secunderabad-100.
Mrs.Vijaya Kumari Associate Professor MRCE Secunderabad-100.
Dr. Sunil Tekale Professor MRCE Secunderabad-100.

Abstract:

Simulation can be used to predict the functionality and behavior of the system . Grid computing uses massive power of idle cycles of PC's .Desktop grids is nothing using the idle cycles of desktop PC's for computing large scale applications. There are many fields which requires large scale massive power such as scientific fields to handle complex and demanding problems. In this paper we mainly discuss about different simulator tools available and how they can efficiently provide quality of services on desktop grids. Desktop Grids are being increasingly used as the execution platform for a variety of applications that can be structured as Bag-of-Tasks (BoT)[1].

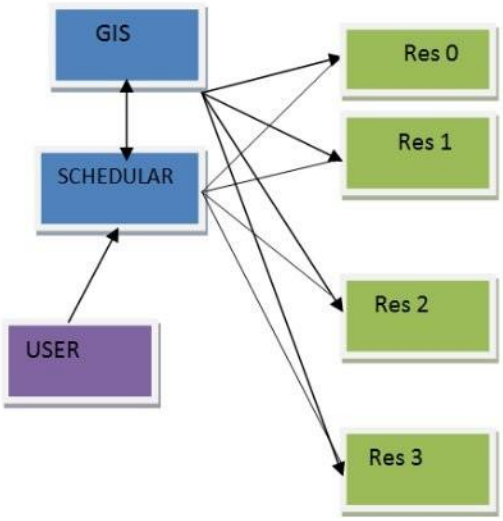
Desktop grids is nothing using the idle cycles of desktop PC's for computing large scale applications. There are many fields which requires large scale massive power such as scientific fields to handle complex and demanding problems. In this paper we mainly discuss about different simulator tools available and how they can efficiently provide quality of services on desktop grids. Desktop Grids are being increasingly used as the execution platform for a variety of applications that can be structured as Bag-of-Tasks

INTRODUCTION

We cannot expect today's human life without computers ,as they became the part of our day to day life. Almost every college, office and every member have computers. In fact modern world is incomplete without Computers.

Desktop Grids are computational grids formed by using resources of idle desktop machines. Most of the computers in offices and personal computers are used only for certain time and are idle for most of the time and also they are not using the whole storage of the system .Grid computing combines all the machines that are idle and form as Virtual Group and uses the group for computing large-scale applications. There are two types of desktop grids one is local and other is individual computers . Local computers are the group of computers in an organization and educational institutes . Second is the individual computers which are used by citizens . This offers the opportunity to resolve the increasing need for computational resources. As most desktop systems are idle for significant periods of time, it should be possible to harvest their idle CPU cycles or other unused resources and apply them towards projects in need of such resources. Apart from providing huge computational power and storage capacity desktop grids also have several challenges in using this volatile and shared platform effectively. The usefulness of desktop grid computing is not limited toMany institutions, ranging from academics to enterprises, hold vast number of desktop machines and could benefit from exploiting the idle cycles of their local machines[2] . Important examples of desktop grids are SETI@home and PrimeGrid, Almere Grid, Condor based grids[4], the WISDOM project is using grid computing to speed the search for a cure for malaria, a disease that effects millions of people all over the developed world, MammoGrid is building a grid for hospitals to share and analyse mammograms in an effort to improve breast cancer treatment.

Desktop grids faces many challenges and the most important challenges are platform is volatile, since users may



reclaim their computer at any time, which makes centralized schedulers inappropriate. Second, desktop grids are likely to be shared among several users, thus we must be particularly careful to ensure a fair sharing of the resources. Fair sharing of resources means there should be balanced share of resource no resource should get more load and no outside of organizational boundaries, it becomes increasingly difficult to guarantee that a resource being used is not malicious in some way[3].PROBLEM DESCRIPTION

In this section, we formally define the problem we target. Our goal is to design a high efficiency secure and fault-tolerant for fair scheduler with quality of service on desktop grid. Our main objective is to provide quality of services and security along with grid simulators. There is more chance of occurring of resources failures in grid computing because grids are in distributed environment. There is possibility of data loss or packet loss while transferring of data from scheduler to resources. There are many security issues such as architecture issues , infrastructure issues and management issues[11]. In the desktop grids it is more because even and individual computer also comes under this category.

ARCHITECTURE

Figure 1:Architecture of Scheduler and Grid Information System with Friewall.

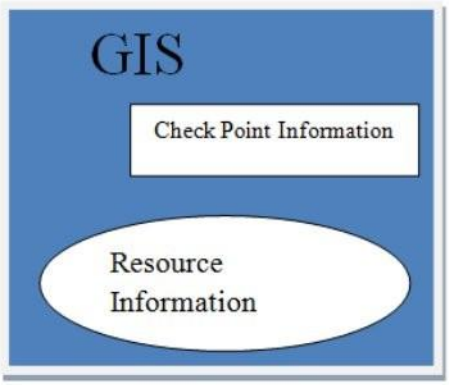


Figure 2: GIS architecture

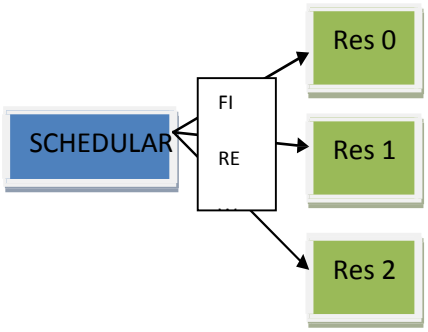


Figure 3: Scheduler with Fire Wall

Algorithm 1 Resource Failure Detection Algorithm used by GIS.

repeat send the messages to all checkpoints up to resource from GIS

if a checkpoint does not respond sending acknowledge back **then** remove it from the list

update GIS entities about the failure

end if

ait for Time GIS seconds**until** simulation is over

Algorithm 2 Resource Failure Detection Algorithm used by Scheduler . Repeat pushthe messages to all resources which are running jobs **if** a resource does not respond by sending acknowledge **then**

Scheduler ask the GIS for a list of resources choose one of them resubmit the jobs.

Architecture of grid is shown in Figure 1, for the same architecture we can add firewalls for each and every flow of data. Firewalls immediately blocks unwanted data flow.

GIS: maintains an up-to-date list of available resources. GIS collects the information from all the checkpoints of all the resources and maintains an up- to-date list.be send to all the checkpoints to all the resources and if message is not send back from the then remove the resource from the resource information maintained in the GIS. Contact a GIS entity for a list of available resources in order to know where to run their jobs. The functionality of this entity can be summarized in Algorithm 2.

[9] (step 4). Hence, *GIS* removes the failed resource from the list. During a routine push, GIS discovers that *R1* has failed. As a result, *scheduler* ask *GIS* for a list of resources (step 5). When *R1* recovers, it registers itself again to *GIS* (step 6). With this approach, *GIS* is able to maintain an up-to-date list of available resources. If the failure only affects some of the machines in a resource, what happens next depends on the allocation policy of this resource. If the resource runs a *space-shared* allocation policy, the jobs that are currently running on the failed machines will be terminated and sent back to users. However, when the resource runs a *time-shared* (round

User Datagram Protocol (UDP) is used by these entities. This is due to the fact that UDP requires a less significant network latency in comparison with a Transmission Control Protocol (TCP), although UDP does not provide retransmission of lost packets. The sequential steps are shown in a box with a number inside. Here is list of steps showing the working of the model .First, *R1,R2,R3* and *R4 resources* register to *GIS* (step 1). Then, *GIS* creates a list of available resources. In order to keep that list up-to-date, *GIS* push the messages to the resources periodically (step 2). When *User* wants to run a job, he/she contacts scheduler and scheduler contacts *GIS* in order to get a list of available resources (step 3). Upon receiving the scheduler's request, *GIS* returns its list. In that moment, scheduler will choose *R1* for example, based on the features of the resource and the job scheduling. When *User* has chosen the resource, he/she submits the job to *R1* and starts a regular pushing mechanism.

In the event of a failure affecting *R1*, *GIS* is able to detect this problem due to the pushing mechanism in place

User Datagram Protocol (UDP) is used by these entities. This is due to the fact that UDP requires a less significant network latency in comparison with a Transmission Control Protocol (TCP), although UDP does not provide retransmission of lost packets. The sequential steps are shown in a box with a number inside. Here is list of steps showing the working of the model .First, *R1,R2,R3* and *R4 resources* register to *GIS* (step 1). Then, *GIS* creates a list of available resources. In order to keep that list up-to-date, *GIS* push the messages to the resources periodically (step 2). When *User* wants to run a job, he/she contacts scheduler and scheduler contacts *GIS* in order to get a list of available resources (step 3). Upon receiving the scheduler's request, *GIS* returns its list. In that moment, scheduler will choose *R1* for example, based on the features of the resource and the job scheduling. When *User* has chosen the resource, he/she submits the job to *R1* and starts a regular pushing mechanism.

policies, the remaining machines are responsible for responding to pushing requests from users and GIS. Moreover, they are required to inform the GIS about such failure. This way, the GIS can have accurate information on the current status of the resource. Simulation Tools Open DSS

Open Dss is a simulation tool used for distribution system, and it stands for open Distribution Simulation Software that was released by EPRI to provide a free, open-source, distribution system simulator.

PowerWorld Simulator Version 16 The PowerWorld Simulator Version 16 is a free software demo of the PowerWorld simulation product.

GridSim

GridSim is one more powerful tool used for the simulation with great features and motivated for experimental implementation of Grid computing.

USING GridSim FOR RESOURCE FAILURES

GridSim:

GridSim allows modeling and simulation of entities in parallel and distributed computing systems such as users, applications, resources, and resource brokers/schedulers for design and evaluation of scheduling algorithms.

Overview of GridSim functionalities:

Incorporates failures of Grid resources during runtime. New allocation policy can be made and integrated into the GridSim Toolkit, by extending from Alloc class. Has the infrastructure or framework to support advance reservation of a grid system. Incorporates a functionality that reads workload traces taken from supercomputers for simulating a realistic grid environment. Incorporates an auction model into GridSim. Incorporates a datagrid extension into GridSim. Incorporates a network extension into GridSim. Now, resources and other entities can be linked in a network topology. Incorporates a background network traffic functionality based on a probabilistic distribution. This is useful for simulating over a public network where the network is congested. Incorporates multiple regional GridInformationService(GIS) entities connected in a network topology. Hence, you can simulate an experiment with multiple Virtual Organizations

(VOs).Adds ant build file to compile GridSim source files[10].

Along with GridSim classes we are using many new classes for implementing resource failures.

UserFailure: as its name suggests, this class implements the behavior of the users of our grid environment. Its functionality can be summarized as follows: (1)creation of jobs; (2) submission of jobs to resources; (3)push the resources used to run its jobs;

(4) on the failure of a job, choose another resource and re-submit the failed job to it; (5) receive successful jobs.

ResourceFailure: based on Grid-Sim's GridResource class, this class interacts with RegionalGISWithFailure to set machines as failed or working. It also interacts with classes implementing AllocPolicyWithFailure to set jobs as failed.

AllocWithFailure: it is an interface class, which provides some functions to deal with resource failures. Each allocation policy implementing this interface will have a different behaviour with regard to the failures.

AvailabilityInfo: This class is used to implement the pushing mechanism. The user and GIS send objects of this class to resources, which in turn send them back, as mentioned previously. When a resource still has some working machines left, it will send these objects back with no delay. However, when all machines are out of order, the resource sends these objects back with some delay with a special tag. This is done to simulate a situation, where if a resource does not reply to the given push before a specified time out, then it is interpreted as not available. This method is used to overcome the same problem in GridSim, i.e. waiting for events that never arrive.

Resource failure statistics

Virtual Group	Resources	Failed Resources
V1	100	5
V2	20	1
V3	60	30
V4	23	23

	Jobs	Failed Jobs
1	200	52
2	89	8
3	100	65
4	10	10

CONCLUSION AND FUTURE WORK

By adding firewall for the scheduler we can provide security by blocking unwanted data and also we provide authentication and authorization before adding resource to the virtual group. We are also providing the usage of simulation tools to improve the Quality of services.

In future we are planning to include cloud for providing high security.

References

[1]: Javier Celaya, Loris Marchal ” A Fair Decentralized Scheduler for Bag-of-tasks Applications on Desktop Grids"

[2]: Derrick Kondo and Filipe Araujo"Characterizing Result Errors in Internet Desktop Grids".

[3]: Domingues, P.Sch. of Technol. & Manage., Polytech. Inst. of Leiria, Portugal.Silva, J.G. ; Silva, L.Sharing Checkpoints to Improve Turnaround Time in Desktop Grid Computing

[4]Universiteit Antwerpen

[5] Paul Townend and Jie Xu"Fault Tolerance within a Grid Environment"Department of Computer ScienceUniversity of Durham, DH1 3LE, United Kingdomp.m.townend@dur.ac.uk jie.xu@dur.ac.uk

[6] Agust´_n Caminero and Anthony Sulistio "Extending GridSim with an Architecture for Failure Detection"Department of Computing Systems 2Dept. of Computer Sc. & Software Eng.The University of Castilla La Mancha, Spain The University of Melbourne, Australiafagustin, blanca, carmeng@dsi.uclm.es fanthony, rajg@csse.unimelb.edu.au 2007

A study on Data Mining Techniques for Online Community System Analysis

Dr.Sunil Tekale Professor MRCE

Mr.Amarnath Associate Professor MRCE

Ch.Mahender Reddy Assistant Professor MRCE

Abstract

In this paper we take into consideration the concepts of using algorithmic and data mining perspective of Online Social Networks (OSNs), with special emphasis on latest hot topics of research area. There are several factors which has made the study of OSNs gain enormous importance by researchers. Few such factors include the availability of huge amount of OSN data, the representation of OSN data as graphs, and so on. Analysis of data in OSNs also has a great prospective for researchers in a variety of disciplines. Hence this paper gives an idea about the key topics of using data mining in OSNs which will help the researchers to solve those challenges that still exist in mining OSNs.

Keywords: Online Social Networks, Data Mining, Structure- based Analysis, Content-based Analysis

1. Introduction

With the advent of Online Social Networks (OSNs), a revolutionary change has occurred in the social interactions of people of this decade. Many popular OSNs such as

Facebook, Orkut, Twitter, and LinkedIn have become increasingly popular. Nowadays, these OSNs allow many easy-to-learn online activities including chatting, online shopping, gaming, tweeting, etc. According to the site *thenextweb.com*, Indian citizens spend one in four minutes online using social networking sites, more than any other Internet activity [1]. In fact, social networking is considered to be the second-

fastest growing activity, behind only entertainment.

However, social media sites provide data which are vast, noisy, distributed and dynamic. Hence, data mining techniques provide researchers the tools needed to analyze such large, complex, and frequently changing social media data. In this section, we introduce some representative research issues in mining social networking sites using data mining techniques as shown in Figure 1.

1.1 Influence Propagation

Nowadays, as OSNs are attracting millions of people, the latter rely on making decisions based on the influence of such sites. For example, influence propagation can help decide which movie to watch, which product to purchase, and so on. Thus, influence propagation has become an important mechanism for effective viral marketing, where companies try to promote their products and services through the word-of-mouth propagations in OSNs. This further motivates the research community to carry out extensive studies on various aspects of the influence propagation problem.

1.2 Community or Group Detection

In general, group detection in OSNs is based on analyzing the structure of the network and finding individuals that correlate more with each other than with other users. Clustering an individual in a particular way can help to further make an assessment about the individual such as what activities, goods, and services, an individual might be interested in.

1.3 Expert Finding

OSNs consist of several experts in a specific domain and other people who join the network to receive help from these experts. These OSNs can be used to search for such experts within a group of people. For example, a topic- related expert can be searched based on the study of the link between authors and receivers of emails.

1.4 Link Prediction

The bulk amount of data available in OSNs can be mined to make predictions about *'who is a friend of whom'* as an individual might be only a few steps away from a desirable social friend but may not realize it. By gathering useful information about an individual, OSNs can infer new interactions among members of an OSN that are likely to occur in the near future.

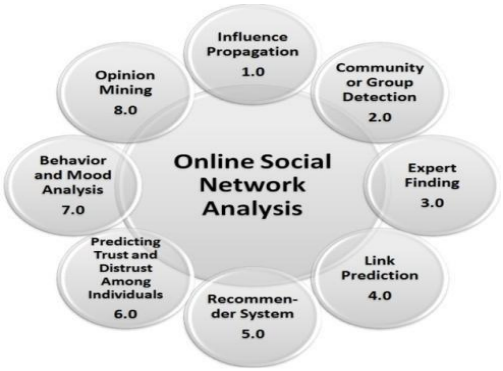


Fig 1. Key Research Issues in Online Social Network Analysis

1.5 Recommender Systems

Recommender systems (RS) provide recommendations to users about a set of articles or services they might be interested in. This facility in OSNs has become very popular due to the easy access of information on the Internet. Few important applications of RS are its use in several websites for recommendation of items such as movies, books, gadgets, etc.

1.6 Predicting Trust and Distrust among Individuals

Due to the continuous expansion of communities in OSNs, the question of trust and distrust among individuals in a community has become a matter of great concern. Past assessments reveal that some users try to either disturb or take undue advantage of the normal atmosphere of such online communities. As a result, there arises a need of assessing each user of an OSN community to predict the level of trust or distrust that can be computed for them.

1.7 Behavior and Mood Analysis

Discovering human behavior or human interaction based on data mining techniques is also an interesting research field that is gaining huge attention in research. Here, human behavior may indicate any human-generated actions such as clicking on a specific advertisement, accepting a friend's request, joining a group or discussion forum, commenting on an image, music, etc, or dating with a person, etc.

1.8 Opinion Mining

OSNs have given rise to various review sites, blog repositories, online discussions, etc where people can express their ideas and opinions, exchange knowledge and beliefs, criticize products and ideas. Data mining of opinions on specific subjects allows the detection of user prospects and needs, and also feelings or reactions of people about certain beliefs, products, decisions or events.

Some other important data mining applications related to OSNs include information and activity diffusion, topic detection and monitoring, marketing research for businesses, data management and criminal detection. These applications are also gaining huge interest in the research community. Thus, it can be concluded that OSN data analysis has a great prospective for researchers in a diversity of disciplines.

2. Current Status of these Research Issues

As the demand and usage of OSNs are increasing on a daily basis, there arises the necessity to critically analyze and understand such networks in an efficient manner. There

is a constant radical change creeping into the OSN research community in the way analysts are interpreting and characterizing OSNs. The current status of the above- mentioned research issues related to OSNs is discussed next:

2.1 Influence Propagation

Domingos and Richardson [6] provided the first algorithmic treatment to deal with influence propagation problem. Then, Kempe et al. [4] studied influence propagation by focusing on two fundamental propagation models, named *Independent Cascade (IC) Model* and *Linear Threshold (LT) Model*, which led to the development of the *Greedy Algorithm* for influence maximization. However, their model is not scalable to large networks. Leskovec et al. [3] dealt with the influence propagation problem from a different perspective namely outbreak detection but their method too faced serious scalability problems [5]. Chen et al. proposed a new propagation model similar to the greedy algorithm but with a better efficient result [7, 8]. Saito et al. [9] were the first to study how to learn the probabilities for the IC model from a set of past propagations. Goyal et al. [10] also had made a study of the problem of learning influence probabilities using an instance of the *General Threshold Model*. Barbieri et al. [11] considered the study on influence propagation from a topic modeling perspective.

2.2 Community or Group Detection

A comparative analysis on various community detection algorithms can be found in [12]. Initial study on community or group detection was focused mainly on the link structure of OSNs while ignoring the content of social interactions, which is also crucial for precise and meaningful community extraction. It is only recently that few researchers have addressed the problem of discovering topically meaningful communities from an OSN. McCallum et al. [14] were the first to have presented the

Author-Recipient-Topic model, a Bayesian network for social network analysis that discovers discussion topics conditioned on the sender-recipient relationships. Pathak et al. [17] have proposed a *Community-Author-Recipient- Topic (CART) model* which uses both link and content information for community detection. Liu et al. [16] also have built a model based on *Topic-Link Latent Dirichlet Allocation (LDA)* but which works

only with document networks. Zhao et al. [13, 15] have addressed topic oriented community detection through social objects and link analysis in social networks. Sachan et al.

have proposed *Topic User Community Model (TUCM)* as well as *Topic User Recipient Community Model (TURCM)* which offer topic modeling capabilities but takes a significantly longer time for result.

2.3 Expert Finding

Study on expert ranking algorithm is usually based on either domain knowledge driven methods or domain knowledge independent methods or both. The expert ranking problem is also researched on email communication relations [19]. Zhang et al. [20] have proposed propagation based approach as well as a mixed approach based on *Probabilistic Latent Semantic Analysis (PLSA)* [21] for expert finding in social networks. Authors in [22] have used the *RarestFirst* and *EnhancedSteiner* algorithms for expert finding while authors in [23] have modified the *RarestFirst* algorithm and found the *Simplified RareFirst* (SRareFirst) algorithm. Smirnova et al. [24] have proposed a user-oriented model for expert finding based on rational user behavior. Jin et al. [25] found the *ExpertRank* algorithm which is based on analyzing closeness and authority for ranking expert in social networks.

2.4 Link Prediction

Liben-Nowell and Kleinberg [26] have dealt with link prediction in social networks but which works with only a static snapshot of a network. Hasan et al. [27] have proposed several classification models for link prediction which provides a comparison of several features by showing their rank of importance as obtained by different algorithms. Fouss et al. [28] have presented a link prediction technique based on a *Markov-chain model* of random walk but which does not scale well for large databases. Zheleva et al. [29] have used a binary classification algorithm in which family relationships were used for link prediction. Tylenda et al. [30] have proposed time-aware and time-agnostic maximum entropy methods for link prediction but have tested data sets only from scientific collaboration networks. Chen et al. [31] have made a detailed study and comparison of four different algorithms for link prediction. Schifanella et al. [32] have proposed a sampling link-prediction algorithm which can help users find friends with similar topical interests.

Papadimi-triou et al. [2] presented a paper on fast and accurate link prediction in social networking systems but which considers only friendship network and no other features for link prediction.

2.5 Recommender Systems

Recommender systems (RS) have developed in parallel with the web. A good survey on various RS can be found in [33]. They were initially based on demographic, content-based and collaborative filtering. Collaborative filtering is the most common technique used for RS [34, 35]. Linden et al. [36] presented their work on item-to-item collaborative filtering for amazon.com recommendations. However, the evolution of RS has shown the importance of hybrid techniques of RS, which merge different techniques in order to get the advantages of each of them. A survey focused on the hybrid RS has been presented in [37] but it does not deal with the role of social-filtering, a technique which has become more popular in the recent years through social networks. Hybrid filtering techniques can use mixed collaborative and content-based filtering [38, 39, 40], or can use mixed collaborative and demographic filtering techniques [41].

Predicting Trust and Distrust among Individuals

A number of disciplines have looked at various issues related to trust. One of the first works on this task was the *EigenTrust* algorithm [42] that aims to reduce the number of inauthentic file downloads in a P2P network. Guha et al. [43] proposed methods of propagation of trust and distrust, each of which is appropriate in certain circumstances. *PowerTrust* [44] is a trust recommendation system that aggregates the positive and negative opinions between the users into the local trust scores, similarly to *EigenTrust*. Other work that studies a social network with positive and negative opinions is presented in [45]. DuBois et al. [46] presented a paper for predicting trust and distrust based on path probability in random graphs. Kim et al. [47] have also proposed a method of predicting trust and distrust of users in online social media-sharing communities. Ortega et al. [48] proposed a novel system intended to propagate both positive and negative opinions of the users through a network, in such way that the opinions from each user about others influence their global trust score.

2.6 Behavior and Mood Analysis

Benevenuto et al. [49] measured the behavior of online social networks' users applying the proxy server-based measurement framework. Schneider et al. [50] also have conducted an in-depth analysis of user behavior based on network traces across several online social networks. Gyarmati et al. [53, 54] crawled the public part of users' profile pages, which contained online status information of the users. Simoes et al. [51] proposed distance, similarity, influence and adjustments-based methods for understanding and predicting human behavior for social communities. Zhang et al. [52] have developed a model called *socioscope* for predicting human-behavior in social networks. Yan et al. [55] also have presented a social network based human dynamics model to study the relationship between the social network attributes of microblog users and their behavior. However, because of the diversity and complexity of human social behavior, no one technique will detect every attributes that arises when humans engage in social behaviors.

2.7 Opinion Mining

Most of works in this research area focus on classifying texts according to their sentiment polarity, which can be positive, negative or neutral [56]. Authors in [57] provided an in-depth survey of opinion mining and sentiment analysis. In [58], the problem was studied further using supervised learning by considering contextual sentiment influencers such as negation (e.g., not and never) and contrary (e.g., but and however). Wilson et al. [59] have studied several different learning algorithms such as boosting, rule learning, and support vector regression that can automatically distinguish between subjective and objective (neutral) language and also among weak, medium, and strong subjectivity. Zhang et al. [60] presented a novel model that unifies topic-relevance and opinion generation by a quadratic combination. Zafarani et al. [61] studied sentiment propagation in social network by making a case study of *LiveJournal* website. In [62], a method was proposed to deal with the problem of product aspects which are nouns and imply opinions using a large corpus. Authors in [63] have studied about several challenges in developing opinion mining tools for social media. Ortigosa et al. [64] developed a hybrid approach for performing sentiment analysis in Facebook with high accuracy.

3 Using Data Mining in OSNs: Research Questions

Data available in OSNs are basically user-generated content which are vast, noisy, distributed, unstructured, and dynamic. These inadvertent characteristics pose challenges to data

mining tasks to develop efficient algorithms and techniques. The following are some key points that need strong consideration while using data mining techniques for OSNA:

analysis of OSNs can yield valuable insights about the underlying social network. Hence researchers need to concentrate on using the concept of structure as well as content for making an analysis of OSNs

A. *Structure-based Analysis versus Content-based Analysis:* As far as research in OSNA is concerned, a good amount of work has been done based on structure analysis of OSNs where the linkage structure is taken into consideration in order to gather interesting characteristics of the underlying network. However, some recent research has shown that content-based

B. *Dynamic Analysis:* Past research on OSNs has mainly treated the network to be static. However, the fact remains that OSNs are dynamic and hence, to improve the quality of the results, a major amount of work yet remains to be done on dynamic analysis of OSNs which can evolve rapidly over time.

C. *Adversarial Networks:* Study of adversarial networks (such as terrorist network) requires attention by researchers in which the relationship among the different adversaries may not be clearly known.

Thus, such kind of complex heterogeneous OSNs require designing of new tools and techniques which can resourcefully analyze the network. In this regard, graph- based data mining can play a major role for such OSN analysis.

4 Conclusion

The analysis of OSN data though has its solid basis in graph theory, yet it is still in its infancy. Researchers still need to focus on these critical social network issues, taking into consideration an algorithmic and data mining perspective for attaining a better solution for the same. There are also strong motivations for efficiently propagating the right information to the right people via OSNs and which has become a research area of increasing importance. Thus, this survey paper is a step forward in the direction of evolving new data mining techniques to address the above mentioned critical online social networking issues.

References

[1]. Josh Ong (August 2012), article retrieved from http://thenextweb.com/in/2012/08/20/social-networking-ites-now-occupy-25-online-time-india/

[2]. Papadimitriou, P. Symeonidis, and Y. Manolopoulos, "Fast and accurate link prediction in social networking systems", The Journal of Systems and Software 85, 2012, pp. 2119–2132

[3]. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. S. Glance, "Cost-effective outbreak detection in networks," in Proc. of the 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'07), 2007, pp. 420–429

[4]. Kempe, J. M. Kleinberg, and ´E. Tardos, "Maximizing the spread of influence through a social network," in Proc. of the 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'03), 2003, pp.137-146

[5]. W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks", in Proc. of the 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'09), 2009.

[6]. M. Richardson and P. Domingos, "Mining knowledge- sharing sites for viral marketing," in Proc. of the 8th ACM SIGKDD Int. Conf on Knowledge Discovery and Data Mining (KDD'02), 2002, pp. 61-70.

[7]. W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks", in Proc. Of the 16th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'10), 2010, pp. 1029-1038.

[8]. W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model", in Proc. of the 10th IEEE Int. Conf. on Data Mining (ICDM'10), 2010.

[9]. K. Saito, R. Nakano, and M. Kimura, "Prediction of information diffusion probabilities for independent cascade model", in Proc. of the 12th Int. Conf. on Knowledge- Based Intelligent Information and Engineering Systems (KES'08), 2008.

[10]. A. Goyal, F. Bonchi, and L. V. S. Lakshmanan, "Learning influence probabilities in social networks," in Proc. of the 3rd ACM Int. Conf. on Web Search and Data Mining (WSDM'10), 2010.

[11]. N. Barbieri, F. Bonchi, and F. Bonchi, "Topic-aware Social Influence Propagation Models", in 2012 IEEE 12th Int. Conf. on Data Mining, 2012, pp. 81-90.

[12]. A. Lancichinetti, and S. Fortunato, "Community detection algorithms: a comparative analysis", arXiv:0908.1062v1 [physics.soc-ph], 2009.

[13]. D. Zhou, E. Manavoglu, J. Li, C. L. Giles, and H. Zha, "Probabilistic Models for Discovering E-Communities", in Proc. of the 15th Int. Conf. on World Wide Web, 2006.

[14]. A. McCallum, A. Corrada-Emmanuel, and X. Wang "Topic and Role Discovery in Social Networks", in Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI), 2005, pp. 786-791.

[15]. Z. Zhao, S. Feng, Q. Wang, J. Z. Huang, G. J. Williams, and J. Fan, "Topic oriented community detection through social objects and link analysis in social networks", In Journal Knowledge-Based Systems 26, 2012, pp. 164–173.

[16]. Y. Liu, and A. Niculescu-Mizil, "Topic-link LDA: joint models of topic and author community", in Proc. of the 26th Int. Conf. on Machine Learning, Montreal, Canada, 2009, pp. 665-672.

[17]. N. Pathak, C. DeLong, and A. Banerjee, "Social Topic Models for Community Extraction", in the 2nd SNA-KDD Workshop'08, 2008.

[18]. M. Sachan, D. Contractor, T. A. Faruquie, and L. V.Subramaniam, "Using Content and Interactions for Discovering Communities in

Social Networks", in Proc. of the 21st Int. Conf. on World Wide Web, 2012 pp. 331-340.

[19]. S. Campbell, P. P. Maglio, A. Cozzi, and B. Dom, "Expertise Identification Using Email Communications", In Proc. of the 12th Int. Conf. on Information and Knowledge Management, 2003, pp.528-531.

[20]. J. Zhang, J. Tang, and J. Li, "Expert finding in a social network", in Proc. of the 12th Database Systems Conf. for Advanced Applications, 2007, pp. 1066–1069.

[21]. J. Zhang, J. Tang, L. Liu, and J. Li, "A Mixture Model for Expert Finding", in Proc. of 2008 Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD2008), 2008, pp. 466–478.

[22]. T. Lappas,K. Liu, and E. Terzi, "Finding a team of experts in social networks", In Proc. of the 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2009, pp. 467–476.

[23]. H. Yin, B. Cui, and Y. Huang, Finding a Wise Group of Experts in Social Networks, Advance Data Mining and Applications, Springer Berlin Heidelberg, 2011, pp. 381– 394.

[24]. E. Smirnova, and K. Balog, "A User-Oriented Model for Expert Finding", in Proc. of the 33rd European Conf. on Advances in Information Retrieval, 2011, pp. 580–592.

[25]. L. Jin, J. Y. Yoon, Y. H. Kim, and U. M. Kim, Based on Analyzing Closeness and Authority for Ranking Expert in Social Network, Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence, Springer Berlin Heidelberg, 2012, pp. 277–283.

[26]. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks", in Proc. of the 12th Int. Conf. on Information and Knowledge Management, ACM, 2004, pp. 556–559.

[27]. A. Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link Prediction using Supervised Learning", in Proc. of the Workshop on Link Discovery: Issues, Approaches and Applications, 2005.

[28]. F. Fouss, A. Pirotte, J. Renders, and M. Saerens, "Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation", in IEEE transactions on Knowledge and Data Engineering, vol. 19, no. 3, 2007, pp.355-369.

[29]. E. Zheleva, L. Getoor, J. Golbeck, and U. Kuter, "Using friendship ties and family circles for link prediction", in Proc. of the 2nd Workshop on Social Network Mining and Analysis (SNA-KDD'2008), 2008, pp. 97–113.

[30]. T. Tylenda, R. Angelova, and S. Bedathur, "Towards Time-aware Link Prediction in Evolving Social Networks", in Proc. of the 3rd Workshop on Social Network Mining and Analysis, 2009, pp.1-10.

[31]. J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy, "Make new Friends, but Keep the Old – Recommending People on Social Networking Sites", in Proc of the 27th Int. Conf. on Human Factors in Computing Systems (CHI'09), 2009, pp. 201-210.

[32]. R. Schifanella, A. Barrat, C. Cattuto, B. Markines, and F. Menczer, "Folks in folksonomies: social link prediction from shared metadata", in Proc. of the 3rd ACM Int. Conf. on Web Search and Data Mining (WSDM'10), 2010, pp. 271-280.

[33]. J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey", In the Journal of Knowledge-Based Systems 46 (2013), pp. 109–132.

[34]. J. Bobadilla, F. Serradilla, and J. Bernal, "A new collaborative filtering metric that improves the behavior of recommender systems", In the Journal of Knowledge Based.

Analysis on Defect Comparison Techniques and Design in Cloud Computing

Mr.CH.Vengaiah Assistant **Professor MRCE Sec-100.**

Mr *Ch.Mahender Reddy Assistant* **Professor MRCE Sec-100.**

Abstract - The "cloud" is a set of different types of hardware and software that work collectively to deliver many aspects of computing to the end-user as an online service. Cloud Computing is the use of hardware and software to deliver a service over a network (typically the Internet). With cloud computing, users can access files and use applications from any device that can access the Internet. The Cloud computing market continues to grow year after year because companies are becoming more aware of the cost saving benefits of adopting the cloud. It is the adoptable technology as it provides integration of software and resources which are dynamically scalable. These systems are more or less prone to failure. Fault-tolerance is the property that enables a system (often computer-based) to continue operating properly in the event of the failure of (or one or more faults within) some of its components. In order to achieve robustness and dependability in cloud computing, failure should be assessed and handled effectively. The main objective of this paper is to discuss about the different techniques and algorithms of fault tolerance.

Keywords² Cloud, Fault tolerance, Load balancing, Priority scheduling, Replication

INTRODUCTION

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., Networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider Interaction[1].

Resource scheduling is the basic and key process for clouds in Infrastructure as a Service (IaaS) as the need of the request processing is must in the cloud. Every server has limited resources so jobs/requests needs to be scheduled. Each application in the cloud computing is designed as a business processes including a set of abstract processes. To allocate the resources to the tasks there need to schedule of the resources as well as tasks coming to the resources. There need to be a Service Level Agreements (SLAs) for Quality of Service (QoS). Till now no algorithm is been introduced which considers reliability and availability. According to the paradigm of cloud there has been a lot of task scheduling algorithms, some are being fetched on the basics of scheduling done on the operating system. The basics of operating system job scheduling is taken and applied to the resources being installed in the cloud environment [2].

According to NIST(National Institute of Standards and Technology), cloud model is composed of five essential characteristics**,**

three service models, and four deployment models [1].

A. Essential Characteristics:

i) *On-demand self-service:* A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

ii) *Broad network access:* Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).

iii) *Resource pooling:* 7 KH SURYLGHU¶V FRP SXWing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, and network bandwidth.

iv) *Rapid elasticity:* Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

v) *Measured service:* Cloud systems automatically control and optimize resource use by leveraging a metering capability1 at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

r

B. Service Models:

i) *Software as a Service (SaaS):* The capability provided to the consumer is to use the provider‟s applications running on a cloud infrastructure2. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

ii) *Platform as a Service (PaaS)*: The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider.3 The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

iii) *Infrastructure as a Service (IaaS):* The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls).

C. Deployment Models:

i) *Private cloud:* The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

ii) *Community cloud:* The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

iii) *Public cloud:* The cloud infrastructure is provisioned for open use by the general

public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

iv) *Hybrid cloud*: The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

II FAULT TOLERANCE IN CLOUD COMPUTING

Cloud Computing is an important computational paradigm which provide on - demand services to users and in low cost. Fault tolerance and reliability are of great importance that provides correct result even in the presence of faulty components. Most of the systems are safety critical and highly reliable. So to achieve reliability in real time computing, the demand for fault tolerance is increased. In real time computing, the capabilities of intensive computing can be an advantage to execute real time tasks. In most of the applications of real time cloud, processing is done on remote cloud computing nodes. Therefore, due to loose control over the computing node, chances of errors increase. Fault tolerance techniques are used to predict these failures and take an appropriate action before failures actually occur. So to achieve reliability in real time computing, the requirement for fault tolerance increases. The reliability of virtual machines changes after every computing cycle i.e. it is adaptive in nature. [43]

A. Basic notion to Fault Tolerance

A client enlists with the service provider (SP) to attain support for the functions of fault tolerance. Service Provider creates the solution for the fault tolerance based on the requirements of client such that the balance between the following aspects is achieved.

REVIEW OF LITERATURE

A fault-tolerant system may be able to tolerate one or more fault-types including -- i) transient, intermittent or permanent hardware faults, ii) software and hardware design errors, iii) operator errors, or iv) externally induced upsets or physical damage. An extensive methodology has been developed in this field over the past thirty years, and a number of fault-tolerant machines have been developed -- most dealing with random hardware faults, while a smaller number deal with software, design and operator faults to varying degrees. A large amount of supporting research has been reported.

Gayathri and Prabakaran discussed some important factors of failures. One important factor is arbitrary node or link failure which results in denial of service. In cloud computing, load balancing is required to distribute the dynamic local workload evenly across all the nodes. It helps to achieve a high user satisfaction and resource utilization ratio by ensuring an efficient and fair allocation of every computing resource. The load balancing should be a good fault-tolerant technique. Identified some of the load balancing algorithms which distribute workload across multiple computers or a computer cluster, network links, central processing units, disk drives, or other resources, to achieve optimal resource utilization, maximize throughput, minimize response time, and avoid overload. When all these issues are addressed naturally the system becomes a fault tolerant one.[8]

Win Win Naing proposed a fault-tolerance management framework for private clouds development. Previous researchers developed Eucalyptus in order to facilitate the creation of private clouds. But Eucalyptus is non fault tolerant system and no VM monitoring is performed thus limiting the support for advanced VM placement policies (e.g., consolidation). Eucalyptus does not also include any self-healing features and strictly distinguishes between cloud and cluster controllers. Therefore, they proposed the fault-tolerance management framework over Eucalyptus by adding new component Cluster Controller Manager (CCM). They choosed CCM as a manager to control the failure occur. They used Zoo Keeper[22] leadership algorithm to select CCM as a leader from many CCs. So, we can develop fault-tolerance management framework for private cloud environment.[9]

Sheheryar and Fabrice proposed a scheme of fault tolerance mechanism for real time computing on cloud infrastructure. It has all the advantages of forward recovery mechanism. It has a dynamic behaviour of reliability configuration. The scheme is highly fault tolerant. The reason behind adaptive reliability is that the scheme can take advantage of dynamic scalability of cloud infrastructure. This system takes the full advantage of using diverse software. In their experiment, they have used three

virtual machines. It utilizes all of three virtual machines in parallel. This scheme has incorporated the concept of fault tolerance on the basis of VM algorithm reliability. Decision mechanism shows convergence towards the result of the algorithm which has highest reliability. Probability of failure is very less in their devised scheme. This scheme works for forward recovery until all the nodes fail to produce the result. The system assures the reliability by providing the backward recovery at two levels. First backward recovery point is TC. Here if all the nodes fail to produce the result, it performs backward recovery. Second backward recovery point is DM. It performs the backward recovery if the node with best reliability could not achieve the SRL. There is another big advantage of this scheme. It does not suffer from domino effect as check pointing is made in the end when all the nodes have produced the result [10].

Singla and Seema proposed priority scheduling algorithm with the functionality of tolerating a fault in the server included in a particular cloud. The algorithm firstly schedules the tasks according to priority and then reallocates tasks from faulty server to another server. This is better than other algorithms as it provides the fault tolerance functionality and also the results depict the total performance of the cloud increases with the proposed algorithm then available scheduling algorithms as it provides more reliability than other algorithms present till now.[2]

Liang Luo et al. have presented an algorithm based on Energy Efficient Optimization Methods. This algorithm is being implemented in Hadoop distributed file system with Energy Management and Regulation also called as GreenHDFS. This algorithm concentrates on usage of the resources that are not fully utilized while execution of the environment. Due to fast advancement in technology the old methods of saving energy has been challenging. The works introduced till now are taken into account with hardware but not with software. They studied the relationship between infrastructure components and power consumption of the cloud computing environment, and discussed the matching of task types and component power adjustment methods, and then presented present a resource scheduling algorithm of Cloud Computing based on energy efficient optimization methods. The experimental results demonstrate that, for jobs that not fully utilized the hardware environment, using their algorithm can significantly reduce energy consumption [3].

Zhongyuan Lee et al. proposed Dynamic priority scheduling algorithm (Service request scheduling) This algorithm is applied on three tier containing service providers, resource providers and consumers. This algorithm gives more optimal then First Come First Serve (FCFS) and Static Priority Scheduling Algorithm (SPSA). The consumer response time for

services has been tried to reduce in this algorithm as running instance is charged as it runs per unit time. The delays in provider side happens but are not counted under the cost charged to the customer so they need to be reduced. In three tiers there needs to be two scheduling: service request scheduling and resource scheduling [4].

Yanbing Liu et al have introduced Loyalty based resource allocation. The trust concept is introduced in architecture and loyalty which improves the successful transaction rate of the system while meeting the requirements. Using Master Slave framework a role based access control is proposed considering the trust of the node and meets the requirements using the services. The unreliability of hardware should be provided by highly reliable software. It assesses the real time condition of the system and allocates resource according to condition. This dynamic feedback mechanism provides stability and reliability of services [5].

Li and Tang has proposed Pareto based optimal scheduling. The cloud banking model is introduced with features like multi dimensional Pareto optimal theory and optimization analysis aiming at improving resource utilization as well as
consumer satisfaction. This algorithm characterizes the user‟s requirements. It takes into consideration resource prices and execution time [6].

Han et al. presented a fault-tolerant scheduling algorithm called QAFT that can tolerate one node's permanent failures at one time instant for real-time tasks with QoS needs on heterogeneous clusters. In order to improve system flexibility, reliability, schedulability, and resource utilization, QAFT strives to either advance the start time of primary copies and delay the start time of backup copies in order to help backup copies adopt the passive execution scheme, or to decrease the simultaneous execution time of the primary and backup copies of a task as much as possible to improve resource utilization. QAFT is capable of adaptively adjusting the QoS levels of tasks and the execution schemes of backup copies to attain high system flexibility. Furthermore, we employ the overlapping technology of backup copies. The latest start time of backup copies and their constraints are analyzed and discussed. They conducted extensive experiments to compare our QAFT with two existing schemes-NOQAFT and DYFARS. Experimental results show that QAFT significantly improves the scheduling quality of NOQAFT and DYFARS.[11]
Patra et al. discussed about the fault taxonomy and need of fault tolerance covering with its various techniques for implementing fault tolerance. Various proposed models for fault tolerance are discussed and compared on the basis of Metrics for fault tolerance in cloud. In the present scenario, there are number of fault tolerance models which provide different fault tolerance mechanisms to enhance the

system. But still there are number of challenges which need some concern for every frame work or model. There are some drawback no one of them can full fill the all aspects of faults. So there is a possibility to overcome the drawbacks of all previous models and try to make a compact model which will cover maximum fault tolerance aspect.[12]

Jasbir Kaur et al. analysed the implementation of fault tolerance in a complex cloud computing environment with a focus on FCFS and SJF along with MPIL method with fault tolerance property. The proposed algorithm works for reactive fault tolerance among the servers and reallocating the faulty servers task to the new server which has minimum load at the instant of the fault. They illustrated this discussion with experiments where exclusive and collaborative fault tolerance solutions are implemented in an autonomic cloud infrastructure that they prototyped. It also includes algorithm comparison between MPI and MPIL. Fault tolerance is carried out by error processing which have two constituent phases. The phases are "effective error processing" which aimed at bringing the effective error back to a dormant state, i.e. before the occurrence of error and "latent error processing" aimed at ensuring that the error does not become effective again. In the end it is concluded that the performance of MPIL is better than the MPI in terms of both energy consumption and checkpoints required. [13]

Sudha and Padmavati proposed fault tolerance in real time cloud computing environment. In the proposed model, the system tolerates the faults and makes the decision on the basis of reliability of the processing nodes, i.e. virtual machines. The proposed technique is based on the execution of design diverse variants on multiple virtual machines, and assigning reliability to the results produced by variants. The system provides both the forward and backward recovery mechanism. The proposed scheme is a good option to be used as a fault tolerance mechanism for real time computing on cloud infrastructure. It has all the advantages of forward recovery mechanism. It has a dynamic behaviour of reliability configuration. The scheme is highly fault tolerant. The reason behind adaptive reliability is that the scheme can take advantage of dynamic scalability of cloud infrastructure. This kind of system takes advantage over the dynamic scalability of cloud infrastructure that‟s why using the adaptive reliability method. And also in this case there is less chances of failure. The main advantage of this scheme is that because of the checkpoints which are made in the end when all the nodes have produced the result this will not cause domino effect. Some new enhancements can be made on this model. The main focus is to include more reliability factors on which decisions are to be made and it will be more effective. Also one resource manager is working i.e. proactive resource manager we can also use

reactive resource manager which will not remove the node but try to resolve the problem which causes node failure.[14]

Pandeeswari and Mohamadi presented RSFTS approach, which proposed a way to allocate resources taking the benefits of semantic technologies. In addition, it deals with the failures using fault tolerance mechanism. Consequently, RSFTS prevents the delay and blocking of entire execution of tasks. So it increases the total and average execution times of tasks and guarantees the completion of execution of tasks. They have tested the semantic models annotating resources from different providers and schemas, and proposing rule examples for implementing different customer and provider policies. Additionally, they have evaluated the semantic approach identifying the most important processes and overheads comparing this approach with the Gossip [23] approach in different situations. As result of their evaluation, they have detected the Gossip approach performs better when the number of resources and users are low due to an important negotiation overhead while the RSFTS approach is a better option when the number of resources and users are big and also Fault tolerance techniques are used to predict the failures and take an appropriate action before failures actually occur when an execution in VM.[15]

Virendra Singh Kushwah et al. investigated about fault-tolerance in load balancing schemes in a cloud environment. [19]

Table I load balancing algorithms for cloud environment

Algorithm Name	Parameters	Merits	Demerits
Honeybee Algorithms[16]	Throughput, Job completion time, Overhead	Achieve Global Load Balancing, Maximize resource utilization, low overhead	Low Priority load
Task Scheduling	Response time,	Minimize the response	Not provider oriented

Honey Bee Algorithm achieves global load balancing through local serve actions. Its whole concept is based on the idea honey bee, how they search their food and then inform others for the same by waggle dance. The strength or power of the waggle dance gives an idea about the amount of food present. In the same way the load balancing is done. As virtual machine is overloaded the user request is forwarded to next less loaded virtual machine [16].

Task Based Scheduling Algorithm, it has two level task scheduling processing. At the first level, it maps tasks to virtual machines and then the virtual machines to host resources. It provides maximum throughput or maximize resource utilization with minimum task response time [17].

Biased Random Sampling achieves load balancing across all system nodes using random sampling of the System domain to accomplish self-organization thus stabilize the load among nodes. A virtual graph is constructed to represent the load on serving and connectivity between them [17].

Equally Spread Current Execution Algorithm, it handles processes with Priorities. It distributes the load randomly by checking the size and transfers the load to those virtual machines which is lightly loaded or handle that task easier and take less time, and give maximize throughput. It is also known as Spread Spectrum technique as the entire load is distributed among nodes by load balancer [18].

Ant Colony Optimization technique takes the idea of the ant behavior, how they collect information and leave the liquid (pheromone) in the path to inform others about the path of good. This algorithm maintains a pheromone table on the basis of resource utilization and node selection method. The ants searches for overloaded node and then traverse it and then go back to fill the under load node so as to make balance or evenly distribute the load [18].

Load balancing is the pre requirements for increasing the cloud performance and for completely utilizing the resources. Load balancing is centralized or decentralized. Several load balancing algorithms are

introduced which differs in their complexity. The effect of the algorithm depends on the architectural designs of the clouds. Today, cloud computing is a set of several data centers, which are sliced into virtual servers and located at different geographical location for providing services to clients. Day-by-day use of cloud computing is increasing which increase the load on the servers providing services to third parties. Due to this the overall performance degrades and there is poor resource utilization. This problem is referred as load balancing which is a major issue now days. To solve this problem various algorithms were proposed as given in Table I above [19]

Peng and Wei proposed a fault tolerance mechanism to detect and then recover from failures. Specifically, instead of simply using a timeout configuration, they designed a trust based method to detect failures in a fast way. Then, a checkpoint based algorithm is applied to perform data recovery. Their experiments showed that their method exhibits good performance and is proved to be efficient. In their paper they proposed a fault tolerance mechanism based on Hadoop. The only support of fault tolerance on native Hadoop is replication on HDFS and re-execution of failed Map or Reduce tasks. However, that increases the cost by re-execution the whole task no matter how far it proceeds. The situation would be worse if a long-last task fails in a large job. To that end, they proposed to first detect failure at a early stage through a trust based method, and then use a checkpoint algorithm for failure recovery. First, instead of simply applying a timeout solution, they assigned a trust value to each node, which decreases if a Reduce task gets a fetch error from it. In that way, they dramatically reduces the cost of failure detection. Second, for their checkpoint algorithm, they employed a non-block method by sending and receiving messages with sequence numbers of specific replica of data blocks. And the data is written to the local storage first, and then combined at the centralized master. Besides, according the metadata information on master, it can choose an appropriate location for rebuilding the missing data block using checkpoint data from other replicas. Besides, they conducted extensive experiments on a Hadoop cluster, and compared their proposed method with the native configuration of Hadoop. Their empirical results indicate that proposed method exhibits good performance and efficiency.[20]

Rejinpaul and Visuwasam have proposed a smart checkpoint infrastructure for virtualized service providers. They have provided a working implementation of the infrastructure that uses Another Union File System (AUFS) to differentiate read-only from read-write parts in the VM image. In this way, read- only parts can be check pointed only once, while the rest of checkpoints must only save the modifications in read-write parts, thus reducing the time needed to make a checkpoint and, as a consequence, the interference on task execution. The checkpoints are compressed (only if this permits saving time) and stored in

a Hadoop Distributed File System. Using this system, the checkpoints are distributed and replicated in all the nodes of the provider. As demonstrated in the evaluation, the time needed to make a checkpoint using their infrastructure is considerably lower by using AUFS. The checkpoint upload time is higher when using HDFS, but this does not increase the time that the VM is stopped, and it is far compensated when resuming tasks. On the other side, the time needed to resume a task execution is comparable to other approaches when only one task is resumed, and significantly lower when resuming several tasks concurrently. This occurs because the checkpoint can be concurrently recovered from different nodes. Furthermore, this made their checkpoint mechanism fault-tolerant, as any single point of failure has been eliminated.[21] Jianfeng Zhao et al. proposed a virtual resources scheduling model and solved it by advanced Non-dominated Sorting Genetic Algorithm II (NSGA II). This model was evaluated by balance load, virtual resources and physical resources were abstracted a lot of nodes with attributes based on analyzing the flow of virtual resources scheduling. NSGA II was employed to address this model and a new tree sorting algorithms was adopted to improve the efficiency of NSGA II. In experiment, verified the correctness of this model. Comparing with Random algorithm, Static algorithm and Rank algorithm by a lot of experiments, at least 1.06 and at most 40.25 speed-up of balance degree can be obtained by NSGA II.[24] Kowsik and Rajakumari has surveyed different types of scheduling algorithms and tabulated their various parameters, scheduling factors and so on. Existing workflow scheduling algorithms does not consider reliability and availability. They presented a novel heuristic scheduling algorithm, called hyper-heuristic scheduling algorithm (HHSA), to find better scheduling solutions for cloud computing systems. The results showed that HHSA can significantly reduce the makespan of task scheduling compared with the other scheduling algorithms. The proposed algorithm uses two detection operators to automatically determine when to change the low level heuristic algorithm and a perturbation operator to fine tune the solutions obtained by each low-level algorithm to further improve the scheduling results in terms of makespan.[25] Simy Antony et al. presented Balance Reduce Algorithm (BAR) based on data locality driven reducing network access thus reducing bandwidth usage and job completion time. This algorithm also handles the machine failure. Initial local task allocation in balanced phase takes place and then job execution time can be reduced by matching initial task allocation in reduced phase. The machine failure is handled by algorithm similar to primary backup approach. [26] Maguluri et al. studied a stochastic model of cloud computing, where jobs arrive according to a stochastic process and request resources like

CPU, memory and storage space. They considered a model where the resource allocation problem can be separate into a routing or load balancing problem and a scheduling problem.

They studied the join-the-shortest-queue routing and power-of-two-choices routing algorithms with MaxWeight scheduling algorithm. It was known that these algorithms are throughput optimal. They have shown that it is heavy traffic optimal when all the servers are identical. They also found that using the power-of-two-choices routing instead of JSQ routing is also heavy traffic optimal. They considered a simpler setting where the jobs are of the same type, so only load balancing is needed. It has been established by others using diffusion limit arguments that the power-of-two-choices algorithm is heavy traffic optimal. [27]

Zhi Yang et al. presented a cost-based resource scheduling paradigm in cloud computing by leveraging market theory to schedule compute resources to meet user's requirement. The set of computing resources with the lowest price are assigned to the user according to current suppliers' resource availability and price. They designed an algorithm and protocol for cost-based cloud resource scheduling. This scheduling paradigm is implemented and evaluated in Java Cloud ware, the pure Java based private cloud platform.

Mingshan Xie et al discussed an algorithm based on Trust Degree. This algorithm takes into consideration the functional characteristics and provides better stability and low risk while completing tasks. It reduces threshold and risks in small and medium enterprises. The trust degree is determined by execution time and reliability. Scheduling logs stores trust degree at any time and sort it decreasingly and then the computer slots are called according to whose trust degree is greater first. This algorithm is stable and reliable. [29]

Xin Lu and Zilong Gu presented load adaptive model based on ant colony algorithm. This algorithm monitors real timely virtual machines on performance parameters and schedules fast resources using any colony algorithm. It is made accordingly to bear load on a load free node to meet the changing load requirements improving resource utilization‟s efficiency. The detection of overload exceeds the threshold limit. This algorithm finds the nearest idle node and allows it to bear some load meeting the performance and resource requirements of load thus achieving the goal of load balancing. [30]

L P Saikia et al. discussed an overview of fault-tolerance techniques in large scale cluster computing systems which is presented by Treaster M. based on survey. These techniques can be grouped into two categories: protection for the cluster management hardware and software infrastructure, and protection for the computation nodes and the long-running applications that execute on them. Cluster management hardware and software fault-tolerance typically makes use of redundancy, due to the relatively small number of components that need to be duplicated for this approach. When a component fails, the redundant components take over the responsibilities of the failed parts. Redundancy can also be used for fault detection by comparing the outputs produced by each replica and looking for discrepancies. Cluster applications are protected from faults using check pointing and rollback recovery techniques. Each process cooperating in the application periodically records its state to a checkpoint files in reliable, stable storage. In the event of a process failure, the application state is restored from the most recent set of checkpoints. There are a variety of protocols that have been developed to determine when processes should record checkpoints and how to restore the application state. Fault tolerance solutions can be implemented in a variety of forms. This includes software libraries, special programming languages, compiler or preprocessor modifications, operating system extensions, and system middleware. Each method has its own tradeoffs in terms of power, portability, and ease of use. [31][32]

Perumalla had studied parallel and distributed simulation on the basis of traditional techniques and recent advances the presented an overview of parallel and distributed simulation systems, their traditional synchronization approaches and a case study using the HLA standard interface and implementation. Recent advances, such as scalability to supercomputing
platforms and novel rollback techniques have been presented. The interaction of parallel simulation with newly emerging hardware architectures is outlined. The future outlook seems to warrant focus on needs from lager scale simulation scenarios to be achieved on high-end computing plat-forms. There is also interest on high-performance simulations on low-end platforms such as using the multi-core architectures, GPGPUs and other co-processor-based systems. However, practical challenges remain to be explored, including: wide spectrum of network latencies, highly dynamic participation by processors and semantics and implementations of always-on presence for large simulation.

Anju and Inderveer discussed the fault tolerance techniques covering its research challenges, tools used for implementing fault tolerance techniques in cloud computing. Cloud virtualized system architecture is also proposed based on HAProxy. Autonomic fault tolerance is implemented dealing with various software faults for server applications in a cloud virtualized environment. When one of the servers goes down unexpectedly, connection will automatically be redirected to the other server. Data replication technique is implemented on virtual machine environment. The experimental results are obtained, that validate the system fault tolerance.

Challenges of Implementing Fault Tolerance in Cloud Computing

Providing fault tolerance requires careful consideration and analysis because of their complexity, inter-dependability and the following reasons.

i) There is a need to implement autonomic fault tolerance technique for multiple instances of an application running on several virtual machines [3.

ii) Different technologies from competing vendors of cloud infrastructure need to be integrated for establishing a reliable system [36].

iii) The new approach needs to be developed that integrate these fault tolerance techniques with existing workflow scheduling algorithms [37].

iv) A benchmark based method can be developed in cloud environment for evaluating the performances of fault tolerance component in comparison with similar ones .

v) To ensure high reliability and availability multiple clouds computing providers with independent software stacks should be used

vi) Autonomic fault tolerance must react to synchronization among various clouds .

Zaipeng Xie et al. surveyed various software fault tolerance techniques and methodologies. The techniques include traditional techniques: recovery blocks (RcB), n-version programming, n self-checking Programming, retry blocks (RtB), n-copy programming and some new techniques: adaptive n-version systems, fuzzy voting, abstraction, parallel graph reduction, rejuvenation. They surveyed and compared various software fault tolerant techniques. First, they summarized traditional techniques with diversity implementations. Then, they addressed some new techniques which either improved the traditional techniques or took a new approach to solve the problem of software fault tolerance. A lot of techniques have been developed for achieving fault tolerance in software. The application of all of these techniques is relatively new to the area of fault tolerance.

Furthermore, each technique will need to be tailored to particular applications. This should also be based on the cost of the fault tolerance effort required by the customer. The differences between each technique provide some flexibility of application. [41]

Priyanka and Geetha proposed a cloud framework to build fault-tolerant cloud applications. They first proposed fault detection algorithms to identify significant components from the huge amount of cloud components. Then, they presented an efficient fault-tolerance strategy selection algorithm to determine the most suitable fault-tolerance strategy for each significant component. Software fault tolerance is widely adopted to increase the overall system reliability in critical applications. System reliability can be enhanced by employing functionally equivalent components to tolerate component failures. Fault-tolerance strategies introduced a three well known techniques are in the following with formulas for calculating the failure probabilities of the fault-tolerant modules. Their work mainly drives toward the implementation of the framework to measure the strength of fault tolerance service and to make an in-depth analysis of the cost benefits among all the stakeholders. An algorithm is proposed to automatically determine an efficient fault-tolerance strategy for the significant cloud components. Using real failure traces and model, they evaluate the proposed resource provisioning policies to determine their performance, cost as well as cost efficiency. The experimental results showed that by tolerating faults of a small part of the most important components, the reliability of cloud applications can be highly improved. In specific, they presented a method for realizing generic fault tolerance approaches as independent modules, validating fault tolerance properties of each mechanism, and matching user‟s requirements with available fault tolerance modules to obtain a comprehensive solution with desired properties. In their proposed component algorithms, the importance value of a component is decided by the number of components that invoke this component, the importance values of these components, how often the current component is invoked by other components, and the component fundamentals. After finding out the importance components, they proposed an efficient fault-tolerance strategy selection algorithm to provide optimal fault-tolerance strategies to the importance components automatically, based on the constraints.[42]

Patel and Singh discussed several fault tolerance techniques that are existing currently in clouds .

are as follows:

i) Self-Healing, in this method divide and conquer technique is used, in which a huge task is distributed into several parts. This division is done for better performance. In this, various instances of an application are running on various virtual machines and failure of all these individual instances are handled automatically.

ii) Job Migration, sometimes it happens that due to some reason a particular machine fails and cannot execute job. On such a failure, a task is migrated to working machine using HA-Proxy. Also, there are algorithms that can automatically determine the fault and migrates batch applications within a cloud of multiple data centers.

iii) Check Pointing, it is a proficient task level fault tolerance technique for large applications. In this method, check pointing is done in system. When a task fails, instead of initiating from beginning it is restarted from the recently checked pointed state. Check pointing is carried out periodically i.e., checkpoints are kept and process is executed from the recent check point, once system governs the fault.

iv) Replication, it means copying. Several replicas of tasks are created and they are run on different resources, for effective execution and for getting the desired result. Hadoop, HA-Proxy, Amazon EC2 tools are there on which replication can be implemented. Also, there are mainly three different types of replication schemes such as Active Replication, Semi-Active Replication and Passive Replication.

v) Task Resubmission, many times it happens that due to high network traffic or due to heavy work load, a task may fail, whenever such failed task is detected, at runtime the task is resubmitted either to the same or different working resource for execution. For these, certain algorithms are designed, which assigns task to resources on the basis of certain properties.

vi) Masking, after occupation of error recovery the new state needs to be identified as a transformed state. If this process applied systematically even in the absence of effective error provide the user error masking [47].

vii) Resource Co-allocation, it refers to the process of allocating resources for further execution of task. Many algorithms are designed, that deals which resource allocation depending on the properties of VM such as workload, type of task, capacity of VM, energy awareness etc.

viii) Timing Check, this is deployed with the help of watch dog. It is a simplest technique with time as a critical function [46]. It keeps the track of task execution, whether the task has been completed in required amount of time or not. Depending on which further action for fault tolerance is taken.

ix) Rescue Workflow, a workflow consists of a sequence of connected steps where each step follows without delay or gap and ends just before the subsequent step may begin. In this technique, it allows the workflow to carry on until it becomes unimaginable to

move forward without catering the failed task.

x) User Specific (defined) Exception handling, in this case, whenever fault is detected, action is predefined by the user, i.e. user defines the particular treatment for a task on its failure.

Several models that are implemented based on above techniques are as follows:

a) **"**AFTRC**"** – It is an Adaptive Fault Tolerance model in Real time Cloud Computing. In this proposed model system tolerates fault proactively and makes decision on the basis of the reliability of the processing nodes [48].

b) **"**LLFT**"** - is a propose model which contains a low latency fault tolerance (LLFT) middleware for providing fault tolerance for distributed applications deployed within the cloud computing environment. This middleware replicates application by the using of semi-active replication or semi-passive replication process to protect the application against various types of faults [40].

c) **"**FTM**"**- is a model to overcome the limitation of existing methodologies and achieve the reliability and flexibility, they propose an inventive perspective on creating and managing fault tolerance .By this particular methodology user can specify and apply the desire level of fault tolerance. FTM architecture this can primarily be viewed as an assemblage of several web services components, each with a specific functionality [49].

d) **"**FTWS**"**- is a Fault Tolerant Work flow Scheduling algorithm for providing fault tolerance by using replication and resubmission of tasked based on based on the priority of the task. This model is based on the fact that work flow is a set of tasks processed in some order based on data and control dependency. Scheduling the workflow along with the task failure consideration in a cloud environment is very challenging. FTWS schedule and replicates the tasks to meet the deadline [50].

e) **"**Candy**"**- is a component based availability model. It is based on the high availability assurance of cloud service is one of the main characteristic of cloud service and also one of the main critical and challenging issues for cloud service provider [51].

f) **"**FT-Cloud**"**- is a component ranking based frame work and its architecture for building cloud application. FT-Cloud occupies the component invocation structure and frequency for identify the component. Also, there is an algorithm to automatically govern fault tolerance stately [52].

IV. CONCLUSION

Fault-tolerance is achieved by applying a set of analysis and design techniques to create systems with dramatically improved dependability. As new technologies are developed and new applications arise, new fault-tolerance approaches are also needed. In the early days of fault-tolerant computing, it was possible to craft specific hardware and software solutions from the ground up, but now chips contain complex, highly-integrated functions, and hardware and software must be crafted to meet a variety of standards to be economically viable. Thus a great deal of current research focuses on implementing fault tolerance using COTS (Commercial-Off-The-Shelf) technology.

Recent developments include the adaptation of existing fault-tolerance techniques to RAID disks where information is striped across several disks to improve bandwidth and a redundant disk is used to hold encoded information so that data can be reconstructed if a disk fails. Another area is the use of application-based fault-tolerance techniques to detect errors.

deep sub-micron VLSI devices to combat increasing noise problems and improve yield by tolerating defects that are likely to occur on very large, complex chips.

REFERENCES

[1] Peter Mell, Timothy Grance, "*NIST Definition of Cloud Computing*", Sept 2011, National Institute of Standards and technology, Gaithersburg, MD 20899-8930.

[2] Nimisha Singla, Seema Bawa "Priority Scheduling Algorithm with Fault Tolerance in Cloud Computing", International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), *3(12), December - 2013, pp. 645-652*

[3] Liang Luo, Wenjun Wu and Dichen Di,Fei Zhang,Yizhou Yan,Yaokuan Mao, „*A Resource Scheduling algorithm of Cloud Computing base d on Energy Efficient Optimization Methods*", IEEE 978-1-4673-2154-9, Vol. 12 (2012).

[4] Zhongyuan Lee, Ying Wang, Wen Zhou, „ *A dynamic priority scheduling algorithm on service request scheduling in cloud computing*", 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, IEEE 978 -1-61284-088-8, Vol. 11 (2011), PP: 4665-4669.

[5] Yanbing Liu, Shasha Yang, Qingguo Lin, and Gyoung-Bae Kim, „*Loyalty-Based Resource Allocation Mechanism in Cloud Computing*", Recent Advances in CSIE 2011, LNEE 125, PP: 233e–238.

[6] Hao Li and Guo Tang, „*Pareto-Based Optimal Scheduling on Cloud Resource'*, ICHCC 2011, CCIS 163, pp. 335–341, 2011.

[7] V.Vinothina, Sr.Lecturer, Dr.R.Sridaran, Dr.PadmavathiGanapathi, „*A Survey on Resource Allocation Strategies in Cloud Computing*", (IJACSA) International Journal of Advanced Computer Science and Applications, www.ijacsa.thesai.or g, Vol. 3, No.6, 2012, PP: 97 -104

[8] Ms.G.Gayathri1, Dr.N.Prabakaran2 "*Achieving Fault Tolerance in Cloud Environment by Efficient Load Balancing*", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 2, Issue 3, May – June , 2013 ISSN 2278-6856

[9] Win Win Naing "*Fault-tolerant Management for Private Cloud System*", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 1, Issue 1, May-June 2012 ISSN 2278-6856

[10] Sheheryar Malik, Fabrice Huet, " *Adaptive Fault Tolerance in Real Time Cloud Computing*", IEEE World Congress on Services, Jul 2011, Washington DC, United States. IEEE, pp.280-287.

[11] Han C.C., Shin K. G. and Wu J., "*A Fault-Tolerant Scheduling Algorithm for Real-Time Periodic Tasks with Possible Software Faults*", IEEE Computers 2003.

[12] Prasenjit Kumar Patra, Harshpreet Singh, Gurpreet Singh, **"***Fault Tolerance Techniques and Comparative Implementation in Cloud Computing***",** International Journal of Computer Applications (0975 – 8887) Volume 64– No.14, February 2013

[13] Jasbir Kaur, Supriya Kinger, "*Efficient Algorithm for Fault Tolerance in Cloud Computing*", International Journal of Computer Science and Information Technologies(IJCSIT), Vol.5 (5) , 2014, 6278-6281

[14] S. Sudha Lakshmi, Sri Padmavati, "*Fault Tolerance in Cloud Computing*", International Journal of Engineering Sciences Research-IJESR, Vol 04, Special Issue 01,2013, issn:2230-8504, e-ISSN-2230-8512.

[15] Pandeeswari.R, Mohamadi Begum "*Rsfts: Rule-Based Semantic Fault Tolerant Scheduling For Cloud Environment***",** Council for Innovative Research International Journal of Computers & Technology, . Volume 4 No. 2, March-April, 2013, ISSN 2277-3061

[16] L. D. Babu and P. Krishna, "*Honey bee behavior inspired load balancing of tasks in cloud computing environments*", in Applied Soft Computing, Vol. 13(5), pp. 2292-2303, (2013)

[17] R. Kaur and P. Luthra (2012), "*Load Balancing in Cloud Computing*", In Proceedings of International Conference on Recent Trends in Information, Telecommunication and Computing, ITC.