Aerial Surveillance using Edge based Approach: A perspective view

1.R.Chiranjeevi, 2.N.Venkatram, 3.R.Kamala

1.M.Tech student, Dept of CSE, Nova's Institute of Technology, Eluru

2.Assoc Professor, HOD of CSE Dept, Nova's Institute of Technology, Eluru

3.Asst Professor, Dept of CSE, Nova's Institute of Technology, Eluru

Abstract

The trend and technology is a very similar auto democratic terminology in the field of computer science. If we consider the fact of image in the perspective of Aerial, many algorithm and technology has already explained to counterpart with advanced human race. In this paper of abstract we like to give emphasis based on Aerial Surveillance where mapping a picture in the various angle or we can tell as parameter like angle of projection, method, and distance and resolution etc. to counterpart our assumption as of to maintain the best to the accuracy. If we consider the Military technology is in the field of air force where people give emphasis of projection to the angel of contact. Hence in this paper we tries to give emphasis on the fact of accuracy using the best algorithmic approach and computational effect of efficiency by using high bandwidth network and image quality in order to ensure the very best solution to the human race.

Keywords: Aerial Surveillance, Edge, Angel of projection, High Bandwidth Network, Dynamic Merging technology.

1. Introduction

Technology and it's introductory is the most base foundation which lead us to next level of research. In this context of paper, the word surveillance is the French word for "watching over"; "sur" means "from above" and "veiller" means "to watch". The inverse (reciprocal) of surveillance is surveillance ("to watch from below").The word surveillance may be applied to observation from a distance by means of electronic equipment (such as CCTV cameras), or interception of electronically transmitted information (such as Internet traffic or phone calls). It may also refer to simple, relatively no- or low-technology methods such as human intelligence agents and postal interception.

There are various Aerial Products has combined years of experience in aerostat manufacturing with extensive knowledge of secure data networking and image sensor expertise to manifesto the research in the very perfect way. The result is the LTAS Series - Aerostat Surveillance Systems and Mast Systems. The LTAS 100 is a completely new concept in aerial imaging, designed for urban environments, while the LTAS 200 and 300

series are more traditional technologies and are designed for borders, ports and areas where lighting conditions require thermal imagers. The technology is rumored to be able to scan objects down to four inches wide, and the military-grade equipment may be able to capture not just aerial views of cities, but also to probe further into the domestic arrangements of the general public.

The technology giant expects to soon announce new mapping technology to rival Google Maps, following Apple's acquisition of Swedish C3 Technologies an advanced 3D mapping business last year.

2. Related work

In the context of this paper, several different processes are needed for the computer analysis of aerial videos and much methodology as adopted to give the best of the technology. In this, we considered first, the static objects in the video frames must be recognized to determine the context of the events. Static objects might include forests, , roads, runways, and buildings, among others. Next the moving objects in the video must be detected, tracked, and identified. Moving objects include vehicles (cars, trucks, tanks, and buses) and people. Given the static objects and moving objects in a set of frames, events are by the actions of the moving objects and their interactions with the static objects. For example, two cars might pull on a road and stop together in a field. People might get out of the cars and approach each other for a meeting. A caravan of trucks might travel in one direction on a dirt road for a period of time and then make a U-turn and proceed in the opposite direction. A vehicle might pull up to a building and disappear into an underground garage or tunnel, then reappear some time later. In all of these cases, both the moving objects and the static objects must be recognized and their interactions noted.

Fig: 2.1 View Of the Surveillance of Dynamic projection

In the fig.2.1 it shows the method of angle and the 3D centric approach towards the destination. As the turn of the century approached yet another national debate on illegal immigration placed the spotlight on America's borders. The resulting attention increased agency budgets and abetted the propagation of technology. As a result, numerous Border Patrol stations along the southwest border installed Remote Video Surveillance Systems that, in some locations, covered several miles of the border. These surveillance systems included both infrared night cameras and closed circuit day cameras. The additional funding also provided a plethora of new technology which included handheld Gamma-Densitometers known as K910B Busters to detect hidden contraband, and fiber optic

Taking the concept to the extended version of telecom technology, Beyond GPRS, EDGE takes the cellular community one step closer to UMTS. It provides higher data rates than GPRS and introduces a new modulation scheme called 8-PSK. EDGE is also being adopted by the TDMA community for their migration to UMTS. Nuntius has solutions for the EDGE handset functions to help reach handset deployment quicker. For more information, contact Nuntius Systems, Inc.Many wireless data applications today can be implemented with 9.6kbit/s data. However, bandwidth-hungry fixed line data applications - web browsing, access to corporate data bases, and so on - would benefit from higher transmission speeds when used over the mobile network. HSCSD will significantly improve performance, especially for time-critical applications.

GPRS will enable cost-effective wireless access to applications that rely upon data bursts, adding packet switching to GSM with a packet-based air interface on top of the current circuit switched mode of operation. GPRS will provide the connectivity needed in packet-switched data networks such as the Internet.

In the context of surveillance, Image stitching has been widely used in many applications in preserve view to implement the best strategy in making a decision for the projection based on the angle and other parameters which is described in this paper of our methodology. In the context we take the basics of computer vision including how an image is formed from a camera, motion models and common image processing methods. This gives an outline of the image stitching techniques in the preview of distance technology toward the image. It presented a simplified flowchart of producing panoramic image from image acquisition to image remapping and finally to image blending. The concept introduces the automatic panoramic image stitching method using invariant features which enables the recent technology migration to panoramic suit. It provides detailed algorithm descriptions about how images can be automatically stitched using SIFT features and RANSAC homographic estimation makes though illustration of feature based image stitching. Theories about feature extraction, feature matching, homograph estimation, image wrapping and image blending are all introduced in this makes comprehensive illustration of two-view geometry. The concept of homographic and infinite homographic . It presents what functions are provided by the library and how these functions can be used to implement camera calibration and image stitching algorithms.

3. Methods

The methodology we follow in this context of surveillance is the trend of the most EDGE technology to minimize the network and overcome the best to the recent trend that has already provided. EDGE, expected to be deployed in 2000-2001, is a major improvement in GSM phase 2+. As a modification to existing GSM networks, EDGE does not require new network elements. EDGE is especially attractive to GSM 900, GSM 1800 and GSM 1900 operators that do not have a license for UMTS, but still wish to offer competitive personal multimedia applications utilizing the existing band allocation. Also, EDGE can co-exist with UMTS, for instance to provide high-speed services for wide-area coverage while UMTS is deployed in urban hot spots. Persistent aerial surveillance is an emerging domain with needs to assess ongoing activity in large areas for tasks such as force protection, traffic management, and urban planning. Automated analysis tools are important as the size of the area monitored and the number of objects to track is difficult to manage manually. Both tracking and activity analysis research in wide-area aerial surveillance (WAAS) video are recent and limited. Here, we introduce a scalable approach to handle the challenges inherent to tracking objects and analysis of traffic activity in such video.

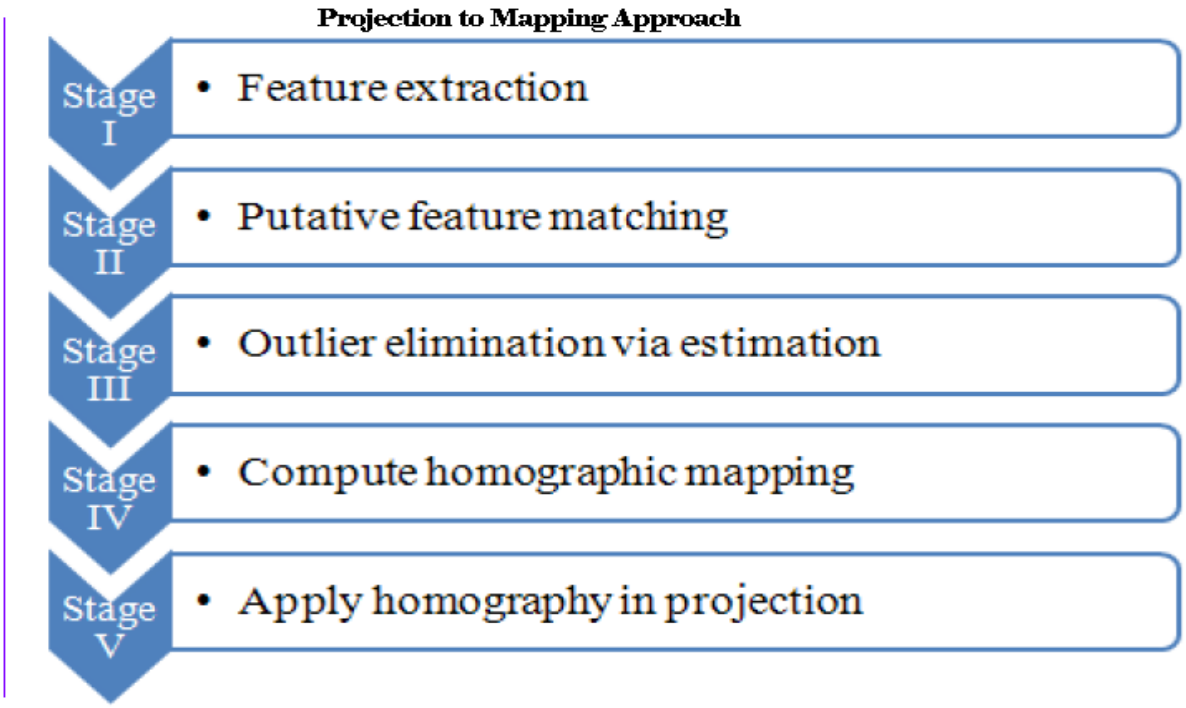


Fig.3.1 Feature Based Mosaic Approach to Projection Based on angel Method.

After features are extracted from each image set to be combined in the mosaic, putative feature matches are computed. The feature matching stage produces these putative matches which should have as high of a correct matching rate as possible, yet undoubtedly will produce some false matches. These false matches are then presumably identified in stage three by an estimation process. During this process the false matches are considered outliers to some sort of likelihood model and the correct matches or inliers are passed on to stage four. The fourth stage takes these estimated true matches and computes the geometric transformation which best maps the corresponding scene points from one image to the same scene points in the other. Many types of transformations are possible including simple Euclidean, similarity and perspective**.**

DSA protocol Based Approach:

In the Direction Specific Algorithm, We try to give emphasis on the path and id associated with approach, follows the following steps.

The Route Cache should support storing more than one source route for each destination.

If a node S is using a source route to some destination D that includes intermediate node N, S should shorten the route to destination D when it learns of a shorter route to node N than the one that is listed as the prefix of its current route to D. However, the cache should still retain the ability to revert to the

older, longer route to N if the shorter one does not work.

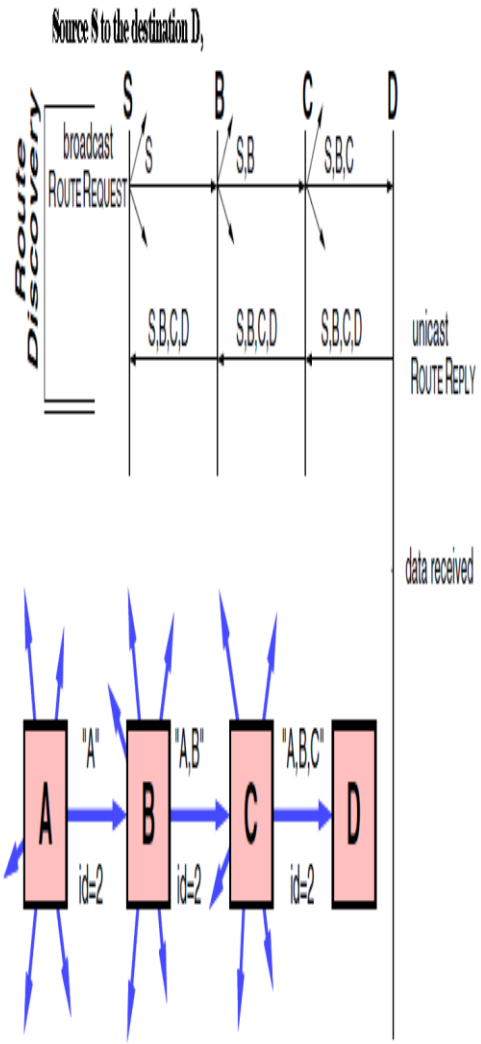


Fig.3.2 Showing the DSA approach to the ID based route.

In a typical streaming application, the execution of a particular task begins after data is received. The execution of a particular task is called id based execution.. At the end of execution of a task, data is produced. The data produced and consumed by the actors are called tokens. The rate at which each actor produces and consumes tokens is called token rate. In the description, each edge forms a channel for each actor to produce and consume tokens.

The technology makes the movement of various methods and also related work, in this paper we propose a new Aerial Surveillance vehicle detection approach which makes the advantages of existing systems avoid their counterpart. In the base of the proposed framework the fig. 3.2 explains the methodology of workflow in the framework. The framework consists of two axis approach trailing phase and implementation phase. In the trailing phase, we take consideration of several features which include projection angle corner features and vehicles colors, size, speed etc... In the implementation phase same feature extraction is also performed as in trailing phase. Afterwards the extracted features are used to classify pixels as vehicle pixel or non vehicle pixel using SVM. In this paper, we do not perform region based classification, which would highly depend on results of color segmentation algorithms such as mean shift. There is no need to generate multi-scale sliding windows either. The distinguishing feature of the proposed framework is that the detection task is based on pixel wise classification. However, the features are extracted in a neighborhood region of each pixel.

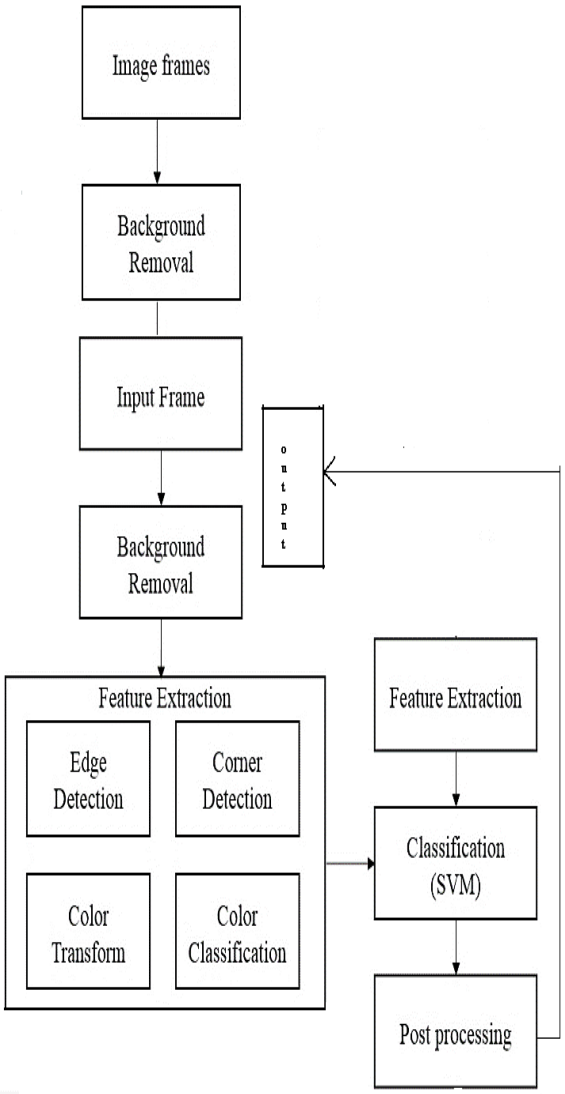


Fig.3.3 Showing the Framework (Proposed Two Vertical Method)

The input image is resized to enlarged image which is de-composed into 2 levels. Some of them are they form a compact representation, they encode edge information which is an important feature for vehicle detection, they capture information from multiple resolution levels and also there exist fast algorithms for computing these features.

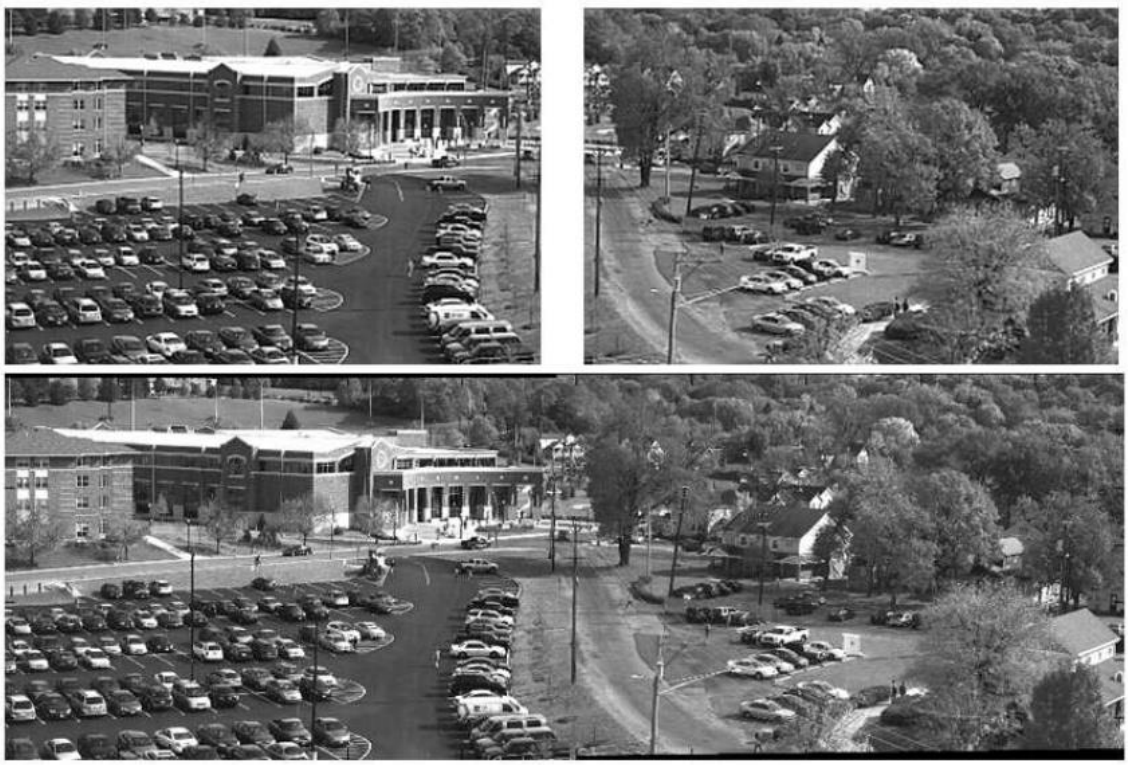


Fig.3.4 Surveillance HD image from High Angle of Projection

In the above fig.3.3. Showing the high crystal HD view in the context to field of Military in the best of the image to give a clear look, how the projection and it's related surrounding can be affected. In this paper, we try to the best to service of the mankind.

4. Conclusion And Recommendations

In the context of the analysis of requirements of feasibility, acceptability, and suitability, we tried to give emphasis on the concept of EDGE and the biased based networking algorithm, which lead to the next level of enhancement of the present solution. It was in this step wherein analysis screened out technology that could potentially enhance the capabilities of captured image under the moving condition, but either exceeded the agency's fiscal abilities, or proved too dangerous to operate. Adopting the light sport aircraft proved unacceptable due to a low survival rate in the event of an accident. Detected patterns include traffic flow direction, unit vs. bidirectional roads, acceleration or deceleration zones, and bidirectional stops.

5. Reference

[1] Air Force Research Laboratory (WPAFB,Dayton). Greene 2007 data collection.Sensor Data Management System.

[2] M.-F. Auclair-Fortier, D. Ziou, C. Armenakis, and S. Wang.Survey of work on road extraction in aerial and satellite images. Technical report, D´epartementdeMath´ematiquesetd'Informatique, Universit´e de Sherbrooke, 1999.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). CVIU, 110, 2008.

[4] S. Srinivasan, H. Latchman, J. Shea, T. Wong, and J. McNair, "Airbornetraffic surveillance systems: Video surveillance of highway traffic," in Proc. ACM 2nd Int. Workshop Video Surveillance Sens. Netw., 2004, pp. 131–135.

[5] A. C. Shastry and R. A. Schowengerdt, "Airborne video registration and traffic-flow parameter estimation," IEEE Trans. Intell. Transp. Syst., vol. 6, no. 4, pp. 391–405, Dec. 2005.

[6] H. Cheng and J.Wus, "Adaptive region of interest estimation for aerial surveillance video," in Proc. IEEE Int. Conf. Image Process., 2005, vol. 3, pp. 860–863.

[7] S. Hinz and A. Baumgartner, "Vehicle detection in aerial images using generic features, grouping, and context," in Proc. DAGM-Symp., Sep. 2001, vol. 2191, Lecture Notes in Computer Science, pp. 45–52.

[8] H. Cheng and D. Butler, "Segmentation of aerial surveillance video using a mixture of experts," in Proc. IEEE Digit. Imaging Comput. —Tech. Appl., 2005, p. 66.

[9] R. Lin, X. Cao, Y. Xu, C.Wu, and H. Qiao, "Airborne moving vehicle detection for urban traffic surveillance," in Proc. 11th Int. IEEE Conf. Intell. Transp. Syst., Oct. 2008, pp. 163–167.

[10] L. Hong, Y. Ruan, W. Li, D. Wicker, and J. Layne, "Energy-based video tracking using joint target density processing with an application to unmanned aerial vehicle surveillance," IET Comput. Vis., vol. 2, no. 1, pp. 1–12, 2008.

[11] R. Lin, X. Cao, Y. Xu, C.Wu, and H. Qiao, "Airborne moving vehicle detection for video surveillance of urban traffic," in Proc. IEEE Intell. Veh.Symp., 2009, pp. 203–208.

[12] J. Y. Choi and Y. K. Yang, "Vehicle detection from aerial images using local shape information," Adv. Image Video Technol., vol. 5414, Lecture Notes in Computer Science, pp. 227–236, Jan. 2009.

[13] C. G. Harris and M. J. Stephens, "A combined corner and edge detector," in Proc. 4th Alvey Vis. Conf., 1988, pp. 147–1

ONLINE MODELING PROACTIVE MODERATION SYSTEM FOR AUCTION FRAUD DETECTION

1.B.Anil Kumar 2.N.Venkataram 3. Y.N.Murali Krishna

1. M.Tech student,Dept of Computer Science,Nova's Institute of Technology,Eluru

2.Associate professor,HOD of Computer science,Nova's Institute of Technology,Eluru

3.Assistant professor,Dept of Computer science,Nova's Institute of Technology,Eluru

ABSTRACT

Online auction and shopping are gaining popularity with the growth of web-based ecommerce. Criminals are also taking advantage of these opportunities to conduct fraudulent activities against honest parties with the purpose of deception and illegal profit. In practice, proactive moderation systems are deployed to detect suspicious events for further inspection by human experts. Motivated by real-world applications in commercial auction sites in Asia, we develop various advanced machine learning techniques in the proactive moderation system. Our proposed system is formulated as optimizing bounded generalized linear models in multi-instance learning problems, with intrinsic bias in selective labeling and massive unlabeled samples an online pro bit model framework which takes online feature selection, coefficient Bounds from human knowledge and multiple instance learning into account simultaneously. By empirical experiments on a real-world online auction fraud detection data we show that this model can potentially detect more frauds and significantly reduce customer complaints compared to several baseline models and the human-tuned rule-based system.

1. INTRODUCTION

Since the emergence of the World Wide Web (WWW), electronic commerce, commonly known as e-commerce, has become more and more popular. Websites such as eBay and Amazon allow Internet users to buy and sell products and services online, which benefits everyone in terms of convenience and profitability. The traditional online shopping business model allows sellers to sell a product or service at a preset price, where buyers can choose to purchase if they find it to be a good deal. Online auction however is a different business model by which items are sold through price bidding. There is often a starting price and expiration time specified by the sellers. Once the auction starts, potential buyers bid against each other, and the winner gets the item with their highest winning bid. Similar to any platform supporting financial transactions, online auction attracts criminals to commit fraud. The varying types of auction fraud are as follows. Products purchased by the buyer are not delivered by the seller. The delivered products do not match the descriptions that were posted by sellers. Malicious sellers may even post non-existing items with false description to deceive buyers, and request payments to be wired directly to them via bank-to-bank wire transfer. Furthermore, some criminals apply phishing techniques to steal high-rated seller's accounts so that potential buyers can be easily deceived due to their good rating. Victims of fraud transactions usually lose their money and in most cases are not

recoverable. As a result, the reputation of the online auction services is hurt significantly due to fraud crimes. To provide some assurance against fraud, E-commerce sites often provide insurance to fraud victims to cover their loss up to a certain amount. To reduce the amount of such compensations and improve their online reputation, e-commerce providers often adopt the following approaches to control and prevent fraud. The identifies of registered users are validated through email, SMS, or phone verifications. A rating system where buyers provide feedbacks is commonly used in e-commerce sites so that fraudulent sellers can be caught immediately after the first wave of buyer com-plaints. In addition, proactive moderation systems are built to allow human experts to manually investigate suspicious

sellers or buyers. Even though e-commerce sites spend a large budget to fight frauds with a moderation system, there are still many outstanding and challenging cases. Criminals and fraudulent sellers frequently change their accounts and IP addresses to avoid being caught. Also, it is usually infeasible for human experts to investigate every buyer and seller to determine if they are committing fraud, especially when the e-commerce site attracts a lot of traffic. The patterns of fraudulent sellers often change constantly to take advantage of temporal trends. For instance, fraudulent sellers tend to sell the "hottest" products at the time to attract more potential victims. Also, whenever they find a loophole in the fraud detection system, they will immediately leverage the weakness. In this paper, we consider the application of a proactive moderation system for fraud detection in a major Asian on-line auction site, where hundreds of thousands of new auction cases are created every day. Due to the limited expert resources, only 20%-40% of the cases can be reviewed and labeled. Therefore, it is necessary to develop an automatic pre-screening moderation system that only directs suspicious cases for expert inspection, and passes the rest as clean cases. The moderation system for this site extracts rule-based features to make decisions. The rules are created by experts to represent the suspiciousness of sellers on fraudulence, and the resulting features are often binary. For instance, we can create a binary feature (rule) from the ratings of sellers, i.e. the feature value is 1 if the rating of a seller is lower than a threshold (i.e. a new account without many previous buyers); otherwise it is 0. The final moderation decision is based on the fraud score of each case, which is the linear weighted sum of those features, where the weights can be set by either human experts or machine-learned models. By deploying such a moderation system, we are capable of selecting a subset of highly suspicious cases for further expert investigation while keeping their workload at a reasonable level. The moderation system using machine-learned models is proven to improve fraud detection significantly over the human-tuned weights [38]. In [38] the authors considered the scenario of building offline models by using the previous 30 days data to serve the next day. Since the response is bi-nary (fraud or non-fraud) and the scoring function has to be linear, logistic regression is used. The authors have shown that applying expert knowledge, such as bounding the rule-based feature weights to be positive and multiple-instance learning, can significantly improve the performance in terms of detecting more frauds and reducing customer complaints given the same workload from human experts. However, offline models often meet the following challenges: (a) Since the auction fraud rate is generally very low (< 1%), the data becomes quite imbalanced and it is well-known that in such scenario even fitting simple logistic regression becomes a difficult problem [27]. Therefore, unless we use a large amount of historical training data, offline models tend to be fairly unstable. For example, in [38], 30 days of training data with around 5 million

samples are used for the daily update of the model. Hence it practically adds a lot of computation and memory load for each batch update, compared to online models. (b) Since the fraudulent sellers change their pattern very fast, it requires the model to also evolve dynamically. However, for offline models it is often non-trivial to address such needs. Once a case is determined as fraudulent, all the cases from this seller will be suspended immediately. Therefore smart fraudulent sellers tend to change their patterns quickly to avoid being caught; hence some features that are effective today might turn out to be not important tomorrow, or vice versa. Also, since the training data is from human labeling, the high cost makes it almost impossible to obtain a very large sample. Therefore for such systems (i.e. relatively small sample size with many features with temporal pattern), online feature selection is often required to pro-vide good performance. Human experts are also willing to see the results of online feature selection to monitor the effectiveness of the current set of features, so that they can understand the pattern of frauds and further add or remove some features. Our contribution In this paper we study the problem of building online models for the auction fraud detection moderation system, which essentially evolves dynamically over time. We propose a Bayesian pro bit online model frame-work for the binary response. We apply the stochastic search variable selection (SSVS), a well-known technique in statistical literature, to handle the dynamic evolution of the feature importance in a principled way. Note that we are not aware of any previous work that tries to embed SSVS into online modeling. We consider the expert knowledge to bind the rule-based coefficients to be positive. Finally, we consider combining this online model with multiple instances learning that gives even better empirical performance. We report the performance of all the above models through extensive experiments using fraud detection datasets from a major online auction website in Asia. The paper is organized as follows. In Section 2 we first summarize several specific features of the application and describe our online modeling framework with fitting details. We review working of our framework in Section 3. In Section 4 we show the experimental results that compare all the models proposed in this paper and several simple baselines.

2. OUR METHODOLOGY

Our application is to detect online auction frauds for a major Asian site where hundreds of thousands of new auction cases are posted every day. Every new case is sent to the proactive antifraud moderation system for pre-screening to assess the risk of being fraud. The current system is featured by:

• Rule-based features: Human experts with years of experience created many rules to detect whether a user is fraud or not. An example of such rules is “blacklist”, i.e. whether the user has been detected or complained as fraud before. Each rule can be regarded as a binary feature that indicates the fraud likeliness.

• Linear scoring function: The existing system only supports linear models. Given a set of coefficients (weights) on features, the fraud score is computed as the weighted sum of the feature values.

• Selective labeling: If the fraud score is above a certain threshold, the case will enter a queue for further investigation by human experts. Once it is reviewed, the final result will be labeled as boolean, i.e. fraud or clean. Cases with higher scores have higher priorities in the queue to be reviewed. The cases whose fraud score are below the threshold are determined as clean by the system without any human judgment.

• Fraud churn: Once one case is labeled as fraud by human experts, it is very likely that the seller

is not trustable and may be also selling other frauds; hence all the items submitted by the same seller are labeled as fraud too. The fraudulent seller along with his/her cases will be removed from the website immediately once detected.

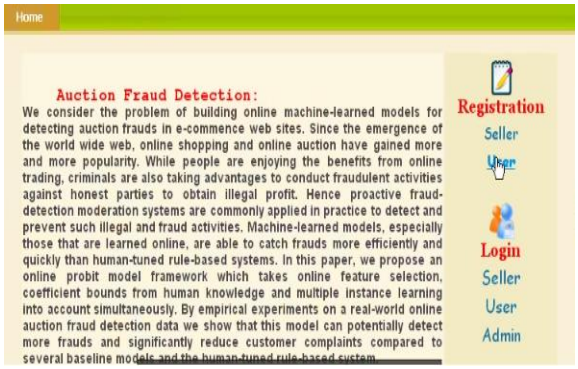
3. OUR WORKING PROCEDURE

• Rule-based features: Human experts with years of experience created many rules to detect whether a user is fraud or not. An example of such rules is "blacklist", i.e. whether the user has been detected or complained as fraud before. Each ule can be regarded as a binary feature that indicates the fraud likeliness.

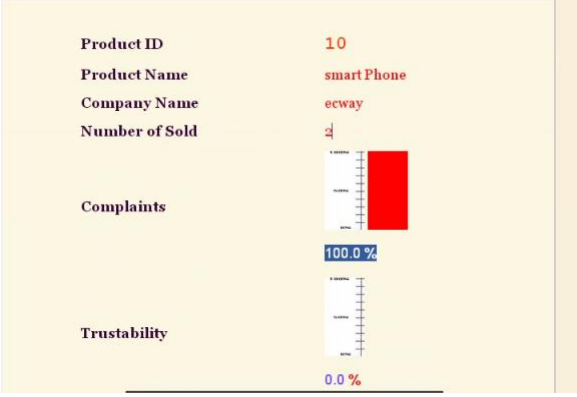
• Selective labeling: If the fraud score is above a certain threshold, the case will enter a queue for further investigation by human experts. Once it is reviewed, the final result will be labeled as Boolean, i.e. fraud or clean. Cases with higher scores have higher priorities in the queue to be reviewed. The cases whose fraud score are below the threshold are determined as clean by the system without any human judgment.

• Fraud churn: Once one case is labeled as fraud by human experts, it is very likely that the seller is not trustable and may be also selling other frauds; hence all the items submitted by the same seller are labeled as fraud too. The fraudulent seller along with his/her cases will be removed from the website immediately once detected.

• User Complaint: Buyers can file complaints to claim loss if they are recently deceived by fraudulent sellers. The Administrator's view the various types of complaints and the percentage of various type complaints. The complaints values of a products increase some threshold value the administrator set the trust ability of the product as un-trusted or banded. If the products set as banded, the user cannot view the products in the website.



4. EXPERIMENTAL RESULTS



REFERENCES

[1] D. Agarwal, B. Chen, and P. Elango. Spatio-temporal

models for estimating click-through rate. In

Proceedings of the 18th international conference on

World wide web, pages 21–30. ACM, 2009.

[2] S. Andrews, I. Tsochantaridis, and T. Hofmann.

Support vector machines for multiple-instance

learning. Advances in neural information processing

systems, pages 577–584, 2003.

[3] C. Bliss. The calculation of the dosage-mortality

curve. Annals of Applied Biology, 22(1):134–167, 1935.

[4] A. Borodin and R. El-Yaniv. Online computation and

competitive analysis, volume 53. Cambridge University

Press New York, 1998.

[5] L. Breiman. Random forests. Machine learning,

45(1):5–32, 2001.

[6] R. Brent. Algorithms for minimization without

derivatives. Dover Pubns, 2002.

[7] D. Chau and C. Faloutsos. Fraud detection in

electronic auction. In European Web Mining Forum

(EWMF 2005), page 87.

[8] H. Chipman, E. George, and R. McCulloch. Bart:

Bayesian additive regression trees. The Annals of

Applied Statistics, 4(1):266–298, 2010.

[9] W. Chu, M. Zinkevich, L. Li, A. Thomas, and

B. Tseng. Unbiased online active learning in data

streams. In Proceedings of the 17th ACM SIGKDD

international conference on Knowledge discovery and

data mining, pages 195–203. ACM, 2011.

[10] C. Chua and J. Wareham. Fighting internet auction

fraud: An assessment and proposal. Computer,

37(10):31–37, 2004.

[11] R. Collins, Y. Liu, and M. Leordeanu. Online selection

of discriminative tracking features. IEEE Transactions

on Pattern Analysis and Machine Intelligence, pages

1631–1643, 2005.

[12] N. Cristianini and J. Shawe-Taylor. An introduction to

support Vector Machines: and other kernel-based

learning methods. Cambridge university press, 2006.

[13] T. Dietterich, R. Lathrop, and T. Lozano-P´erez.

Solving the multiple instance problem with

axis-parallel rectangles. Artificial Intelligence,

89(1-2):31–71, 1997.

[14] Federal Trade Commission. Internet auctions: A guide

for buyers and sellers. http://www.ftc.gov/bcp/

conline/pubs/online/auctions.htm, 2004.

WWW 2012 – Session: Security 2 April 16–20, 2012, Lyon, France

THE ENTERPRISE CLOUD MANAGEMENT IN SUB CLOUD

1.M.Venkata padma sirisha 2.N.Venkataram 3 R.Kamala

1. M.Tech student,Dept of Computer Science,Nova's Institute of Technology,Eluru

2.Associate professor,HOD of Computer science,Nova's Institute of Technology,Eluru

3.Assistant professor,Dept of Computer science,Nova's Institute of Technology,Eluru

Abstract

In The use of cloud computing has increased rapidly in many organizations. General firms have already adopted the cloud computing and successfully applied to 'smart working' for security & remote-services. In addition, government organizations have recently introduced this system, which is gaining more users.There could be a problem regarding service efficiency such as private cloud lacking the capacity or resource for security. In order to solve these temporary difficulties, ways to support H/W resources (cpu, memory, network, etc) are necessary. For this reason, so-called 'multi-clouds' is needed to cooperate between simple 'single cloud's, providing accessible resources. Unlike the former studies which have regarded 'single cloud' as 'the solution' for every service and data capacity problem, this paper specifically focused on how to deal with the big data services that cannot be dealt with the that system. Thus this paper displays research on application of 'rain computing' which is 'cloud-of-cloud', the concept of multi-clouds.

Keywords: Cloud computing, single cloud, multi-clouds, cloud storage, data integrity, data intrusion, service availability.

1. Introduction

The use of cloud computing has increased rapidly in many organizations. Subashini and Kavitha argue that small and medium companies use cloud computing services for various reasons, including because these services provide fast access to their applications and reduce their infrastructure costs. Cloud providers should address privacy and security issues as a matter of high and urgent priority.

Dealing with "single cloud" providers is becoming less popular with customers due to potential problems such as service availability failure and the possibility that there are malicious insiders in the single cloud. In recent years, there has been a move towards "multi-clouds", "intercloud" or "cloud-of-clouds".

This paper focuses on the issues related to the data security aspect of cloud computing. As data and information will be shared with a third party, cloud computing users want to avoid an untrusted cloud provider. Protecting private and important information, such as credit card details or a patient's medical records from attackers or malicious insiders is of critical importance. In addition, the potential for migration from a single cloud to a multi-cloud environment is examined and research related to security issues in single and multi-clouds in cloud computing are surveyed.

The remainder of this paper is organized as follows. Section 2 describes the beginning of cloud computing and its components. In addition, it presents examples of cloud providers and the benefits of using their services. Section 3 discusses security risks in cloud computing. Section 4 analyses the new generation of cloud computing, that is, multi -clouds and recent solutions to address the security of cloud computing, as well as examining their limitations. Section 5 presents

suggestions for future work. Section 6 will conclude the paper.

2. Background

NIST describes cloud computing as "a model for enabling 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".

2.1 Cloud Computing Components

The cloud computing model consists of five characteristics, three delivery models, and four deployment models. The five key characteristics of cloud computing are: location-independent resource pooling, on-demand self-service, rapid elasticity, broad network access, and measured service. These five characteristics represent the first layer in the cloud environment architecture (see Figure1).

The three key cloud delivery models are infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). In IaaS, the user can benefit from networking infrastructure facilities, data storage and computing services. In other words, it is the delivery of computer infrastructure as a service. An example of IaaS is the Amazon web service. In PaaS, the user runs custom applications using the service provider's resources. It is the delivery of a computing platform and solution as a service. An example of PaaS is GoogleApps. Running software on the provider's infrastructure and providing licensed applications to users to use services is known as SaaS. An example of SaaS is the Salesforce.com CRM application. This model represents the second layer in the cloud environment architecture.

Cloud deployment models include public, private, community, and hybrid clouds. A cloud environment that is accessible for multi-tenants and is available to the public is called a public cloud. A private cloud is available for a particular group, while a community cloud is modified for a specific group of customers. Hybrid cloud infrastructure is a composition of two or more clouds (private, community, or public cloud). This model represents the third layer in the cloud environment architecture.

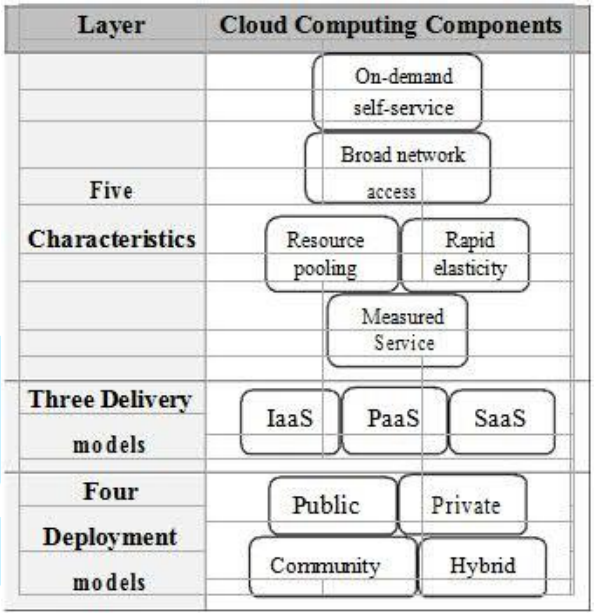


Figure 1: Cloud Environment Architecture.

Kamara and Lauter present two types of cloud infrastructure only, namely private and public clouds. The infrastructure that is owned and managed by users is in the private cloud. Data that is accessed and controlled by trusted users is in a safe and secure private cloud, whereas the infrastructure that is managed and controlled by the cloud service provider is in a public cloud. In particular, this data is out of the user's control, and is managed and shared with unsafe and untrusted servers.

2.2 Cloud Service Providers Examples

In the commercial world, various computing needs are provided as a service. The service providers take care of the customer's needs by, for example, maintaining software or purchasing

expensive hardware. For instance, the service EC2, created by Amazon, provides customers with scalable servers. As another example, under the CLuE program, NSF joined with Google and IBM to offer academic institutions access to a large-scale distributed infrastructure.

There are many features of cloud computing. First, cloud storages, such as Amazon S3, Microsoft SkyDrive, or NirvanixCLoudNAS, permit consumers to access online data. Second, it provides computation resources for users such as Amazon EC2. Third, Google Apps or versioning repositories for source code are examples of online collaboration tools .

Cloud service providers should ensure the security of their customers' data and should be responsible if any security risk affects their customers' service infrastructure. A cloud provider offers many services that can benefit its customers, such as fast access to their data from any location, scalability, pay-for-use, data storage, data recovery, protection against hackers, on-demand security controls, and use of the network and infrastructure facilities.

Reliability and availability are other benefits of the public cloud, in addition to low cost. However, there are also concerning issues for public cloud computing, most notably, issues surrounding data integrity and data confidentiality. Any customer will be worried about the security of sensitive information such as medical records or financial information.

3. Security Risks in Cloud Computing

Although cloud service providers can offer benefits to users, security risks play a major role in the cloud computing environment. Users of online data sharing or network facilities are aware of the potential loss of privacy. According to a recent IDC survey, the top challenge for 74% of CIOs in relation to cloud computing is security. Protecting private and important information such as credit card details or patients' medical records from attackers or malicious insiders is of critical importance. Moving databases to a large data centre involves many security challenges such as virtualization vulnerability, accessibility vulnerability, privacy and control issues related to data accessed from a third party, integrity, confidentiality, and data loss or theft. Subashini and Kavitha present some fundamental security challenges, which are data storage security, application security, data transmission security, and security related to third-party resources.

In different cloud service models, the security responsibility between users and providers is different. According to Amazon, their EC2 addresses security control in relation to physical, environmental, and virtualization security, whereas, the users remain responsible for addressing security control of the IT system including the operating systems, applications and data.

According to Tabakiet, the way the responsibility for privacy and security in a cloud computing environment is shared between consumers and cloud service providers differs between delivery models. In SaaS, cloud providers are more responsible for the security and privacy of application services than the users. This responsibility is more relevant to the public than the private cloud environment because the clients need more strict security requirements in the public cloud. In PaaS, users are responsible for taking care of the applications that they build and run on the platform, while cloud providers are responsible for protecting one user's applications from others. In IaaS, users are responsible for protecting operating systems and applications, whereas cloud providers must provide protection for the users' data.

Ristenpartet claim that the levels of security issues in IaaS are different. The impact of security issues in the public cloud is greater than the impact in the private cloud. For instance, any damage which occurs to the security of the physical infrastructure or any failure in relation to the

management of the security of the infrastructure will cause many problems. In the cloud environment, the physical infrastructure that is responsible for data processing and data storage can be affected by a security risk. In addition, the path for the transmitted data can be also affected, especially when the data is transmitted to many third-party infrastructure devices.

As the cloud services have been built over the Internet, any issue that is related to internet security will also affect cloud services. Resources in the cloud are accessed through the Internet; consequently even if the cloud provider focuses on security in the cloud infrastructure, the data is still transmitted to the users through networks which may be insecure. As a result, internet security problems will affect the cloud, with greater risks due to valuable resources stored within the cloud and cloud vulnerability. The technology used in the cloud is similar to the technology used in the Internet. Encryption techniques and secure protocols are not sufficient to protect data transmission in the cloud. Data intrusion of the cloud through the Internet by hackers and cybercriminals needs to be addressed and the cloud environment needs to be secure and private for clients.

We will address three security factors that particularly affect single clouds, namely data integrity, data intrusion, and service availability.

3.1 Data Integrity

One of the most important issues related to cloud security risks is data integrity. The data stored in the cloud may suffer from damage during transition operations from or to the cloud storage provider. Cachinet give examples of the risk of attacks from both inside and outside the cloud provider, such as the recently attacked Red Hat Linux's distribution servers. Another example of breached data occurred in 2009 in Google Docs, which triggered the Electronic Privacy Information Centre for the Federal Trade Commission to open an investigation into Google's Cloud Computing Services. Another example of a risk to data integrity recently occurred in Amazon S3 where users suffered from data corruption.

Cachinet argue that when multiple clients use cloud storage or when multiple devices are synchronized by one user, it is difficult to address the data corruption issue. One of the solutions that they propose is to use a Byzantine fault-tolerant replication protocol within the cloud. Hendricks state that this solution can avoid data corruption caused by some components in the cloud. However, Cachinet claim that using the Byzantine fault-tolerant replication protocol within the cloud is unsuitable due to the fact that the servers belonging to cloud providers use the same system installations and are physically located in the same place.

Although this protocol solves the problem from a cloud storage perspective, Cachinet argue that they remain concerned about the users' view, due to the fact that users trust the cloud as a single reliable domain or as a private cloud without being aware of the protection protocols used in the cloud provider's servers. As a solution, Cachinet suggest that using Byzantine fault -tolerant protocols across multiple clouds from different providers is a beneficial solution.

3.2 Data Intrusion

According to Garfinkel, another security risk that may occur with a cloud provider, such as the Amazon cloud service, is a hacked password or data intrusion. If someone gains access to an Amazon account password, they will be able to access all of the account's instances and resources. Thus the stolen password allows the hacker to erase all the information inside any virtual machine instance for the stolen user account, modify it, or even disable its services. Furthermore, there is a possibility for the user's email(Amazon user name) to be hacked, and since Amazon allows a lost password to be reset by

email, the hacker may still be able to log in to the account after receiving the new reset password.

3.3 Service Availability

Another major concern in cloud services is service availability. Amazon mentions in its licensing agreement that it is possible that the service might be unavailable from time to time. The user's web service may terminate for any reason at any time if any user's files break the cloud storage policy. In addition, if any damage occurs to any Amazon web service and the service fails, in this case there will be no charge to the Amazon Company for this failure. Companies seeking to protect services from such failure need measures such as backups or use of multiple providers. Both Google Mail and Hotmail experienced service down-time recently. If a delay affects payments from users for cloud storage, the users may not be able to access their data. Due to a system administrator error, 45% of stored client data was lost in LinkUp (MediaMax) as a cloud storage provider.

Garfinkel argues that information privacy is not guaranteed in Amazon S3. Data authentication which assures that the returned data is the same as the stored data is extremely important. Garfinkel claims that instead of following Amazon's advice that organizations encrypt data before storing them in Amazon S3, organizations should use HMAC technology or a digital signature to ensure data is not modified by Amazon S3. These technologies protect users from Amazon data modification and from hackers who may have obtained access to their email or stolen their password.

4. Multi-Clouds Computing Security

This section will discuss the migration of cloud computing from single to multi-clouds to ensure the security of the user's data.

4.1 Multi-Clouds: Preliminary

The term "multi-clouds" is similar to the terms "interclouds" or "cloud-of-clouds" that were introduced by Vukolic. These terms suggest that cloud computing should not end with a single cloud. Using their illustration, a cloudy sky incorporates different colors and shapes of clouds which leads to different implementations and administrative domains.

Recent research has focused on the multi-cloud environment which control several clouds and avoids dependency on any one individual cloud.

Cachin identify two layers in the multi-cloud environment: the bottom layer is the inner-cloud, while the second layer is the inter-cloud. In the inter-cloud, the Byzantine fault tolerance finds its place. We will first summarize the previous Byzantine protocols over the last three decades.

4.2 Introduction of Byzantine Protocols

In cloud computing, any faults in software or hardware are known as Byzantine faults that usually relate to inappropriate behavior and intrusion tolerance. In addition, it also includes arbitrary and crash faults. Much research has been dedicated to Byzantine fault tolerance (BFT) since its first introduction. Although BFT research has received a great deal of attention, it still suffers from the limitations of practical adoption and remains peripheral in distributed systems.

The relationship between BFT and cloud computing has been investigated, and many argue that in the last few years, it has been considered one of the major roles of the distributed system agenda. Furthermore, many describe BFT as being of only "purely academic interest" for a cloud service. This lack of interest in BFT is quite different to the level of interest shown in the mechanisms for tolerating crash faults that are used in large -scale systems. Reasons that reduce the adoption of BFT are, for example, difficulties in design, implementation, or understanding of BFT protocols.

As mentioned earlier, BFT protocols are not suitable for single clouds. Vukolic argues that one of the limitations of BFT for the inner-cloud is that BFT requires a high level of failure independence, as do all fault-tolerant protocols. If Byzantine failure occurs to a particular node in the cloud, it is reasonable to have a different operating system, different implementation, and different hardware to ensure such failure does not spread to other nodes in the same cloud. In addition, if an attack happens to a particular cloud, this may allow the attacker to hijack the particular inner-cloud infrastructure.

4.3 DepSky System: Multi-Clouds Model

This section will explain the recent work that has been done in the area of multi-clouds. Bessani present a virtual storage cloud system called DepSky which consists of a combination of different clouds to build a cloud-of-clouds. The DepSky system addresses the availability and the confidentiality of data in their storage system by using multi-cloud providers, combining Byzantine quorum system protocols, cryptographic secret sharing and erasure codes.

4.3.1 DepSky Architecture

The DepSky architecture consists of four clouds and each cloud uses its own particular interface. The DepSky algorithm exists in the clients' machines as a software library to communicate with each cloud (Figure 2). These four clouds are storage clouds, so there are no codes to be executed. The DepSky library permits reading and writing operations with the storage clouds.

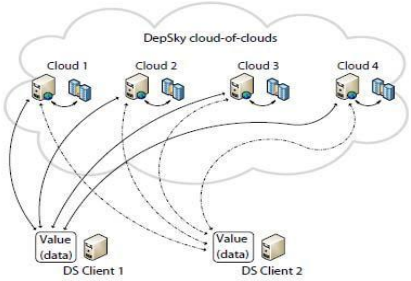


Figure 2:DepSky Architecture.

DepSky Data model. As the DepSky system deals with different cloud providers, the DepSky library deals with different cloud interface providers and consequently, the data format is accepted by each cloud. The DepSky data model consists of three abstraction levels: the conceptual data unit, a generic data unit, and the data unit implementation.

DepSKy System model. The DepSky system model contains three parts: readers, writers, and four cloud storage providers, where readers and writers are the client's tasks. Bessani explain the difference between readers and writers for cloud storage. Readers can fail arbitrarily (for example, they can fail by crashing, they can fail from time to time and then display any behavior) whereas, writers only fail by crashing.

Cloud storage providers in the DepSky system model. The Byzantine protocols involve a set of storage clouds (n) where n = 3 f +1, and f is maximum number of clouds which could be faulty. In addition, any subset of (n – f) storage cloud creates byzantine quorum protocols.

4.4 Analysis of Multi-Cloud Research

Moving from single clouds or inner-clouds to multi-clouds is reasonable and important for many reasons. According to Cachinet Services of single clouds are still subject to outage". In addition, Bowers showed that over 80% of company management "fear security threats and loss of control of data and systems". Vukolic assumes that the main purpose of moving to interclouds is to improve what was offered in single clouds by distributing reliability, trust, and security among multiple cloud providers. In addition, reliable

distributed storage which utilizes a subset of BFT techniques was suggested by Vukolic to be used in multi-clouds. A number of recent studies in this area have built protocols for interclouds. RACS (Redundant Array of Cloud Storage) for instance, utilizes RAID -like techniques that are normally used by disks and file systems, but for multiple cloud storage. Abu-Libdeh assume that to avoid "vender lock -in", distributing a user's data among multiple clouds is a helpful solution. This replication also decreases the cost of switching providers and offers better fault tolerance. Therefore, the storage load will be spread among several providers as a result of the RACS proxy.

HAIL (High Availability and Integrity Layer) is another example of a protocol that controls multiple clouds. HAIL is a distributed cryptographic system that permits a set of servers to ensure that the client's stored data is retrievable and integral. HAIL provides a software layer to address availability and integrity of the stored data in an intercloud.

Cachin present a design for intercloud storage (ICStore), which is a step closer than RACS and HAIL as a dependable service in multiple clouds. Cachin develop theories and protocols to address the CIRC attributes (confidentiality, integrity, reliability and consistency) of the data stored in clouds.

As mentioned before, Bessani present a virtual storage cloud system called DepSky consisting of a combination of different clouds to build a cloud-of-clouds. Bessani discuss some limitations of the HAIL protocol and RACS system when compared with DepSky. HAIL does not guarantee data confidentiality, it needs code execution in their servers, and it does not deal with multiple versions of data. None of these limitations are found in DepSky, whereas the RACS system differs from the DepSky system in that it deals with "economic failures" and vendor lock-in and does not address the issue of cloud storage security problems. In addition, it also does not provide any mechanism to ensure data confidentiality or to provide updates of the stored data. Finally, the DepSky system presents an experimental evaluation with several clouds, which is different from other previous work on multi-clouds.

There are a number of studies on gaining constancy from untrusted clouds. For instance, similar to DepSky, Depot improves the flexibility of cloud storage, as Mahajan et al. believe that cloud storages face many risks. However, Depot provides a solution that is cheaper due to using single clouds, but it does not tolerate losses of data and its service availability depends on cloud availability. Other work which implements services on top of untrusted clouds are studies such as SPORC and Venus. These studies are different from the DepSky system because they consider a single cloud (not a cloud -of-clouds). In addition, they need code execution in their servers. Furthermore, they offer limited support for the unavailability of cloud services in contrast to DepSky.

4.5 Current Solutions of Security Risks

In order to reduce the risk in cloud storage, customers can use cryptographic methods to protect the stored data in the cloud. Using a hash function is a good solution for data integrity by keeping a short hash in local memory. In this way, authentication of the server responses is done by recalculating the hash of the received data which is compared with the local stored data. If the amount of data is large, then a hash tree is the solution. Many storage system prototypes have implemented hash tree functions, such as SiRiUS and TDB. Mykletun and Papamanthou claim that this is an active area in research on cryptographic methods for stored data authentication. Cachinet argue that although the previous methods allow consumers to ensure the integrity of their data which has been returned by servers, they do not guarantee that the server will answer a query without knowing what that query is and whether the data is stored correctly in the server or not. Proofs of Retrievability (PORs) and Proofs of Data Possession (PDP) are protocols introduced by

Juels and Kaliski and Ateniese to ensure high probability for the retrieval of the user's data. Cachinet suggest using multiple cloud providers to ensure data integrity in cloud storage and running Byzantine-fault -tolerant protocols on them where each cloud maintains a single replica. Computing resources are required in this approach and not only storage in the cloud, such a service provided in Amazon EC2, whereas if only storage service is available, Cachin suggest working with Byzantine Quorum Systems by using Byzantine Disk Paxos and using at least four different clouds in order to ensure users' atomicity operations and to avoid the risk of one cloud failure.

As mentioned earlier, the loss of availability of service is considered one of the main limitations in cloud computing and it has been addressed by storing the data on several clouds. The loss of customer data has caused many problems for many users such as the problem that occurred in October 2009 when the contacts, photos, etc. of many users of the Sidekick service in Microsoft were lost for several days.

Bessani use Byzantine fault-tolerant replication to store data on several cloud servers, so if one of the cloud providers is damaged, they are still able to retrieve data correctly. Data encryption is considered the solution by Bessani to address the problem of the loss of privacy. They argue that to protect the stored data from a malicious insider, users should encrypt data before it is stored in the cloud. As the data will be accessed by distributed applications, the DepSky system stores the cryptographic keys in the cloud by using the secret sharing algorithm to hide the value of the keys from a malicious insider.

In the DepSky system, data is replicated in four commercial storage clouds (Amazon S3,Windows Azure, Nirvanix and Rackspace); it is not relayed on a single cloud, therefore, this avoids the problem of the dominant cloud causing the so-called vendor lock-in issue. In addition, storing half the amount of data in each cloud in the DepSky system is achieved by the use of erasure codes. Consequently, exchanging data between one provider to another will result in a smaller cost. The DepSky system aims to reduce the cost of using four clouds(which is four times the overhead) to twice the cost of using a single cloud, which is a significant advantage.

DepSky uses a set of Byzantine quorum system protocols in order to implement the read and write operations in the system, so it needs only two communication round trips for each operation to deal with several clouds. The use of several clouds needs a variety of locations, administration, design and implementation, which are the requirements of the Byzantine quorum systems protocols. Executing codes in servers is not required in the DepSky system (storage clouds) in contrast to other Byzantine protocols that need some code execution. After using these protocols, the DepSky system aims to deal with data confidentiality by decreasing the stored amount of data in each cloud.

4.6 Limitation of Current Solutions

The problem of the malicious insider in the cloud infrastructure which is the base of cloud computing is considered by Rocha and Correia. IaaS cloud providers provide the users with a set of virtual machines from which the user can benefit by running software on them. The traditional solution to ensure data confidentiality by data encryption is not sufficient due to the fact that the user's data needs to be manipulated in the virtual machines of cloud providers which cannot happen if the data has been encrypted. Administrators manage the infrastructure and as they have remote access to servers, if the administrator is a malicious insider, then he can gain access to the user's data. Van Dijk and Juels present some negative aspects of data encryption in cloud computing. In addition, they assume that if the data is processed from different clients, data encryption cannot ensure privacy in the cloud.

Although cloud providers are aware of the malicious insider danger, they assume that they

have critical solutions to alleviate the problem. Rocha and Correia determine possible attackers for IaaS cloud providers. For example, Grosse propose one solution is to prevent any physical access to the servers. However, Rocha and Correia argue that the attackers outlined in their work have remote access and do not need any physical access to the servers. Grosse propose another solution is to monitor all access to the servers in a cloud where the user's data is stored. However, Rocha and Correia claim that this mechanism is beneficial for monitoring employee's behavior in terms of whether they are following the privacy policy of the company or not, but it is not effective because it detects the problem after it has happened.

Rocha and Correia classified four types of attacks that can affect the confidentiality of the user's data in the cloud. These four types of attacks could occur when the malignant insider can determine text passwords in the memory of a VM, cryptographic keys in the memory of VM files, and other confidential data. In addition, they argue that the recent research mechanisms are not good enough to consider the issue of data confidentiality and to protect data from these attacks. This does not mean that these mechanisms are not useful; rather that they do not focus on solving the problems that Rocha and Correia address in their research. Some of the solutions are mechanisms and are used as part of cloud computing solutions, while different types of solutions focus on solving the whole data confidentiality issue intrinsic to cloud computing. Rocha and Correia suggest trusted computing and distributing trust among several cloud providers as a novel solution to solving security problems and challenges in cloud computing. The idea of replicating data among different clouds has been applied in the single system DepSky. Rocha and Correia present the limitations of this work which occurs due to the fact that DepSky is only a storage service like Amazon S3, and does not offer the IaaS cloud model. On the other hand, this system provides a secure storage cloud, but does not provide security of data in the IaaS cloud model. This is because it uses data encryption and stores the encrypted key in the clouds by using a secret sharing technique, which is inappropriate for the IaaS cloud model.

Table 1 details the security risks addressed in the previous research and the mechanisms that have been proposed as a solution for these security risks in the cloud computing environment. Security risk issues in cloud computing have attracted much research interest in recent years.

It is clear from the table that in the past more research has been conducted into single clouds than into multi-clouds. Multi–clouds can address the security issues that relate to data integrity, data intrusion, and service availability in multi-clouds. In addition, most of the research has focused on providing secure "cloud storage" such as in DepSky. Therefore, providing a cloud database system, instead of normal cloud storage, is a significant goal in order to run queries and deal with databases; in other words, to profit from a database-as-a-service facility in a cloud computing environment.

Table 1 illustrates that in 2009, 67% of the research on security in cloud computing addressed the issue of a single cloud, whereas 33% of the research in the same year addressed the issue of multi-clouds. In 2010, 80% of research focused on single clouds while only 20% or research was directed in the area of multi-clouds.

Ref	Year	Addressed Security Risks				Type of cloud		Type of service	
		Securitv Cl	itvDat integr	ionDa intrus	Servic availabilit	Singl e	Multi	Cloud	Cloud

							cloud	clouds	storage	database
[5]	2011	√	√			Multi shares+ secret sharing algorithm				
[8]	2011	√	√	√	√	DepSky,(Byzantine + secret sharing + cryptography)		√	√	
[42]	2011	√ survey	√				√		√	
[3]	2010	√				RAID-like techniques+ introduced RACS		√	√	
[11]	2010	√	√			ICStore ,(client-centric distributed protocols)		√	√	
[17]	2010	√			√	SPORC, (fork)	√			
[22]	2010	√								
[25]	2010	√				cryptography	√		√	
[30]	2010					Depot, (FJC)	√		√	
[48]	2010	√	√			Venus	√		√	
[49]	2010	√ survey	√		√		√		√	
[51]	2010	√					√		√	
[52]	2010	√	√				√		√	
[10]	2009	√	√		√	HAIL (Proofs + cryptography)		√	√	
[12]	2009	√ survey	√					√	√	
[16]	2009	√	√			encrypted cloud VPN	√		√	
[41]	2009	√					√		√	
[43]	2009	√	√		√	TCCP	√		√	
[55]	2009	√	√			homomorphic token + erasure-coded	√		√	

[7]	2007		√			PDP schemes				
[19]	2007	√					√		√	

Table 1. Related Work on Cloud Computing Security.

5. Future Work

For future work, we aim to provide a framework to supply a secure cloud database that will guarantee to prevent security risks facing the cloud computing community. This framework will apply multi -clouds and the secret sharing algorithm to reduce the risk of data intrusion and the loss of service availability in the cloud and ensure data integrity.

In relation to data intrusion and data integrity, assume we want to distribute the data into three different cloud providers, and we apply the secret sharing algorithm on the stored data in the cloud provider. An intruder needs to retrieve at least three values to be able to find out the real value that we want to hide from the intruder. This depends on Shamir's secret sharing algorithm with a polynomial function technique which claims that even with full knowledge of (k – 1) clouds, the service provider will not have any knowledge of vs (vs is the secret value). We have used this technique in previous databases-as-a-serves research. In other words, hackers need to retrieve all the information from the cloud providers to know the real value of the data in the cloud. Therefore, if the attacker hacked one cloud provider's password or even two cloud provider's passwords, they still need to hack the third cloud provider (in the case where k = 3) to know the secret which is the worst case scenario. Hence, replicating data into multi-clouds by using a multi-share technique may reduce the risk of data intrusion and increase data integrity. In other words, it will decrease the risk of the Hyper-Visor being hacked and Byzantine fault-tolerant data being stolen from the cloud provider.

Regarding service availability risk or loss of data, if we replicate the data into different cloud providers, we could argue that the data loss risk will be reduced. If one cloud provider fails, we can still access our data live in other cloud providers. This fact has been discovered from this survey and we will explore dealing with different cloud provider interfaces and the network traffic between cloud providers.

6. Conclusion

It is clear that although the use of cloud computing has rapidly increased; cloud computing security is still considered the major issue in the cloud computing environment. Customers do not want to lose their private information as a result of malicious insiders in the cloud. In addition, the loss of service availability has caused many problems for a large number of customers recently. Furthermore, data intrusion leads to many problems for the users of cloud computing. The purpose of this work is to survey the recent research on single clouds and multi-clouds to address the security risks and solutions. We have found that much research has been done to ensure the security of the single cloud and cloud storage whereas multi-clouds have received less attention in the area of security. We support the migration to multi-clouds due to its ability to decrease security risks that affect the cloud computing user.

7. References

[1] (NIST), http://www.nist.gov/itl/cloud/.

[2] Abraham, G. Chockler, I. Keidar and D. Malkhi, "Byzantine disk paxos: optimal resilience with Byzantine shared

memory", Distributed Computing, 18(5), 2006, pp. 387-408.

[3] H. Abu-Libdeh, L. Princehouse and H. Weatherspoon, "RACS: a case for cloud storage diversity", SoCC'10:Proc. 1st ACM symposium on Cloud computing, 2010, pp. 229-240.

[4] D. Agrawal, A. El Abbadi, F. Emekci and A. Metwally, "Database Management as a Service: Challenges and Opportunities", ICDE'09:Proc.25thIntl. Conf. on Data Engineering, 2009, pp. 1709-1716.

[5] M.A. AlZain and E. Pardede, "Using Multi Shares for Ensuring Privacy in Database-as-a-Service", 44th Hawaii Intl. Conf. on System Sciences (HICSS), 2011, pp. 1-9.

[6] Amazon, Amazon Web Services. Web services licensing agreement, October3,2006.

[7] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson and D. Song, "Provable data possession at untrusted stores", Proc. 14th ACM Conf. on Computer and communications security, 2007, pp. 598-609.

[8] Bessani, M. Correia, B. Quaresma, F. André and P. Sousa, "DepSky: dependable and secure storage in a cloud-of-clouds", EuroSys'11:Proc. 6thConf. on Computer systems, 2011, pp. 31-46.

[9] K. Birman, G. Chockler and R. van Renesse,"Toward a cloud computing research agenda", SIGACT News, 40, 2009, pp. 68-80.

[10] K.D. Bowers, A. Juels and A. Oprea, "HAIL: A high-availability and integrity layer for cloud storage", CCS'09: Proc. 16th ACM Conf. on Computer and communications security, 2009, pp. 187-198.

[11] C. Cachin, R. Haas and M. Vukolic, "Dependable storage in the Intercloud", Research Report RZ, 3783, 2010.

[12] C. Cachin, I. Keidar and A. Shraer, "Trusting the cloud", ACM SIGACT News, 40, 2009, pp. 81-86.

[13] C. Cachin and S. Tessaro, "Optimal resilience for erasure-coded Byzantine distributed storage", DISC:Proc. 19thIntl.Conf. on Distributed Computing, 2005, pp. 497-498.

[14] M. Castro and B. Liskov, "Practical Byzantine fault tolerance", Operating Systems Review, 33, 1998, 173-186.

[15] G. Chockler, R. Guerraoui, I. Keidar and M. Vukolic, "Reliable distributed storage", Computer, 42, 2009, pp. 60-67.

[16] Clavister, "Security in the cloud", Clavister White Paper, 2008.

[17] A.J. Feldman, W.P. Zeller, M.J. Freedman and E.W. Felten, "SPORC: Group collaboration using untrusted cloud resources", OSDI, October2010, 1-14.

[18] S.L. Garfinkel, "Email-based identification and authentication: An alternative to PKI?", IEEE Security and Privacy, 1(6), 2003, pp. 20-26.

[19] S.L. Garfinkel, "An evaluation of amazon's grid computing services: EC2, S3, and SQS", Technical Report TR-08-07, Computer Science Group, Harvard University, Citeseer, 2007, pp. 1-15.

[20] E. . Goh, H. Shacham, N. Modadugu and D. Boneh, "SiRiUS: Securing remote

untrusted storage",NDSS: Proc. Network and Distributed System Security Symposium, 2003, pp. 131–145.

[21] G.R. Goodson, J.J. Wylie, G.R. Ganger and M.K. Reiter, "Efficient Byzantine-tolerant erasure-coded storage",DSN'04: Proc.Intl. Conf. on Dependable Systems and Networks,2004, pp.1-22.

[22] E. Grosse, J. Howie, J. Ransome, J. Reavis and S. Schmidt, "Cloud computing roundtable", IEEE Security & Privacy, 8(6), 2010, pp. 17-23.

[23] J. Hendricks, G.R. Ganger and M.K. Reiter, "Low-overhead byzantine fault-tolerant storage", SOSP'07: Proc. 21st ACM SIGOPS symposium on Operating systems principles, 2007, pp. 73-86.

[24] Juels and B.S. Kaliski Jr, "PORs: Proofs of retrievability for large files", CCS '07: Proc. 14th ACM Conf. on Computer and communications security, 2007, pp. 584-597.

[25] S. Kamara and K. Lauter, "Cryptographic cloud storage", FC'10: Proc. 14thIntl.Conf. on Financial cryptograpy and data security,2010, pp. 136-149.

[26] H. Krawczyk, M. Bellare and R. Canetti, "HMAC: Keyed-hashing for message authentication", Citeseer, 1997, pp. 1-11.

[27] P. Kuznetsov and R. Rodrigues, "BFTW 3: why? when? where? workshop on the theory and practice of byzantine fault tolerance", ACM SIGACT News, 40(4),2009, pp. 82-86.

[28] L. Lamport, R. Shostak and M. Pease, "The Byzantine generals problem", ACM Transactions on Programming Languages and Systems, 4(3), 1982, pp. 382-401.

[29] P.A. Loscocco, S.D. Smalley, P.A. Muckelbauer, R.C. Taylor, S.J. Turner and J.F. Farrell, "The inevitability of failure: The flawed assumption of security in modern computing environments", Citeseer, 1998, pp. 303-314.

[30] P. Mahajan, S. Setty, S. Lee, A. Clement, L. Alvisi, M. Dahlin and M. Walfish, "Depot: Cloud storage with minimal trust", OSDI'10: Proc. of the 9th USENIX Conf. on Operating systems design and implementation, 2010, pp. 1-16.

[31] U. Maheshwari, R. Vingralek and W. Shapiro, "How to build a trusted database system on untrusted storage", OSDI'00: Proc. 4thConf. on Symposium on Operating System Design & Implementation, 2000, p. 10.

[32] D. Malkhi and M. Reiter, "Byzantine quorum systems", Distributed Computing, 11(4),1998, pp. 203-213.

[33] J.-P. Martin, L. Alvisi and M. Dahlin, "Minimal byzantine storage", DISC '02: Proc. of the 16thIntl. Conf. on Distributed Computing, 2002, pp. 311-325.

[34] H.Mei, J. Dawei, L. Guoliang and Z. Yuan, "Supporting Database Applications as a Service", ICDE'09:Proc. 25thIntl.Conf. on Data Engineering, 2009, pp. 832-843.

[35] R.C. Merkle, "Protocols for public key cryptosystems", IEEE Symposium on Security and Privacy, 1980, pp. 122-134.

[36] E. Mykletun, M. Narasimha and G. Tsudik, "Authentication and integrity in

outsourced databases", ACM Transactions on Storage (TOS), 2,2006, pp. 107-138.

[37] C. Papamanthou, R. Tamassia and N. Triandopoulos, "Authenticated hash tables", CCS '08: Proc. 15th ACM Conf. on Computer and communications security, 2008, pp. 437-448.

[38] M. Pease, R. Shostak and L. Lamport, "Reaching agreement in the presence of faults", Journal of the ACM, 27(2), 1980, pp. 228-234.

[39] R. Perez, R. Sailer and L. van Doorn, "vTPM: virtualizing the trusted platform module", Proc. 15th Conf. on USENIX Security Symposium,2006, pp. 305-320.

[40] RedHat, https://rhn.redhat.com/errata/RHSA-2008-0855.html.

[41] T. Ristenpart, E. Tromer, H. Shacham and S. Savage, "Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds", CCS'09: Proc. 16thACM Conf. on Computer and communications security, 2009, pp. 199-212.

[42] F. Rocha and M. Correia, "Lucy in the Sky without Diamonds: Stealing Confidential Data in the Cloud", Proc. 1stIntl. Workshop of Dependability of Clouds, Data Centers and Virtual Computing Environments, 2011, pp. 1-6.

[43] N. Santos, K.P. Gummadi and R. Rodrigues, "Towards trusted cloud computing", USENIX Association, 2009, pp. 3-3.

[44] D. Sarno, "Microsoft says lost sidekick data will be restored to users", Los Angeles Times, October 2009.

[45] F. Schneider and L. Zhou, "Implementing trustworthy services using replicated state machines", IEEE Security and Privacy, 3(5),2010, pp. 151-167.

[46] G. Brunette and R. Mogull (eds), "Security guidance for critical areas of focus in cloud computing", CloudSecurityAlliance, 2009.

[47] Shamir, "How to share a secret", Communications of the ACM, 22(11), 1979, pp. 612-613.

[48] Shraer, C. Cachin, A. Cidon, I. Keidar, Y. Michalevsky and D. Shaket, "Venus: Verification for untrusted cloud storage", CCSW'10: Proc. ACM workshop on Cloud computing security workshop, 2010, pp. 19-30.

[49] S. Subashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing", Journal of Network and Computer Applications, 34(1), 2011, pp 1-11.

[50] Sun, http://blogs.sun.com /gbrunett/entry/ amazon_ s3_silent_data_corruption.

[51] H. Takabi, J.B.D. Joshi and G.-J. Ahn, "Security and Privacy Challenges in Cloud Computing Environments", IEEE Security & Privacy, 8(6), 2010, pp. 24-31.

[52] M. Van Dijk and A. Juels, "On the impossibility of cryptography alone for privacy-preserving cloud computing", HotSec'10: Proc. 5thUSENIX Conf. on Hot topics in security, 2010, pp.1-8.

[53] J. Viega, "Cloud computing and the common man", Computer, 42, 2009, pp. 106-108.

[54] M. Vukolic,"The Byzantine empire in the intercloud", ACM SIGACT News, 41,2010, pp. 105-111.

[55] Wang, Q. Wang, K. Ren and W. Lou, "Ensuring data storage security in cloud computing", ARTCOM'10: Proc. Intl. Conf. on Advances in Recent Technologies in Communication and Computing, 2010, pp. 1-9.

CREATION OF NEW ENHANCED ANONYMIZATION NETWORK USING DISTRIBUTED NYMBLE MANAGERS

1.G.RamaKrishna, 2. N.Venkatram, 3. M.Nirupa

1.M.Tech student, Dept of CSE, Nova's Institute of Technology, Eluru

2.Assoc Professor, HOD of CSE Dept, Nova's Institute of Technology, Eluru

3.Asst Professor, Dept of CSE, Nova's Institute of Technology, Eluru

ABSTRACT

Anonymization networks are creates with the help of anonymization routers. Anonymization routers are important resources for control the attackers in internet services. Anonymization routers are provides the successful solution in blocking of misbehaving users. Existing system anonymization networks are contains single Nymble manager related credential system. It is not effective under blocking of misbehaving users. Some issues and risks are present in current nimble system. Those risks are showing the problems in performance like scalability and efficiency. Present systems are not possible to detect all dimensions of attacker's detection.

Overcome above all systems problems we introduce new credentials systems with group of Nymble managers. In different nymble managers itself we place different properties for filter the attackers. Using different levels of properties start the detection. In detection environment we use the sequential patterns procedure. Patterns consist of different rules and constraints. One constraint is depends on another constraint. First constraint is authorized we allow in second constraint. We control the more number of attackers. This systems save the sensitive data in different e-commerce applications. This system gives the accurate and scalable solutions in detection of attackers.

KEYWORDS: Distributed Nymble managers, Anonymization Networks, Cryptographic techniques.

INTRODUCTION

All internet applications are offer high quality service providers. We use the Service providers for identification of unnecessary or unauthentication users. Many users collect the service by internet environment. All Different locations of users are forward the requests to web server. In communication system some intermediate routers are present. In those intermediate routers apply the anonymization techniques for detection of unauthentication users here. Some kinds of web servers are unable to detect all attackers.

Some problems and disadvantages are present in wireless ad-hoc networks. Incorrect user's detection is possible in some

cases. It is not possible to detect the attacks all cases here. In detection environment some properties are available in sender and receiver also. In two components start the communication here. In communication stage path is present. There is no perfect path identification, how the attacker detection is available here. It's impossible task for detection of misbehavior user.

Now in proposed distributed nymble systems add the new properties for detection of unauthentication users. It is improved system under detection of attackers. In following chapters implementation steps are available here.

II.RELATED WORK

Protection is main vital role in accessing of data under allow process. Using network security concepts we trade the applications. Many number of application owners are requires new advanced features for control the different attackers. Previous existing approaches are failing under detection of attackers or control the attackers. Attackers are entering in different locations, all locations of attackers detection is not possible with previous anonymization approaches. Previous anonymization approaches are not defining total features for detection of various attacks.

First in existing approach, symmetric encryption and message authentication code schemes are introduced here. These are verification approaches based on hash functions. Hash functions related verification possible in particular time periods only. Some other new time intervals it's not possible to detect the attacks. Present approaches are controls the less amount attacks in less dimensions only [1][2].

Second approach, trusted third party related approaches are introduced here. Trusted third party users are verification managers. Verification managers start the verification based on signatures. It is one of the robust technique in detection of attacks, but performance is increases. This approach computational cost is high [2]. It is related low speed attack detection approach.

After some days new approaches are introduced in market. Those approaches are blacklist and independent blacklist. Blacklists are separate for every location. One attacker is entering in network, we detect and store into blacklist. Same attackers it may chance to enter into another location once again start the process for detection. It is computationally expensive in detection of attackers [1][2].

Next new approach Group signature related concept. All group signatures are present into group manager itself. Any user requires the signature information without different users permissions directly all users are signatures are displayed here. Any user can possible to capture the different user's signatures also. Attackers are also entering with authorized user's signatures information. Whenever previous case is occur traffic it may chance to generate in network [3][4].

Next approach, Reputation related approach for controlling of attackers. Reputation system requires different levels for detection of attacks. First level training approach starts here. Using training approach

decide as a attackers or normal users. This is status related content. Using status related approach every time create the new pattern of network. Reputation system is called refinement system. Using refinement system impossible to detects all different types of attacks [4][1][3]. It is one of the good approach for reducing the attackers.

Next judgment double spending approach related concept. Any misbehavior attacker identifies into different e-commerce systems. Same user repeatedly enters into same web pages we detect as an attacker. This type of attacker's detection purpose it does not require any proof in network people [5][6].

Some previous systems cryptography related approaches also control some attackers without modifying of content. Cryptography performs the data operations like plain text to cipher text data. Cipher text data is completely unreadable. Sender transfers the data using outsourcing operations. In outsourcing we get the some problems like data manipulation problems [7] [6].

Another anonymization approaches, hiding of IP addresses control some of the attackers. Same attackers it may chance to generate the problems in other networks also in implementation part. Using Hiding of IP addresses related approach is not possible to detect all routers traffic. It is somewhat better approach in detection of attackers.

The above existing approaches some percentage of attackers we detect very easily. Using those above approaches 100% attackers' detection is not possible. Now we propose new techniques in below sections or chapters for detection of attackers.

III.PROBLEM STATEMENT

The goal of present paper minimizes the potential leakage of data. Now in this paper we introduce the Distributed Nymble system for reducing the threats. New Distributed Nymble system prevents all different categories of attacks, distribute the data as a sensitive data. We create the new distributed nymble system design. In Design many number of levels are available for verification without missing of any attacks. Every level filters some number of attacks for detection. It is not possible to filter, and then shift the requests to second levels. Filter the attacks in second level. This procedure applies repetitively until detection of all kinds of attacks here. New system experimentally shows the accurate and scalable solution in internet application. This approach is called n –round approach for detects all misbehavior related attacks.

IV.PROPOSED SYSTEM ARCHITECTURE

Distributed Nymble System architecture consists of different components. Those components are

1. User or client component
2. Pseudonym manager component
3. Distributed Nymble managers component
4. Server component

Every component performs some steps of works. All components are combined and control all different types of attacks.

User and Client Component:

Each and every user registers into pseudonym manager and different nymble managers also. Two categories of managers give the different kinds of credentials. All credentials use in sequential manner. Every user collects the sequential pattern. Nymble managers are not provided any random number of conditions. All users systems are connected by server. Those privacy servers control the attackers.

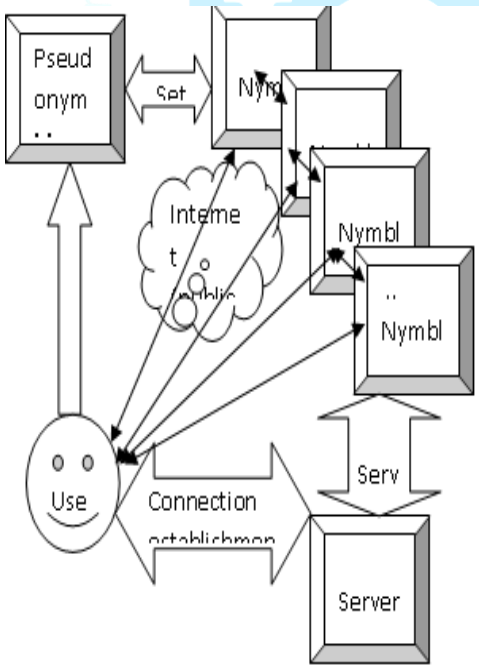


Fig1: Proposed System Architecture

Pseudonym manager component:

Every user request first we forward to pseudonym manager. Pseudonym manager verifies the IP address. IP address is not authorized black the IP address. Those IP addresses store into blacklist. This is unique IP address verification approach. IP address is unique start the linkability to maintain the communication process. Pseudonym manager performs the mapping operation based on verification approach.

Distributed Nymble manager's component:

Distributed nymble managers follow the sequential verification approach. Once authorized in first nymble manager we allow in another nymble manager here. Every nymble manager maintains the constraints. One constraint is depends on another constraint procedure. This is entirely dependent based approach. Every stage places the constraints without loss of sensitive data content. In transmission of data without leakage also control the data here.

Server component:

Different Nymble managers are gives the status related to different number of users. Every server how many number of attackers are detect here that kind of complete information update in server. Servers every time active for detection of all attackers are also possible. There is no possibility to missing of attackers with the help of sequential approach.

V.CRYPTOGRAPHY RELATED APPROACHES

Cryptographic functions are used in verification approaches. Those cryptographic approaches are digital certificate generation. In digital certificate every time update the new features for detection and authorized

with new technologies. Certificate is generating with the help of hash functions and symmetric encryption technique also. These things implement proposed system architecture then only it works as a security model. It gives the accurate attackers is possible without leakage.

VI.EXPERIMENTAL AND PERFORMANCE EVOLUTION IN DIFFERENT ATTACKS DETECTION

Here we show the comparison performance results in between of Single Nymble Manager and Distributed Nymble Manager. We save the more amounts of sensitive data without leakage compare to existing to proposed system architecture here.

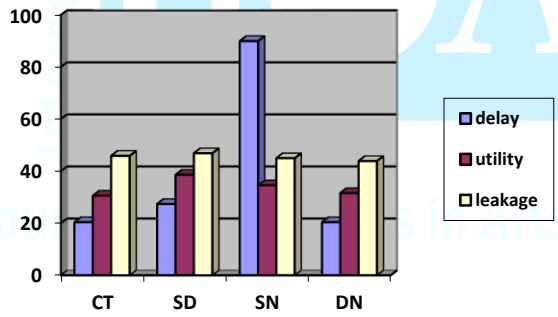


Fig2: Performance Graph in between of Different Approaches

VII. CONCLUSION AND FUTUREWORK

Here in this paper we have proposed distributed nimble system as a new credential system. Proposed credential system follows sequential pattern. This system controls all different locations of attacks. New system is efficient and scalable under detection of attacks and processing of attacks also in implementation. In present system some new properties are added, these properties are control the different attacks here. This system practically works efficiently. Our work increases the detection of more possible attacks in wireless sensor network environment.

VIII.REFERENCES

[1] Omid Ardakanian, Nam Pham, Enhancements for Nymble System,2010
[2] Peter Lofgren and Nicholas Hopper, BNymble, 2011.
[3] Yenumula Sankara Rao, Pradeep ENHANCING SECURITY BY BROADENING NYMBLE SYSTEM IN ANONYMIZING NETWORKS, 2012
[4] Elli Androulaki, Seung Geol Choi, Steven M. Bellovin, and Tal Malkin, Reputation Systems for Anonymous Networks, 2008
[5] Extending Nymble-like Systems, Ryan Henry and Ian Goldberg
[6] Peter C. Johnson1, Apu Kapadia1,2, Patrick P. Tsang1, and Sean W. Smith1,Nymble: Anonymous IP-Address Blocking,2011
[7] Matteo Maffei Kim Pecina Manuel Reinert, Security and Privacy by Declarative Design, 2011
[8] 2011, Leonardo Aguilera, An implementation of accountable anonymity
[9] Ryan Henry, Nymbler: Privacy-enhanced Protection from Abuses of Anonymity, 2010

Distributed Wireless Network Routing: An Eaves

1.T AnandBabu 2.N.Venkatram 3.M.Nirupa

1.M.Tech student,Dept of Computer Science,Nova's Institute of Technology,Eluru

2.Associate professor,HOD of Computer science,Nova's Institute of Technology,Eluru

3.Assistant professor,Dept of Computer science,Nova's Institute of Technology,Eluru

Abstract

The time and technology in the current trend of change is a dollar currency. In this context of our research paper, we tries give emphasis on the technology which is already there, but not to the simplified one. In wireless network routing is the most fundamental factor and if that to a distributed we need mechanism to control the efficiency and its operational activities such as packet dropping, data loss, authentication and others which involved is the most standard encryption and decryption algorithm. In this paper, we address the concept of Floyd-Warshall Algorithm to give the thread based security mechanism to protect the data and give the very best distributed optimization to the networking. In the Routing we use the concept of wireless in the networking i.e. the FWKernel Algorithm; which provides the thread based ID in the distributed base and ensure the efficiency and optimality in the passage of acknowledgement.

Keywords: *Dynamic Routing, Bellman-Ford Algorithms,* Floyd-Warshall Algorithm, FWKernel Algorithm

.

1. Introduction

An important concept in network security is trust, interpreted as a relation. Among entities that participate in various protocols to maintain the pooper way of communication. Trust relations are based on evidence related to the previous interactions of entities within a protocol. If we consider the history of trustworthy let us consider the below mentioned example as our introductory evidence to make our research to next extend.

In May 2003, spammers hijacked an unused block of IP address space owned by Northrop Grumman, and used the address space to send spam email [2]. It took two months for Northrop Grumman to resolve the situation and get the rogue routing announcements blocked.

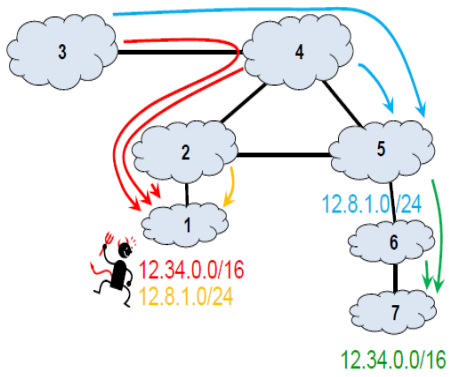


Figure 1.1: Malicious node 1 announces address prefixes that it does not own.

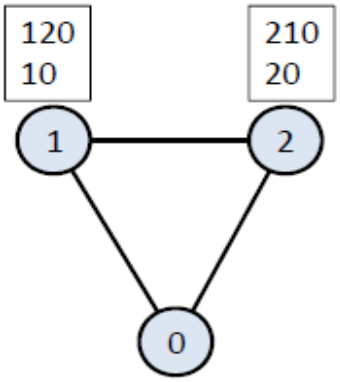
For about 18 minutes on April 8, 2010, Chinese Telecom rerouted traffic destined to about 15% of the address space through servers in China. This incident reportedly [3] also affected traffic destined to US government and military sites, including those for the US Senate, four branches of the military, the secretary of defense, and NASA. The network traffic was forwarded further to the destination. Such forwarding can be done transparently without the owner's knowledge, as described in [1].

Several cryptographic countermeasures have been proposed still now after the incident. However, deployment of these new protocols is complicated by the fact that participation of all in the Internet is required to achieve significant security benefits. Since the Internet currently consists of approximately 35000 independently administered, many of which are small regional ISPs or enterprise networks, speedy deployment of a new version of BGP cannot be expected. We argue that for any practical solution that will be eventually deployed, incentives must be provided to early adopters, and security benefits must exist even for small scale deployments.

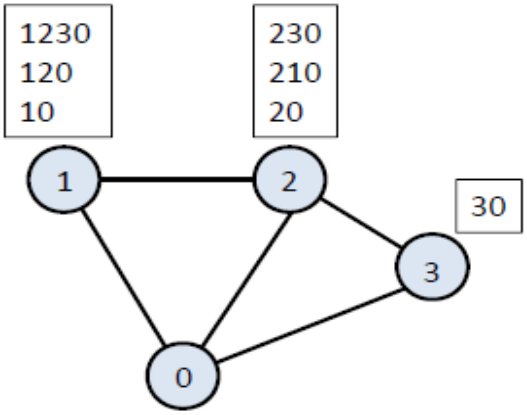
2. Related work

In the medium of ether the data and its cryptographic technology to its strong robustness to medium injection by third part is most important and crucial in today's world of technology in the field of networking. Hence, in this related technology work field of this paper we tires take the consideration fig.1.1 of introduction part and other related aspect to make our methodology i.e. the proposed algorithm the best to suite to market of networking in all the aspect which will lead to our future work of enhancement. Although Autonomous Systems are free to choose their route preferences and route export policies, certain policy combinations can lead to permanent oscillations in the routing system. In these oscillations, routers exchange control plane messages in a cyclical fashion indefinitely. An example of such a policy choice is illustrated in in Figure 2.1(a). The two BGP speakers represented by nodes 1 and 2 are configured as follows. To reach node 0, node 1 prefers the indirect route through node 2 over the direct route. Similarly, node 2 prefers the indirect route through node 1. An oscillation may occur as follows. Initially the nodes select the direct route 10 and 20, respectively, and the nodes simultaneously update each other about the route they are using. Subsequently, they both switch to the two indirect routes 210 and 120, creating a transient loop between nodes**.** This problem is significant because routers must already process route updates for some 350; 000 address prefixes in the Internet, causing

a heavy load even without oscillations. Oscillations can also cause unacceptable delays or packet drops in the data plane. As illustrated above, loops in the data plane may arise and data packets may be dropped. Path changes during oscillations can also interfere with TCP that monitors end-to-end delays and expects regular timing of acknowledgements; oscillations may lead to severe performance degradation for the user.



(a) Permanent oscillations may occur.



(b) The network will stabilize with every node picking its most preferred route.

Fig.2.1 Describing Presence of a dispute wheel is necessary but not sufficient for oscillations.

In the control plane, oscillations increase the number of route updates, which may overload the routers that are not able to process the messages at high enough speed.

3. Methods

Technology is such a terminology in the field of Computer Science which means the "*Change*", i.e. in this paper we tries to give emphasis on such algorithm that gives best suite solution the distributed wireless networking routing . By considering the aspect of threading to process the data faster and to make data safer we need various concepts which is reflected in the algorithm. We have analyzed by taking a small example in order to understand the system better and easy. Network operators need to carefully balance the load in the network in order to efficiently use the existing link capacity. In addition, they need to handle planned equipment maintenance and unplanned failures gracefully, without noticeable disruptions to the users. This challenging task is further complicated by the fact that traffic patterns change significantly during the day, and network operators need to satisfy strict service level agreements (SLAs) which specify, e.g., the maximum average delay that the network traffic can experience.

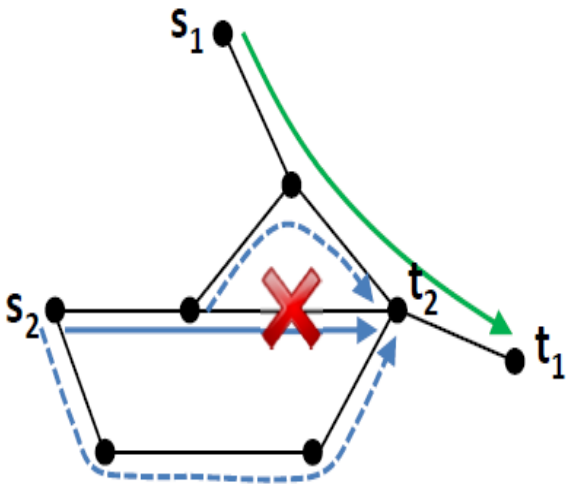


Fig: 3.1 showing the Pictorial flow of Nodes in the Network Traffic

Figure 3.1 illustrates the advantages and disadvantages of two possible techniques that can be used to protect against failures local path protection and global path protection. In local path protection, the failure is repaired locally by sending the traffic on an alternate route between the two endpoints of the failed link. The figure illustrates the disadvantage of this approach congestion in the neighborhood of the failure where the rerouted traffic shares a link with another flow. Global path protection sends the traffic on an alternate end-to-end path.

These spurious updates can be caused by several router-level mechanisms that delay the propagation of update messages (to reduce overhead and improve stability) or limit visibility into the alternate routes (to improve scalability), including as follows.

Route flap damping: Route flap damping temporarily suppresses a route if it appears unstable. As a result, a router may temporarily select a less-preferred route.

MRAI timers: The Minimum Route Advertisement Interval (MRAI) timer paces BGP update messages. Delaying message delivery can cause a router to temporarily select a lower-ranked alternate route.

Router queuing mechanisms: The BGP message queues between routers delay the delivery of updates. These queues, coupled with optimizations that stop generating new messages when the queue grows large, can lead to delays in selecting the highest-ranked route.

Cluster routers: Large routers are distributed, with BGP sessions terminating on different processor blades. To improve scalability, these blades do not exchange full information with each other, which may lead to a temporary selection of a less-preferred route.

Proposed router extensions: Extensions to the BGP route-selection process were proposed to improve router reliability or to reduce convergence time. This changes the timing of routing decisions.

All spurious updates share two common properties:

(i) a router can only send spurious updates for a short time after receiving information changing its most preferred route, and

(ii) Spurious updates are based on routes that have been recently available (including spurious withdrawals because \no route" is always available). DPVP allows any spurious update with these properties.

We argue that such model is general enough to capture all spurious updates, but at the same time we show that the model is not overly broad. Just as local routing policies can affect global convergence, the local

engineering decisions that cause spurious updates can also trigger oscillations and slow convergence exponentially. Eliminating all sources of spurious updates would require major changes to router design and the BGP protocol. Some of these mechanisms are important for reducing protocol overhead and improving scalability, making it unappealing to eliminate them entirely. Protocol designers, router designers, and network operators could strive to reduce the frequency and duration of spurious updates. However, it is not clear that such a quest is warranted or plausible. Rather than advocating for a world free of spurious updates, we argue for a better understanding of their consequences**.**

The algorithm uses a duplicate copy of the graph that is identical to the original graph when the algorithm starts. To avoid global memory communication, the use of this identical graph P and the original graph G are interchanged. The algorithm performs a series of iterations, where P initially holds the results for the first iteration, its use is swapped so that the algorithm thinks P is now G for the second iteration. This swapping occurs for every iteration until the complete result is formulated. Utilizing two data structures such as this ensures that read-after-write inconsistencies are eliminated. The intention is that threads calculating the results will fit into a sub-block of the graph that is dynamically assigned to the thread block. The kernel for this algorithm is executed $^{j}V_{j}$ times, with the use of G and P alternating with each execution.

Threads within a block are executed in a group of 32, known as a warp. As each half-warp share the row G[k; j]; j = blockIdx:x _ 16; blockIdx:x _ 16 + 1; :::; blockIdx:x _ 16 + 15 and one element G[i; k] where i = blockIdx:y _ 16; blockIdx:y _ 16 + 1; :::; blockIdx:y _ 16 + 15 the row G[k; j] and the column G[i; k] are copied into shared memory for every thread block.

Threads whose corresponding row address, blockIdx:y, is 0, copy the row data, and threads with 0 for their column address, blockIdx:x, copy the column data. The data for both row and column is copied from G. Once this is complete, the standard Floyd-Warshall computation is performed on the data that has been copied into shared memory. The result is then written back to P as shown in line nine of Algorithm 13.

Algorithm Floyd-Warshall (G; P)

- Allocate and copy G; P to device
- for i = 0 to $^{j}V_{j}$
- if i%2 6= 0 then
- FWKernel(G; P)
- else
- Allocate FWKernel(P;G)

Algorithm FWKernel (G; P)

- threadID= obtainThreadID(From Network)
- row[16] **Shared // Shared Path
- column[16] **Shared// Allocate
- if threadIdx.y = 0 then
- row[threadIdx.x] = G[threadID]
- if threadIdx:x = 0 then // Path to the optimized way
- column[threadIdx.y] = G[threadID]
- P[threadID]=min(G[threadID], row[threadIdx.x] + column[threadIdx.y])

- Close the threadID. // Connection

Where ${}^{j}P_{j}$ is the number of cores on the GPU. Clearly, this provides a slight asymptotic speed-up over the standard CPU version of Floyd-Warshall that runs in O(n3) time.

The algorithm has been optimized to ensure that read and write operations from global memory are coalesced, so as to access as much data as possible in one processor cycle, thus reducing the latency on memory accesses. Ensuring that the efficiency of the memory is as great as possible. As with the algorithm described this algorithm can be expanded to work across multiple GPUs, or simply with larger graphs than the GPU can store at one time.

4. Conclusion and Future Enhancements

In this paper we tries to give emphasis on the dynamic routing model where we First, we introduced a new model, that provides a convenient and accurate framework for understanding the dynamic properties of interdomain routing. Using this model, we resolved a decade-long open question that asks for the necessary and sufficient conditions under which routing policies are safe. Policy induced oscillations are responsible for temporary losses of connectivity, creation of transient loops leading to packet losses, and frequent route changes have negative impact on TCP performance.

Second, we studied the security of interdomain routing. We started with the observation that current proposals of secure interdomain routing protocols cannot yield measurable security benefits in small scale deployments, and network operators do not have incentives for early adoption. We identified a combination of mechanisms that provides significant and measurable security benefits to participants and non-participants alike even if the proposed solution is only deployed by a small group of participants. If deployed in the Internet, our solution can significantly reduce the occurrence of route black holing and data interception caused either by malicious attacks or missed configurations which lead to next level of research.

5. Reference

[1] A. Pilosov and T. Kapela. Stealing the Internet: An Internet-scale man in the middle attack, 2008. DEFCON 16, Las Vegas, NV, USA.

[2] L. Benkis. Practical bgp security: Architecture, techniques and tools. http://www.renesys.
com/tech/notes/WP_BGP_rev6.pdf.

[3] 2010 Report to Congress of the U.S.-China Economic and Security Review Commission,
2010.
http://www.uscc.gov/annual_report/2010/annual_report_full_10.pdf.

[4] T.H. Cormen, C.E. Leiserson, and R.L. Rivest, Introduction to Algorithms. MIT Press, 1990.

[5] P. Erdo¨s and A. Re´nyi, “On Random Graphs,” Publicationes Math. Debrecen, vol. 6, 1959.

[6] M. Faloutsos, P. Faloutsos, and C. Faloutsos, “On Power-Law Relationships of the Internet Topology,” Proc. ACM SIGCOMM’99, pp. 251-262, 1999.

[7] FreeS/WAN, http://www.freeswan.org, 2008.

[8] I. Gojmerac, T. Ziegler, F. Ricciato, and P. Reichl, “Adaptive Multipath Routing for Dynamic Traffic Engineering,” Proc. IEEE Global Telecommunications Conf. (GLOBECOM), 2003.

[9] C. Hopps, Analysis of an Equal-Cost Multi-Path Algorithm, Request for comments (RFC 2992), Nov. 2000.
[10] C. Kaufman, R. Perlman, and M. Speciner, Network Security PRIVATE Communication in a PUBLIC World, second ed. Prentice Hall PTR, 2002.
[11] J.F. Kurose and K.W. Ross, Computer Networking—A Top-Down Approach Featuring the Internet. Addison Wesley, 2003.
[12] V.I. Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions, and Reversals," Soviet Physics Doklady, vol. 10, no. 8, pp. 707-710, 1966.
[13] S.-H. Liu, Y.-F. Lu, C.-F. Kuo, A.-C. Pang, and T.-W. Kuo, "The Performance Evaluation of a Dynamic Configuration Method over IPSEC," Proc. 24th IEEE Real-Time Systems Symp.: Works in Progress Session (RTSS WIP), 2003.
[14] W. Lou and Y. Fang, "A Multipath Routing Approach for Secure Data Delivery," Proc. IEEE Military Comm. Conf. (MilCom), 2001.
[15] W. Lou, W. Liu, and Y. Fang, "SPREAD: Improving Network Security by Multipath Routing," Proc. IEEE Military Comm. Conf. (MilCom), 2003.
[16] G. Malkin, Routing Information Protocol (RIP) Version 2 Carrying Additional Information, Request for comments (RFC 1723), Nov. 1994.

Improving The Efficiency of Forward Error Correction Coding In Reducing The Network Packet Loss

1.K.Srinivasulu, 2.N.Venkatram, 3.M.Sarada

1.M.Tech student, Dept of CSE, Nova's Institute of Technology, Eluru

2.Assoc Professor, HOD of CSE Dept, Nova's Institute of Technology, Eluru

3Asst Professor, Dept of CSE, Nova's Institute of Technology, Eluru

Abstract – In this paper we explores a method for measuring the performance of FEC coding combining with interleaving in reducing the packet loss in IP networks. In order to evaluate the performance of FEC data can be transfer from the source to destination and creates the packet loss voluntarily, at the destination the lost packets can recovered using FEC decoder. The performance of the FEC coding can be measured using an analytical method stated in this paper. Here we use the single multiplexer network model for transmission of the data from multiple sources to destinations. In this a unified approach provides an integrated framework for exploring the compromises between the various key parameters i.e. channel coding rates, interleaving depths, block lengths. It provides the selection of various optimal coding strategies with various QOS requirements and system constraints.

Index Terms- FEC Coding, interleaving, packet loss rates, multiplexer network model, multi session and single session.

INTRODUCTION

The packet transport service provided by representative packet-switched networks, including IP networks, is not reliable and the quality-of-service (QoS) cannot be guaranteed. The Packets may be lost on their route, Switching nodes requires more processing power as the packet switching protocols are more complex, Switching nodes for packet switching require large amount of RAM to handle large quantities of packets, A significant data transmission delay occurs - Use of store and forward method causes a significant data transmission.

Packets may be discarded due to excessive bit errors and failure to pass the cyclic redundancy check (CRC) at the link layer, or be discarded by network control mechanisms as a response to congestion somewhere in the network.

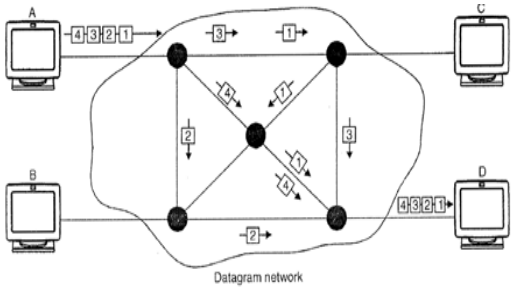


Fig1: Packet switched networks (datagram)

Forward error correction (FEC) coding has often been proposed for end-to-end recovery from such packet losses. FEC can be defined as A class of methods for controlling errors in a one-way communication system. FEC sends extra-information along with the data, which can be used by the receiver to check and correct the data. Codes that include the unmodified input in the output are **systematic**, while those that do not are **nonsystematic**.

A novel technique based on forward error correction (FEC) has been proposed that allows the destination to reconstruct missing data packets by using redundant parity packets that the source adds to each block of data packets.

Example: an analog to digital converter that samples three bits of signal strength data for every bit of transmitted data. If the three samples are mostly all zero, the transmitted bit was probably a zero, and if three samples are mostly all one, the transmitted bit was probably a one. The simplest example of error correction is for the receiver to assume the correct output is given by the most frequently occurring value in each group of three.

Forward Error Correction is particularly well suited for satellite transmissions, where bandwidth is reasonable but latency is significant.

The use of FEC in applications provides a double-edged sword. From an end user's perspective, FEC can help recover the lost packets in a timely fashion through the use of redundant packets, and generally adding more redundancy can be expected to improve performance provided this added redundancy does not adversely affect the network packet loss characteristics.

On the other hand, from the network's perspective, the widespread use of FEC schemes by end nodes will increase the raw packet-loss rate in a network because of the additional loads resulting from transmission of redundant packets. Therefore, in order to optimize the end-to-end performance, the appropriate trade-off, in terms of the amount of redundancy added, and its effect on network packet-loss processes, needs to be investigated under specific and realistic modelling assumptions.

Types of FEC: The two main categories of FEC are block coding and convolutional coding.1) Convolutional codes work on bit or symbol streams of arbitrary length. This code can be turned into a block code, if desired. Convolutional codes are most often decoded with the Viterbi algorithm, though other algorithms are sometimes used. 2) Block codes work on fixed-size blocks (packets) of bits or symbols of predetermined size. There are many types of block codes, but the most notable is Reed-Solomon coding because of its widespread use on the Compact disc, the DVD, and in computer hard drives. Block and convolutional codes are frequently combined in **concatenated** coding schemes in which the convolutional code does most of the work and the block code (usually Reed-Solomon) "mops up" any errors made by the convolutional decoder.

Advantages: EDAC has a number of advantages for the design of high reliability digital systems:

1) Forward Error Correction (FEC) enables a system to achieve a high degree of data reliability, even with the presence of noise in the communications channel. Data integrity is an important issue in most digital communications systems and in all mass storage systems.

2) In systems where improvement using any other means (such as increased transmit power or components that generate less noise) is very costly or impractical, FEC can offer significant error control and performance gains.

3) In systems with satisfactory data integrity, designers may be able to implement FEC to reduce the costs of the system without affecting the existing performance.

This is accomplished by degrading the performance of the most costly or sensitive element in the system, and then regaining the lost performance with the application of FEC.

In general, for digital communication and storage systems where data integrity is a design criterion, FEC needs to be an important element in the trade-off study for the system design. The introduction of the Per FEC line of FEC encoders and decoders makes powerful FEC implementation a realistic goal for most digital communication and storage systems. More than ever before, FEC is available for a wide range of applications.

Most telecommunication systems used a fixed channel code designed to tolerate the expected worst-case bit error rate, and then fail to work at all if the bit error rate is ever worse. However, some systems adapt to the given channel error conditions: hybrid automatic repeat-request uses a fixed FEQ method as long as the FEQ can handle the error rate, then switches to ARQ when the error rate gets too high; adaptive modulation and coding uses a variety of FEQ rates, adding more error-correction bits per packet when there are higher error rates in the channel, or taking them out when they are not needed.

Averaging noise: To reduce the errors FEC could be said to work by "averaging noise"; since each data bit affects many transmitted symbols, the corruption of some symbols by noise usually allows the original user data to be extracted from the other, uncorrupted received symbols that also depend on the same user data. Because of this "risk-pooling" effect, digital communication systems that use FEC tend to work perfectly above a certain minimum signal-to-noise ratio and not at all below it. This *all-or-nothing tendency* becomes more pronounced as stronger codes are used that more closely approach the theoretical limit imposed by the Shannon limit. Interleaving FEC coded data can reduce the all or nothing properties of transmitted FEC codes. However, this method has limits. It is best used on narrowband data.

This explores a framework for using Forward Error Correction (FEC) codes with applications in public and private IP networks to provide protection against packet loss. The framework supports applying FEC to arbitrary packet flows over unreliable transport and is primarily intended for real-time, or streaming, media. This framework can be used to define Content Delivery Protocols that provide FEC for streaming media delivery or other packet flows. Content Delivery Protocols defined using this framework can support any FEC scheme (and associated FEC codes) that is compliant with various requirements defined in this document. Thus, Content Delivery Protocols can be defined that are not specific to a particular FEC scheme, and FEC schemes can be defined those are not specific to a particular Content Delivery Protocol.

Interleaving:

Interleaving in computer science is a way to arrange data in a non-contiguous way in order to increase performance. It is used in: Time-division multiplexing (TDM) in telecommunications, Computer memory and Disk storage. Interleaving is mainly used in data communication, multimedia file formats, radio transmission (for example in satellites) or by ADSL. The term multiplexing is sometimes used to refer to the interleaving of digital signal data. Interleaving was used in ordering block storage on disk-based storage devices such as the floppy disk and the hard disk. The primary purpose of interleaving was to adjust the timing differences between when the computer was ready to transfer data, and when that data was actually arriving at the drive head to be read.

Interleaving was very common prior to the 1990s, but faded from use as processing speeds increased. Modern disk storage is not interleaved.

A method for making data retrieval more efficient by rearranging or renumbering the sectors on a hard disk or by splitting a computer's main memory into sections so that the sectors or sections can be read in alternating cycles. i.e. Iinterleaving was used to arrange the sectors in the most efficient manner possible, so that after reading a sector, time would be permitted for processing, and then the next sector in sequence is ready to be read just as the computer is ready to do so. Matching the sector interleave to the processing speed therefore accelerates the data transfer, but an incorrect

interleave can make the system perform markedly slower.

Here we explores a brief study of the overall effectiveness of packet-level FEC coding, employing interlaced Reed-Solomon codes, in combating network packet losses and provide an information-theoretic methodology for determining the optimum compromise between end-to-end performance and the associated increase in raw packet-loss rates using a realistic model-based analytic approach. Intuitively, for a given choice of block length we expect that there is an optimum choice of redundancy, or channel coding rate, since a rate too high (low redundancy) is simply not powerful enough to effectively recover packet losses while a rate too low (high redundancy) results in excessive raw packet losses due to the increased overhead which overwhelms the packet recovery capabilities of the FEC code

1. Single Source Model

Single multiplexer model- The performance of the network is limited by a single bottleneck node, the network can often be modelled in terms of the single multiplexer. The single multiplexer is queuing system which consist of three parts one is the arrival process of the packets from N different sources S_i at arrival rate λ_i, second buffer that to hold up to the k packets and last one is an output link with averaging packet service rate μ, assume that packet service times are independent and identically Distributed (i.i.d) with an exponential distribution and average packet service time T=1/ μ, the normalized load to system is $\rho = \lambda / \mu$.

The packet arrives to the single multiplexer from single source or multiple sources. The single source corresponds to the per-flow control from the traffic (assigns fixed bandwidth). Whereas multiple source having no per-flow control is applied. Here the packet shares the bandwidth of the output and buffer.

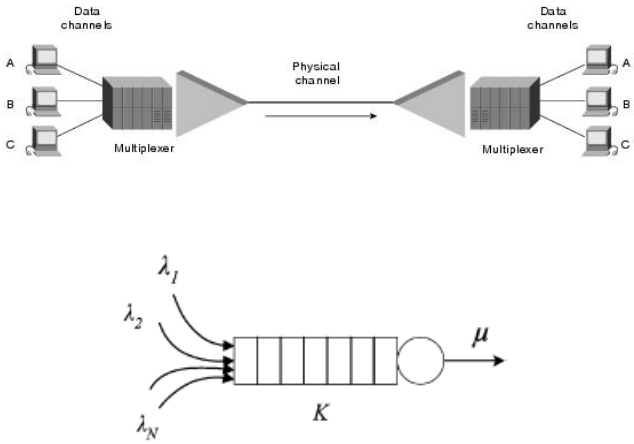


Fig. 2 single multiplexer network model

Source Model- Assume that the packet arrival process for each source S_i is renewal process. The packet interval times are i.i.d with probability density function a(t). erlang inter arrival time distribution is

$$a(t) = \frac{\beta(\beta t)^{h-1}}{(h-1)!} e^{-\beta t}, t \geq 0$$

For erlang distribution, the average arrival rate is $\lambda_i = \beta / h$ when h=1, a(t) becomes exponential and arrival process is Poisson when $h \to \infty$, the variance of a(t) converges to 0 and the interval time becomes deterministic with period $1/\lambda_i$

System model for FEC performance Evaluation–

Consider communication system illustrated in Fig.1 suppose there are N homogeneous and independent data sources sharing the single multiplexer and each source generates packets with average rate λ_i. The FEC coder for each source applies the Reed Solomon codes. *RS(n, k)* to the packets from the source, which means for every block of k source information packets it creates additional n-k parity packets to the network. The channel coding rates is given by $R_C = k/n$. because of this channel rate, the packet arrival rate into the network will increase $\lambda'_i = \lambda_i / R_c$.

The random variable N_p denotes the number of lost packets within a block. If $N_p \le n-k,$ then assume that all the lost packets are recovered by the channel decoder. Assume that $p(j,n)$ denotes the block error distribution, the expected number of lost packets within a block is

$$E[N_p] = \sum_{j=n-k+1}^{n} j * P(j,n)$$

And the expected number of the lost information packets within a block is

$$E[N_I] = E[N_p]*(k/n) = \left(\sum_{j=n-k+1}^{n} j * P(j,n) \right) *(k/n)$$

Finally the effect information packet loss rate after channel decoding is

$$\mathrm{PLR_{eff}} = \mathrm{E[N_I]/k} = \left(\sum_{j=n-k+1}^{n} j * P(j,n) \right) / n$$

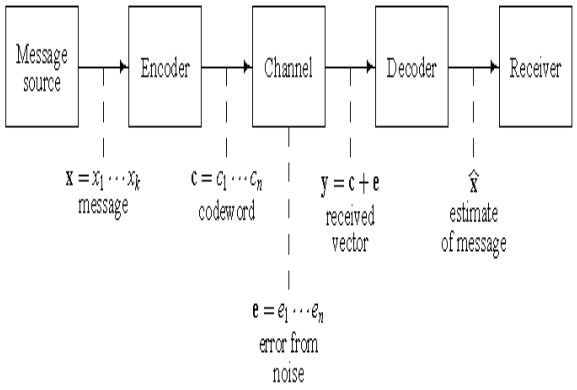


Fig3. Communication Channel

Evaluation Metric: Packet loss probability or Frame loss probability

The frame loss metric is more suitable than the packet loss probability for evaluation of FEC performance the frame lost probability FLR_{eff} is given by

$$\mathrm{FLR_{eff}} = \left(\sum_{j=n-k+1}^{n} P(j,n) \right)$$

The differences between residual packet loss probabilities and frame-loss probabilities with decoding and without decoding denoted by FLR_{eff} – FLR_{wo} respectively.

A single-multiplexer model for this bottleneck node is widely used to analyze the associated queuing-related

Packet losses, e.g., losses due to buffer overflows and excessive delays. Since the correlation level of the packet-loss process has great impact on the FEC efficacy, this dependence using the autocorrelation function of the packet-loss process

Autocorrelation function of packet loss processes- For this packet loss process we use the autocorrelation function to characterize the dependence between the packet loss events over the time, let $\{Y_i\}$ random sequence represent the packet loss process with 1 denoting the loss and 0 denoting the reception if $\{Y_i\}$ is stationary then autocorrelation function of $\{Y_i\}$ is given as

$$\rho y(l) = \frac{E[(Y_{i+1} - \mu_y)(Y_i - \mu_y)]}{E[(Yi - \mu y)^2]}$$

Where l is the lag and μ y is the expectation of the sequence $\{Y_i\}$. The performance of FEC in recovering network packet losses. Redundant parity packets was proposed to reconstruct lost data packets and the corresponding performance evaluation indicated that residual packet-loss rates can be reduced up to three orders of magnitude. ***FEC performance with single source FEC without Interleaving -*** Suppose there is only one user for multiplexer the key quantity in evaluating the residual packet loss rate after FEC decoding is $p(j,n)$, the block-error distribution for

an arbitrary number n of consecutive packets. Queuing system there are two types of queuing system denoted as M/M/1/K queue: finite buffer queue with Poisson intervals and exponential service times and extension of G/M/1/K queue: the finite buffer with general independent and identically distributed interval times and exponential service times.

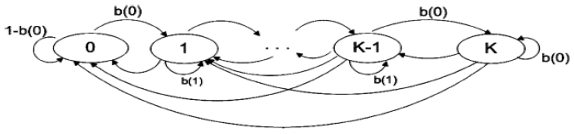


Fig. 4 state transition of arrival-epoch system size

Analysis of Block error Distribution: suppose there is only one source sharing multiplexer and the packet inter arrival times are i.i.d with arbitrary probability density function $a(t)$, single multiplexer can be modeled as a standard G/M/1/K queue. Discrete-time marcov chain: let Xn be the no.of packets in the buffer just before the nth packet arrives at the system because the memoryless property of the exponential service time { Xn} lets consider the above figure 4 with state space S={0,1,2,…K}.

Numerical example: The Fig. 4 shows the effective packet loss rates $\mathrm{PLR}_{\mathrm{eff}}$ computed according to

$$\mathrm{PLR}_{\mathrm{eff}} = \mathrm{E}[\mathrm{N_I}]/\mathrm{k} = \left(\sum_{j=n-k+1}^{n} j * P(j,n) \right) / n$$

With different coding block size n=63,127,255 and 511 as a function of coding rates R_C = *k/n.* having the Poisson arrivals (h=1), ρ =0.8, k=10.

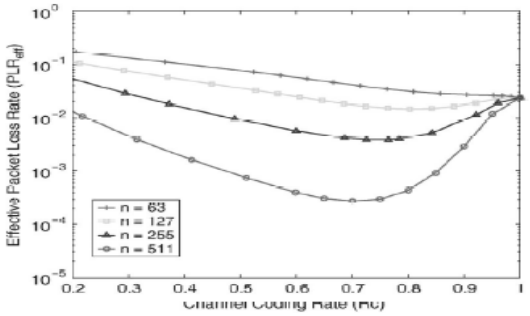


Fig5: effective packet loss rates With different coding block size n=63,127,255 and 511 as a function of coding rates R_C = k/n. having the Poisson arrivals (h=1), ρ =0.8, k=10

The above fig shows the FEC performance with

different numbers of sources multiplexed, where the load from each source is fixed at λi = 0.02 with buffer size K = 10 and coding block size n=63 . It shows that, with an increase in the number of sources N, the effective packet-loss rates increase due to the increased system

load. Suppose now the load from each source is again λi = 0.02 and the required effective packet-loss rate is 10^{-6}.

FEC with Block Interleaving- FEC performance is often limited by the bursty nature of typical packet-loss processes, and block interleaving techniques are frequently used to reduce the burstiness of the packet-loss processes in networks, thereby improving FEC performance. In this section, we analyse the efficacy of interleaving in reducing the burstiness of network packet-loss processes and in improving the FEC performance. *Interleaving Operation:* Before packets being transmitted into the network, packets are filled into an M1*M2 Row wise.

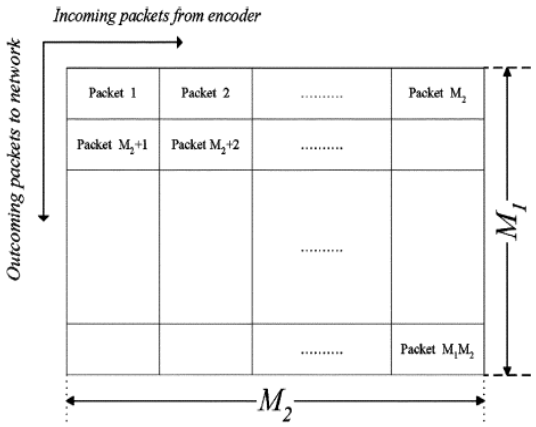
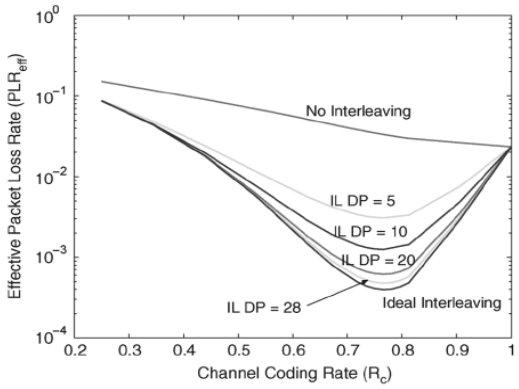
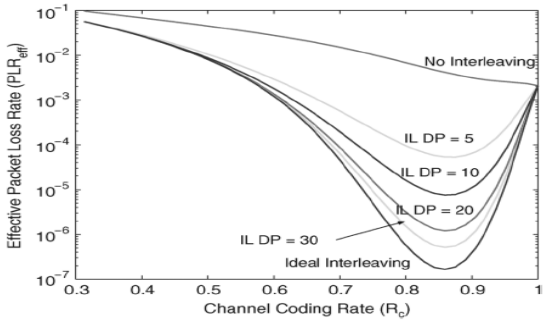


Fig 5: Illustration of block interleaving operation (interleaving depth=M1).

The above fig shows the case of deterministic arrivals(h=∞) with all other system parameters the same as in Fig. 6 Compared to Fig. 5 shows that for more deterministic source arrivals an increased coding rate Rc is required to achieve the optimum performance. Figure demonstrate that with interleaving the performance of FEC coding can be greatly improved, and interleaving with even larger depth can achieve increasingly improved performance.



(a)



(b)

Fig6: (a) Evaluate interleaving FEC performance Poisson arrival (b) Evaluate interleaving FEC performance deterministic arrival.

2. FEC Performance with Multiple Sources (N>1)

Here, we proceed to explore the FEC performance in case of multiple sources sharing the multiplexer. In order to facilitate the analysis, here assume in the packet arrival process seen by the multiplexer from each source is Poisson.

FEC Performance without Interleaving: In order to evaluate the FEC performance for one of the N sources, the block-error distribution P(j,n). For a single isolated source is required. Here we discuss different method to compute P(j,n) for the N*M/M/1/K. queue, which can be extended easily to incorporate the analysis for interleaving.

Analysis of Block-Error Distribution for a Single Source: Assume the packets arriving at the single-multiplexer come from N independent sources: S1, S2……Sn. indicates that, for N homogeneous sources with a fixed overall load ρ, the loss process of a single source becomes less and less correlated with increasing N

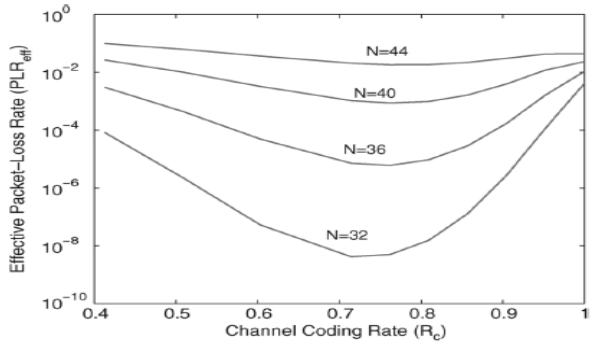


Fig7. FEC performance with N homogeneous sources; Poisson arrivals, load from each source fixed at ρi =0.02, K=10 block size n=63

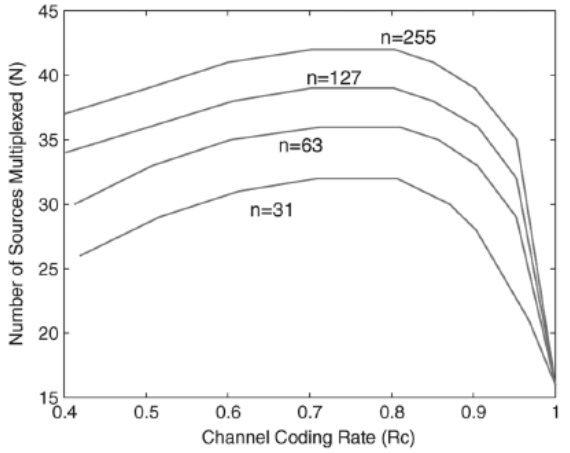


Fig8. Multiplexing gain achieved by FEC coding with different coding block sizes n; Poisson arrivals, the effective packet-loss rate fixed at PLReff= 10-6, load from each source fixed at ρi=0.02,K=10

Now look at the FEC performance in improving the Statistical multiplexing gain. Suppose the FEC coder for each homogeneous source applies an RS(n,k)code to the packets from the corresponding source coder. The channel coding rate remains Rc = k/n. As a result of the channel coding, the packet arrival rate into the single-multiplexer will increase to $\lambda^1 = \lambda i/Rc$. We assume that the average load from each source is fixed while the total load $\rho = \lambda/\mu$ changes with varying N.

FEC Performance with Interleaving: Now we suppose the packets from each homogeneous source are interleaved with the same interleaving depth before being transmitted into the network. The algorithm for computing the block-error distribution P(j,n) for a single source can be extended to include the interleaving procedure, as provided in previous Section. It can be expected that, compared to the case of a single source (N=1), the need for interleaving will be significantly reduced in a multiplexing environment (N ≥ 2), due to the already reduced packet-loss correlation as a result of the natural interleaving effect of multiplexing. Fig. 8 shows the to the already reduced packet-loss correlation as a result of the natural interleaving effect of multiplexing Depths, where the number of sources is N = 3 and the total system load is fixed at ρ =0.8 (Scenario 1) with buffer size K = 10. As expected, when, in order to optimize the FEC performance, an interleaving depth M ≥ 10 is required,

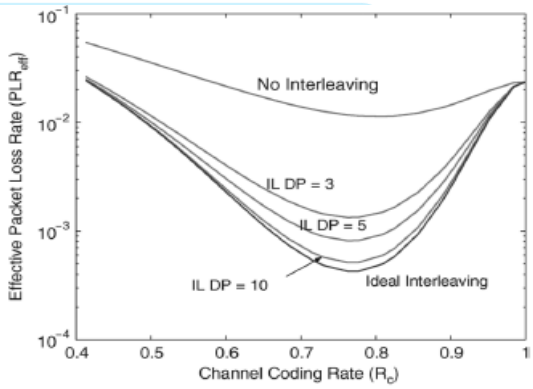


Fig9. Effect of interleaving on FEC performance with N=3 sources; Poisson arrivals, total load fixed at ρ = 0.8, K = 10, n=63.

FEC performance with different numbers of sources N for a given total load ρ=0.8. It shows that when the number of sources N increases, the need for interleaving depth decreases, which means reduced latency associated with the interleaving/de-interleaving operation. When N ≥ 14, interleaving makes an insignificant difference in FEC performance, because in this case the packet-loss process of each source is nearly independent.

3. Potential of FEC

Channel Model for Packet Transmission over Networks: Consider a channel model for packet transmission over a general packet-switched network. Assume a packet has M-bits. It is either transmitted and received by the receiver, or is lost due to network congestion or buffer overflow. For a received packet, bit errors may be introduced. Then packet transmission over networks can be modeled for coding purpose in terms of serial bit by bit transmission of M-bit symbols either over a binary symmetrical channel (BSC) with crossover probability ρ (state 0) or over a binary erasure channel (BEC) (state1),

Both of which are illustrated in Fig. 10 where ǿ is used to indicate the erasure symbol. A lost packet corresponds to the entire codeword symbol of m bits being erased, while a received packet means each of the m bits is sequentially transmitted over the BSC. This channel model belongs to the class of Block Interference Channels

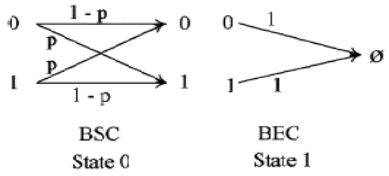


Fig10. Component channels of BIC corresponding to packet delivery and loss.

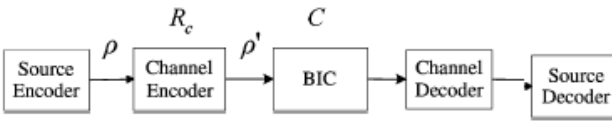


Fig11. Simplified communication system model.

Block Interference Channels (BIC), introduced by McEliece and Stark. Let S € {0,1} represent the state space of the BIC. If the state transitions are independent, then the Shannon capacity of the BIC is given as,

$$C = Es\{Cs\}; \text{ bits/transmission}$$

Where Cs is the capacity of the component channel s €

S, and the expectation is over the state space S. It

Follows that

$$C = (1-\rho) * (1-H(p)); \text{ bits/transmission},$$

Where ρ is the probability of being in the loss state and H(p) is the binary entropy function,

$$H(p) = -p \log p-(1-p) \log (1-p); 0 \le p \le 1.$$

4. *Information Theoretic Bound On FEC*

Suppose the interleaving is ideal, and consequently the packet-loss process seen by the channel decoder is independent. If we consider the interleaver and the de-interleaver as components of the coding channel, then the channel, consisting of the interleaver, the single-multiplexer and the de-interleaver, can be modeled as a BIC with independent state transitions

Here we consider only the packet losses caused by the buffer overflows, and assume no bit errors, i.e., the BSC crossover probability p=0. Let PL be the packet loss rate of the single-multiplexer, so p= PL. Then

$$C=(1-p)*(1-H(p))=1- PL$$

Assume the source creates packets at rate λ and the

packet service rate is μ. Then the normalized system load before coding is ρ = λ/μ. The channel encoder applies channel coding (not necessarily RS codes) with coding rate Rcto the source traffic.

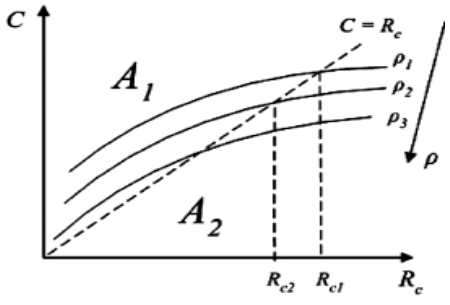


Fig12. Schematic illustration of the functional relationship C =1 – f (ρ / Rc), for different values

of ρ with

ρ1 ≤ ρ 2≤ ρ3.

Then the normalized system load after coding will

increase to ρ1 = ρ/Rc. Given the buffer size K, the average raw packet loss rate PL depends only on the load ρ1, as expressed by PL = f(ρ1) = f(ρ / Rc)

Where the function f can be determined by queuing analysis of the single-multiplexer model C = 1 - PL = 1 – f(ρ / Rc)

For a given load, we can plot the functional relationship of C with Rc. The channel coding theorem establishes that any rate less than the channel capacity can be supported with arbitrary low error probability. In other words, with regards to our model discussed here, as long as the channel coding rate Rc is smaller than the BIC capacity C, the source rate can be supported with arbitrarily high reliability.

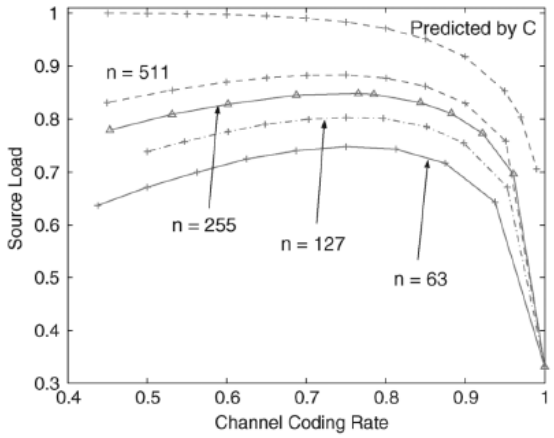


Fig13. The upper bound on the source loads ρ max predicted by the channel capacity considerations, compared to the maximum source loads ρ max that can be supported at a fixed effective packet-loss rate $PLR_{eff} = 10^{-5}$ using FEC coding; Poisson arrivals, ideal interleaving,

For the M/M/1/K model shows the upper bound on the source loads that can be supported as predicted by the preceding channel capacity considerations. It shows that with increasing coding block size the end-to-end performance achieved by FEC coding approaches that predicted by channel capacity.

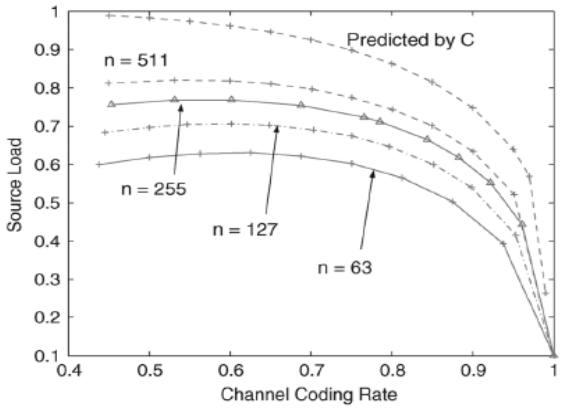


Fig14. The upper bound on the source loads ρ max predicted by the channel capacity considerations, compared to the maximum source loads ρ max that can be supported at a fixed effective packet-loss rate $PLR_{eff} = 10^{-5}$ using FEC coding; Poisson arrivals, ideal interleaving, using FEC coding; Poisson arrivals, ideal interleaving, K = 5.

The case of a smaller buffer (K = 5). The two figures indicate that, generally, the system with a larger buffer has a larger capacity. Note that in these two figures the capacity C is the capacity of the single-multiplexer combined with an ideal interleaver/de-interleaver, and not the capacity of the single-multiplexer itself. Actually, the capacity of the single-multiplexer channel can be greater than C described here since the capacity of the memory less interleaved channel is generally lower than the capacity of the original channel.

5. Conclusion

As the above analysis on the efficiency of FEC in packet losses in networks based on a single-multiplexer network model and explored potentiality of the FEC in recovering the packet losses occurred due to congestion at a bottleneck node of a packet-switched network, provided that the coding rate and other coding parameters are appropriately chosen. A discrete-time Markov chain model is used to analyse the efficacy of interleaving in improving the FEC

performance and determined how much interleaving depth is required for FEC to approach the optimum performance.

6. Future scope

The implementation complexity of FEC coding and the corresponding coding/decoding delay also need to be considered, which an issue particularly important for real-time applications. One objective for future work is the analysis of the additional delay caused by the FEC coding, perhaps combined with interleaving/de-interleaving. Likewise, the application of FEC for network transport is limited by the time-varying and often uncertain error characteristics of the channel, which makes the appropriate choice of FEC coding rate difficult to determine. In real-world applications, FEC coders are required which can adapt the channel code rate to the time-varying channel conditions.

7. References

[1] O. J. Boxma, "Sojourn times in cyclic queues—The influence of the slowest server," in Proc. 2nd Int. MCPR Workshop on Computer Performance and Reliability, Rome, Italy, May 1988.

[2] D. Y. Eun and N. B. Shroff, "Network decomposition: Theory and practice," IEEE/ACM Trans. Networking, vol. 13, no. 3, pp. 526–539, Jun. 2005.

[3]D. Gross and C. M. Harris, *Fundamentals of Queuing Theory*, 3rd ed. New York: Wiley, 1998.

[4]I. Cidon, A. Khamisy, and M. Sidi, "Analysis of packet loss processes in high-speed networks," *IEEE Trans. Inf. Theory*, vol. 39, no. 1, pp. 98–108, Jan. 1993.

[5] A Model-Based Approach to Evaluation of the

Efficacy of FEC Coding in Combating Network

Packet Losses Xunqi Yu*, Student Member, IEEE*, James W. Modestino*, Fellow, IEEE*, Ragip urceren*, Member, IEEE*, and Yee Sin Chan*, Member, IEEE*

[6] V. Parthasarathy, J. W. Modestino, and K. S. Vastola, "Reliable transmissions of high-quality video over ATM networks," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 361–374, Mar. 1999.

[7] A. Jean-Marie, P. Dube, D. Artiges, and E. Altman, "Decreasing loss probabilities by redundancy and interleaving: A queueing analysis," in *Proc. ITC 18*, Berlin, Germany, Sep. 2003

ENHANCED APPROPRIATE DATA COLLECTION USING VIRTUAL TREE BASED SENSOR NETWORKS

1. M.Praveen Kumar, 2.N.Venkatram, 3.M. Sarada

1.M.Tech student, Dept of CSE, Nova's Institute of Technology, Eluru

2.Assoc Professor, HOD of CSE Dept, Nova's Institute of Technology, Eluru

3.Asst Professor, Dept of CSE, Nova's Institute of Technology, Eluru

ABSTRACT

Wireless networks are not depends on cables and wires in between of different devices here. It is utilizes the less amount of bandwidth compare to wired networks. Network or network topology creates with the help of dynamic nodes of environment procedure. In network topology start the transmission of packets using different time intervals and resources. Data transmission start based on multi hop network using multi paths. Previous existing system related multi paths or multi channels of different networks are not efficient for transmission of total packets in receiver side. Tree based wireless sensor networks are failure for transmission of packets. Different networks contains different number of issues are present for transmission of packets. Those problems related to QoS parameters content.

The above all systems related algorithms or networks limitations we overcome with the help of virtual tree based networks. These virtual tree based networks provide the solutions for variable data load also. In virtual tree we use the coordinated spanning tree algorithm in implementation. These algorithms work as efficient algorithms in data transmission and collecting data also. Here we check the performance related to different number of QoS parameters. All performance shows the result as increased performance. At last we compare the previous algorithms and proposed algorithms. Proposed algorithms are best we provided in implementation related to delay, throughput and latency.

KEYWORDS: Wireless Sensor Networks, Coordinated Spanning tree algorithms, QoS parameters.

I.INTRODUCTION

Virtual tree based wireless sensor networks contains major objectives. Those major objectives are high data rate transmission with minimum cost and resources. Wireless sensor networks provide the services in different applications like remote data transmission applications. Here we consider the concept of fast data collection in receiver side. Data collection tasks are possible with the help of different layers.

Previous wireless sensor networks are spanning the complete networks with the help of time division multiplexing concepts (TDMA) in single channel. It is not give the proper solution for collecting efficiently. After some days we offer multiple channels for data transmission. Multi channels tasks also some interference problems are available here.

The above all interference problems we overcome using virtual tree based wireless sensor networks for data transmission. In virtual tree based network adding the coordinated spanning tree algorithms. It controls interference and gives the optimal solution in fast data collection.

II.RELATED WORK

Receiver side collecting data is a major operation in wireless broad area sensor networks. Collecting data like maximized utility content possible with the help of life time networks. In collecting data stage itself contains performance issues and risks here. In present related work we work different concepts and identify the some analysis results. These analysis results are not efficient and effective here.

Authors are discussed about many number approaches here. First existing approach related to TDMA (Time division multiple access). Time intervals based data transmission is available here. Data transmission we start in between of sender to receiver here. Here it is the single channel communication. Using this approach it's not possible to collect the maximized data in receiver. Receiver collects the data as a limited data only in implementation process. This is not best approach to collect efficient data. Some links are handling high load, same load it's not possible in all links here.

The above approach problems we overcome in multiple channels here. Time intervals solutions are not good. Multi channels start the data transmission as a parallel channel environment. Here using multiple channels increases the data collection possibility here. In this present network controls the packet errors, but it is not possible to deliver the total data in destination point or receiver side. Compare to previous approach it's possible to increases the performance levels here. It is somewhat fast approach to collect the data in receiver side. In some positions some of the interference problems are generate here.

The above approach interference problems we control using correlation in between multiple channels. Using correlation mechanism may chance to adjustment of power and remaining resources also in network environment here. Receiver gets the quality results without interference in new network environment. In detection of interference we implements the approximation related logarithmic approaches. These algorithms are not works properly for controlling the interference problems here. It is not gives 100% of data delivery content in receiver side here.

Reduce the interferences and time intervals we introduce the new scheduling algorithms for gets the optimal results of content. In network environment we know the details of link or path capacities. These all details identifies based on training procedure.

After we the details of capacities we allocate the packets in distributed number of channels efficiently as a load balancing results here. Load balanced channels there is possibility to generate the interference problems. All packets it may chance to deliver the content in receiver side here. Sometimes some channel failures problems are generated here.

The above the channel failure problems we overcome using disjoint algorithms implementation next proposed system approaches here. Using disjoint number of routes generate the mapping in between of different routes. These disjoint routes combination also works in certain area limits only here. This approach is called minimum distance of minimum spanning tree based approach. In some locations of coverage areas some problems are available. It takes more amount of time deliver the content in receiver side. Some problems are available here that is called delay.

Those disjoint routes of problems we overcome in topologies also. All topologies problems are identifies in implementation and control in present algorithms. Different networks are all problems we control in implementation part here. Finally here we control all routing topologies of different networks problems like interferences. Deliver the original data in receiver side effectively in implementation process.

III.PROBLEM STATEMENT

Previous different approaches are implemented here. Those approaches are scheduling, channel assignment, power control, and impact routing and channel models. These all approaches are control the static data deliver maintainance in implementation process. These approaches are not works for variable data transmission content. Here some new interference problem is generated here. These problems are showing the effect in performance parameters. Those parameters are called throughput, delay and some other parameters.

The above problem we overcome using virtual tree related approaches. In all virtual routes allocate the capacities also virtual in implementation part here. This is good control the mechanism with different routes. These types of routes we use in high data rate transmissions environment. These types of virtual tree route related algorithms are provides the efficient results in implementation part. This is one of the new data collection methodologies here. It's give the best performance related different parameters here.

IV. PROPOSED SYSTEM MODEL

Data transmission is required network design. Once data transmission load it may chance to change automatically network design also change here. This is called virtual network design creation. It gives the successful solution for transmission of total number of packets in destination point here. This is new collecting data design using virtual channels mechanism. Using proposed networks possible to increase the life time and high data rate transmission also in implementation part.

Here we use the adoptable based approach for scheduling and transmission data purpose in network channels here. This mechanism gives the light weight solution

with help of virtual carrier's mechanism. These approaches are performs efficient scheduling for transmission of data. In proposed design architecture we consider different parameters. Those parameters are

1. Minimized length of route
2. Minimized latency
3. Minimized energy utilization
4. Maximized utilization
5. Reduce the end-end delay, zitter

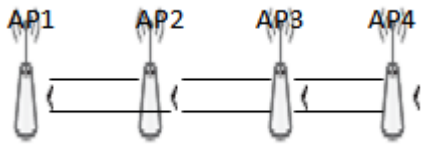


Fig1: Variable packet assignment in multiple channels

Different steps are required to implement new virtual channel assignment in network design

1. Creation of network design
2. Assignment of packets in multiple channels
3. Changes the packets allocation (channel variability)
4. Coordinated spanning tree construction
5. Performance Evolution

4.1 Creation of network design:

Create the number of nodes with different distances and costs here. This is called system model creation. Using different nodes generate the different number of paths. Every path of distance and cost we define here as a first network design content creation. We have the knowledge of each and every path capacity here in implementation part.

4.2 Assignment of packets in multiple channels

In all number of paths select only minimized length of the paths in implementation process. After select the number of paths allocates the packets as a load balanced manner for transmission. In scheduling time we use the different scheduling algorithms for resource allocation efficiently for transmission. Load is matching with capacity. Once change the load possible to create the virtual carriers in implementation process.

4.3 Changes the packets allocation (channel variability)

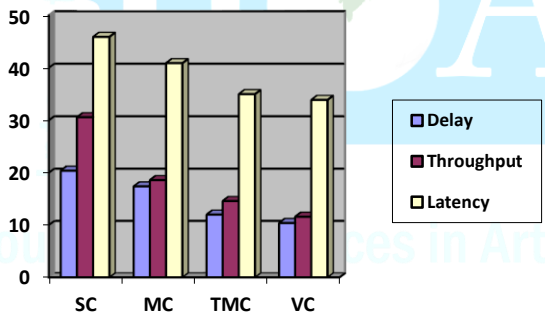
Previous modules are not works properly then we offer another module that is called channel variability. This module provides the optimal solution for any kind of load. This is implementation related virtual carriers specification. Dynamic load is accepts all new channels implementation in proposed system. This is possible with the help of dynamic programming. This same procedure works until destination point in all different links of environment procedure.

4.4 Coordinated spanning tree construction

One tree route or path is not gives the best result. Some other disjoint routes are participated in implementation part content. All disjoint routes placement also possible with different relay of content identification. These relays are give the good analysis process in implementation part here.

4.5 Performance Evolution

All packets data collected in destination point or receiver side effectively in implementation part. The new modules of architecture give the accurate solution for transmission of total packets. Here in present architecture we satisfy all QoS parameters in transmission. Some parameters of performance results show into bar chart or bar graph.



In this bar chart we have the details of four approaches. Those approaches are single channel, multiple channels; Tree based multiple channel and virtual channel. Compare to all previous approaches virtual channel packets assignment or scheduling is best in transmission and collecting data tasks.

VI. CONCLUSION AND FUTURE WORK

Existing Tree based data collection techniques are not provides the optimal for high load. It works for low data rate transmission purpose only in implementation. It gives the NP- hard problem related solution. Now new channel variability related solutions also we satisfies with new virtual carrier's algorithms and different scheduling algorithms also. It shows the optimal solution for transmission. All QoS parameters we increase in transmission of data and fast data collection also possible here in implementation.

VI. REFERENCES

[1] S.Thivyan Ravi Kumar, B.Maheswari, Dr. S.Karthik, DYNAMIC AND EFFICIENT TREE-BASED DATA COLLECTION IN WIRELESS SENSOR NETWORKS, 2013

[2] DIPAK B. KHADSE, MANISH NARNAWARE, COMPARATIVE ANALYSIS OF VARIOUS DATA AGGREGATION TECHNIQUES IN WSN, 2013

[3] S.Thivyan Ravi Kumar1, B.Maheswari2,, Dr. S.Karthik3, A SURVEY ON ENHANCING CONVERGECAST IN WIRELESS NETWORKS, 2013

[4] V.REDYA JADAV, D.SRAVAN, A.DEVILAL, Tree-Based Wireless Sensor Networks in Fast Data Collection, 2013

[5] Xuelian Cai, Energy-efficient Relay MAC with Dynamic Power Control in Wireless Body Area Networks, 2013

[6] Safieh Khodadoustan , Majid Hamidzadeh, Tree of Wheels: A New Hierarichal and Scalabale Topology for Underwater Sensor Networks, 2011

[7] Siddhartha Chauhan, Lalit Kumar Awasthi, Cluster Based Task Scheduling in Wireless Sensor Network, 2011

[8] Jayachandran.J 1 and Ramalakshmi.R Fast Data Collection with Reduced Interference and Increased Life Time in Wireless Sensor Networks, 2013

[9] Wan Du and Mo Li ,Harnessing Mobile Multiple Access Efficiency with Location Input, 2013
[10] Abhinava Sadasivarao* Sharfuddin Syed* Ping Pan* , Open Transport Switch - A Software Defined Networking Architecture for Transport Networks,2013

ENHANCED USEFUL FREQUENT ITEM SETS EXTRACTION USING PARALLEL PARTITIONING ALGORITHMS AND DIFFERENT FUNCTIONS

1 Anil Kumar Chandu,2. N.Venkatram,3.P.V.Kishore Kumar

1.M.Tech student, Dept of CSE, Nova's Institute of Technology, Eluru

2.Assoc Professor, HOD of CSE Dept, Nova's Institute of Technology, Eluru

3.Asst Professor, Dept of CSE, Nova's Institute of Technology, Eluru

ABSTRACT

Best Frequent pattern detection is the major problem data mining domain. Frequent pattern are extract the useful information. It is major important step in data mining approach. Frequent association rule association is good statistics results calculation mechanism here. In all real time application every time we require updated transaction content. Those updated transaction itself perform the association rule mining algorithms. Most Existing algorithms start the scan sequential or parallel. Sequential and parallel association rule mining algorithms are not giving the quality results. All current algorithms have some limitations like I/O cost, computation cost and CPU cycles utilization.

The above all algorithms limitation we overcome with new parallel partitioning algorithms. Those parallel partitioning algorithms implement in multiple processors parallel manner here. After processing in multiple processors extracts the frequent item sets content here. These are not quality. Those frequent item sets content apply the suggested function and airthematic function. These all functions show the quality frequent pattern items information. Those all frequent item sets are sufficient and meaningful. These are optimal frequent item patterns content. In this paper at last we show the performance levels about existing and proposed system algorithms.

KEYWORDS: Association Rule mining, parallel approaches, partitioning approaches, suggested and airthematic functions.

I.INTRODUCTION

Data mining extract the frequent item sets useful data from large databases using different association rule mining techniques. These association rule mining technique apply in business and science applications. All applications every time expect the new patterns creation process. In pattern data is available as a summarized data part content specification process here. Total process we call as a discovery pattern environment.

Existing sequential and parallel algorithms are failing in extraction of patterns data in large databases content.

Present patterns is not gives the better solution for takes the decision. These are not useful patterns in present cases. Performance and efficiency wise also we show the problems in implementation.

Now we create the efficient frequent item sets patterns and give the meaningful information. Those algorithms are parallel partitioning related approaches. After extraction of frequent item sets apply the different functions. Those functions are gives the optimization frequent item sets content. We show the experimental results in bar graph or bar chart.

II.RELATED WORK

Association rule mining technique also related to data integration and data cleaning. Here a large database converts into small databases here. In large databases discover the interesting pattern using some data mining algorithms. Those all previous algorithms have some limitations in computation, storage and time complexities. Different algorithms are applying different rules in large datasets. Algorithms are shows the different patterns, those patterns are not interesting patterns. Every time based trend changes patterns also it may chance to change. Now in below paragraphs of content we discuss about different existing algorithms issues.

Some existing association rule mining algorithms scan the transactions in single databases only. These algorithms generate the local user's analysis decisions only. It is not sufficient information for getting the good benefits here. New algorithms start the processing in relationship databases or global databases. Apply the association rule mining techniques in global databases for better analysis here.

Existing Apriori algorithm start the database scan for identification of frequent item sets. Identification of all frequent item sets apriori algorithm use the multiple passes here. Its show the more number of candidate set generation. It takes more amount of time for processing and storage also. These all results show based on threshold content here.

Next another existing approach FP-Tree (Frequent Pattern Growth) algorithm. This algorithm gives the results within two steps only. It is not show the candidate set results content here. FP-tree algorithm consists of two steps. Those steps are prefix tree and item set tree index. It is reduce the some number of frequent item sets, this is related to compress the frequent item sets. It is one of the good pruning technique. It has some limitations. Those limitations are first one allow the duplicate data. There is no sufficient interaction in between of different frequent item sets here. It is not possible for incremental mining process. The above two techniques are expects the more amount of I/O cost.

Reduces the I/O cost and increase the performance using parallel association rule mining algorithms. That is called shared memory multiple processors. Database scan procedure applies into directly in multiple processors. We complete to identification of frequent item sets very easily here. Next Tree based incremental association rule mining technique performs in large datasets. It is discover the candidate sets in single pass. We

reduce the number of steps in extraction of frequent item set results. Here we have some more practical issues are present. It is not work into all practical cases environment here.

Some of the authors are propose the new algorithm that is called frequent updated frequent pattern mining. Compare to previous algorithm we reduce the some number of links in communication environment. We get here fewer amounts of frequent items. Those all frequent items are occupy less amount of data storage that is called as a space here. Here in present system there is no perfect instructiveness. It is not possible to reduce all unnecessary number of frequent item sets here.

New algorithm we propose here that is called interactive association rule mining algorithm. Interactive mechanism possible reduces more number of unnecessary of frequent item sets of content here. Interactiveness calculates the weight of each and every frequent item. After calculation of all frequent items weight. In all frequent items weight select only highest weight of frequent item sets only here.

Some other different authors are proposing the different association rule mining algorithms. Those algorithms like continuous incremental association rule mining algorithm and clustering based algorithms. These algorithms give the cluster based results. These results are not contains the any sufficient pattern. These are static pattern. There is no perfect updating procedure based on time intervals here.

Lastly new authors are proposing another new technique periodic updating related incremental association rule mining. It gives best and powerful meaningful frequent patterns. The above all algorithms are low performance metrics in different complexities.

III.PROBLEM STATEMENT

In previous algorithms, in some of the algorithms we use the multiple passes, two passes and single pass. Previous all algorithms different risks and limitations are present here. Those all previous algorithms metrics values calculates using I/O cost and CPU cycles. The above all algorithms show the less performance wise related metrics here.

The above all algorithms performance issues we overcome with proposed system method. New algorithm we propose as a parallel association rule mining. In parallel association rule mining algorithm implement new suggested function. Suggested function gives the rating related results. In rating frequent items apply the airthematic function for good relationship. New parallel association algorithm show the good test accuracy related frequent item sets as a final result in implementation process. Here we use the less amount of I/O cost and CPU cycles.

IV.PROPOSED SYSTEM MODEL

Fig1 consist of all proposed system architecture steps. First we consider the transaction database as input. In transaction database apply the extraction operation. All frequent items related data extraction

possible after partitioning only here. Total transactions we divide and allocate into different processors. In all processors process the data and identifies the efficient frequent patterns content. In those frequent items itself apply different extraction function. Those new extraction functions give the better results in implementation part. Those new extraction functions are suggested function and airthematic function. Suggested function identifies the similar characteristics of frequent item sets content and allocates the rating. In rating content apply the airthematic function and identifies relationship related frequent items as a minimized frequent item sets. Those minimized frequent item sets are optimized frequent item sets here.

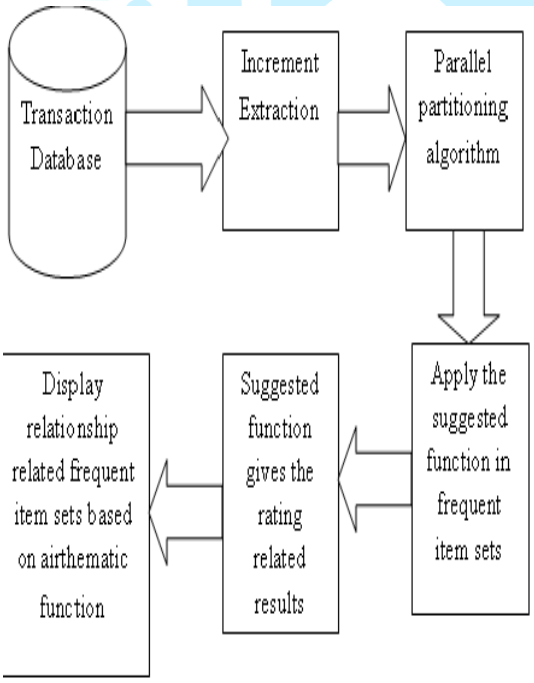


Fig1: Proposed System Architecture

Proposed system architecture divides into number of modules. Those numbers of modules are

1. Transaction database creation
2. Increment extraction using parallel partitioning algorithm
3. Apply the suggested function and airthematic function
4. Display the frequent item set extraction.

4.1Transaction Database Creation

Different applications generate the different number of transactions related database here. These transactions generate with different users interactions content. Interactions depend on based on user's opinion here.

4.2 Increment Extraction using Parallel partitioning algorithm

Transaction database divides into number of parts. Each and every part store in one processor. Parallels all processors start the process here. We complete the frequent item sets extraction very easily here in implementation part. Here we show the result with reduce the time and cost and space.

4.3 Apply the suggested function and airthematic function

Parallel partitioning algorithm provides some item sets. These are not necessary item sets. Now we apply suggested function for identification of similar frequent item sets content. We provide here combination of different frequent item sets content. We start the task to reduce the memory and all details here. After identification combination based results we apply airthematic function and display only

necessary. Those frequent item sets are meaningful and semantic.

4.4 Display the frequent item set extraction

Different Frequent association mining techniques we apply and extract the required results. Those results are personalization frequent item sets content here. All personalization frequent items combination is good, efficient with performance also.

V.EXPERIMENTAL EVOLUTION

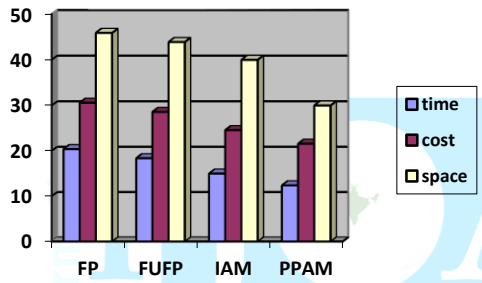


Fig2: performance evolution in different algorithms

In above graph we show the performance in between four different algorithms here. Those algorithms are FP, FUFP, IAM, and PPAM. In all different algorithms show the performance in different parameters. Those parameters are time, cost and space. Compare to all previous algorithm proposed system algorithm is best.

VI. CONCLUSION

Here in this paper we discuss about all existing frequent association algorithms and new algorithms also. Previous all existing algorithms are not possible to reduce the frequent item sets. Some item sets are unnecessary item sets are present, because those problems we got the some performance issues here. Using proposed frequent item set association rule mining reduce the frequent item sets and show the better performance levels related to cost, time and space complexities.

VI.REFERENCES

[1] A.Muralidhar, Pattabhiraman.V, Bench Marking Frequent Item set Mining Models and Algorithms: Current State of the Art, 2013

[2] Hany Mahgoub, IARMMD: A Novel System for Incremental Association Rules Mining from Medical Documents, 2013.

[3]T.SathishKumar,V.Kavitha,Dr.T.Ravichandran, Efficient Tree Based Distributed Data Mining Algorithms for mining Frequent Patterns, 2010

[4] Vinay Kumar Khare,Vedant Rastogi, Mining Positive and Negative Sequential Pattern in Incremental Transaction Databases,2013

[5] Chien-Min Lin, Yu-Lung Hsieh, Kuo-Cheng Yin, Ming-Chuan Hung, Don-Lin YangADMiner: An Incremental Data Mining Approach Using a Compressed FP-tree, 2013

[6] Sotiris Kotsiantis, Dimitris Kanellopoulos, Association Rules Mining: A Recent Overview, 2006

[7] Pradnya A. Shirsath, Vijay Kumar Verma, A Recent Survey on Incremental Temporal Association Rule Mining, 2013

[8] Sanjay Patel* and Dr. Ketan Kotecha, Incremental Frequent Pattern Mining using Graph based approach, 2013

[9] Frans Coenen, Paul Leng, and Shakil Ahmed, Data Structure for Association Rule Mining:T-Trees and P-Trees, 2004

[10] Osmar R. Za¨_ane Mohammad El-Hajj Paul Lu,Fast Parallel Association Rule Mining Without Candidacy Generation*?,2012*

[11] Jyoti Jadhav, Lata Ragha, Vijay Katkar, Incremental Frequent Pattern Mining, 2012

Enhanced Noisy Reduction Using EHC & SHR Algorithms

1.R.Nagaraju, 2. N.Venkatram, 3. P.V.Kishore Kumar

1.M.Tech student, Dept of CSE, Nova's Institute of Technology, Eluru

2.Assoc Professor, HOD of CSE Dept, Nova's Institute of Technology, Eluru

3.Asst Professor, Dept of CSE, Nova's Institute of Technology, Eluru

ABSTRACT

Noisy reduction speed up and optimal results possible in present EHC and SHR algorithm. These two algorithms are good learning algorithms in noisy. These algorithms are competitive algorithms related learning. Algorithms are applicable in pattern recognization, structure formation. Existing many algorithms, K-mediods, K-means, CURE approach, Fuzzy C-means are not detect accurately in all dimensions. These algorithms are utilizes the more amount of space and time. Algorithms are show the some overhead problems related to computation. These algorithms are not efficient learning algorithms in noisy.

In this paper we propose new algorithms, those algorithms are EHC (Exhaustive Hierarchical Clustering) and SHR (Subset Hierarchical Refinement). These algorithms are effective pruning algorithms in noisy reduction. These algorithms are defining the different rules in processing and identify the best closest algorithms in implementation. Closest results are identifying based on minimum distance techniques and threshold based techniques. These two techniques are not gives effective results then we perform refinement approach to remove the total noisy. Algorithms are gives the results like accurate noisy reduction with less computational overhead. Here we show the comparison with existing and proposed system algorithms. At last proposed system algorithms are best in time and space complexities wise in implementation.

KEYWORDS: Learning algorithms, pattern recognization, Refinement Algorithms.

I.INTRODUCTION

Many numbers of applications and organizations are required related to noisy reduction techniques. These noisy reduction techniques apply in large databases environment. Now a day every day databases of data is increases here related organization. Data mining related techniques are required for analysis in large databases. Analysis purpose we offer clustering techniques. Clustering techniques are good decision making techniques. Clustering algorithms are attractive algorithms. Noisy Reduction is the benefit task in cluster analysis. Noisy reduction improvement we show in

implementation with new different algorithms.

Previous clustering algorithms are not detect maximize. Those algorithms are K-mediods, CURE, and Fuzzy C-means. K-mediods are works based on centroid distance. It works in limited distance only. That's why remove the noisy in less number of dimensions only here. CURE approach removes the in large distance but it's not cover all distances. Remaining algorithms also fail in noisy reduction accurately. These algorithms show the disadvantages related time and space.

In this paper we propose the new algorithms, those algorithms are Exhaustive hierarchical clustering and Subset hierarchical refinement. These algorithms are removes the noisy maximize. In remaining sections we show the perfect comparision in between of existing and proposed systems of algorithms.

II. RELATED WORK

Clustering is the fundamental approach in detection of patterns and pattern recognition. Some issues are available in processing of clustering analysis. Those problems are time and space complexity. We identify these problems in handling of noisy data. Now in this paper we enhance the performance and controls noisy data in processing and extraction. In processing new clustering approaches is not allowing noisy data. We reduce the overhead problems.

Previous approaches CURE, K-means, K-Mediods, and Fuzzy C-Means are not handle noisy data efficiently. The above algorithm does not recognize the structures effectively. First approach CURE starts the cluster analysis. Clustering analysis starts based on data points. All data points are arranged into different shapes or structures. Using centers or centroid start the identification of data points and place into shapes or structures. It's works in local centers only. It's not detects the noisy in global centers of data points. CURE is not possible reduces the noisy data points global centers of environment process.

Next approach, K-means start the cluster analysis based on Euclidian distances. K-means algorithms control the noisy in less number of dimensions. It does not cover all dimensions content. It's work in limited distance only for identification of dissimilarities content and remove noisy data also. After remove noisy data in less number of dimensions, remaining locations of noisy data store in database. In processing of data and extraction we spend the more amounts of time and cost. Its show the high overhead problems here.

Reduce the overhead problems with new clustering approach that is called as a k-mediods algorithm. It's works in high dimensional content and any distance of work in implementation. High dimensional content show the clustering process in different clustering process. In different clusters apply the overlap or intersection operation. All clusters are overlap slightly and remove the less amount of noisy only. It's also inefficient techniques under control of noisy.

Next approach, Fuzzy clustering we introduced for reducing the noisy in different clusters also. This is completely related to partitioning approach. Here many numbers of clusters are available here. It's possible to reduce the clusters and reduce the noisy also. This complete approach everything depends on subset based on partitioning. It works for small data only. This is completely works on membership function. All data points of clusters information allow and store into subset. Its fuzzy clustering also allows the noisy. We store the more number of data points of content. It takes more amount of processing I/O cost and CPU cycles of data. Some limitations are available here, those limitations are its not work in large databases of content here. It's possible to perform only one single time.

Next existing approach, Fuzzy C-means clustering is applicable in large databases environment also. We find out centers based on dataset points of content. One center data is not possible to identify the total noisy content, update the center value till identification of noisy reduction. Every time same steps perform in implementation process. Fuzzy C-means is available as a iterative approach. It works as a hierarchical clustering approach. It works in large distance. All distances of results also possible to identify in implementation process. Present Fuzzy c-means approach works based on cuts. Certain distances only we start the searching process. It is not possible to remove all the distances.

The above all approaches have some limitations, issues and risks. Those all problems we overcome in following sections.

III. PROBLEM STATEMENT

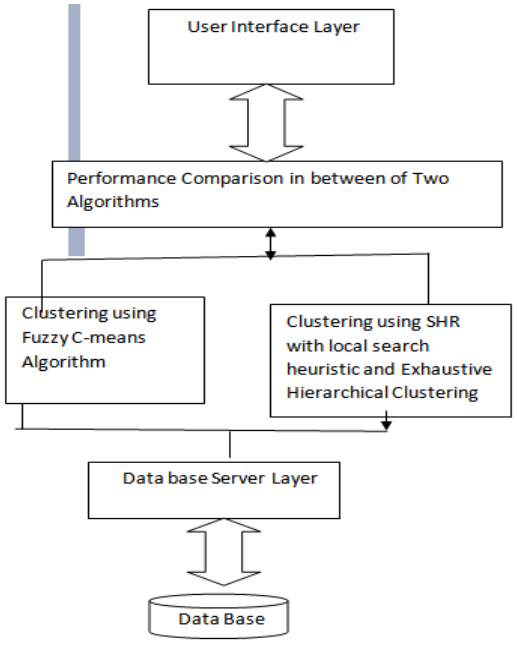
Enhance the noisy reduction using subset hierarchical reduction technique with local search heuristic and exhaustive hierarchical clustering approaches. These two algorithms remove the noisy data points in cluster analysis using pruning technique. Algorithms gives the efficient and closest necessary results. Here we use iterative approach for reducing the unnecessary data points of content until gets the accurate result. These algorithms are save the memory, I/O cost and CPU cycles. It gives the good noisy reduction direction. These algorithms work as good decision making and learning. Here we show the comparison with different clustering techniques. All clustering techniques of performance represents in the form of time complexity and space complexity. We conclude Combination of two algorithms approach is better and accurate.

IV.SYSTEM MODEL

The overall system architecture contains different steps. All steps of content are organized here. Data points of content are present in database. All clusters data points content is present in under control of database server. In database server of all clusters start the analysis process with the help of different kinds of algorithms. After that next process step that is called performance comparison in between of different algorithms. In implementation we identify the reduced time complexity and space complexities. In two algorithms Subset hierarchical refinement with local search heuristic is best.

All performance related results show into user interface layer content with different clusters. Different algorithms are extracts different features and reduce the noisy reduction with different parameters.

First user interface layer creates for display of two different algorithms of data point's contents here. This is initialization process to start the work here. Next Process step in implementation transaction flow of clusters in processing of clusters content. Complete transactions related database operations content store in database environment. In transaction flow of operations perform each and every algorithm with all steps of content here. In Existing algorithm its remove the approximate noisy reductions of data points content here. Its not remove the all noisy data points. Now in proposed system using SHR remove the total noisy data points of content. It's give the accurate solutions in implementation part.



V.PROPOSED SYSTEM ALGORITHM

5.1 Exhaustive Hierarchical Clustering Algorithm for noisy reduction:

Consider the cluster data points of content as an input. All data points of information identifies as a cluster distance. Cluster distance data points divides as two bounds content distances. Those distances are maximum and minimum or lower and upper bound content.

Original data directly we are not store in heap. After remove the upper bound noisy content only considers the necessary data points that are called as a minimized data points. Those data points are called minimum distance data points content. Minimum distance points store in heap memory environment process. These data points are occupies less amount of space only. Second rule we introduce here that is called threshold based approach. Using threshold also some other noisy data points also we reduce here in implementation process. The above all steps are gives the efficient results in noisy reduction. It's show the performance saving of time complexity and space complexity.

Pseudo Code:

Input: Cluster data points, number of entries with minimum distance and threshold

Output: minimized cluster distance points

Step1: initially heap is empty

Step2: in total cluster data points divides into maximum and minimum distances

Step3: remove the upper bounds points as a noisy here

Step4: apply the threshold operations remove the more number of noisy data points content.

Step5: after removing of all noisy data points remaining data points store into heap

Step6: calculation of time and space complexities as a performance here.

5.2 Subset Hierarchical Refinement algorithms with local search Heuristic

The above algorithm is not possible to remove the noisy then changes the distance of cluster processing or cluster analysis procedure. Choose some other new distances for start the clustering analysis work in implementation content. Every time update the cluster analysis distance points and cluster processing here.

This is one of the good adaptation procedures in processing. Using proposed algorithms possible to remove all dimensions of noisy in implementation process. This is one of the good noisy reduction directions in execution process.

These above algorithms are not possible to remove all kinds of noisy results then we offer another algorithm that is called as a further candidate pruning noisy reduction technique.

5.3 Further Candidate Pruning Noisy Reduction Technique

This is one of the good replacement procedure and its offer only less amount distance cluster points or smaller distance points only. In all possible clusters areas remove the noisy data here. These final cluster data points are final subset point's information. It's give the accurate results with less computation overhead and operational cost specification process.

VI. EXPERIMENTAL RESULTS

The above algorithms we implement with different cluster data points information here. All existing algorithms are not accurate in overhead and computation of time and cost and space complexities. These algorithms are not accurate under noisy reduction. New algorithm practically we got results efficiently in noisy reduction with low amount of operational cost, time and space complexities in output generation.

Existing and present algorithms show the performance in between of cost, time and space parameters. Proposed algorithm is somewhat effective and efficient related to performance wise.

VII. CONCLUSION AND FUTUREWORK

Comparison of different clustering algorithms performance at last we show the best performance algorithm that is called Exhaustive hierarchical clustering with noisy reduction. These are removing the accurate noisy reduction techniques. After remove the noisy data points, remaining data points occupies less amount of space and time. It's give the better results in computation overhead wise and performance.

In Future work also integration of different clustering algorithms. This way gives the good research scope in future related to space and time wise.

VIII.REFERENCES

[1] SUDIPTO GUHA~, RAJEEV RASTOGI~, and KYUSEOK SHIMS, CURE: AN EFFICIENT CLUSTERING ALGORITHM FOR LARGE DATABASES+,pp 35-58, 2001

[2] Levent Ertöz Michael Steinbach Vipin Kumar, Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data,2011, pp: 1-2

[3] V.Venkateswara Rao, Dasu Dasari Efficient Clustering Techniques in presence of Noise, 2011

[4] Raghuvira Pratap A, K Suvarna Vani, J Rama Devi, Dr.K Nageswara Rao, An Efficient Density based Improved K-Medoids Clustering algorithm, 2011.

[5] Osonde Osoba, Bart Kosko, Noise-enhanced clustering and competitive learning algorithms,2012

[6] Metin KAYA, An Algorithm for Image Clustering and Compression, 2005

[7] Ronen Talmon, Israel Cohen, Transient Noise Reduction Using Nonlocal Diffusion Filters, 2011

[8] Ahmed K. Farahat, Mohamed S. Kamel, Enhancing Document Clustering Using Hybrid Models for Semantic Similarity, 2010

[9] Abdessalam H. Elhabbash, Enhanced k-means Clustering Algorithm,2011

[10] Vadeerat Rinsurongkawong, and Christoph F. Eick, Correspondence Clustering: An Approach to Cluster Multiple Related Spatial Datasets, 2010

Evolution of Record linkage Using Proposition Logic and Generational Evolutionary Algorithm

1.Y.NagaDurgaAravind,2. N.Venkatram, 3. M.Venu Gopal

1.M.Tech student, Dept of CSE, Nova's Institute of Technology, Eluru

2.Assoc Professor, HOD of CSE Dept, Nova's Institute of Technology, Eluru

3.Asst Professor, Dept of MATHEMATICS,Lingaiah`s Institute of Technology,Vijayawada

ABSTRACT

In multiple databases redundant and noisy records problems are available. Data cleaning is major approach for removing the noisy and redundant data. All databases convert as a quality databases. Previously some approaches are detects the duplicates. Those approaches are statistical record linkage and decision tree [1][2][3]. These all approaches are require more number of training phase and testing phases for removing the duplicates. It's not gives the accurate duplicate detection. This approach utilizes more amount of operational cost.

In this paper we introduce new approaches and possible to reduces the operational cost also. These new approaches are gives the accurate solution in extraction of records from multiple databases. First record linkage related statistical approach removes the duplicates using pair matching of records. Pair's matching approach show the result as unique records content. Here we use the less number of training phases approach. These unique records filter as a truth records or guaranteed unique records we find out using propositional logic. Its reduces the irrelevant records and show the result as a minimized records content. Combine of all records show it's as fitness records using generational evolutionary approach. These results are not fixed, every time adding the new records and show the results as a refinement manner. These are good performance records [3][4].

KEYWORDS: Data mining, Propositional Logic, Record Linkage, evolutionary approach.

I.INTRODUCTION

All government organizations are maintaining the different number of web databases. In present web databases any user forward the query large amount of data is displayed here. Same data it may chance to present in two or more number of databases. All researchers are interested in this field. Same real world entities detection is the major approach in multiple databases. This complete detection is possible with the help of record linkage. Record linkage is one good process technique in detection of duplicates.

Present record matching techniques apply into different kinds of query results. Those query results are positive, negative and both. In all types of query results for removing the duplicates it's very complex. It takes more amount of training cost. After some days some new predefined rules are implemented for extracting the high quality databases maintainace. Its gives partial optimal solution results with high operational cost.

To overcome such problems, issues and risks propose the new method that is called proposition logic [7][8]. This is one of the good unsupervised classification techniques. Here we perform the classification two or more number of times in detection of duplicates. It's gives the truth records. Those results are minimized results with minimized cost. All minimized results we arranges with generational evolutionary approach. Its gives good fitness and effective results records with less operational cost.

II. RELATED WORK

This complete paper related to data integration and data cleaning. Resolve the distributed databases problems and create quality databases environment. Creation of quality databases using many number data mining many approaches. Every previous approach some limitations and risks are present here. Every approach itself calculates the overhead and operational cost [1][2][3][4].

First and starting time of creation databases there is no importance to constraints. Whenever constraints are missing data entry problems are generated. Data entry problems are gives the ambiguity solutions. It's not possible to detect all duplicates. It's show less amount of statistical performance in duplicate detection [1]. These problems we observed in many number of real time applications. All real times applications identify the problems like high amount of operational cost utilization in detection of all duplicates.

After some number days next approach to detects the duplicates in distributed databases. That new approach is called as a merge list. Merge list we apply the approximate query processing environment. Compare to previous [2] approach this present approach gives the best results. It's not gives the best performance in detection of duplicates.

Next approach is called N-grams. It is completely incremental string environment approach. After enter one string it's not detects all duplicates, in first string add another string. Using two strings start the detection of duplicates [3]. This is process we continue until to detect all duplicates. In this approach we spend more amounts of operational cost utilization and overhead problem.

After some number of days for reducing the operational cost we introduce new format that is called as a distance based query. Distance based query also related to approximate query processing and remove the duplicates. No need to detect all duplicates in all distances. User required area itself start the detection of duplicates. User gets the less features, he will not satisfy.

Distance based metrics related duplicates it's not gives the better results [3][4][2].

Next approach related to probabilistic with statistical approach and naive Bayesian with statistical approach. These two approaches works based on similarity functions. Every record of similar records we find out here. Count the similar records and allocates as a weights. In all types of weights apply the threshold and categorizes the records. Below threshold records considers as a non duplicate and above threshold information consider as a duplicate records. Using these two approaches reduces the some operational cost [2][3][4][1].

Next approach is called as a rule based prediction. This rule based prediction also works based similarity record weight. Using different kinds of rules different clusters we generate here. In total number of clusters which cluster is best cluster no user is not identify the effectively. This is also one good information retrieval techniques under detection of duplicates. Its level by level approach in detection of duplication. Here also we spend the more amount of operational cost here [6]. The above all approaches are related supervised clustering. These all approaches are training approaches.

After completion of training approaches we choose the different number of testing approaches. Combine of training and testing approaches are called unsupervised clustering. Here in training approach only detection of all duplicates we spend more amount of operational cost. Second time classification detects the number duplicates and gives the results high duplicate ratio. Unsupervised is the one good learning approaches in detection of duplicates here.

The above all problems and limitations we overcome in new approach. That new approach we discuss in following contents.

III.PROBLEM STATEMENT

Previously decision trees are used. Decision tree starts align records based on conditions or constraints. This is one of the good classification approach. Classification is works as a training approach. Many times perform the classifications approach in present record linkage with decision trees. These approaches are requires more number of training phases. In classification time we spend more amount of operational cost.

In this paper we propose new statistical record linkage with new generational evolutionary approach. Using these approaches reduces the number of training phases and automatically reduces the operational cost. Using normal query process extracts the records. In these displayed records some records are available as a duplicate. We remove duplicate records using record linkage technique. Unique records we align with the help of different kinds of prepositions in generational evolutionary algorithm. Proposition related results show as a better and effectiveness results.

IV.SYSTEM MODEL

4.1 OUTLINE OF RECORD LINKAGE:

User forwards the query for retrieving of results here. Results are extracted from multiple databases. It may chance same results are present in two or more number of databases. Those results are duplicate results or records. Using cleaning approaches remove the duplicate records of content. All records we display as a unique records. We align all unique records as a index format. Index based records are aligns using similarity function. Similarity function starts the calculation of each and every record weight based on matching environment process. It is one of the evolutionary processes. It's not gives the proper alignment result, no user is not satisfied with results.

Similarity function related approach completely statistical approach. In statistical approach applies the probability related mechanism.

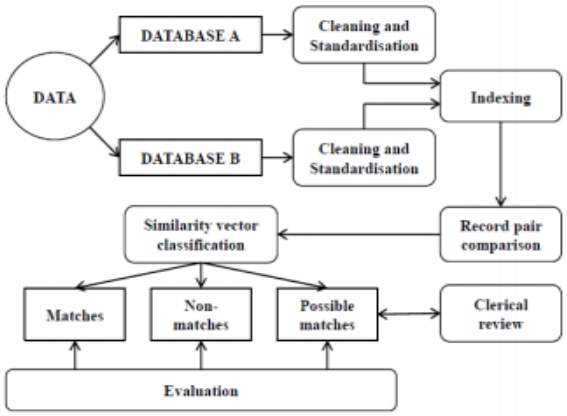


Fig4. 1: Outline related to record linkage

In this present record linkage whenever to start the calculation of statistical performance or statistical weight, there is no communication in between of one record to another record. This is called as a correlation problem. All problems and risks we overcome using proposition related generation evolutionary approach.

4.2 proposition related generational evolutionary approach

Using similarity function we identify the unique records weight. In similarity record weight it may chance to present the duplicates, no author is not gave the guarantee there is no duplicates. We give guarantee to remove the maximized duplicates using proposition logic. Proposition logic gives truth records in alignment with understandable content. It's show high duplicate ratio in detection of duplicate records.

These kinds of truth results we show with the help of different experiments. These experiments give the good proper alignment and performance results.

V. IMPLEMENTATION APPROACH WITH PREPOSITION LOGIC

We use the proposition logic for mapping of records. Using propositional logic aligns the records with fewer amounts of operational cost and time consuming. Propositional logic works as good inference rule logic. Using inference rules formulate the records with good relationship. Relationship of records creates as a structure.

Relationship generate with the help of different number of symbols and operators. Operators are use in between two or more number of records. This is called one verification approach in alignment of records. Verification approach gives the result as true or false information. This is one of the good mathematical component

procedures in relationship of records. We increase the efficiency in record alignment.

5.1 Basic concepts related to proposition Logic in record alignment

In basic concepts we follow some steps. Those steps are

1. Collection of similar records with good primitive symbols and operators.

2. Align the records with good meaningful representation

3. Apply the different inference rules in implementation process.

5.2 Identification of Truth records with Different Inference Rules

5.2.1 Apply the conjunction in between of Different Records

In between of two or more number of records apply the conjunction operation identifies the relationship records. These type results are deeper results and display highest features of content or vectors of content. Deep web data extraction is possible with the help of "AND" operation. It does not allow any false records information. Its show the minimization records in alignment. That's why we spend the less amount of operational cost in implementation process.

5.2.2 Apply the disjunction in between of Different Records

In between of two or more number of records apply the disjunction operation. Disjunction operation performs based on "OR". These operation related results also we discard in implementation of results extraction process. These types of conditions related results we reduce in extraction. Its gives the solution with less operational cost specification process.

VI.TEST ACCURACY IN DUPLICATE DETECTION

Record linkage and proposition logic reduces the number of duplicates records and provides the truth records. Compare to previous all approaches its show the good results. Those test accuracy operations are called Precision, Recall and F-measure.

Precision= Number of Correctly Identified Duplicate Pairs / Number of Identified Duplicate pairs

Recall= Number of Correctly Identified Duplicate Pairs / Number of True Duplicate Pairs

F-Measure = 2*P*R / P+R

These three operations are provides the good metric results in maximization of duplicate detection. It's providing the meaningful in alignment with different proposition logic operations. It's give high accuracy with less amount of operational cost.

VII. PERFORMANCE EVOLUTION

Different approaches related duplicate detection works based on operational cost. Compare to all previous concepts, present concepts show the less amount operation cost and overhead results. Those results show into output with help of experiment.

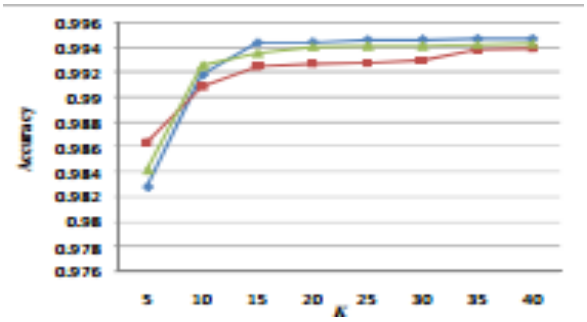


Fig2: Test Accuracy evolution with operational cost

In X-axis consider the operational cost and y-axis considers accuracy. Here in graph three iterations are present here. Every iteration automatically increases the accuracy. Those performance levels store in each and every line.

VIII. CONCLUSION

We test different approaches in record linkage environment. We have done the analysis of operational cost with different approaches. Previous all approaches utilizes the more amount of operational cost. Now in this paper we discuss about different approaches like propositional logic and generational evolutionary approach. Its gives good accuracy duplicates detection with less amount of operational cost. It's show the good optimal solution.

IX. REFERENCE PAPERS

[1] Duplicate Record Detection: A Survey, Ahmed K. Elmagarmid, Panagiotis G. Ipeirotis, Vassilios S. Verykios, 2006

[2] An Iterative Two-Party Protocol for Scalable Privacy-Preserving Record Linkage, 2012 Dinusha Vatsalan and Peter Christen

[3] A Comparison of Fast Blocking Methods for Record Linkage, Rohan Baxter, Peter Christen, Tim Churches, 2003

[4] Towards Duplicate Detection for Situation Awareness Based on Spatio-Temporal Relations, Norbert Baumgartner1, Wolfgang Gottesheim2, Stefan Mitsch2, Werner Retschitzegger2, and Wieland Schwinger2, 2010

[5] A Comparative Study of Duplicate Record Detection Techniques, Osama Helmi Akel, 2012

[6] Improving the Accuracy of Duplicate Detection at Scale, Aubrey Barnard and Allison Terrell,2009.

[7] A manual and introductory tutorial, David Martin, James Procter, Andrew Waterhouse, Saif Shehata and Geoff Barton, 2013

[8] Formal Veri_cation of Data Provenance Records, Szymon Klarman1, Stefan Schlobach1, Luciano Sera_ni2,2011

[9] Efficient Record Linkage in Large Data Sets, Liang Jin, Chen Li, and Sharad Mehrotra

[10] Efficient Private Record Linkage, Mohamed Yakout, Mikhail J. Atallah, Ahmed Elmagarmid, 2013

Frequent itemset generation using dynamic itemset counting & hashing technique

1.Ch.Rambabu 2.N.Venkatram 3. M.Sarada

1.M.Tech student,Dept of CSE,Nova's Institute of technology Eluru

2.Assoc Professor,HOD of CSE Dept,Nova's Institute of Technology,Eluru

3.Assistant Professor,Dept of CSE,Nova's Institute of Technology,Eluru

Abstract:

It proposes a new optimization algorithm called APPRIORI-EFFICIENT based on the insufficient of Apriori. APRIORI-EFFICIENT algorithm presents optimizations on 2-items generation, transactions compression and so on. APRIORI-EFFICIENT uses hash structure to generate L2, uses an efficient horizontal data representation and optimized strategy of storage to save time and space. The performance study shows that APRIORI- EFFICIENT is much faster than Apriori.

Keywords- Association rule mining,Frequent itemsets,Apriori algorithm,Candidate itemsets,Apriori using indexed tree,Aprori using incremental candidate sets ,Dynamic itemset counting

1.INTRODUCTION

Data analysis and data mining tools become very important to perform business analysis and intellegence with the help of data mining tools the business organisations finds the customer interests, customer behaviour, business growth and business trends. Association rule mining is one of the popular datamining technique which is used to find interesting relationships or corelation among the items involved in transactional records of transactional database.The association rule mining generates frequent itemsets and association rules by extracting business transactional data from database using these association rules decissions can be taken for market basket analysis,catalog design,store layout and cross marketing which improves business sales of organisation. Apriori is fundamental association rule mining algorithm to genetrate frequent itemsets,association rules but it does not provides performance.It generates large number of candidate sets to produce frequent itemsets.Association Rule Mining is a data mining technique which is well suited for mining Market-basket dataset. The research described in the current paper came out during the early days of data mining research and was also meant to demonstrate the feasibility of fast scalable data mining algorithms. Although a few algorithms for mining association rules existed at the time, the Apriori and Apriori TID algorithms greatly reduced the overhead costs associated with generating association rules.

2.PROBLEM STATEMENT

Association rule mining used to find associations or interesting relationships among the items involved in several transaction records of sales database.There are several algorithms proposed to implement this association rule mining as a part of

research.Apriori is an algorithm that is fundamental algorithm for generating frequent itemsets and association rules.The existing system identifies several limitations in this apriori algorithm implementation like finding more candidate itemsets and modified the apriori efficient .This apriori efficient uses hashing technique with linear probing to reduce candidate itemsets.The basic apriori algorithm generates a large number of candidate itemsets. Though apriori efficient reducess number of candidate sets,it is not efficient as the size of database increases.Because of more number of candidate sets there is wastage of space as well as time that is system is inefficient interms of time complexity and space complexity.

3. APRIORI – EFFICIENT ALGORITHM

APRIORI-EFFICIENT algorithm generates L_2 directly from one scan over the database without generating C_1, L_1 and C_2. And it replaces the hash tree by a hash table to reduce searching cost. It also uses an efficient horizontal data representation and optimization strategy of storage to save time and space. It is relatively simpler and easier to implement.

3.1 Generate L_2 directly by using perfect hash function

Let $I=\{X_1, X_2, \ldots\ldots, X_N\}$, where X_i is the identification of the i-th data item in the database and N is the number of distinguishable data items. $P(X_i)$ denotes the function that maps the identification of X_i to its position, that is, $P(X_i)=i$. Then $C_1=I=\{X_1, X_2, \ldots\ldots, X_N\}$,

let $C_1^2=C_1*C_1=\{(X_i,X_j) \mid X_i \in C_1, X_j \epsilon C_1, i<j\}$. All the possible 2-subsets in C_1^2 with first item being group name X_g is defined as

$$G_2(X_g)=\{ (X_i,X_j) \mid X_i,X_j \in C_1^2 \text{ and } X_i=X_g \} \quad (1\leq g\leq N-1)$$

The set of the second items in $G_2(X_g)$ is defined as

$$S_2(X_g)=\{ (X_j) \mid (X_g,X_j) \in G_2(X_g) \} \quad (2\leq g\leq N)$$

It can be seen that $\text{InterGroupOffset}=|G_2(X_g)|=|S_2(X_g)|=N\text{-item}_1=N-P(X_g)$, and $\text{IntraGroupOffset}=\text{item}_2-\text{item}_1 = P(S_2(X_g))-P(G_2(X_g))$. This relationship among the 2-itemsets in C_1^2 and the hash address are shown in the table 3.

TABLE III. 2-ITEMSET HASH TABLE

2itemsets	Group Name Item$_1$	P(i_1)	Item$_2$	P(i_2)	Inter Group Offset	Intra Group Offset	Hash Address
X_1,X_2	X_i	1	X_2	2	$\lvert G_2(X_0)\rvert=0$	1	1
X_1,X_3			X_3	3		2	2
.....			.....	.....		.....	.....
X_1,X_N			X_N	N		N-1	N-1
$\{X_2,X_3\}$	X_2	2	X_3	3	$\lvert G_2(X_1)\rvert=P(X_N)-P(X_1)=N-1$	1	
X_2,X_4			X_4	4		.....	.....
.....			.....	.....		.....	.....
X_2,X_N			X_N	N		.....	.....
.....	.....	.....			.....	.....	.....
X_{N-1},X_N	X_{N-1}	N-1	X_N	N	$\lvert G_2(X_1)\rvert+\lvert G_2(X_2)\rvert+\ldots+\lvert G_2(X_{N-2})\rvert$	N-(N-1)	N*(N-1)/2

Details of how to generate L_2 directly by using perfect hash function in Figure 1.

Apriori uses an efficient candidate generation procedure, such that only the frequent itemsets at a level are used to construct candidates at the next level. However, it requires multiple database scans, as many as the longest frequent itemset. The APRIORI –

EFFICIENT algorithm need to scan database only once in order to obtain L_1 and L_2. Thus, it saving the system time and improve the operating efficiency of the program.

3.2. Prunning Strategy

Two different pruning techniques are exploited. Dataset global pruning which transforms a generic transaction t, read from D_k into a pruned transaction t_1, and Dataset local pruning which further prunes the transaction, and transforms t_1 into t_2 before writing it to D_{k+1}.

Dataset global pruning is based on the following argument: t may contain a frequent k-itemset I only if all the (k-1)-subsets of I belong to L_{k-1}. Since searching L_{k-1} for all the (k-1)-subsets of any I $\subseteq$ t may be very expensive, a simpler technique, whose pruning effect is smaller, was adopted. Note that the (k-1)-subsets of a given k-itemset I $\subseteq$ t are exactly k, but each item belonging to t should only appear in k-1 of these k itemsets. Therefore, we derive a necessary (but weaker) condition to keep a given item in t: item t_j is retained in t' if it appears in at least k-1 frequent itemsets of L_{k-1}. Dataset local pruning has already been adopted by DHP.

D

TID	Items
T1	A C D
T2	B C E
T3	A B C E
T4	B E

Minsup=50%

↓

C_1

	TID	Support
P(A)=1	A	2
P(B)=2	B	3
P(C)=3	C	3
P(D)=4	D	1
P(E)=5	E	3

L_1

Itemset	Support
A	2
B	3
C	3
D	3

T1{AC},{AD},{CD}
T2{BC},{BE},{CE}
T3{AB},{AC},{AE},{BC},{BE},{CE}
T4{BE}
H(x,y)=(P(x)-1)*(N-1)+P(y)-P(x)*(P(x)-1)/2-1

AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
1	2	1	1	2	0	3	1	2	0

↓

L_2

Itemset	Support
AC	2
BC	2
BE	3
CE	2

Figure 1: Mining L

3.2. Direct Candidate Generation Method

For any itemset X in R, if X's number of occurrences in R is not equal to K * (K-1)/2, then X will not frequent. R represent k-itemset which is generated by joining any two itemsets in L_{k-1}. The pseudo code of direct candidate generation is depicted as follow.

Procedure: apriori_gen_change(L_{k-1})
Input: L_{k-1}
Output: C_k
RESULT = Φ;

```
for each itemset I1 ϵLk-1                    // I1
ϵLk-1
{
    for each itemset I2 {        // I2 ϵLk-1
        c=I1 join I2;
        if (length(c)>k) then
            continue;
        Else {
            If(exists(RESULT,c)) then
                c.count++;
            Else {
                Add c to RESULT;
                c.count=1;
                }
             }
         }
}
For each itemset c ϵ RESULT {
        If c.count==k*(k-1)/2 then
         Add c to Ck;
        }
return Ck;
```

4.PERFORMANCE EVOLUTION

The authors generated synthetic data sets involving transactions to evaluate the performance of algorithms. The transactions mimic the transactions in the "real" world. Transaction may contain more than one large itemsets. Transaction sizes are typically clustered around a mean and a few transactions have many items. To model the phenomenon that large itemsets often have common items, some fraction of items in subsequent itemsets are chosen from the previous itemset generated. Each itemset in a transaction has a weight associated with it, which corresponds to the probability that this itemset will be picked. It should be noted that the paper was written at a time when data mining was coming of age and not many datasets available for association rule mining were publically available. Although the authors do justify their use of synthetic datasets for validation, it should be noted that some later studies revealed [3] that the performance of association rule mining algorithms on even meticulously created synthetic datasets may not be the same as their performance on real world datasets. Such a risk is associated with using synthetic datasets in experimental validation. The Apriori Algorithm however does represent a big leap in terms of performance over previous algorithms like SETM or AIS. To be fair, the authors do point out to another paper where real world datasets were used, however these were limited in scope.

5.CONCLUSION

We have presented a new algorithm to mine all frequent itemsets, named APRIORI-EFFICENT algorithm. This algorithm reads transaction database by scanning only one time and does not generate candidate sets. We have observed that compared to Apriori, FP-Growth shows much higher performance gain on sparse as well as dense datasets.

6.REFERENCES

1. R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In VLDB '94: Proceedings of the 20th International Conference on Very Large Data Bases, pages 487--499, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc.
2. Jiawei Han, Jian Pei, Yiwen Yin: Mining Frequent Patterns without Candidate Generation. SIGMOD Conference 2000: 1-12
3. Zijian Zheng, Ron Kohavi, and Llew Mason, Real World Performance of Association Rule Algorithms, KDD 2001.
4. Agrawal.R, Imielinski.T, Swami.A Mining Association rules between Sets of Items in large Databases[C]. In Proceedings of the ACM-SIGMOD

Conference on Management of Data,1993:207-216
5. J. Dong, M. Han, BitTableFI: an efficient mining frequent itemsets algorithm, Knowledge-Based Systems 20(2007) 329-335
6. S. Zhang, J. Zhang, C. Zhang, EDUA: an efficient algorithm for dynamic database mining, Information Sciences 177(2007) 2756- 2767.
7. T. Hu, S.Y. Sung, H. Xiong, Q. Fu, Discovery of maximum length frequent itemsets, Information Sciences 178(2008) 68-87.
8 Introduction to Data Mining,Pang-Ning Tan,Vipin Kumar,Michael Steinbach

Generalization: A Data Mining Solution to Privacy Protection and Data Quality

1. R.Chetan Babu 2.N.Venkatram 3.M.Srinivasarao

1.M.Tech student,Dept of CSE,Nova's Institute of technology Eluru

2.Assoc Professor,HOD of CSE Dept,Nova's Institute of Technology,Eluru

3.Assistant Professor,Dept of CSE,Nova's Institute of Technology,Eluru

ABSTRACT: *The Well-known privacy-preserved data mining modifies existing data mining techniques to randomized data. In this paper, we investigate data mining as a technique for masking data, therefore, termed data mining based privacy protection. This approach incorporates partially the requirement of a targeted data mining task into the process of masking data so that essential structure is preserved in the masked data. The idea is simple but novel: we explore the data generalization concept from data mining as a way to hide detailed information, rather than discover trends and patterns. Our work demonstrated another positive use of data mining technology: not only can it discover useful patterns, but also mask private information. We consider the following privacy problem: a data holder wants to release a version of data for building classification models, but wants to protect against linking the released data to an external source for information. We adapt an iterative bottom-up generalization from data mining to generalize the data. The generalization space is specified by a hierarchical structure of generalization. A key is identifying the best generalization to climb up the hierarchy at each iteration.*

Keywords: *Annealing, Classification, Cross mining, Data mining, Generalization, Hierarchical Structure, Privacy Protection, Pruning.*

1. INTRODUCTION

The increasing ability to accumulate, store, retrieve, cross-reference, mine and link vast number of electronic records brings substantial benefits to millions of people. For example, cross-mining personal records on chemical exposure and death records could help identify cancer-causing substances. These advances also raise responsibility and privacy concerns because of the potential of creating new information. An example given in [1] is that a sensitive medical record was uniquely linked to

a named voter record in a publicly available voter list through the shared attributes of Zip, Birth date, Sex, Indeed, since "the whole is greater than the sum of the parts", protection of individual sources does not guarantee protection when sources are cross-examined. A relevant research topic is finding ways to safeguard against inferring private information through record linkage while continuing to allow benefits of information sharing and data mining.

Information becomes sensitive when they are specific to a small number of individuals. Data mining, on the other hand, typically makes use of information shared by some minimum number of individuals to ensure a required statistical significance of patterns. As such, sensitive information are to be discarded for reliable data mining. This observation motivates us to apply the requirement of an intended data mining task to identify useful information to be released, therefore, sensitive information to be masked. This approach, called *data mining based privacy protection* turns data mining from a threat into a solution to privacy protection.

2. THE PROBLEM

Consider that the data holder wants to release a person-specific data R(D1,…,Dn,C) to the public. A record has the form <v1,…,vn,cls>, where vi is a domain value from the attribute Di and cls is a class in C. Suppose that R shares some attributes with an external source E, denoted ∩E. If a value on R∩E is so specific that the probability of having this value by chance is negligible, each linking from a record in R to some information in E through this value has a good chance of identifying a real life fact. The data holder protects against such linkages by requiring a minimum number of records linkable through each value on R∩E.

Definition 1 (Anonymity) The virtual identifier, denoted VID, is the set of attributes shared by R and E. a(vid) denotes the number of records in R with the value vid on VID. The anonymity of VID, denoted A(VID), is the minimum a(vid) for any value vid on VID. If a(vid)=A(VID), vid is called an anonymity vid. R satisfies the anonymity requirement<VID,K> if A(VID)≥K, where K is specified by the data holder. We transform R to satisfy the anonymity requirement by generalizing specific values on VID into less specific but semantically consistent values. The generalization increases the probability of having a given value on VID by chance therefore, decrease the probability that a linking through this value represents a real life fact. The generalization space is specified through a taxonomical hierarchy per attribute in VID, provided by either the data holder or the data recipient. A hierarchy is a tree with leaf nodes representing domain values and parent nodes representing less specific values. R is generalized by a sequence of generalizations, where each generalization replaces all child values with their parent value p in a hierarchy. Before a value c is generalized, all values below c should be generalized to c first.

Definition 2 (Generalization) A generalization, written {c}→p, replaces all child values {c} with the parent value p. A

Relationship	Race	Work Class	A(vid)	C
C1	b2	a3	4	0Y4N
C1	b2	c3	4	0Y4N
C1	b2	d3	3	0Y3N
C1	c2	a3	3	2Y1N
C1	c2	b3	4	2Y2N
d1	c2	b3	4	4Y0N
d1	c2	e3	2	2Y0N
d1	d2	b3	3	2Y1N
d1	d2	e3	2	2Y0N

Relationship Race Workclass

generalization is valid if all values below c have been generalized to c. A vid is generalized by {c}→p if the vid contains some value in {c}.

Example:

Consider

VID=(Relationship,Race,Workclass)

And the hierarchies and vids in Figure 1. We have compressed all records having the same value on VID into a single row with the distribution of the Y/N class

label and the count a(vid). Initially, the generalizations at e1,f1,e2,f2,f3 are valid, A(V ID)=2, and d1c2e3 and d1d2e3 are anonymity vids. The requirement of K=3 can be satisfied by applying {c2,d2}→f2, which generalizes the vids d1c2e3 and d1d2e3 into a single vid d1f2e3 with a(d1f2e3)=4.

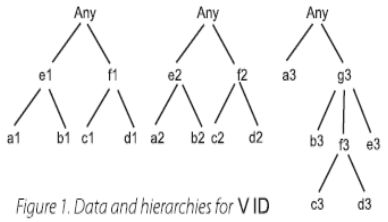


Figure 1. Data and hierarchies for V ID

Definition 3 (Anonymity for Classification) Given a relation R, an anonymity requirement <VID,K>, and a hierarchy for each attribute in VID, generalize R, by a sequence of generalizations, to satisfy the requirement and contains as much information as possible for classification. The anonymity requirement can be satisfied in more than one way of generalizing R, and some lose more information than others with regard to classification. One question is how to select a sequence of generalizations so that information loss is minimizes. Another question is how to find this sequence of generalizations efficiently for a large data set.

3. IINFORMATION-PRIVACY METRIC:

To minimize the information loss for achieving a given K-anonymity, our criterion is to favor the generalization having the minimum information loss for each unit of anonymity gain:

Minimize: IP(G)=I(G)/P(G).

IP(G) is ∞ if P(G)=0. If P(G)=0 for all (valid) generalizations G, we compare them based on I(G). This metric also maximizes the anonymity gain for each unit of information loss. We use I(G)/P(G) instead of I(G)-P(G) because differentiating semantically different quantities makes little sense. Unlike the "penalty" metric in [7] that focuses on information distortion alone, IP(G) takes into account both information and anonymity. The anonymity consideration helps focus the search on the privacy goal, therefore, has a look-ahead effect. However, this presents a new challenge to scalability because the effect on anonymity is only available after applying a generalization.

4. BOTTOM-UP GENERALIZATION:

Algorithm1 describes our bottom-up generalization process. In the ith iteration, we generalize R by the "best" generalization G_{best} according to IP metric. This algorithm makes no claim on efficiency because Line 2 and 3 requires computing IP(G) for all candidate generalizations G. let us look at this computation in more details.

Consider a candidate generalization G:{c}→p in an iteration. $|R_c|$ and freq(R_c , cls) can be maintained after each iteration. $|R_p|$ and freq(R_p,cls) can be obtained by aggregating $|R_c|$ and freq(R_p,cls). Therefore, I(G) can be easily computed, i.e., without accessing vids. In fact, any metric on a single attribute (plus the class label) can be computed this way. A(VID) is available as a result of applying the previous generalization. Computing A_G(VID), however, depends on the "effect" of G, which is only available after applying G, and requires accessing vids. This is a new challenge to scalability. Our insight is that most generalizations G do not affect A(VID), therefore, A_G(VID)=A(VID). In fact, if a generalization G fails to generalize all anonymity vids, G will not affect A(VID). For such G, P(G)=0 and IP(G)=∞, and our metric does not need A_G(VID) therefore, we can focus on "critical generalizations" as de-fined below.

Algorithm1: The bottom-up generalization

1. While R does not satisfy the anonymity requirement do
2. for all generalization G do
3. compute IP(G);
4. end for;
5. find the best generalization G_{best};
6. generalize R by G_{best};
7. end while;
8. output R;

Definition 4 (Critical generalization) : G is critical if $A_G(VID)>A(VID)$.

A critical generalization G has a non-zero P(G) and a finite IP(G), where as a non-critical generalization G has a zero P(G) and infinite IP(G). Therefore, so long as one generalization is critical, all non-critical generalizations will be ignored by the IP metric. If all generalizations are non-critical, the IP metric will select the one with minimum I(G). In both cases, $A_G(VID)$ is not needed for a non-critical generalization G. Based on this observation, we optimize Algorithm 1 by replacing Line 2 and 3 with

2: for all critical generalization G do

3: compute $A_G(VID)$;

5. PRUNING STRATEGIES:

We store all distinct vids in a tree structure, called Taxonomy Encoded Anonymity(TEA) index. Each level of the tree represents the current generalization of a particular attribute, and each path represents a particular vid with a(vid) stored in the leaf node. In addition, the TEA index links up the vids according to the generalizations that generalize them. Each time a generalization is applied, the TEA index is updated by adjusting the vids linked to this generalization. The purpose of this index is to prune the number of candidate generalizations to no more than |VID| at each iteration, where |VID| is the number of attributes in VID.

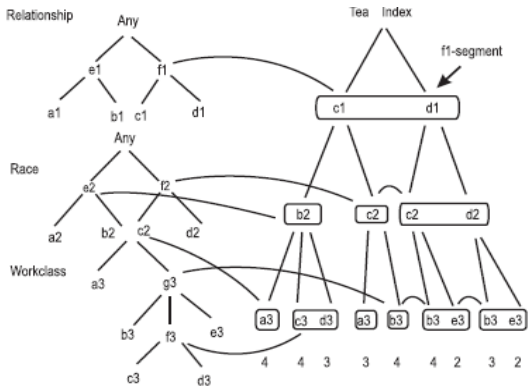


Figure 2. The TEA index for V ID

5.1 Step 1: Pruning generalizations: This step finds all generalizations satisfying the "only if" condition in Theorem 1, denoted Cand. We start at the leaf nodes for the anonymity vids in the TEA index, walk up their paths synchronously one level at a time. At each level, we check if every anonymity vid is generalized by some size-k segment of the same generalization G,k>1. If not, no critical generalization exist at the current level. If yes, we add G to Cand. We then move up to the next level in the TEA index.

5.2 Step 2: Finding the best generalization: This step finds the best generalization by computing IP(G) for every (valid) generalization G. A(VID) and I(G) are available or easily computed from the result of the previous iteration. For every G not in Cand,G is non-critical (Theorem 1), so IP(G)=I(G). So, we focus on computing $A_G(VID)$ for $G \in$ Cand. We present a method that examines only the vids actually generalized by G, not all vids. Let A_G be the minimum a(vid) for the new vids produced by applying G. Let A_G be the minimum a(vid) for all old vids not generalized by G. $A_G(VID)=\min\{A_G, A_G\}$. To compute A_G, we apply G to the TEA index.

To compute A_G, we keep track of the number of vids not generalized by G such that a(vid)=i, stored in O[i], for $1 \le i \le K$. K is typically a few hundreds, so this is a small cost. Before applying G,O[i] is available from the previous iteration. Each time a vid having a(vid)=i is generalized by G, O[i] stores the correct value. Now,

if O[i]>0 for 1≤i≤K, let A_G be the smallest such i. If O[i]=0 for 1≤i≤K, we consider two cases: if A_G≤K, then A_G(VID)=A_G; if A_G>K, then A_G(VID)>K, but such A_G(VID) is never used in our metric. The cost in this step is proportional to the number of vids generalized by G, not all vids.

5.3 Step 3: Applying the best generalization: This step applies the best generalization Gbest found. If Gbest is in Cand, we just make the effect of G permanent. If Gbest is not in Cand we apply Gbest to the TEA index. In this case Cand must be empty, otherwise Gbest must come from Cand.

6. CONCLUSION

We have investigated data mining as a technique for masking data, called data mining based privacy protection. The idea is to explore the data generalization concept from data mining as a way to hide detailed information, rather than discover trends and patterns. Once the data is masked, standard data mining techniques can be applied without modification. Our work demonstrated another positive use of the data mining technology: not only can it discover useful patterns, but also mask private information.

In particular, we presented a bottom-up generalization for transforming specific data to less specific but semantically consistent data for privacy protection. We believe that the framework of bottom-up generalization is amenable to several extensions that will make it more practical: incorporating different metrics, handling data suppression where a value is taken out entirely, and partial generalization where not necessarily all child values are generalizing numeric attributes without a pre-determined hierarchy. We plan to investigate these issues further.

7. REFERENCES

[1] R.Agrawal and R.Srikant. Privacy preserving data mining. In SIGMOD, 2000

[2] C.F.Clark. The introduction of statistical noise to utility company data on a micro data tape with that data matched to annual housing survey data. In Draft Project Report, Bureau of The Census, Washington D.C., 1978

[3] W.A.Fuller. Masking procedures for micro data disclosure limitation. Journal of Official Statistics, 9(2):383-406,1993.

[4] B.Greenberg. Disclosure avoidance research at the census bureau. In Proceedings of the Bureau of the Census, Washington, D.C., 1990.

IMPROVING PRIVACY PRESERVATION USING HIGHLY CORRELATED ATTRIBUTES

1.K.Raghavendra 2. N.Venkatram 3. R.Kamala

1.M.Tech student, Dept of CSE, Nova's Institute of Technology, Eluru

2.Assoc Professor, HOD of CSE Dept, Nova's Institute of Technology, Eluru

3.Asst Professor, Dept of CSE, Nova's Institute of Technology, Eluru

ABSTRACT

Different data engineering approaches are mine data and publish as a privacy data without leakage and data loss. All present internet applications require new technologies for privacy data maintenance. Different engineering approaches are tuning the databases with good privacy preservation [5][6][7][8]. Present approaches are uses the triggers here. Triggers related privacy control is not gives the effective solution. Its shows some risks in data publishing environment. Triggers maintaianace creation purpose we use the different approaches like pseudonyms, k-anonymity, l-diversity, random perturbation, condensation approach. These all above approaches are not gives desired privacy results in implementation part.

Now in this paper we propose the new proposal for increase the privacy levels in implementation part. In datasets we maintain the different attributes. In attribute selection we apply the partitioning technique. In different number of partitions identifies the highly correlated attributes in implementation. These high correlated attributes are possible to increase the privacy levels and utility values. Experimentally we show the performance compare to normal correlation or random correlation attributes to high correlation attribute. Highly correlated attributes are gives the efficient privacy guarantee and trust results [5][6][10].

KEYWORDS OR INDEX TERMS: Correlation, partitioning techniques. Decision tree classification, frequent association rule mining.

1. INTRODUCTION

Many numbers of researchers are interest control intrusions without misuse. Sensitive data maintenance is required in organizations. In data collection of all fields we expect some privacy related techniques. Different real time applications of databases whenever servers are access the data we implement different analysts. Analyst is called methodology [11][12][13].

Previous methodologies are non disclosures. These non disclosures are not detects the attacks data and duplicate data. It's may chance to data loss in publish of

records like medical records and health insurance databases. After some days some new privacy algorithms are gave the good solutions in publishing of data. Data publish for only truthful users.

Present in this paper we propose some new statistical disclosure measure. This statistical measure calculates using some partitioning techniques like decision tree classification. Every partition considers one class. In all attributes identifies highly correlated attributes. These new correlated attributes gives the efficient solution.

Experimental show the comparison with existing and new techniques. Proposed system techniques are works as a efficient technique in privacy probability calculation.

2. RELATED WORK OR LITERATURE REVIEW

Much number of organizations contains different web databases. Any user can access any kind of data directly we publish. In published data some other users data also present here. Some privacy problems are generating for sensitive data. Some people are think like negatively whenever we are not providing the privacy. In published data some intrusions or attackers are misuse the data. Now we prevent those misuse data using some privacy techniques [1].

Data publishing records show into number of tables. Tables of all records we are provide the privacy. Every record contains different attributes are present. Select one of the attribute that is called as a personal identifier. Replace the personal identifier with pseudonyms. It's not gives sufficient privacy in all number of dimensions. Some problems are generating like lack of privacy errors here [10].

Next privacy technique called k-anonymity. Published data show into number of tables. Every record display as a individual record. We provide the security for individual records. It's possible to provide the privacy for limited records. In extraction of records we extract some location data. It's possible to provide the privacy in limited number of locations. It's somewhat relevant in data privacy environment. K-anonymity also contains some limitations.

After some days next privacy technique was introduced that is related generalization. That privacy technique is called suppression. Suppression techniques minimize the number of attributes. Tuples values also changes with roundup values specification in implementation part. Those all attributes and tuples are displayed as a less informative content [11]. Suppression techniques work based on dependency principle in implementation part. It's not possible to control the sensitive data maintenance. Its suppress the less number of attributes and tuples data in publishing of data.

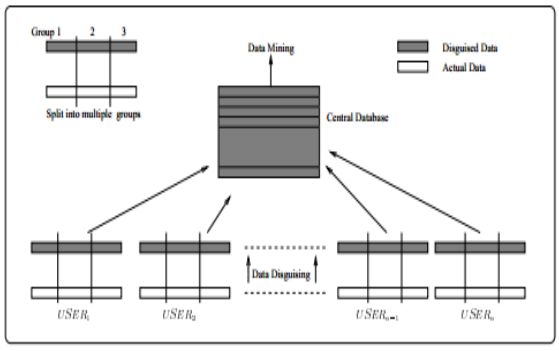


Fig1: Privacy Preserving Data Mining

L-diversity and K-anonymity provide the privacy related to different number of attributes. It's possible to control the privacy related quasi identifiers data. There is no specific privacy involvement in sensitive attributes. In sensitive attributes may chance to leakage of data. The above all approaches are provides less privacy in preserving to publish the data [3] [4].

Next another privacy technique, Randomization applies into number of tuples values in column specification. According to present tuples select one random value. Those random values add as a noisy. These intrusions it's not possible to misuse the data and save data. This is one of the perturbation techniques to control the data with privacy rules [7] [8].

Next another privacy approach, condensation approach we apply covariance into different columns at a time. Using covariance it's possible to provide the data without misuse by attackers [12]. In group of columns at a time provide the privacy and generate the covariance matrix. Covariance matrix gives the solution in multiple dimensions, but at last some dimensions limitations are present here.

Previous approaches generalization, bucketizations are not follows the probability approach. It's not possible to provide the privacy in sensitive attributes. Present paper covers the partitioning techniques. These techniques apply the probability approach and calculate the sensitive values with less correlation. It's also gives the less privacy and save the less sensitive data maintenance.

3. PROBLEM STATEMENT

Attribute Disclosure Protection and membership disclosure protection provide privacy for present tuples and attributes. It's not possible to provide privacy for update number of attributes and tuples in published data. That is called for future attributes and tuples. In updated records previous techniques are not gives the sufficient privacy. Some limitations are available related to cost and time processing.

We propose new privacy techniques. Those privacy techniques are correlation, summarization and frequent data mining approaches. These all approaches apply in different tables and show the good privacy accuracy results. We implement some new correlation coefficients and calculate the measures of privacy. Present calculation of correlation measures is not sufficient, then changes to attributes to maintain the correlation in between of attributes. All attributes measure correlation we verify select one best correlation attribute pair related to different domains.

4. SYSTEM MODEL

Decision tree we apply in partitioning of attributes recursively for updating of measure of correlation. Present Decision tree approach cover all n number of dimensions of table records. This is no chance to miss the sensitive data environment in implementation of approach [13]. Decision tree approach is the classification approach. Tables of data classified into different partitions. Different partitions are showing the different classes. Calculate the each and every class of correlation attributes privacy.

In total classes frequent which class is occur, that class is best class like we find out in implementation part.

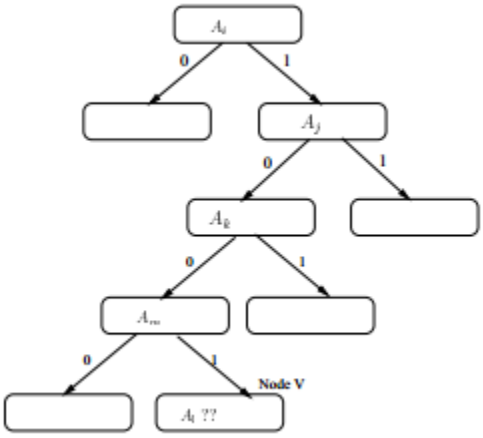


Fig2: Decision Tree related table Partitions

In normal decision tree add the improved id3 algorithm and show the good prediction results in sensitive data maintenance. Improved id3 algorithm follows the different steps. First step completely related training phase. Training phase start selection of attributes and verifies the each and every attribute accuracy in prediction stage. In total number of accuracy levels choose one of the best class attributes. Any new attributes are generating compare with highest accuracy attributes here. Every time process the different types of attributes. All attributes of information completely we arrange in decision tree. Decision tree approach covers all number of attributes. In total number of attribute filter one dominant attribute pair information like final output content.

In different pair attribute classes sometimes one attribute show the measure is high, another attribute measure is low. Different attribute pair classes of single attribute we select and create the new attribute pair using summarization technique. Summarization techniques are save the sensitive data compare to all previous approaches.

Different attribute pairs are provide the better privacy results. In total number of attribute pairs of measure of quality display as a output. Data owner apply threshold value in different number of attribute pairs in implementation part. Above threshold measure related all attributes are good privacy and sensitive attributes. Attribute sensitive value is available below threshold then detects as a low level privacy. This is basic prediction of good sensitive attribute approach.

4.1 IMPROVED ID3 ALGORITHM USING DECISION TREE

Pseudo code:

1. Create the different classes with different attributes.
2. Divide the attribute with partitioning technique
3. All samples are allocated in different attributes.
4. Verify the classes from parent to leaf node
5. Perform the training process select the best performance gain attributes pair
6. Using performance gain attributes start the testing process

7. In different attribute pair highest single attribute pair gain value
8. After verification return the result of best results with good prediction accuracy
9. These all types of results extraction using summarization approach.

5. EXPERIMENTAL EVOLUTION

Choose any kind of dataset with different attributes with tuples. We try to train the datasets with different privacy techniques. Calculate privacy measure for each and every attribute specifically in implementation. In present datasets apply membership disclosure protection, identity disclosure protection, and attribute disclosure protection. First two protection functions are not show the privacy result with probability. Attribute disclosure protection show the probability with privacy values specification. Compare to previous all privacy protection function; in this paper we propose some new correlation coefficient, decision tree classification. These all techniques filter efficient attributes in privacy protection.

Those dominate attributes privacy implement in any real time applications. Those real time applications also show the result as a better privacy guaranteed application here.

6. CONCLUSION AND FUTURE WORK

Previous all random grouping techniques are not gives the accurate solutions in privacy probability. Its provide the privacy in less number of dimensions itself. In some dimensions we got the problems in implementation part. Those all the problems are solved in present paper. Its cover all number of dimensions in calculation of privacy results. Calculation of privacy starts based on partitioning algorithms. These partitioning operations are performing using decision tree classification and free pattern mining operation in implementation part. In all number of partitions classes choose the best partition classes. These best partition classes gives the best accurate in privacy calculation in implementation.

7. REFERENCES

1. Privacy Preserving for High-dimensional Data using Anonymization Technique, Neha V. Mogre, Prof. Girish Agarwal, Prof. Pragati Patil,2013

2. Privacy-Preserving Data Publishing, Bee-Chung Chen, Daniel Kifer, Kristen LeFevre and Ashwin Machanavajjhala,2009.

3. Privacy-Preserving Data publishing: A Survey of Recent Developments, 2010, BENJAMIN C. M. FUNG

4. Towards Privacy Preserving Data Publishing¤,Xiaoxun Suny,*2011*

5. Privacy Preserving in Knowledge Discovery and Data Publishing, *B.Lakshmana Rao1, G.V Konda Reddy2, G.Yedukondalu3*

6. The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing, 2008

7. Privacy Preservation Using Randomized Attribute Selection Based On Knowledge Hiding, P.Vijayakumar, 2,Dr. R.Manicka chezian, 2013.

8.On the Anonymization of Sparse Dimensional data, Gabriel Ghinita #1, Yufei Tao *27 Panos Kalnis #',2008.

9. An Efficient Approach for Data Privacy in Distributed Environment Using Nearest Neighbor Search Anonymization, L. Madhuridevi, J. JesuVedhaNayahi, V.Kavitha,2012.

10. Movement Data Anonymity through Generalization, Anna Monreale2,3, Gennady Andrienko1, Natalia Andrienko1, Fosca Giannotti2,4, Dino Pedreschi3,4, Salvatore Rinzivillo2, StefanWrobel1, 2009.

11. Privacy-Preserving Data Mining, 2009

12. Privacy Preserving Data Mining, 2013

13. Privacy-Preserving Classi_er Learning, Justin Brickell and Vitaly Shmatikov,2013

14. Privacy-Preserving Data Mining Using Multi-Group Randomized Response Techniques, Zhijun Zhan and Wenliang Du,2011

15. Slicing: A New Approach for Privacy Preserving Data Publishing, 2012.

KNOWLEDGE STABILITY IN CLOUDS

1.T.Nagababu 2.N.Venkataram 3 M.Venugopal

1. M.Tech student,Dept of Computer Science,Nova's Institute of Technology,Eluru

2.Associate professor,HOD of Computer science,Nova's Institute of Technology,Eluru

3.Assistant professor,Dept of S&H, Lingayas Institute of technology,Vijayawada

Abstract:-Cloud computing is a model for enabling 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 (from NIST). Offering strong data protection to cloud users while enabling rich applications is a challenging task. We explore a new cloud platform architecture called Data Protection as a Service, which dramatically reduces the per-application development effort required to offer data protection, while still allowing rapid development and maintenance.

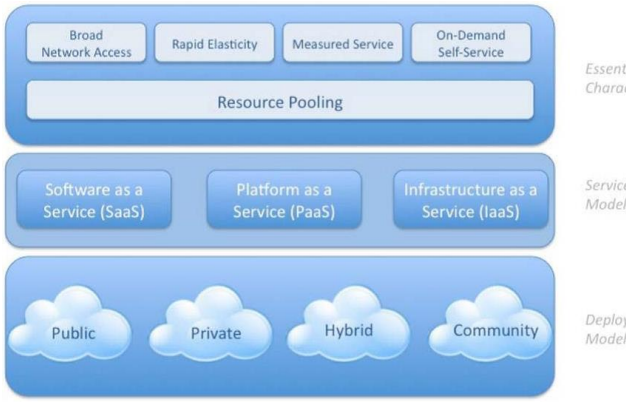
Introduction:

Although cloud computing promises lower costs, rapid scaling, easier maintenance, and service availability anywhere, anytime, a key challenge is how to ensure and build confidence that the cloud can handle user data securely. A recent Microsoft survey found that "58 percent of the public and 86 per-cent of business leaders are excited about the possibilities of cloud computing. But more than 90 percent of them are worried about security, availability, and privacy of their data as it rests in the cloud." This tension makes sense: users want to maintain control of their data, but they also want to benefit from the rich services that application developers can provide using that data. So far, the cloud offers little platform-level support or standardization for user data protection beyond data encryption at rest, most likely because doing so is nontrivial. Protecting user data while enabling rich computation re-quires both specialized expertise and resources that might not be readily available to most application developers. Building in data-protection solutions at the platform layer is an attractive option: the platform can achieve economies of scale by amortizing expertise costs and distributing sophisticated security solutions across different applications and their developers.

SECURITY AND PRIVACY CHALLENGES

Five Characteristics

- On-demand self-service
- Ubiquitous network access
- Location independent resource pooling
- Rapid elasticity
- Measured service



Four Cloud Deployment Models

• Private cloud

A private cloud implementation aims to avoid many of the objections regarding cloud computing security. Because a private cloud setup is implemented safely within the corporate firewall, a private cloud provides more control over the company's data, and it ensures security, albeit with greater potential risk for data loss due to natural disaster.

•Ex: Enterprise owned or leased

• Community cloud

A **community cloud** in computing is a collaborative effort in which infrastructure is shared between several organizations from a specific community with common concerns (security, compliance, jurisdiction, etc.), whether managed internally or by a third-party and hosted internally or externally.

• Shared infrastructure for specific community

• Public cloud

A **public cloud** is a set of computers and computer network resources based on the standard cloud computing model, in which a service provider makes resources, such as applications and storage, available to the general public over the Internet. Public cloud services may be free or offered on a pay-per-usage model

• Sold to the public, mega-scale infrastructure

• Hybrid cloud

A hybrid cloud is an integrated cloud service utilizing both private and public clouds to perform distinct functions within the same organization. All cloud computing services should offer certain efficiencies to differing degrees but public cloud services are likely to be more cost efficient and scalable than private clouds. Therefore, an organization can maximize their efficiencies by employing public cloud services for all non-sensitive operations, only relying on a private cloud where they require it and ensuring that all of their platforms are seamlessly integrated.

• Composition of two or more clouds

Threats to Cloud Computing

1. Abuse and Nefarious Use of Cloud Computing
2. Insecure Application Programming Interfaces
3. Malicious Insiders
4. Shared Technology Vulnerabilities
5. Data Loss/Leakage
6. Account, Service, and Traffic Hijacking
7. Unknown Risk Profile

Abuse and Nefarious Use

- Password and key cracking
- DDOS
- Launching dynamic attack points
- Hosting malicious data

- Botnet command and control
- Building rainbow tables
- CAPTCHA solving
- Exploits exist already

Insecure Interfaces and APIs

• Could expose more functionality than intended

• Policy could be circumvented

• Credentials may need to be passed –is the interface secure?

Malicious Insiders

• Particularly poignant for cloud computing

• Little risk of detection

• System administrator qualifications and vetting

Process for cloud services provider may be different that of the data owner

Shared Technology Issues

• Underlying architecture (CPU cache, GPU, etc.) not intended to offer strong isolation properties Virtualization hyper visor used to mediate access between guest OS and physical resources Exploits exist (Blue Pill, Red Pill It's impossible to develop a single data-protection solution for the cloud because the term means too many different things. Any progress must first occur in a particular domain—accordingly, our work focuses on an important class of widely used applications that includes e-mail, personal financial management, social net-works, and business tools such as word processors and spreadsheets. The following criteria define this class of

Applications:

• provide services to a large number of distinct end users, as opposed to bulk data processing or workflow management for a single entity;

• use a data model consisting mostly of sharable units, where all data objects have access control lists (ACLs) with one or more users; and

• Developers could run the applications on a separate computing platform that encompasses the physical infrastructure, job scheduling, user authentication, and the base software environment, rather than implementing the platform themselves. Overly rigid security is as detrimental to cloud ser-vice value as inadequate security. A primary challenge in designing a platform-layer solution useful to many applications is ensuring that it enables rapid development and maintenance. To ensure a practical solution, we considered the following goals relating to data protection as well as ease of development and maintenance:

• Integrity: The user's stored data won't be corrupted.

• Privacy: Private data won't be leaked to any unauthorized entity.

• Access transparency: Logs will clearly indicate who or what accessed any data.

• Ease of verification: Users will be able to easily verify what platform or application code is running, as well as whether the cloud has strictly enforced their data's privacy policies.

• Rich computation: The platform will allow efficient, rich computations on sensitive user data.

• Development and maintenance support: Because they face a long list of challenges—bugs to find and fix, frequent software upgrades, continuous usage pattern changes, and user demand for high performance—developers will

receive both development and maintenance support. Any credible data protection approach must grapple with these issues, several of which are often overlooked in the literature.

DATA PROTECTION AS A SERVICE

Currently, users must rely primarily on legal agreements and implied economic and reputational harm as a proxy for application trustworthiness. As an alternative, a cloud platform could help achieve a robust technical solution by

• making it easy for developers to write maintain-able applications that protect user data in the cloud, thereby providing the same economies of scale for security and privacy as for computation and storage; and enabling independent verification both of the plat-form's operation and the runtime state of applications on it, so users can gain confidence that their data is being handled properly. Much as an operating system provides isolation between processes but allows substantial freedom inside a process, cloud platforms could offer transparently verifiable partitions for applications that compute on data units, while still allowing broad computational latitude within those partitions. DPaaS enforces fine-grained access control policies on data units through application confinement and information flow checking. It employs cryptographic protections at rest and offers robust logging and auditing to provide accountability. Crucially, DPaaS also directly addresses the issues of rapid development and maintenance. To truly support this vision, cloud platform providers would have to offer DPaaS in addition to their existing hosting environment, which could be especially beneficial for small companies or developers who don't have much in-house security expertise, helping them build user confidence much more quickly than they otherwise might.

WHAT ABOUT ENCRYPTION?

In the realm of data protection, developers often view encryption as a kind of a silver bullet, but in reality, it's just a tool—albeit a powerful one—to help achieve data protection properties. Although full-disk encryption(FDE) and computing on encrypted data have recently gained attention, these techniques have fallen short of answering all of the security and maintenance challenges mentioned earlier. FDE encrypts entire physical disks with a symmetric key, often in disk firmware, for simplicity and speed. Al-though FDE is effective in protecting private data in certain scenarios such as stolen laptops and backup tapes, the concern is that it can't fulfill data protection goals in the cloud, where physical theft isn't the main threat. At the other end of the spectrum, Craig Gentry recently proposed the first realization of fully homomorphic encryption (FHE), which offers the promise of general computation on cipher texts. Basically, any function in plaintext can be transformed into an equivalent function in cipher text: the server does the real work, but it doesn't know the data it's computing. Naturally, this property gives strong privacy guarantees when computing on private data, but the question of its practicality for general cloud applications still remains. FDE versus FHE A comparison of FDE and FHE in the cloud computing setting reveals how these encryption techniques fall short of addressing the aforementioned security and maintenance challenges simultaneously Key management and trust. With FDE, the keys reside with the cloud platform, generally on or close to the physical drive: the cloud application user isn't involved in key management. While user data is encrypted on the physical disk, it is always accessible in the clear to any layer above it. Consequently, FDE doesn't prevent online attacks from leaking the data to an unauthorized party, which is far more common in the cloud setting than physical attacks. With FHE, untrusted applications can't easily learn or leak data. Users typically own and manage FHE encryption keys, while applications

compute on encrypted forms of user data without actually "seeing" the data. This raises questions about how users can store their keys securely and reliably, especially in the presence of sharing. After all, the point of the cloud is to avoid maintaining local state. Sharing: Collaboration is often cited as a "killer feature" for cloud applications. Fine-grained access control is necessary to let a data owner selectively share one or more data objects with other users. With FDE, users must fully trust the cloud provider to enforce correct access control because the key granularity (the whole disk) doesn't line up with access control granularity (a single data unit). With FHE, because the user—or a third-party cloud provider employed by the user—manages the encryption keys, the best way of providing access control isn't clear yet. To offer fine-grained encryption-based access control, we might need to define key management on a per data object granularity basis or over collections of data objects. How-ever, to support homomorphic operations across multiple encrypted objects, those objects must still be encrypted under the same public key Aggregation. Many cloud applications require per-forming data mining over multiple users' data for tasks such as spam filtering or computing aggregate statistics. Because users fully trust the cloud provider, performing such data aggregation is relatively easy with FDE. Current FHE techniques don't readily allow computing on multiple users' data encrypted under different keys. Therefore, it isn't clear yet how to support such data aggregation applications with FHE; similarly, offline aggregation across users' data isn't possible. One solution might be to escrow keys to the cloud provider, but that would eliminate many of FHE's benefits, making its cost harder to justify Performance. According to a recent survey, 49 percent of users abandon a site or switch to a competitor after experiencing performance issues. And the need for speed is only increasing: in 2000, a typical user was willing to wait 8 seconds for a webpage to load before navigating away; by 2009, that number dropped to 3 seconds. When FDE is implemented in disk firmware, its sym-metric encryption can run at the disk's full bandwidth, effectively avoiding a slowdown. Although researchers have made significant advances in improving FHE's performance since Gentry's original proposal, it has a long way to go before becoming efficient enough to deploy at scale. In Gentry's estimation, implementing something like a Google search with FHE would require roughly 1 trillion times more computation than the one without FHE Ease of development. Because FDE is hidden behind an abstraction of the physical disk, it typically has no impact on application development. In theory, FHE could also be relatively automatic: it works on an abstraction of the program as a circuit and transforms that circuit. In practice, however, performing this translation for arbitrary programs—especially when marshaling data—could be quite complex. At a minimum, programming tools would need to evolve dramatically. FHE doesn't allow developers to input data-driven judgments into the development cycle. Specifically, application developers can't look at the data, making debugging, A/B testing, and application improvements more difficult Maintenance. Bugs are inevitable. However, availability is a primary cloud goal, so the need to debug quickly is a top priority. Systems often fail for some unforeseen reason, requiring someone to step in and manually take action. Determining the nature of the problem might require detecting unusual activity or understanding exactly what went wrong, which isn't easy with FHE. If the application writer can't inspect application state meaning-fully, debugging could be a real challenge. The DPaaS approach moves key management and access control to a middle tier—the computing platform— to balance rapid development and easy maintenance with user-side verifiability. Splitting the difference

Although FDE offers excellent performance and ease of development, it does little to protect privacy at the required granularity. FHE, on the other hand, pushes the privacy envelope in the other direction by removing data visibility entirely from both the server and application developer. However, having a remote machine see and compute on sensitive data isn't automatically a privacy violation. FHE's guarantees go beyond what's necessary to protect data, and in so doing, it incurs significant performance and development costs. We believe the DPaaS approach is better suited for the target applications because it falls between the two. It keeps the "natural" granularity of FHE by keying on units of sharable data and maintains the performance of FDE by using symmetric encryption. It moves key management and access control to a middle tier—the computing platform—to balance rapid development and easy maintenance with user-side verifiability.

A WAY FORWARD

In an OS, processes and files are the primary units of access control, and the OS provides suitable isolation for them. Applications can do what they like within these boundaries. In a cloud setting, the unit of access control is typically a sharable piece of user data—for example, a document in a collaborative editor. Ideally, the system offers some analogous confinement of that data, restricting its visibility only to authorized users and applications while allowing broad latitude for what operations are done on it. This can make writing secure systems easier for programmers because confinement makes it more difficult for buggy code to leak data or for compromised code to grant unauthorized access to data. A malicious program might find different ways to exfilt rate data, such as employing a side channel or covert channel, but the priority here is to sup-port benign developers, while making all applications and their actions on users' sensitive data more easily auditable to catch improper usage. One of the main concerns people and organizations have about putting data in the cloud is that they don't know what happens to it. Having a clear audit trail of when data is accessed—and by whom or what—bolsters confidence that data is being handled appropriately. Confinement can be effective for most normal user accesses, but administrative access that's outside the normal flow of user access and involves human administrators (for example, for debugging and analysis) can especially benefit from auditing. Verifiable platform support Bugs need to be fixed. Data needs to be updated and migrated as schemas change. Offline computation is valuable for data aggregation across users or for pre-computation of expensive functions. To reduce the risk of unaudited backdoor access, all these functions should be subject to the same authorization flows and platform-level checks as normal requests, albeit with a separate, appropriate policy. Platform providers should build support for confinement and auditing into the platform in a verifiable way. This approval has many advantages:

• Application developers don't have to reinvent the wheel;

• Application code is independent of ACL enforcement;

• Third-party auditing and standards compliance are easier; and

• The verifiable platform extends to virtualized environments built atop it.

Finally, the cost of examining the platform is amortized across all its users, which means significant economies of scale for a large-scale platform provider.

Design space and a sample architecture

Figure 1 illustrates an example architecture for exploring the DPaaS design space.

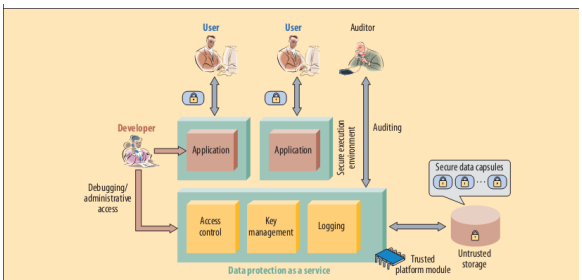


Figure 1. Sample architecture for data protection as a service illustrates how it's possible to integrate various technologies, such as application confinement, encryption, logging, code attestation, and information flow checking to realize DPaaS.

Here, each server contains

a trusted platform module(TPM) to provide secure and

verifiable boot and dynamic root of trust. This example architecture demonstrates at a high level how it's potentially

possible to combine various technologies such as application confinement, encryption, logging, code attestation,

and information flow checking to realize DPaaS. **Trusted Platform Module** (TPM) is both the name of a published specification detailing a secure crypto processor that can store cryptographic keys that protect information, as well as the general name of implementations of that specification, often called the "TPM chip" or "TPM Security Device". The TPM specification is the work of the Trusted Computing Group.

Disk encryption is a technology which protects information by converting it into unreadable code that cannot be deciphered easily by unauthorized people. **Disk encryption** uses disk encryption software or hardware to encrypt every bit of data that goes on a disk or disk volume. Disk encryption prevents unauthorized access to data storage. The term "full disk encryption" (or **whole disk encryption**) is often used to signify that everything on a disk is encrypted, including the programs that can encrypt bootable operating system partitions. But they must still leave the master boot record (MBR), and thus part of the disk, unencrypted. There are, however, hardware-based full disk encryption systems that can truly encrypt the entire boot disk, including the MBR**.**

3. Third Party Auditor

In this module, Auditor views the all user data and verifying data and also changed data. Auditor directly views all user data without key. Admin provided the permission to Auditor. After auditing data, store to the cloud.

4. User Module

User store large amount of data to clouds and access data using secure key. Secure key provided admin after encrypting data. Encrypt the data using TPM. User store data after auditor, view and verifying data and also changed data. User again views data at that time admin provided the message to user only changes data.

References

1. C. Dwork, "The Differential Privacy Frontier Extended

Abstract," Proc. 6th Theory of Cryptography Conf.(TCC 09),

LNCS 5444, Springer, 2009, pp. 496-502.

2. C. Gentry, "Fully Homomorphic Encryption Using Ideal

Lattices," Proc. 41st Ann. ACM Symp. Theory Computing

(STOC 09), ACM, 2009, pp. 169-178.

3. E. Naone, "The Slow-Motion Internet," Technology Rev.,

Mar./Apr. 2011; www.technologyreview.com/files/54902/

GoogleSpeed_charts.pdf.

4. A. Greenberg, "IBM's Blindfolded Calculator," Forbes,

13 July 2009; www.forbes.com/forbes/2009/0713/

breakthroughs-privacy-super-secret-encryption.html.

5. P. Maniatis et al., "Do You Know Where Your Data Are?

Secure Data Capsules for Deployable Data Protection,"

Proc. 13th Usenix Conf. Hot Topics in Operating Systems

(HotOS 11), Usenix, 2011; www.usenix.org/events/hotos11/

tech/final_files/ManiatisAkhawe.pdf.

6. S. McCamant and M.D. Ernst, "Quantitative Information

Flow as Network Flow Capacity," Proc. 2008 ACM SIGPLAN

Conf. Programming Language Design and Implementation

(PLDI 08), ACM, 2008, pp. 193-205.

7. M.S. Miller, "Robust Composition: Towards a Unified Ap-proach to Access Control and Concurrency Control," PhD

dissertation, Dept. of Philosophy, Johns Hopkins Univ.,

2006.

8. A. Sabelfeld and A.C. Myers, "Language-Based Information-Flow Security,"IEEE J. Selected Areas Comm., Jan. 2003, pp.

5-1 9.

9. L. Whitney, "Microsoft Urges Laws to Boost Trust in

the Cloud," CNET News, 20 Jan. 2010; http://news.cnet.

com/8301-1009_3-10437844-83.html.

Packet Classification Methods to Counter Jamming Attacks in Adhoc Networks

P Samanth N.Venkatram M.Nirupa

M.Tech student, Dept of CSE, Nova's Institute of Technology, Eluru

Assoc Professor, HOD of CSE Dept, Nova's Institute of Technology, Eluru

Asst Professor, Dept of CSE, Nova's Institute of Technology, Eluru

Abstract:

Jamming attacks though is not a new phenomenon leads to disruptions in the communications channel which is a serious concern in Reactive protocols driven Adhoc networks. The jamming models are categorised as both external and internal with the later being more serious nature because the "always-on" strategy employed in external model has several risk factors to the jammer's identity. External model involves the jammer spending a significant amount of energy to jam frequency bands of interest. The continuous presence of these unusually high interference levels makes this type of attacks easy to detect. In an internal threat model a jammer is assumed to be aware of network details and the implementation details of network protocols at any layer in the network stack. The jammer exploits his internal knowledge for launching selective jamming attacks in which specific packets of "high priority" are targeted. Although RREQ,RREP,RERR, RREP-ACK are primary Message Formats in reactive protocols, the adversary selectively targets RREQ and RREP packets in the network to launch jamming attacks. Existing approaches concentrated on using commitment schemes that are cryptographic primitives to hide the RREQ and RREP packets from the purview of the adversary. These approaches being successful, we propose to use them along with intrusion detection techniques for identifying compromised access points to increase overall network security significantly by marginalizing the working boundaries of an adversary, thus risking exposure. A resultant network prototype validates our claim.

Keywords—Selective Jamming, Denial-of-Service, Wireless Networks, Packet Classification.

I INTRODUCTION

Ad hoc networks are an integral part in mission critical communication for the military, utilities, and industry. An adversary may attempt to attack a victim ad hoc network to prevent or hijack some or all of the victim's communication. Such attacks have been considered as potential threats in ad hoc wireless networks at several levels. A number of researchers have considered DoS where the attackers are internal participants in the victim ad hoc network (see e.g. [1]). Internal threat model of Ad hoc networks requires the cooperation of participant nodes for their operation and are especially susceptible to such peer based attacks.

In this paper, we address the problem of jamming under an internal threat model. We consider a sophisticated jammer who is aware of network configurations and the implementation details of network protocols at any layer in the network stack. The jammer exploits his internal knowledge for launching selective jamming attacks in which specific

packets of "high importance" such as RREQ and RREP are targeted[9]. For example, a jammer can target route-request/route-reply messages at the routing layer to prevent route discovery, or target and convenience, WLAN is regularly used in daily life. An introduction of WLANs was done by Gast (2005) and Mark (2005). They presented basic wireless LAN technology, why the technology had emerged, how it works, the architecture of WLANs, and the types of WLANs.Because of the popularity of WLANs, security research must be done in various types of WLANs. Experiments were done by Varadarajan , Kumar, and Reddy (2011) about improving WLAN performance under DoS attacks. DoS attacks on the physical layer were analyzed and expanded to the security of the physical layer of the sensor network model. This research was done by using the ant system. By using Receiver Operating Characteristics (ROC) on nodes, DoS 8 attacks can be predicted by formulating the classification of jammers under various attack scenarios. This approach can help improving detecting DoS attacks in WLANs.Research in this thesis was focuses on two types of WLANs: client-server and ad-hoc networks.

Jamming Attacks The DNS is a hierarchical tree structure whose root node is known as the root domain. A label in a DNS name directly corresponds with a node in the DNS tree structure. A label is an alphanumeric string that uniquely identifies that node from its brothers. Labels are connected together with a dot notation, ".", and a DNS name containing multiple labels represents its path along the tree to the root. Labels are written from left to right. Only one zero length label is allowed and is reserved for the root of the tree. This is commonly referred to as the root zone. Due to the root label being zero length, all FQDNs end in a dot [RFC 1034].A study into DoS attacks and defense was done by Raymond and Midkiff (2008). Since WSNs are used in monitoring medical uses, homeland security, industrial automation, and military applications, security of WSNs must be guaranteed. Defeating many threats of DoS attacks on WSNs can be done by encryption and authentication, but some other techniques still need to be found to prevent from special DoS attacks, especially Denial of Sleep attacks, which are still critical threats in WSNs.

Detection of Jamming WLANs are built upon a shared medium that makes it easy to launch jamming attacks. These attacks can be easily accomplished by sending radio frequency signals that do not follow any MAC protocols. Detection of jamming attacks can be done in multiple ways. One of the most efficient ways is to jump channels. Because communication between two legitimate nodes is done through a specific frequency, the frequency can be changed if necessary. While a jammer is attacking the wireless network, there are other effective ways to continue legitimate communication in the network. Engaging the jammer on the jammed channel and continuing communication in another channel was introduced by Beg, Ahsan, and Mohsin (2010). When the nodes detected the jamming in the wireless network, they jumped to another channel to continue legitimate communication. In the experiments, both 10 and 20 nodes experiments were done, and in both scenarios, after channels were jumped, the network resumes communications as normal. In both scenarios, the amount of packets dropped reduced immediately. The research concluded that channel jumping will decrease the throughput of the network. Also, it was easier to detect jamming through intermitted channel jumping. Concluded, channel jumping was a superior method of combating network interference, rather than changing network protocols (Jeung, Jeong, and Lim, 2011).The research concluded that channel jumping will decrease the throughput of the network. Also, it was easier to detect jamming through intermitted channel jumping. Concluded, channel jumping was a superior method of combating network interference, rather than changing network protocols (Jeung, Jeong, and Lim, 2011).In order to prevent from multi-channel jamming attacks, a cross-layer jamming detection method was developed (Chiang and Hu, 2011). Cross-layer jamming detection is a tree-based approach. A jamming detection algorithm was utilized in all legitimate nodes; when the communication process began, all the nodes had the ability to report jamming attacks in different layers, and only the reports which were generated by nodes with jamming detection algorithm were accepted by the system in order to avoid error. Research was also done about multi-channel jamming attacks by Jiang and Xue (2010). The difference from the jamming detection algorithm

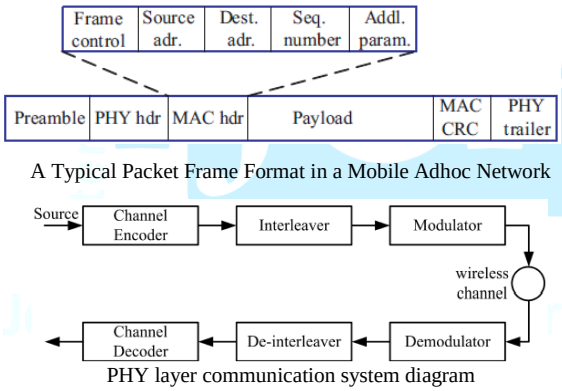
was that it focused on network restoration and design of traffic rerouting.

III PRELIMINARIES

The following lists basic terminologies required for understanding of Adhoc network implementations.

Sequence Name	Packets in Sequence
Data-Ack	TCP-DATA, TCP-ACK
ARP	ARP-REQ, ARP-RESP
TCP-Startup	TCP-SYN, TCP-SYN-ACK, TCP-ACK
AODV	AODV-RREQ, AODV-RREP(unicast)
DNS-Lookup	UDP-DATA, UDP-DATA

The types of packet sequences are shown in the following table.



A Typical Packet Frame Format in a Mobile Adhoc Network

PHY layer communication system diagram

Fig 5 (f) Impact of monitor node ratio on detection rate

VI CONCLUSION

We addressed the problem of selective jamming attacks in adhoc networks in an internal threat model in which the jammer is part of the network under attack, thus being aware of the protocol specifications and shared network details. We showed that the jammer can classify transmitted packets at will in a real time network scenario by decoding the first few symbols of an ongoing transmission. We evaluated its impact of selective jamming attacks on network protocols such as TCP and routing. Our results show that a selective jammer can significantly impact performance with very low effort. We developed cryptographic primitives scheme that uses commitment schemes, cryptographic puzzles, and all-or-nothing transformations (AONTs) with physical layer characteristics. We analyzed the security of our schemes and quantified their computational and communication overhead. The source of the problem lien in the Access point validation of the jammer which is addressed using the router Minrank strategy preventing Denial Of Service based authentication attempts of the jammer, thus improving the network conditions. As discussed in the introductory part jamming gain estimations do help to improve performance more which can be an interesting future research.

VIII REFERENCES

[1] Hu, Y.-C., Perrig, A. A survey of secure wireless ad hoc routing. IEEE Security & Privacy Magazine. v. 02, n. 3, (May–Jun.2004),pp. 28–39.

[2] T. X. Brown, J. E. James, and A. Sethi. Jamming and sensing of encrypted wireless ad hoc networks. In Proceedings of MobiHoc, pages 120–130, 2006.

[3] D. Thuente andM. Acharya. Intelligent jamming in wireless networks with applications to 802.11 b and other networks. In Proceedings of the IEEE Military Communications Conference MILCOM, 2006.

[4] M. Wilhelm, I. Martinovic, J. Schmitt, and V. Lenders. Reactive jamming in wireless networks: How realistic is the threat? In Proceedings of WiSec, 2011.

[5] M. Cagalj, S. Capkun, and J.-P. Hubaux. Wormhole-based antijamming techniques in sensor networks. IEEE Transactions on Mobile Computing, 6(1):100–114, 2007.

[6] A. Chan, X. Liu, G. Noubir, and B. Thapa. Control channel jamming: Resilience and identification of traitors. In Proceedings of ISIT, 2007.

[7] T. Dempsey, G. Sahin, Y. Morton, and C. Hopper. Intelligent sensing and classification in ad hoc networks: a case study. Aerospace and Electronic Systems Magazine, IEEE, 24(8):23–30, August 2009.

[8] O. Goldreich. Foundations of cryptography: Basic applications. Cambridge University Press, 2004.

[2] Alejandro Proano and Loukas Lazos. Packet-Hiding Methods for Preventing Selective Jamming Attacks. IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING ,VOL. 9, NO. 1, JAN-FEB 2012

Packet Data Transmission in TCP Networks and it's Security Concern: A foreword Approach

1.A VamsiKrishna 2.N.Venkatram 3.M.Sarada

1.M.Tech student,Dept of Computer Science,Nova's Institute of Technology,Eluru

2.Associate professor,HOD of Computer science,Nova's Institute of Technology,Eluru

3.Assistant professor,Dept of Computer science,Nova's Institute of Technology,Eluru

Abstract

Technology makes things easier, smarter and efficient and the most challenging one in the communication media. Communication is one of the integral parts of science that has always been a focus point for exchanging information among parties at locations physically apart. The technology must enable data-transfer i.e. the packet that will make (Mobile or Internet Network) internet access both attractive and useful. In order to make the transmission effective and attractive, many technologies which still focuses on the term of effectiveness? In this paper, packet data transmission to the destination uses the approach of segment and the cumulative encryption technology, which prevents all the link factors towards the loss and retransmission of data. As of wireless packet data networks will provide flexible access to a vast array of data applications by many users, each requiring a share of the network resources. Thus, In this paper we bring foreword a major issue in wireless packet data networks is the development of medium access control (MAC) protocols that make efficient use of available network and mobile resources. Controlling the transmission power level is one approach that provides improvements in both these areas. The capacity of the network is improved by optimizing spatial packing of source-destination pairs, known as spectral reuse, which leads to the further high end data EDGE technology modification and revelation.

Keywords: protocol, MAC, Packet Data, TCP

1. Introduction

The increased demand in the wireless technology has helped to simplify networking by enabling multiple computer users to simultaneously share resources in a home or business without additional or intrusive wiring. With the increasing demand a user of a wlan will have high demands on the system and will not accept too much degradation in performance to achieve mobility and flexibility. A wireless lan is based on a cellular architecture where the system is subdivided into cells, where each cell is controlled by a base station. For a non-power-controlled MAC to best use the

network resources in a single shared data channel scenario, it must provide a mechanism for avoiding collisions among mobile nodes competing for the channel. A power-controlled MAC adds to the complexity of this mechanism by also requiring that nodes send with only enough power to reach their intended destination.

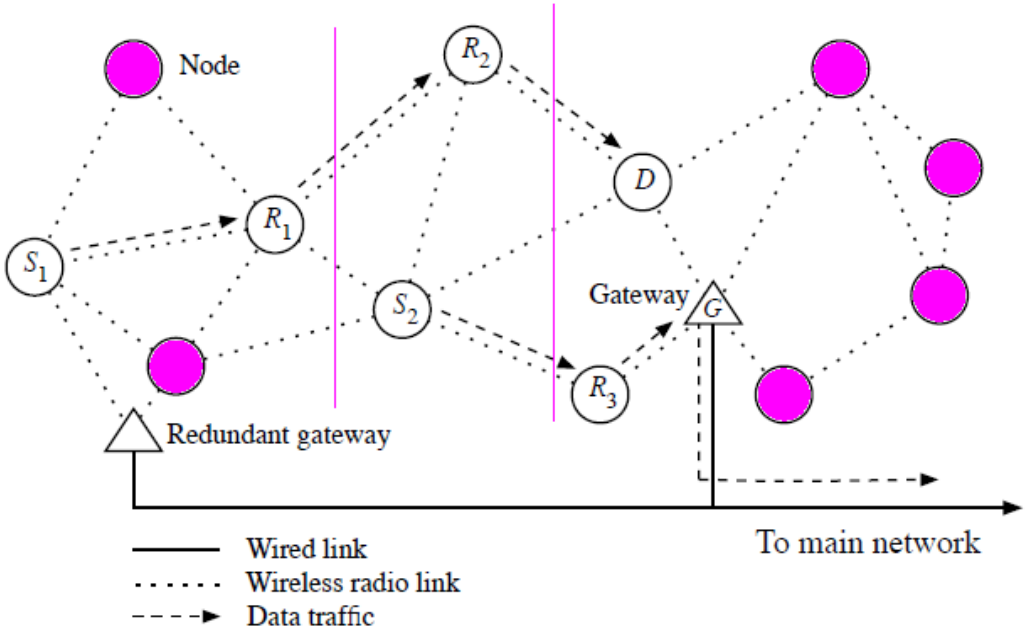


Fig. 1.1 Showing the Node *S* 1 sends data to node *D* via cooperating nodes *R*1 and *R*2, while node *S* 2 sends data out of the network via node *R*3 and the gateway *G*

In the above fig. 1.1 the source node transmits a packet to a neighboring node with which it can communicate directly. The neighboring node in turn transmits the packet to one of its neighbors, and so on until the packet is transmitted to its ultimate destination. Each link that a packet is sent over is referred to as a hop; the set of links that a packet travels over from the source to the destination is called a route or path. Routes are discovered by running a distributed routing protocol on the network.

2. Related Work

The approach of this report is to address two issues reliability and the efficiency of the data in wireless environment and how the data security is managed in cellular systems. After thorough research, collecting information about various techniques from the evolutionary cellular systems, we need to take the effectiveness in every aspect to the high end reliability. In the methodology of hub network the fundamental change that we make in the existing approach is the following: unlike current protocols that use the reception of control packets as an on-off trigger for transmission/deferral by hidden

and exposed stations, our approach is to use the signal strength of a received control message to bound the transmission power of these stations. Finally, a transport protocol study depicts the deficiencies of current transport techniques' abilities to regulate traffic in wireless ad hoc networks.

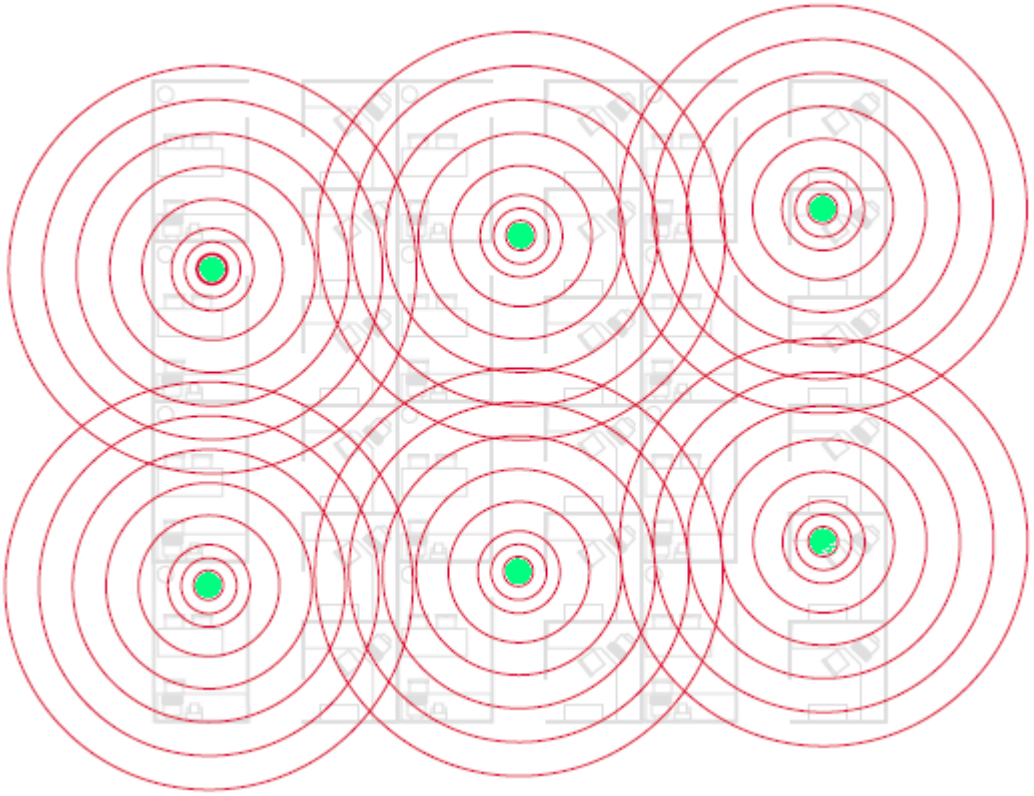


Fig. 2.1 Behavior of spectrum roaming among access points with non-overlapping frequencies allows for virtually unlimited coverage range.

In the fig. 2.1; a basic wireless infrastructure with a single access point is called a Basic Service Set (BSS). When more than one access point is connected to a network to form a single sub-network, it is called an Extended Service Set (ESS). A wireless LAN NIC may decide to "re-associate" itself with another access point within range because the load on its current access point is too high for optimal performance. These capabilities can have a positive impact on overall network performance. Hence of the increasing the performance in terms of non-centric collision, which keeps rise to loss of strength lead to decrease in reliability. Depending on environmental specifics, automatic

downshifting by the access point or client allows compatibility adjustment to prevailing radio frequency conditions. At any one moment, an 802.11b network can be running at 11 Mbps, 5.5 Mbps, 2 Mbps, or 1 Mbps 22 Mbps wireless networking products. And depending on where each wireless device is in a home or office, each of those devices can be transmitting at any one of these speeds.

3. Methods

As of the technology and demand increasing, it leads to last during the beginning of the commercialization of the Internet, organizations and individuals connected without concern for the security of their system or network. Overtime, they realized that some form of security was required to prevent others from exploiting the connected resources. Deployment of Wireless LAN as a medium of communication to connect mobile devices with the wired infrastructure has paved a new path in the networking technology.

Organizations are rapidly deploying 802.11 standard based wireless infrastructures. For most WLAN users, there are three basic issues:

a) Data Compromise is any form of disclosure to unintended parties of information. Data compromise can be inappropriate access to payroll records by company employees, or industrial espionage whereby marketing plans are disclosed to a competitor.

b) Denial of Service is an operation designed to block or disrupt normal activities of a network or facility. This can take the form of false requests for login to a server, whereby the server is too distracted to accommodate proper login requests.

c) Unauthorized access is any means by which an unauthorized party is allowed access to network resources or facilities. Unauthorized access can lead to compromise, for example, if access is gained to a server with unencrypted information, or destruction in the case that critical files, although encrypted on the server, may be destroyed.

In this paper, we considered the step by step approach to correspond to the situation shown in Figure 3.1, where some source node ‘i’ is sending to a destination node ‘j’ and a potential interfering transmitter l wishes to transmit. The steps also reference the pseudo code algorithm that represent the idle, transmitting, and receiving phases respectively in order to achieve the best way of transmission.

Conventional collision avoidance methods had an “on/off model,” wherein a node can either transmit or receive. In this paper, we implement in the two main component phase A request-power-to-send (RPTS)/acceptable-power-to-send (APTS) handshake between the data sender and receiver, which is used to determine the minimum transmission power that will result in a successful packet reception at the receiver. In this paper, we tries to give focus on collisions occurs as a result of other transmitters sending with enough power that a receiver observes more noise power than can be tolerated, based on the strength of the

desired signal. Therefore, by comparing the current desired signal power to the total power observed from other sources, a receiver can calculate the amount of additional noise from interfering stations that can be tolerated.

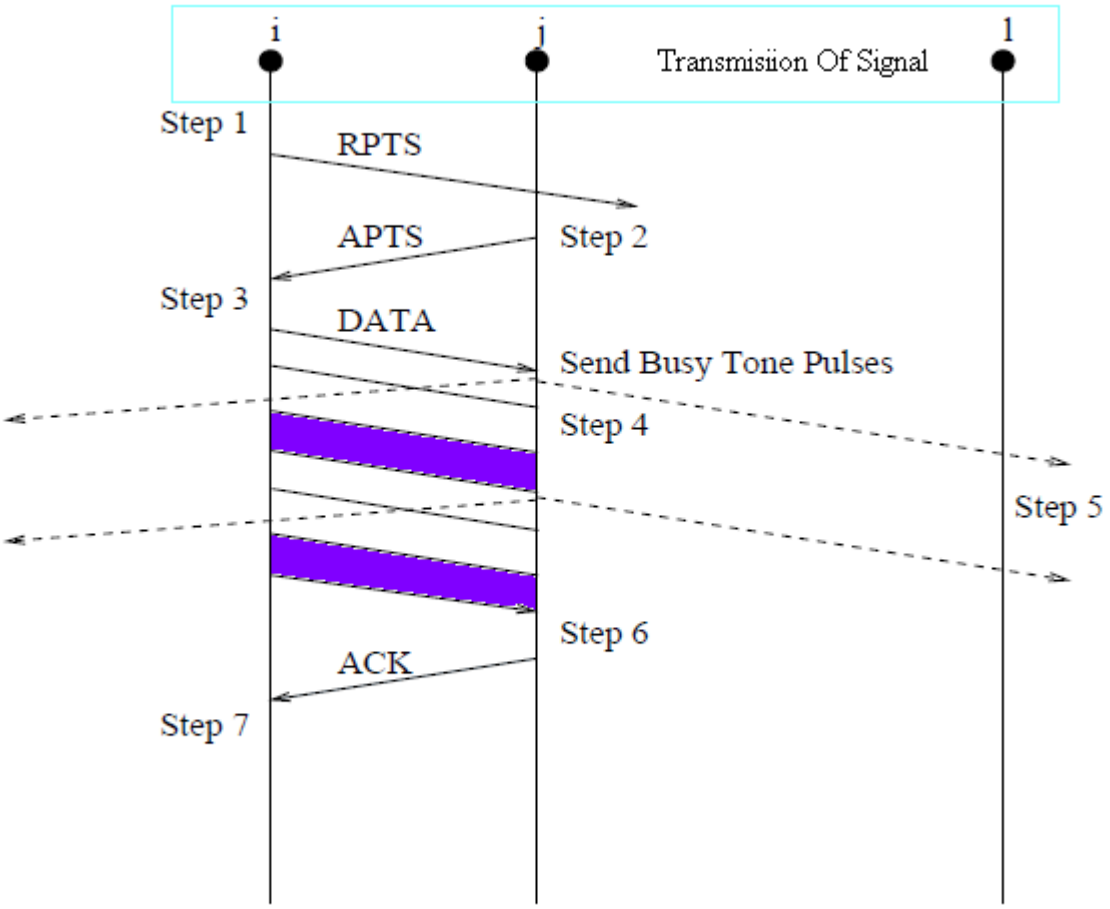


Fig. 3.1 Showing the Packet Transmission through the periodic step of acknowledgement.

In this paper, the noise tolerance advertisement or busy tone is periodically pulsed by each receiver in the busy tone channel, where the signal strength of the pulse indicates the tolerance to additional noise. The packet handshake sequence on the data channel is RPTS-APTS-DATA-ACK and on the busy tone channel busy tone pulses are periodically sent while the data is received to protect the data packet. As of wireless devices need to be small and wireless networks are bandwidths limited, some of the key challenges in wireless networks are:

a) Data Rate Enhancements.
b) Low power networking.
c) Security.
d) Radio Signal Interference.

e) System Interoperability.

Algorithm for protocol Based Approach for effective Transmission of Packet

1. Start IDLE:
2. do{
3. handle route ad(Packet p)
4. {
5. for each Route r in p do
6. handle update(r);
7. }
8. Set Pr BT to max BT power observed while sensing
9. }while(no Data Generated and no RPTS Received)
10. if(Data Generated){
11. Goto TX State
12. }
13. else{
    a. handle update(Route r) {
    b. update metric(r);
    c. if (r.seq == curr[r.dest].seq && r.metric < curr[r.dest].metric) {
    d. curr[r.dest] = r;
    e. curr[r.dest].best time = now;
    f. schedule triggered update(r);
14. }
    a. else if (r.seq > curr[r.dest].seq) {
15. Goto RX State
16. }
17. old[r.dest] = curr[r.dest];
18. curr[r.dest] = r;
19. curr[r.dest].first time = now;
20. curr[r.dest].best time = now;
21. Start TX:
22. lookup route(IPAddress dst) {
23. // use old route if we haven't yet advertised current route
24. if (curr[dst].first time + 2×curr[dst].wst > now)
25. return old[dst].next hop;
26. else
27. return curr[dst].next hop;
28. }
29. backoff = rand(1,Max backoff) ∗ aSlot Time
30. Set Pr BT to max BT power observed
31. Pt bound = min{C/Pr BT, Pt Max}
32. if(Pt bound < Pt Min) {
33. Goto START
34. }
35. Pt = Pt bound
36. Pn S = get noise power()
37. Send RPTSh SenderID, ReceiverID, Packet len, Pt, Pn Si on Data Channel at Pt
38. Wait for APTS on Data Channel
39. if(Time-out waiting for APTS) {
40. Max backoff
41. Goto START
42. }
43. Pt = extract Pt desired(APTS)
44. if(Pt > Pt bound){
45. Goto START
46. }
47. Send data at Pt on Data Channel
48. Wait for ACK on Data Channel
49. if(Time-out waiting for ACK){
50. Max backoff
51. Goto START
52. }
53. else{
54. Max backoff
55. }
56. Goto IDLE State

In the above algorithm there are three types of variables used in the algorithm: fixed variables or those that store values

determined from off-line calculations, variables that store measured values (primarily from the receiver), and variables for storing the result of calculations made from measured and fixed variables.

3.1 Performance Evaluation

In the effectiveness of packet data, this phenomenon is particularly evident over the 250-m connectivity range for data packet, as demonstrated in Figure 3.2, where the fraction of total packets received by destinations in five distance ranges (0-50, 50-100, 100-150, 150-200, and 200-250 m) from their sources is shown for 100 flows sending 1, 4, 16, and 64 packets per second. A perfectly fair protocol would result in a linearly increasing number of packets sent to each range since the number of destinations within each range increases as 2_r, where r is the distance from the source node.

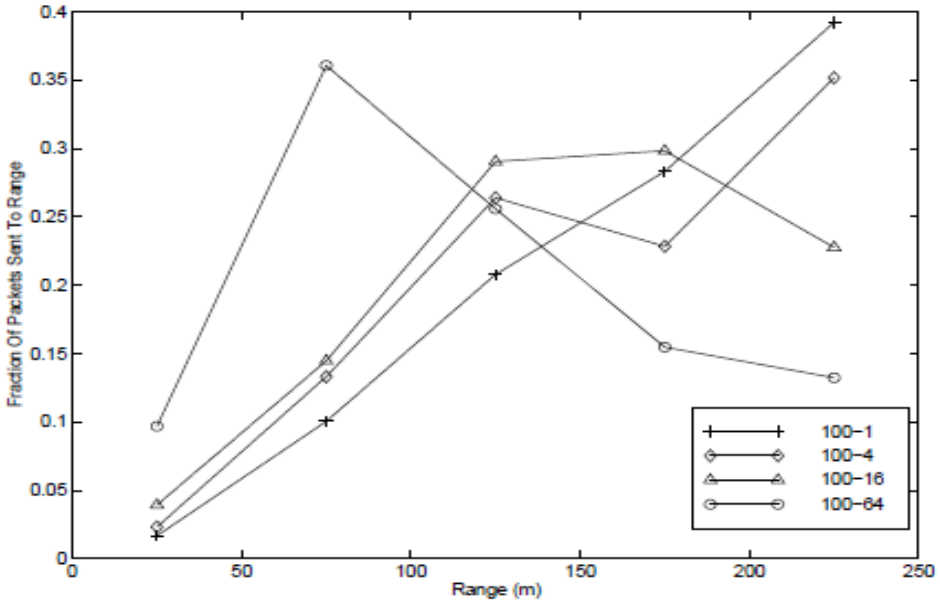


Fig.3.2 Destination range distribution for 802.11 showing the packet Transmission

The transmissions are sent at a fixed power level, there is less noise protection for destinations further from their sources, resulting in a greater number of lost packets at greater network loads. Therefore, at high loads the aggregate background noise will cause receivers that obtain less signal power from their corresponding transmitter to have more packet corruptions.

4. Conclusion

Technology in the domain of networking is changing environment making things simpler. The MAC mechanisms for

discovering the power needed to reach the intended destination and avoid collisions with other receivers were defined, and their effectiveness was tested with various network configurations. However, this method also limits the protocols ability to fully exploit the network resources. This is a trade that must be evaluated based on the constraints of the applications employed in the mobile network. In order to increase the optimality we need to reduce the number of hubs between source and destination. However further, these routing packets must be sent over the entire transmission range such that the improvements in spectral reuse and energy savings of implementing power control. This paper leads to controlling the maximum transmission range that the MAC layer is allowed to transmit over based on the density of the network and location of neighboring nodes.

5. References

[1] IEEE 802.11Working Group. http://grouper.ieee.org/groups/802/11/index.html.

[2] www.intersil.com and www.ti.com for information on IEEE 802.11g.

[3] V. Bhuse, A. Gupta, and L. Lilien, "DPDSN: Detection of Packet Dropping Attacks for Wireless Sensor Networks," Proc. Fourth Trusted Internet Workshop, 2005.

[4] M. Kefayati, H.R. Rabiee, S.G. Miremadi, and A. Khonsari, "Misbehavior Resilient Multi-Path Data Transmission in Mobile Ad-Hoc Networks," Proc. Fourth ACM Workshop Security of Ad Hoc and Sensor Networks (SASN '06), 2006.

[5] R. Mavropodi, P. Kotzanikolaou, and C. Douligeris, "Secmr—A Secure Multipath Routing Protocol for Ad Hoc Networks," Ad Hoc Networks, vol. 5, no. 1, pp. 87-99, 2007.

[6] F. Ye, H. Luo, S. Lu, and L. Zhang, "Statistical En-Route Filtering of Injected False Data in Sensor Networks," Proc. IEEE INFOCOM, 2004.

[7] S. Zhu, S. Setia, S. Jajodia, and P. Ning, "An Interleaved Hop-by- Hop Authentication Scheme for Filtering False Data in Sensor Networks," Proc. IEEE Symp. Security and Privacy, 2004.

[8] H. Yang, F. Ye, Y. Yuan, S. Lu, and W. Arbaugh, "Toward Resilient Security in Wireless Sensor Networks," Proc. Sixth ACM Int'l Symp. Mobile Ad Hoc Networking and Computing (MobiHoc '05), 2005.

[9] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: A media access protocol for wireless LAN's," in Proceedings of ACM SIGCOMM, Oct. 1994.

[10] IEEE Std 802.11 - 1997, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 1997.

[11] C. Fullmer and J. Garcia-Luna-Aceves, "Floor acquisition multiple access (FAMA) for packet-radio networks," in Proceedings of ACM SIGCOMM, Sept. 1995.

Secure Routing in the perspective of authentication scheme for mobile Communication

1.D Ashok Kumar 2.N.Venkatram 3.M.Venugopal

1. M.Tech student,Dept of Computer Science,Nova's Institute of Technology,Eluru

2.Associate professor,HOD of Computer science,Nova's Institute of Technology,Eluru

3.Assistant professor,Dept of S&H, Lingayas Institute of technology,Vijayawada

Abstract

In the recent trend of communication; where today's world is moving in the higher perspective of human race to make the next level of technology domain. Technology makes things easier, smarter and efficient and the most challenging one in the communication media. Communication is one of the integral parts of science that has always been a focus point for exchanging information among parties at locations physically apart. Similarly, the term `mobile' has completely revolutionized the communication by opening up innovative applications that are limited to one's imagination. Today, mobile communication has become the backbone of the society. All the mobile system technologies have improved the way of living. It's main plus point is that it has privileged a common mass of society. In addition, the existence of distinct routing protocols presents difficulties in standardization of routing models. Accordingly, it affects the design of authentication schemes because we always take the specific routing procedures into consideration. Therefore, in this paper, we bring forward the problem of designing an adaptable authentication scheme for both AODV and DSR protocols and try to introduce the first scheme in order for the other to come up with more valuable opinions.

Keywords: AODV, Authentication (OTP), TGT (Ticket Granting Control)

1. Introduction

Authentication is one of the most crucial modules for any given system. Since the time of wireless telegraphy, radio communication has been used extensively. Our society has been looking for acquiring mobility in communication since then. Initially the mobile communication was limited between one pair of users on single channel pair. The range of mobility was defined by the transmitter power, type of antenna used and the frequency of operation. With the increase in the number of users, accommodating them within the limited available frequency spectrum became a major problem. To resolve this problem, the concept of cellular communication was evolved. The present day cellular communication uses a basic unit called cell. Each cell consists of small hexagonal area

with a base station located at the center of the cell which communicates with the user. To accommodate multiple users Time Division multiple Access (TDMA), Code Division Multiple Access (CDMA), Frequency Division Multiple Access (FDMA) and their hybrids are used. Numerous mobile radio standards have been deployed at various places such as AMPS, PACS, GSM, NTT, PHS and IS-95, each utilizing different set of frequencies and allocating different number of users and channels. A researcher firstly identifies the specific security requirements, and then certain signature is constructed against each requirement. In general, signatures for one application are not adaptable to others. However, we observe that the adaptability of signature schemes is preferable in some cases where standard operation procedures are not available. The adaptability of signature schemes will then be able to deal with the variability of the application domain, and in turn provide generic security for different procedures with the help of a single signature scheme.

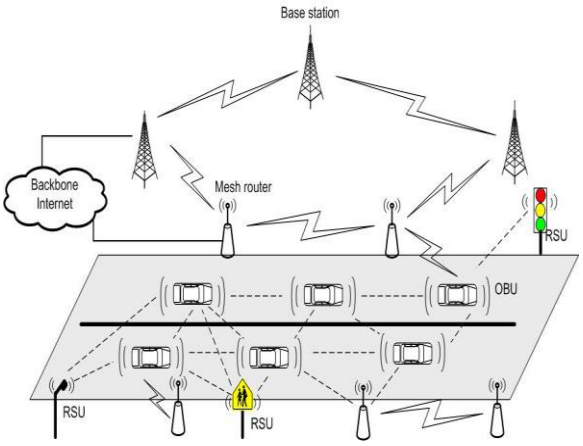


Fig 1.1 The basic radio transmission techniques (Simplex, Half Duplex, Full Duplex)

When a user moves from one cell to the other, to keep the communication between the user pair, the user channel has to be shifted from one BS to the other without interrupting the call, i.e., when a cell moves into another cell, while the conversation is still in progress, the master cell automatically transfers the call to a new FDD channel without disturbing the conversation. This process is called as handoff.

2. Related Work

Communication in networks implies transmitting data packets along some certain paths, routes. How to find the path (routing) thus enable data transmission is the fundamental step of network communication. Routing is conducted using routing protocols which use metrics to evaluate what path will be the best for a packet to travel. A metric is a standard of measurement, such as path bandwidth, which is used by routing algorithms to determine the optimal path to a destination. To enable the process of path determination, routing algorithms initialize and maintain routing tables, which contain route information. Cellular telephone systems must accommodate a large number of users over a large geographic area with limited frequency spectrum, i.e., with limited number of channels. If a single transmitter/ receiver is used with only a single base station, then sufficient amount of power may not be present at a huge distance from the base station. For a large geographic coverage area, a high powered transmitter therefore has to be used. But a high power radio transmitter causes harm to environment. Mobile communication thus calls for replacing the high power transmitters by low power

transmitters by dividing the coverage area into small segments, called cells. Each cell uses a certain number of the available channels and a group of adjacent cells together use all the available channels.

2.5G Mobile Networks:

2.5G networks also brought into the market some popular application, a few of which are: Wireless Application Protocol (WAP), General Packet Radio Service (GPRS), High Speed Circuit Switched Dada (HSCSD), Enhanced Data rates for GSM Evolution (EDGE) etc.

Third Generation Networks

3G networks enable network operators to offer users a wider range of more advanced services while achieving greater network capacity through improved spectral efficiency. Services include wide-area wireless voice telephony, video calls, and broadband wireless data, all in a mobile environment. Additional features also include HSPA data transmission capabilities able to deliver speeds up to 14.4Mbit/s on the down link and 5.8Mbit/s on the uplink.

WCDMA

It supports two basic modes of operation: FDD and TDD. In the FDD mode, separate 5-MHz carrier frequencies with duplex spacing are used for the uplink and downlink, respectively, whereas in TDD only one 5-MHz carrier is time shared between the uplink and the downlink. WCDMA uses coherent detection based on the pilot symbols and/or common pilot. WCDMA allows many performance- enhancement methods to be used, such as transmit diversity or advanced CDMA receiver concepts. In this transmission technology, code specific frequency may not authenticate as of GSM technology.

3. Methods

Authentication is a process to verify a person's identity for security purposes and authentication factor is the piece of information used to achieve it. With an increase in the need for secure environments, the authentication modules of systems (including operating system) are moving from traditional one-factor authentication (usually based on a static password) to multi-factor authentication. People are now expecting efficient group communication in education, entertainment, and industries enabled by mobile ad hoc networks. The mobile communication has provided global connectivity to the people at a lower cost due to advances in the technology and also because of the growing competition among the service providers. We would review certain major features as well as standards of the mobile communication till the present day technology like Digital modulation formats were introduced in this generation with the main technology as TDMA/FDD and CDMA/FDD. The 2G systems introduced three popular TDMA standards and one popular CDMA standard in the market. Using the security requirements as a scale, we studied some existing secure routing protocols and justified their performance according to our security

requirements. We noticed that the security of the existing proposals is not established from a realistic point of view. Some of the assumptions, such as the pre-establishment of security associations, and the requirement of time synchronization, generally conflict to the characteristics of mobile ad hoc networks. Hence of both level follows authentication.

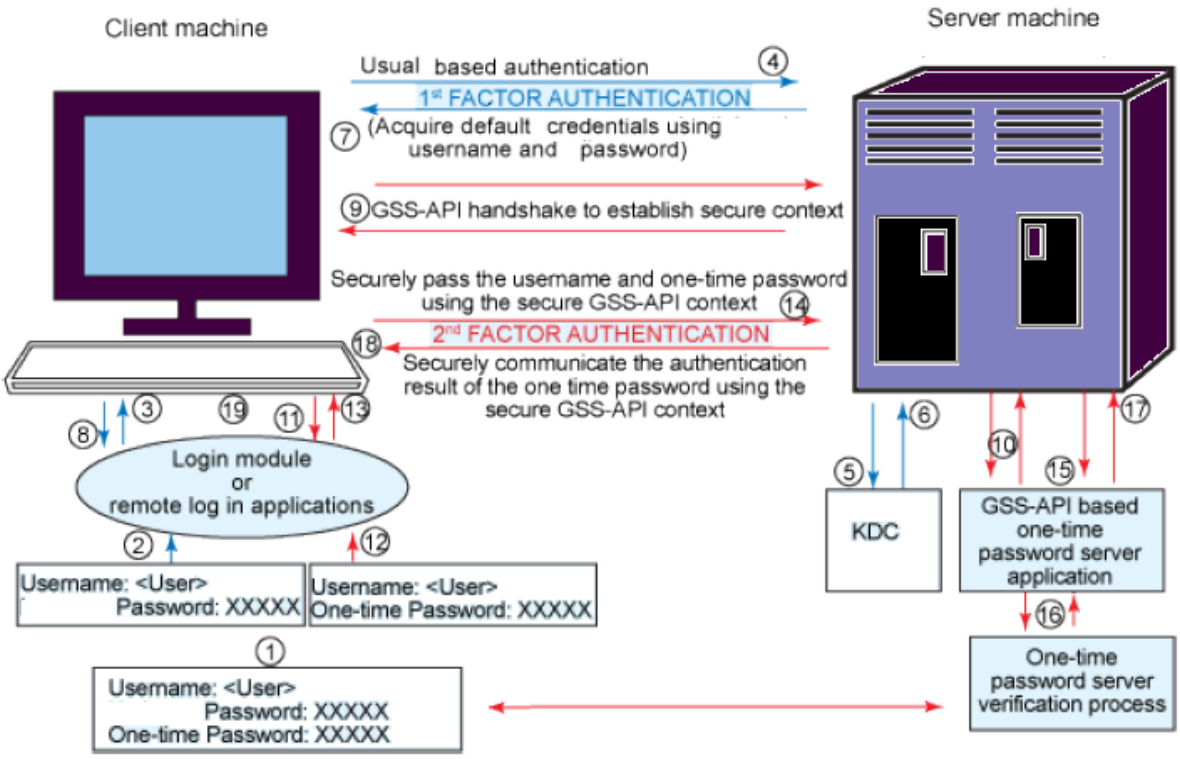


Fig: 3.1 Architecture Showing the Authentication of Mobile communication

The architecture involves following steps to exhibit the authenticated operation.

Steps 1 - Prompt the user to enter the user name and password.

Steps 2, 3, 4, 5, 6, 7 - Use the User Name and the corresponding password (which the user has to remember) to acquire the credential (TGT-Ticket Granting Ticket). If the password entered is incorrect, the authentication fails and the user is not allowed any access. On successful acquisition of the ticket (TGT), the first-factor authentication is completed. This is similar to any regular login module that based on authentication.

Steps 9, 10, 11 - Use the above-acquired credential and establish secure GSS-API context with the GSS-API based OTP application server residing on the system (assuming the OTP server has been using GSS-API).This involves a handshake between the client login module and the GSS-API-base done-time password application server. Note that here both the login module and the OTP application server are GSS-API-

based applications running on the communication medium.

The main theme behind every communication technology is to be effective i.e. Efficient and secure authentication with integrity. Efficient secure features provide protocol algorithm for two significant mobile ad hoc routing protocols, AODV and DSR. We intend to provide authentication, integrity and non-repudiation for AODV and DSR routing operations. It is generally recognized that the security deployment for mobile network routing protocol is difficult because of the following reasons.

- No central control exists in network. In a pure ad hoc environment, there is no trusted third party in the network.
- Nodes are resource constraint. The security deployment, such as signature generation and verification, will somehow consume the limited resources, which in turn affects the performance of the node.
- Routing protocols are distinct. Accordingly, an authentication scheme designed for certain types of routing protocols will not be applicable to others. On the other hand, design a general authentication scheme without considering the nature of protocols will result in huge waste in routing operation overhead.

Cryptographic primitives which provides authentication, integrity and non-repudiation are especially suitable for the mobile ad hoc scenario. Digital signatures, which have been long used as an authentication method, offer the above three properties. However, the deployment of a digital signature enabled authentication scheme in mobile ad hoc networks is not straightforward. The way leading to secure mobile ad hoc environments is full of disturbances.

Performance Analysis:

Following factors analyses the performance.

- ✓ Transmitted power
- ✓ Received power
- ✓ Area and shape of the cell
- ✓ Mobility of users

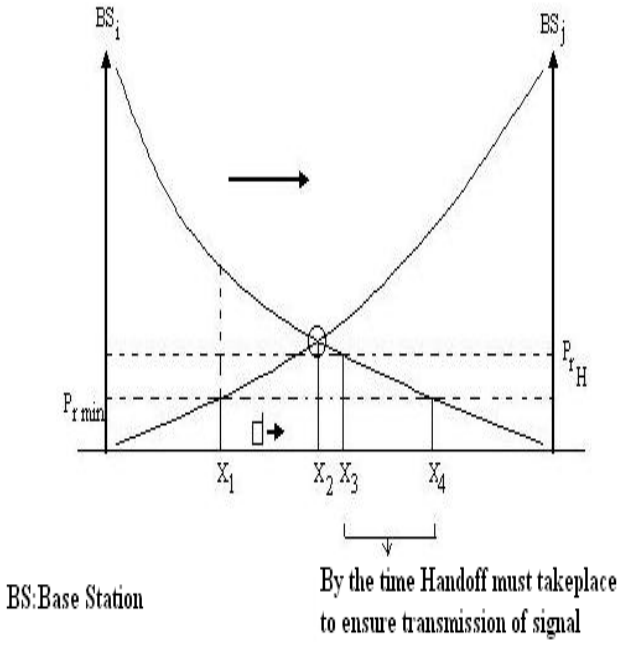


Fig.3.2 Association of cell tower when moving from i-th cell to j-th cell

Let us consider rectangular cell with sides R1 and R2 inclined at an angle _ with horizon, as shown in the Figure 3.2. Assume N1 users are having handoff in horizontal direction and N2 in vertical direction per unit length.
The number of crossings along R1 side is: (N1cosθ+N2sinθ) R1 and the number of crossings along R2 side is: (N1sinθ + N2cosθ) R2

Then the handoff rate Λ_H can be written as

$$\Lambda_H = (N1\cos\theta + N2\sin\theta)\ R1 + (N1\sin\theta + N2\cos\theta)\ R2$$

It follows the angle based approach for coupling while changing the network in order to ensure the authentication.

4. Conclusion & Future Devlopement

Tremendous changes are occurring in the area of mobile radio communications, so much so that the mobile phone of yesterday is rapidly turning into a sophisticated mobile device capable of more applications than personal computers/ laptops were capable of only a few years ago. Rapid development of the Internet with its new services and applications has created fresh challenges for the further development of mobile communication systems. As a highly dynamic, infrastructure less network, how to find peer nodes and establish links, namely routing, become the major issue to be solved. In our routing paradigm, shorter key size and signature size can reduce the data transmission overhead. In turn, the network performance can be enhanced by this means. Thus, the future work lead to the generation based speed and device based service.

5. Reference

[1] D. Boneh and H. Shacham. Group Signatures with Verifier Local Revocation. In proceedings of the 11th ACM conference on Computer and Communications Security (CCS), pp.168-177, 2004.

[2] J. Baek and Y. Zheng. Identity-Based Threshold Signature Scheme from the Bilinear airings. International Conference on Information Technology: Coding and Computing (ITCC04), Vol.1, pp. 124-128, April, 2004.

[3] M. Rahnema, "Overview of the GSM system and protocol architecture," IEEE Commun. Mag., vol. 31, no. 4, pp. 92–100, Apr. 1993.

[4] B. Mallinder, "An overview of the GSM system," in Proc. 3rd Nordic Seminar Digital Land Mobile Radio Commun., Copenhagen, Denmark, 1998, pp. 12–15.

[5] A. Aziz and W. Diffie, "Privacy and authentication for wireless local area networks," IEEE Personal Commun., vol. 1, no. 1, pp. 24–31, 1993.

[6] M. S. Hwang, Y. L. Tang, and C. C. Lee, "An efficient authentication protocol for GSM networks," in Proc. AFCEA/IEEE Euro-Comm, 2000, pp. 326–329.

[7] S. Suzuki and K. Nakada, "An authentication technique based on distributed security management for the global mobility network," IEEE J. Sel. Areas Commun., vol. 15, no. 8, pp. 1608–1617, Oct. 1997.

[8] C. H. Lee, M. S. Hwang, and W. P. Yang, "Enhanced privacy and authentication for the global system for mobile communications,"

Wireless Netw., vol. 5, no. 4, pp. 231–243, 1999.

[9] L. Buttyan, C. Gbaguidi, S. Staamann, and U.Wilhelm, "Extensions to an authentication technique proposed for the global mobility network," IEEE Trans. Commun., vol. 48, no. 3, pp. 373–376, Mar., 2000.

[10] K. F. Hwang and C. C. Chang, "A self-encryption mechanism for authentication of roaming and teleconference services," IEEE Trans. Wireless Commun., vol. 2, no. 2, pp. 400–407, Mar. 2003.

[11] C. C. Lee, M. S. Hwang, and W. P. Yang, "Extension of authentication protocol for GSM," IEE Proc., Commun., vol. 150, no. 2, pp. 91–95, 2003.

[12] L. Harn and W. J. Hsin, "On the security of wireless network access with enhancements," in Proc. ACM Workshop Wireless Security, 2003, pp. 88–95.

[13] A. Peinado, "Privacy and authentication protocol providing anonymous channels in GSM," Comput. Commun., vol. 27, no. 17, pp. 1709–1715, 2004.

[14] C. C. Chang, J. S. Lee, and Y. F. Chang, "Efficient authentication protocol of GSM," Comput. Commun., vol. 28, no. 8, pp. 921–928, 2005.

[15] C. Tang and D. O.Wu, "An efficient mobile authentication scheme for wireless networks," IEEE Trans. Wireless Commun., vol. 7, no. 4, pp. 1408–1416, Apr. 2008.

[16] M. Al-Fayoumi, S. Nashwan, S. Yousef, and A. R. Alzoubaidi, "A newhybrid approach of symmetric/asymmetric authentication protocol for future mobile networks," in Proc. Wireless Mobile Comput., Netw. Commun., 2007, pp. 29–29.

[17] V. Kalaichelvi and R. M. Chandrasekaran, "Secure authentication protocol for mobile," Proc. Comput., Commun. Netw., pp. 1–4, 2008.

[18] K. P. Kumar, G. Shailaja, A. Kavitha, and A. Saxena, "Mutual authentication and key agreement for GSM," in Proc. ICMB, 2006, p. 25.

[19] K. Ammayappan, A. Saxena, and A. Negi, "Mutual authentication and key agreement based on elliptic curve cryptography for GSM," in Proc. ADCOM, 2006, pp. 183–186.

[20] 3rd Generation Partnership Project; Technical Specification Group SA; 3G Security, "Security architecture, version 4.2.0, release 4," 3GPP, TS 33.102, 2001.

[21] U. S. Department of Commerce/National Institute of Standard and Technology, "Specification for the Advanced Encryption Standard (AES)," FIPS PUB 197, Nov. 2001 [Online]. Available: http://csrc.nist.gov/encryption/aes