Evaluation of a Knowledge Based Authentication Mechanism through Persuasive Cued Click Points

Y.Sravana Lakshmi

Computer Science and Engineering

BVSR Engineering College

Chimakurthy, India

Abstract: Authentication is the first line of defence against compromising confidentiality and integrity. Password protection is a security process that protects information from unauthorized user. An important usability goal for knowledge-based authentication systems is to support users in selecting passwords of very high security, in the sense of being from an expanded effective security space. We use persuasion to influence user choice in click-based graphical passwords, encouraging users to select more random, and hence more difficult to guess, click-point. This paper presents an integrated evaluation of the persuasive cued click-points graphical password scheme, including usability and security evaluations, and implementation considerations. It reflects a proposed system which uses graphical password using knowledge based authentication mechanism. Users select their images for any click points and for one click point display the next image having a specific relational order. The use of inclusive exclusive principal can be applied to find the minimum or maximum number of images required to form the image sequence reducing shoulder surfing problem.

Keywords: Cued Click-Points; Graphical password authentication; Inclusive- Exclusive principal; Shoulder surfing; Knowledge-based authentication.

1. INTRODUCTION

Beginning around 1999, a multitude of graphical password schemes have been proposed, motivated by the promise of improved password memorability and thus usability, while at the same time improving strength against guessing attacks. Like text passwords, graphical passwords are knowledge-based authentication mechanisms where enter a shared secret as evidence of their identity. Graphical passwords are of three types:

- Click based graphical password scheme
- Choice based graphical password scheme
- Draw based graphical password scheme

According to person psychology, humans are able to memorize pictures simply. Users are creating unforgettable passwords like text and symbols passwords that are easy for crack hackers to guess, but strong system-assigned passwords are difficult for users to remember. Computer security systems should also consider the human factors such as ease of a use and accessibility. Present secure systems undergo because they typically ignore the importance of human factors in security. Graphical passwords are intended to capitalize on this human characteristic in hopes that by reducing the memory load on users, coupled with a larger full password space offered by image, Thumb impression, digital signatures, mobile passwords, more secure passwords can be produced and users will not resort to insecure practices in order to extent.

The problems of knowledge-based authentication, typically text-based passwords, are well known. Users often create memorable passwords that are easy for attackers to guess, but strong system-assigned passwords are difficult for users to remember. A password authentication system should encourage strong passwords while maintaining memorability. Rather than increasing the burden on users, it is easier to follow the system's suggestions for a secure password a feature lacking in most schemes.

The approach to create the first persuasive click-based graphical password system, Persuasive cued click-points (PCCP) and conducted user studies evaluating usability and security. This systematic examination provides a comprehensive and integrated evaluation of PCCP, covering both usability and security issues, to advance understanding as in prudent before practical deployment of new security mechanisms. The images act as memory cues to aid recall. Example systems include Pass points and Cued Click-Points.

In Passpoints, passwords consist of a sequence of five click-points on a given image. Users may select any pixels in the image as click-points for their password. To log in, they repeat the sequence of clicks in the correct order, within a system-defined tolerance square of the original click points. Although Paasspoints is relatively usable security weaknesses make passwords easier for attackers to predict. Hotspots are the areas of the image that have higher likelihood of being selected by users as password click points. Attackers who gain knowledge of these hotspots through harvesting sample passwords can build attack dictionaries and more successfully guess Passpoints passwords. Users also tend to select their click points in predictable patterns. By adding a persuasive feature to CCP, PCCP, encourages users to select less predictable passwords, and makes it more difficult to select passwords where all five click points are hotspots. Specifically, when users create a password, the images are slightly shaded except for a viewport. The viewport is positioned randomly, rather than specifically to avoid known hotspots, since such information might allow attackers to improve guesses and could lead to the formation of new hotspots.

The viewports size is intended to offer a variety of distinct points but still cover only an acceptable small fraction of all possible points. Users must select a click-point within this highlighted viewport and cannot click outside of the viewport, unless they press the shuffle button to randomly reposition the viewport. While users shuffle as desired, this significantly slows password creation. The viewport and shuffle button appear only during password creation. During later password entry, the images are displayed normally, without shading or the viewport, and users may click anywhere on the images.

Password capture attacks occur when attackers directly obtain passwords by intercepting user entered data, or by tricking users into revealing their passwords. For systems like PCCP, CCP, and Passpoints, capturing one login instance allows fraudulent access by a simple replay attack. An alternative to increasing the number of images is to use larger images but crop them differently for each user. Hotspot analysis would be more difficult for attackers because the coordinates of hotspots could not be directly applied across accounts. If furthermore, each user receives a different pool of images, an attacker would need to collect these data on a peer-user basis when launching an attack.

To overcome all these existing defects we provide an authentication scheme in which the user choice of selecting the password scheme plays a vital role. This method also provides a more secure password scheme. The use of persuasive technology persuades the user choicer of selecting the password. The graphical passwords use a click based authentication scheme. The persuasive cued click point's method uses the concept of persuading the user to select the password. Here the prediction of password is difficult for the hackers as it is generated in a random manner.

2. PROPOSED SYSTEM

The user will provide an option of selecting the hotspots in an image. The successive selection of the exact hotspots will enable the user to move the next successful images. For login into the system, the user will provide an option of selecting the hotspot in the continuous 5 images. After reaching the successful login attempt the user will be allowed to access the application. Storing the images in a secure database through file stream data type is one of the options used to secure the images instead of storing the images in the server. Users will be provided an option of selecting the images to create the authentication page which is not included in the existing system. Possibility of monitoring the hotspots by the nearby user is possible. To avoid the same the password with matrix formation is one of the complex password schemes in the world. This option is been added in this paper.

Advantages:

Integrated evaluation of the password scheme in graphical manner.
Users will be provided an option of selecting the images to create the authentication page.

2.1 Persuasive Cued Click Points

For creating persuasive Cued Click points persuasive feature is added to CCP.PCCP encourages users to select less probable passwords. For password generation PCCP uses requisites like viewport & shuffle. When users making a secrete word, the images are a little monochromic except for a viewport for to avoid known hotspots the viewport is positioned casually. The most useful benefit of PCCP is hackers have to improve their presumptions. Users have to choose a clickable area within the highlighted view port and cannot click outside of the viewport

unless they press the shuffle button to randomly reposition the viewport. At the time of password creation users may shuffle as often as desired but it slows the process of password generation. Only during the password generation, the viewport & shuffle buttons are displayed. After the secrete word generation process, graphical images are presented to users casually without the viewport& shuffle button. Then user has to choose exact clickable area on particular image. Now a day's PCCP is a best technology but has security problems. Using this method HOTSPOT problem is reduced, but this method is difficult to remember the exact clickable area.

Figure 1. Persuasive Cued Click-Points. During password creation, users select a click-point from the highlighted viewport or press the shuffle button to relocate the viewport.

3. SYSTEM DESIGN

The system designed consist of three modules such as registration module, picture selection module and system login module (see Figure 2).
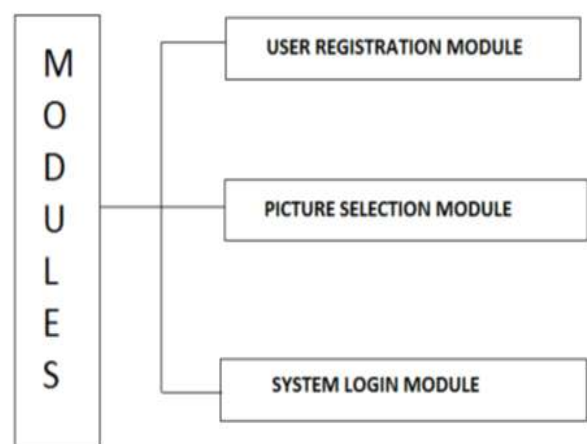
Figure 2. System design modules

3.1 User Registration Flow Chart

Below flowchart (see figure 3) shows the user registration procedure, this procedure include both registration phase (user ID) and picture selection phase. The process flow starts from registering user id and tolerance value. Once user completes all the user details then precede to next stage, which is selecting click points on generated images, which ranges from 1-5. After done with all these above procedure, user profile vector will be created.
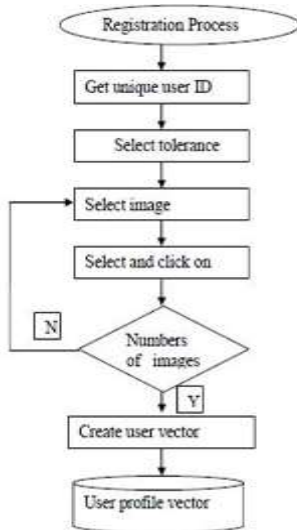

Figure 3. User registration flow chart

3.2 Login Flow Chart

In this login procedure (see figure 4), first user enters the unique user ID as same as entered during registration. Then images are displayed normally, without shading or the viewport, and repeat the sequence of clicks in the correct order, within a system-defined tolerance square of the original click-points. After done with all these above procedure, user profile vector will be opened.
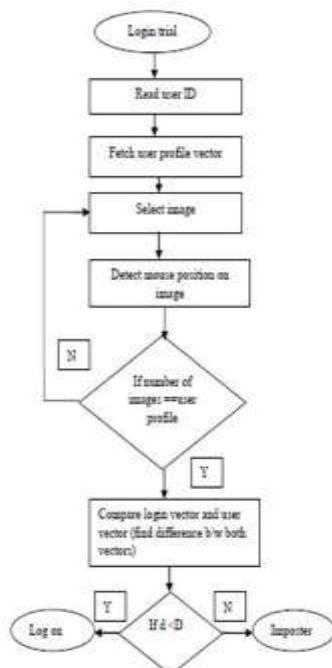

Figure 4. Login phase flowchart

4. MODULES

4.1 Authentication Scheme Module

In this module, the user will be permitted to provide their valid credentials to login in to the system. Before finalizing the validation of the user, they need to cross two level of boundaries. The first check will be the matrix validation. Once the user provides valid data, he will be allowed for the second check. In the second check users will be checked for the valid hotspot of the images. He will be permitted to view his profile page. If more than a stipulated time, the user has tried the login. In that case, the user will be blocked permanently and it can be overcome by the admin.
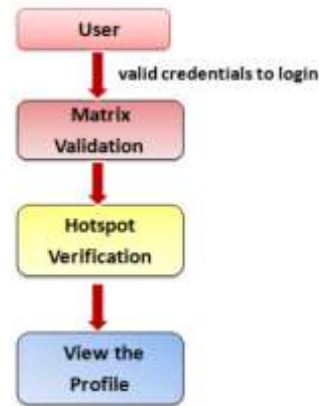

Figure 5. Authentication Scheme Module.

4.2 Matrix Verification Module

In this module, the user needs to enter the correct username. After entering, an 8*8 matrix is formed and it's displayed to the user. The user needs to verify the color combinations in the matrix. Based on the intersection of the input value provided by the user during registration, the matrix will be manipulated and the user needs to provide the exact intersection points are valid; he will be authorized user, so that he will be allowed for the second level of validation.
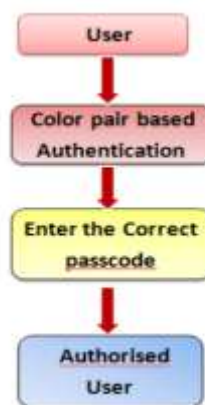

Figure 6. Matrix Verification Module.

4.3 Hotspot Verification Module

In this module, users will be provided with an image. If the user clicks the correct hotspots, then he will be the authorized user to access the application. If the user clicks the fake hotspots, then he will not be the authorized user to access the application.
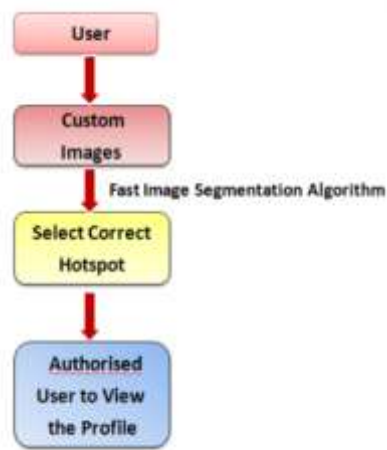
Figure 7. Hotspot Verification Module

5. GRAPHICAL PASSWORD METHOD USING KNOWLEDGE BASED AUTHENTICATION

Graphical password considers there are five images. There are five click points on one image. The next image displayed is based on the relation on the location of the previously entered click point creating a path through an image set. Users select their images for any click points and in return for once click point another click point and again it will display next image having relational order of second image click point and so on. In PCCP limitations all images could be reused at each stage in the password for every user. This strategy has the highest probability of collision where a user clicks on an incorrect click-point.

Figure 8. PCCP passwords as a choice-dependent path of images.

Using PCCP implementation, there is a possibility that images are reused for a given user. While this poses a potential usability concern, the likelihood of this happening is correspondingly low with enough images. In this technique reused images are not used for different users. If any image sequence is selected by a user then same image sequence is not available for any other user.

This means same image sequence is not used by two different users for the authentication. It is proposed to select different image sequences according to a specific relational order. Due to this when the user clicks on the image, the possibility of any collision is minimized. This probability can be reduced or nearly eliminated if the overlap of images is reduced between password stages by increasing the number of images in a users set. An alternative for increasing the no. of images is to use larger images. We can make a different set of images for each user, because hotspot analysis would be more difficult as suggested in PCCP.

6. EVALUATION OF PERFORMANCE AND RESULTS

6.1 Evaluation of Performance

In this evaluation of performance chapter we evaluated the usability of PCCP: with CCP through performance measures. The distributions contain all user-chosen click points for the given scheme for graphical passwords that were, at smallest amount, effectively re-entered at least once during login. In the Figure 9, this random distribution would appear as a curved diagonal line. In comparison, the PCCP graph shows that in the worst case, half of all click-points are confined within the most popular 0.00075 percent of hotspots within the distribution, while in the best case 0.00200. This indicates that CCP click-points have a flatter distribution and thus an attack dictionary based on hotspots should be less effective for CCP than for the other schemes. This analysis focused on individual click-points, not complete passwords. A key feature in PCCP and CCP is that creating a harder to guess password is the path of smallest resistance, likely making it more useful than schemes where secure behavior adds an extra problems on users. The approach has proven successful at reducing the formation of hotspots and patterns, thus growing the good space.
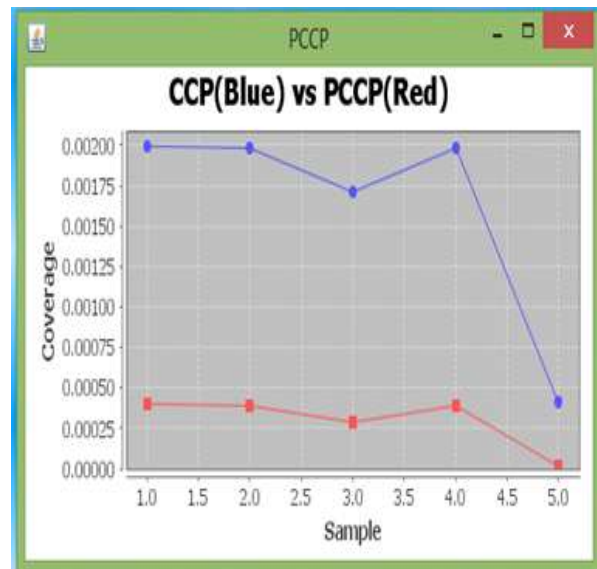
6.2 Results

Figure 9. Cumulative frequency distribution of hotspot coverage for CCP&PCCP

Table 1. Input parameters of CCP and PCCP for calculating coverage ratio

Images	Height	Width	Click Area	View Port Height	View Port Width	PCCP Coverage Ratio	CCP Coverage Ratio
Image 1	177.0	284.0	10*10	100	100	0.00039575	0.00198934
Image 2	168.0	300.0	10*10	100	100	0.00039368	0.00198413
Image 3	210.0	278.0	10*10	100	100	0.00028341	0.00171292
Image 4	168.0	300.0	10*10	100	100	0.00039368	0.00198413
Image 5	387.0	620.0	10*10	100	100	0.00001737	0.00041677

Equations used in coverage ratio calculation:

Image click ratio (10 10) (height width)

View port ratio (100 100) (height width)

PCCP coverage ratio image click ratio view port ratio

CCP coverage ratio image click ratio

In this Figure .9, the performance of the CCP and the performance of PCCP measure the coverage results in the number of input samples. The number of input samples are shown in the X axis and the coverage of the result are measured in Y axis. The CCP and PCCP measure the coverage result in the number of input samples; it shows that the proposed CCP coverage is most important result in the samples. CCP's segments were the longest and within range of the random distributions. Given that no other spatial patterns are apparent for PCCP. This proposed work suspects that these shorter segments are an artifact of the viewport positioning algorithm, which vaguely favored more focal areas of the image.

Image click ratio is measured based on the width and height in the original image. The height and width was changed according to the image was used in the application. In this proposed work we, measured the image click ratio for each and every image used in the project work. If we taken 5 images the image ratio is measured to five images. The viewport is situated randomly, rather than explicitly to avoid known hotspots, since such statistics might allow hackers to improve guesses and could lead to the formation of new hotspots. The viewports size is anticipated to proposal a variety of distinct points but still cover only an acceptably small fraction of all possible points. View port ratio is calculated based on the rectangle boundary was selected in the image; their corresponding height and width are selected measure the view port.

7. TASKS

Ease of login is the most frequently examined task, but is only one of many. Ideally, usability should be explored along several dimensions. For usability, essential elements to measure and report include: time to create a password, and time to login: memorability (typically through success rates and number of errors made during login over an extended period): and interference, by testing with a normal password load (as opposed to with only one password at a time).

Password Initialization

Authentication systems require initialization. A graphical password can either be assigned or user selected. In PCCP the middle ground between allowing user choice and system assigned passwords led to passwords nearly indistinguishable from random on the measures examined. Further works needed to evaluate the effect on long-term memorability.

Login

Login should be quick and simple since it is the most common task completed by users of an authentication system. Memorability issues are important when discussing login performance, as memorability is a main factor determining login success. Most graphical password studies to date required users to remember only one password at a time, whereas in real life users must remember many passwords and may get them confused. With authentication, interference occurs when remembering a password for one system impairs the users memory of a password for another system. This may be of particular concern with graphical passwords since exposure to similar images from multiple concurrent passwords or from password resets may aggravate the problem.

Password reset and password change

The tasks of resetting or changing passwords are not typically examined during usability testing of new graphical password schemes, but these are often required in practice when users forget passwords. The process may involve the user interacting only with the system, or may require contact with help desk personnel. System configuration and design of password reset and password change mechanisms can impact memorability, interference, and security of the system. For example, if the users are presented with the same, or similar, images as in previous graphical passwords, they may be more likely to confuse the memories of passwords or to reuse passwords. This suggests that reuse of password images should be avoided, and also argues against images being uploaded by users. Most authentication systems must allow password changes. The usability and security concerns are similar to password reset, except users can complete the task themselves without requiring a temporary password, entering their current graphical password as authentication.

Portable login

Unless restricted to specific environments users of graphical password systems may need to log in from different physical devices or locations. Usability issues to consider include whether the system is suitable for access from devises having different screen sizes or resolutions, and whether local bandwidth constraints impact performance. Moreover, portable login may require a modified login process or completion of additional tasks; these should also be considered and tested.

Shoulder-surfing

`Shoulder-surfing is targeted attack exacerbated by the visual aspect of graphical passwords. As users enter login information, an attacker may gain knowledge about their credentials by direct observation or external recording devices such as video cameras. High resolution cameras with telephoto lenses and surveillance equipment make shoulder-surfing a real concern if attackers target specific users and have access to their geographic location. Several existing graphical schemes believed to be resistant or immune to shoulder-surfing have significant usability drawbacks, usually in the time and effort required to log in, making them less suitable for everyday authentication. Multi-touch tabletop interfaces support novel approaches offering shoulder-surfing resistant properties. For some graphical

[passwords, multiple successful logins must be observed to deduce the full password. Passwords in other schemes can be recovered from one successful login.

8. SECURITY

An authentication system must provide adequate security for its intended environments; otherwise it fails to meet its primary goal. A proposed system should at minimum be evaluated against common attacks to determine if it satisfies security requirements. We classify the types of attacks on knowledge-based authentication into two general categories; guessing and capture attacks. In successful guessing attacks, attackers are able to either exhaustively search through the entire theoretically password space, or predict higher probability passwords so as to obtain an acceptable success rate within a manageable number of guesses.

Password capture attacks involve directly obtaining the password, or part thereof, by capturing login credentials when entered by the user. Shoulder-surfing, phishing, and some kinds of malware are common forms of capture attacks. In shoulder-surfing, credentials are captured by direct observation of the login process or through some external recording device such as a video camera. Phishing is a type of social engineering where users are tricked into entering their credentials at a fraudulent website recording user input. Malware uses unauthorized software on client computers or servers to capture keyboard, mouse, or screen output, which is then parsed to and login credentials..

The security can be achieved by reducing the hotspots and the shoulder surfing problem.

Hotspots: One of the main goals of this work is to prevent hotspot problem. For achieving this, we divide the image into block in the form of square matrix. The matrix size up to 6*6 so that we can get more blocks. Once an image is dividing into more blocks there is a less chance for hotspot issue in the generated block.

Table 2. Hotspot percentage for both IPCCP and PCCP

	In PCCP	In IPCCP
Probability of selected point to be a hotspot in percentage	13%	8%

Both PCCP and IPCCP through put are very good in the case of hotspot removal. But our proposed work gives some more good result.
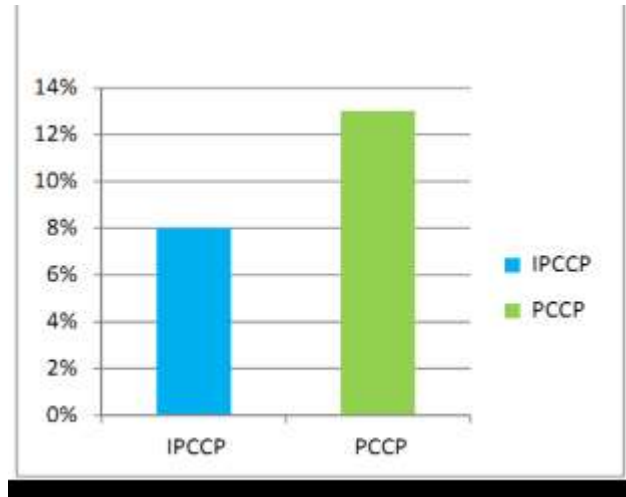
Figure 10. Comparison of hotspot occurrence percentage of two schemes.

Shoulder surfing problem: IPCCP provides more security by reducing shoulder surfing problem. Earlier in PCCP, it uses a single click method for selecting click points in the login phase. Where as in IPCCP, it uses both single click and double clicks method randomly for the selection of click points in login phase. So that it is very hard to predict the exact click point method used in the password.

9. RESULTS AND ANALYSIS

The study and analysis of normal point (CCP) and persuasive cued click point (PCCP) was designed to explore ways of increasing the efficiency of tolerance value and also conducted lab study for comparison between login success rate and security success rate of existing CCP's and proposed PCCP's.

Efficiency of the tolerance value

Initially eight participants are considered for the experiment. Each participant has a unique password which includes clicking on 5 click points in 5 different images. Each image contain of different characters, among which the participant needs to click on any one point of his choice to make it a click point in the series. Similarly the participant select a click point each of the images. Then, the participant logs in with that password, meantime the other participants are made to stand in a group behind the participant who is entering the password and are made to peek in over the shoulder of the participant and observe his password. The first participant has logged out once, the other [participants are asked to enter the same password which they have observed of the first participant.

Tolerance value: It is the value which defines the degree of closeness to the actual click point.

Tolerance region: The area describes an original click point is accepted as correct since it is unrealistic to expect user to accurately target an exact pixel.

Success rate: It is the rate which gives the names of successful trails for a certain number of trails, the success rates calculated as the number of trails given completed without errors or restarts.

Table 3. Results of tolerance value efficiency of the PCCP method

No	Tolerance value	Success rate	Percentage of success rate	Security (in percentage)
1	5	7/8	87.5	12.5
2	4	6/8	75.8	37.5
3	3	3/8	37.8	62.5
4	2	2/8	25	75
5	1	0/8	0	100

10. CONCLUSION

Published research in the area of graphical passwords currently lacks consistency, making it different to compare or reproduce results. A closer look at individual systems has typically revealed less security than promised; matching historical early experience in other areas usually repaired with maturity. The main purpose of authentication schemes is to allow system access only by legitimate users. To thoroughly evaluate the security of a graphical password proposal, and to facilitate comparison with alternatives, all standard threats and known attacks should be analyzed, with convincing arguments o how a scheme precludes them.

We expect tomorrow's ideal graphical password systems may have many of the following desirable characteristics, reecting lessons learned from proposals to date.

1. Theoretical password space meeting the security policy of the intended domain.

2. Avoidance of exploitable reductions in security due to user choice of passwords, e.g., through persuading password choice towards attar distributions.

3. At least mild resistance to different types of capture attacks including shoulder surfing and key logging, through variable response design.

4. Cues aiding memorability, design features minimizing password interference.

5. Usability as close as possible to, or better than, text passwords.

6. Implicit feedback to legitimate users, when passwords are multi-part.

7. Leveraging of pre-existing user specific knowledge where possible, rather than having users memorize entirely new and/or random information.

The major advantage of persuasive cued click point scheme is its later password and it helps in reducing number of hotspots in the image compared to existing click based graphical password systems. Therefore it provides better security. Randomness of the system is very high in comparison to both single image multi point based technique and multi image signal point based techniques. A common security goal is password based authentication systems is to maximize the effective password space. In this paper we have brought the color matrix formation method, so that even though the hotspots viewed by the neighboring user while entering the account will not be easy for the users to enter into the others user account. The key feature of PCCP is that creating a harder to guess password is the path of least resistance, likely making it more effective than schemes where secure behavior adds an extra burden on users. The approach has proven effective at reducing the formation of hotspots and patterns, thus increasing the effective password space.

11. REFERANCES

[1] S.Chiasson, R. Biddle, and P.van Oorschot, "A Second Look at the usability of Click-Based Graphical Passwords," Proc. ACM symp. Usable Privacy and Security (SOUPS), July 2007.

[2] S.Chiasson, A.Forget, R. Biddle, and P.van Oorschot, "Influencing Users towards Better Passwords: Persuasive Cued Click-Points," British HCI Group Ann. Conf. People and Computers: Culture, Creativity, Interaction, Sept. 2008.

[3] S.Chiasson, A. Forget, E.Stobert, P.van Oorachot, and R.Biddle, "Multiple Password Interference in text and Click-Based Graphical Passwords," Proc. ACM Conf. Computer and Comm. Security CCS, Nov.2009.

[4] P.C.van Oorschot and J.Thorpe, "Exploiting Predictability in Click-Based Graphical Passwords," J.Computer Security, vol.19, no.4, pp.669-702, 2011.

[5] Sonia Chiasson, Elizabeth Stobert, Alain Forget, Robert Biddle, and Paul C.van Oorschot "Persuasive Cued Click Points: Design, Implementation, and Evaluation of a Knowledge Based authentication Mechanism", IEEE Transactions on Dependable and Secure Computing, Vol.9, No.2, March/April 2012.

[6] R.Biddle, S.Chiasson, and P.van Oorschot, "Graphical Passwords: Learning from the First Twelve Years," to be published in ACM Computing Surveys, vol.44, no.4, 2012.

[7] E.Stobert, A.Forget, S.Chiason, P.van Oorschot, and R.Biddle, "Exploring Usability effects of Increasing security in Click-Based Graphical Passwords," Proc. ANN. Computer Security Applications Conf. (ACSAC), 2010.

[8] A.E. Dirik, N.Memon, and J-C. Birget, "Modeling user choice in the passpoints graphical password scheme," in Proc. Symp. Usable Privacy Security, 2007, pp.20-28.

[9] (2010, Feb.). The Science behind Passfaces [Online].

Available:http://www.realuser.com/published/SceienceBehindPassfaces.pdf

AUTHORS PROFILE

Y.Dasradh Ram Reddy working as Associate Professor in C.S.E department BVSR Engineering College, Chimakurthy, A.P, India. He has done his M.Tech and PhD in Computer Science Engineering. He has around 12 years of teaching experience for UG and PG students.

Y.Sravana Lakshmi is currently studying M.Tech in C.S.E in BVSR Engineering College, Chimakurthy, A.P, and India.

Strategic Management of Intellectual Property: R&D Investment Appraisal Using Dynamic Bayesian Network

L . O. Nwobodo

Enugu State University of Science and

Technology

Enugu State, Nigeria

H.C. Inyiama

Nnamdi Azikiwe University[NAU]

Awka, Anambra State, Nigeria

Abstract :
Company executives are under increasing pressure to proactively evaluate the benefits of the huge amounts of investment into intellectual property (IP). The main goal of this paper is to propose a Dynamic Bayesian Network as a tool for modeling the forecast of the distribution of Research and Development (R&D) investment efficiency towards the strategic management of IP. Dynamic Bayesian Network provides a framework for handling the uncertainties and impression in the qualitative and quantitative data that impact the effectiveness and efficiency of investments on R&D. This paper specifies the process of creating the graphical representation using impactful variables, specifying numerical link between the variables and drawing inference from the network.

KEY TERMS: IP Dynamic Bayesian Network, R&D Investment Efficiency.

1 INTRODUCTION

Intellectual property (IP), which is now being regarded [1][2] as constituting the core assets of companies, is centered around innovation. The very basis for innovation is developing new products and services, which logically means that more IP will be generated and thus increases the need for protection [3].

The research and development (R&D) Investment of major manufactures has reached an annual level of several billion dollars [4]. Questions are being raised as to whether this investment amount is efficient from the perspective of its effectiveness, such as whether such investment is made efficiently and whether the results of R&D are appropriately contributing to company profits.

Essentially, activities for R&D and intellectual properties must be dealt with in the management cycle shown in figure 1. It is important to realize that R&D strategy and the IP strategy exactly represent the business strategy of a company. There are many

companies that spend amounts for R&D that exceed their amount of capital investment. In response to these moves, institutional investors have started to show an interest in the content of R&D investment and the effectiveness and efficiency of such investment. However, due to uncertainties and imprecise data, it is extremely difficult to proactively demonstrate explicitly the efficiency and effectiveness of R&D investment.

In an automated Bayesian Network decision support system, the probability judgment analysis make over time can be captured, compared to actual data when it becomes available, and then provide feedback on a timely basis [5]. Based on the work by AIpart and Raiffa [6], this feedback should improve analysts' probability assessment which should lead to improved future performance.

[7] observed: "whether or not they recognize it, virtually all decisions that investors make are exercises in probability: For them to succeed, it is critical that their probability statement combines the

historical record with the most recent data variable. And that is Bayesian analysis in action".

Hence this paper proposes a Bayesian Network model for the forest of the distribution of R&D investment efficiency.
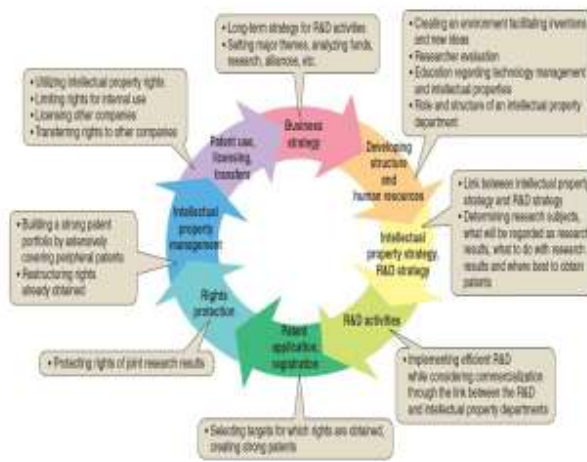
Figure 1: Ideal management cycle for activities for R&D and intellectual properties [4]

2. BAYESIAN NETWORKS

Bayesian networks provide a formalism for reasoning about partial beliefs under conditions of uncertainty. These parameters are combined and manipulated according to the rules of probability theory [8][9]. Let us consider n discrete random variables x_1, x_2---- x_n, a discrete acyclic graph with n nodes, and suppose the jth node of the graph is associated to the x_j variable. Then the graph is a Bayesian network, representing the variables $x_1,x_2 \ldots \ldots x_n$, if

$$P(x_1, x_2, --------x_n) = \pi \, p(x_j / parents(x_j)),$$

j

Where parents (x_j) denotes the set of all variables X_i such that there is an arc from node x_i to x_j in the graph. The probability terms in the product are described by conditional probability Tables (CPT) which may be set by hand or learned from data. Standard algorithms such as junction tree [9] [10] exist to perform inference networks.

Dynamic Bayesian Networks (DBNs) allow the modeling of entities in a changing environment where the values of variables change over time [10] [11]. Functionally, DBNs capture the process of variable values changing over time by representing multiple copies of network modes with one copy for each time step [10]. Visually, they may be displayed using two copies of each recurrent node representing the current t, and previous, t-1 states. A Bayesian network is a tool to help expert represent uncertainty, ambiguous or incomplete knowledge. Bayesian networks use probability theory to represent uncertain knowledge. A Bayesian network consists of two parts-a qualitative graphical structure of the relationships in the model and a quantitative structure represented by the probability distributions that are indicated by the graph. In a Bayesian network, historical information can provide a framework or baseline model to develop prior distribution. New quantitative information, qualitative information, or evidence can be added to the network as appropriate to develop posterior probabilities.

Decision makers in many different contexts combine qualitative data and qualitative information. Bayesian networks have been applied in a wide variety of decision-making contexts. Some examples are venture capital financing [12], auditing [13], medical diagnosis [14], and software design [15], among many.

3 DEVELOPING THE BAYESIAN NETWORK

There are three important steps in building the Bayesian network. The first step is developing the graphical model. This step includes the relevant variables and specifying whether they are independent, or not. The second step is the specification of the numerical relationship between the variables that are not independent. The third step is making inferences or decisions based on new evidence.

3.1 Graphical Representation

As mentioned in the previous section, the first step in construction the Bayesian network is the graphical model. The graphical model is a directed, acyclic graph were nodes represent variables and directed arc (arrows) represent the conditional probability relationship assumed in the model. The variables and the arcs between the variables are the inputs to the graph.

In this paper, as part of R&D strategies, the forecast of R&D investment efficiency is to be made using the variables: infrastructure, capability of science and technology, Risk, R&D efficiency management performance, total R&D personal capability, total R&D expenditure. These are variables that are assumed by management (for the sake of analysis) that impacts Return on Investment (ROI) on IP R&D.

The relationship among the IP variables maybe represented by the Bayesian network of figure 2. The network consists of four discrete variables: Capability of Science and Technology (CST), Risk (RSK), R&D Efficiency (RDE), Management

Performance (MP), Total R&D Personnel Capability (PC),Total R&D Expenditure (RDEP), and Investment Efficiency (IE).
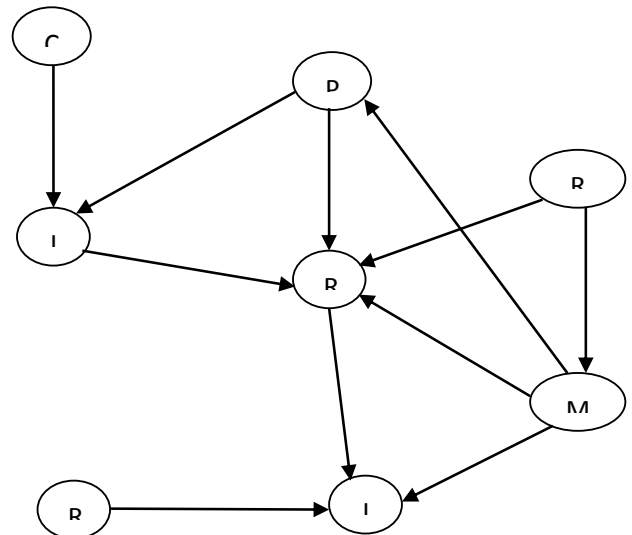
Figure 2: Bayesian Network for forecasting R&D Investment efficiency.

In the network, a node with arcs leading out only indicates a marginal probability distribution. For example CST, RDEXP, RSK are marginal probability distribution. A node with arcs leading into it indicate a conditional relationship. For example, the factors (variables) that determine the value for infrastructure (IF) is the capability of science and technology (CST) and total R&D Personnel Capability (PC). The node infrastructure (IF) is a conditional node.

The absence of a directed arc from a node is also meaningfully because the absence indicates independence assumption. The absence of a directed arc denotes conditional independence between nodes. Thus the lack of an arc from CST to RSK signifies that it is assumed that RSK is independent of CST;

the lack of an arc from IF to MP signifies that it is assumed that MP is independent of IF.

The analyst benefits in several ways from the graphical construction of a Bayesian network. First, the construction of the graphical portion of the network helps the analyst clarify and refine his view of the relationships among the variables [5] Next, the analyst may not always have a good understanding of how a decision is reached. They fully understand which variables are used, but how the variables are combined and heighted to come up with a decision is not always well understood or systematic [5]. In [16], almost all analyst agree that qualitative information was important, but when questioned about how it was incorporated in a decision, most analyst could not be specific.

3.2 Determining Numerical links between variables

Each node in the Bayesian network is a variable that is described either as a constant value, a probability distribution, or as a function of other variables [17]. In a Bayesian network, the primary focus is on determining the probability distribution of the relevant nodes. A Bayesian network model is represented at two levels, qualitative and quantitative. At the qualitative level a directed graph is used (as done here in figure 2) in which nodes represent variables and directed arcs describe the conditional independence relations embedded in the model. At the level, conditional probability distributions are specified for each variable in the network. Each variable has a set of possible values called its state space that consists of mutually exclusive and exhaustive values of the variable.

There are two primary ways to find probability distribution, for the nodes in the network. One way is historical data. The other is to use subjective probability judgments. The two methods can also be combined.

For the model in this paper capability science and technology (CST) is specified as having two states: "High" and "Low; infrastructure (IF) has two states: "Good" and "Bad", total R&D Personnel Capability (PC) has two states: "High" and "Low", total R&D Expenditure (RDEXP) has two states: "High" and "Low"; Management Performance (MP) has two states: "Good" and "Bad"; R&D efficiency (RDE) has two stets: "High" and "Low"; Risk (RSK) has two states: "High" and Low" and investment efficiency (IE) has two states: "High" and "Low".

A fundamental assumption of a Bayesian network is that when the conditionals for each variable is multiplied, the joint probability distribution for all variables in the network is obtained.

Suppose a sequence of the variables in a Bayesian network is picked such that for all directed arcs in the network, the variable at the tail of each arc precedes the variable at the head of the arc in the sequence. Since the directed graph is acyclic, there always exist one such sequence. In figure 1, one such sequence is CST IF RDE MP RSK IE.

From figure 2, this shows that the model makes the assumption:

$P(CST, IF, RDE, PC, RDEXP, MP, RSK, IE) =$

$P(CST) \otimes P(IF/CST,PC)$ $P(RDE/IF, PC, RDEXP, MP)$ $P(PC/MP \otimes)$ $P(RDEXP)$ $\otimes$

P(MP/RDEXP) $\otimes$ P(RSK)$\otimes$

P(IE/RDE,MP,RSK),

Where $\otimes$ denotes point wise multiplication of tables.

For each variable, a table of conditional probability distribution has to be specified, one for each configuration of states of its parents.

For the model, tables 1(a) – (h) gives the tables of conditional distributions-P(CST), P(IF/CST,PC) P(RDE/IF, PC,RDEXP,MP), P(PC/MP), P(RDEXP), P(MP/RDEXP), P(RSK), P(IE/RDE,MP,RSK).

(a)

P(CST)	High Low
	0.25 0.75

(b)

P(IF/CST,)PC	Good Bad
High, High	0.80 0.2
High, Low	0.06 0.40
Low, High	0.25 0.75
Low, Low	0.50 0.50
High, High	0.01 0.90

(c)

P(RDE/IF, PC,RDEXP,MP)	High Low
High, High, High, Good	0.85 0.35
High, High, High, Bad	0.60 0.40
High, High, Low, Good	0.80 0.20
High, High, Low, Bad	0.55 0.45
High, Low, High, Good	0.67 0.33
High, Low, High, Bad	0.70 0.30
High, Low, Low, Good	0.68 0.32
High, Low, Low, Bad	0.66 0.34
Low, High, High, Good	0.50 0.50
Low, High, High, Bad	0.49 0.51
Low, High, Low, Good	0.40 0.60
Low, High, Low, Bad	0.75 0.25
Low, Low, High, Good	0.30 0.70
Low, Low, High, Bad	0.20 0.80
Low, Low, Low, Good	0.10 0.90
Low, Low, Low, Bad	0.25 0.75

(d)

P(PC/MP)	High Low
Good	0.15 0.85
Bad	0.70 0.30

(e)

P(RDEXP)	High Low
	0.65 0.35

(f)

P(MP/RDEXP)	Good Bad
High	0.10 0.90
Low	0.70 0.30

(g)

P(RSK)	High Low
	0.70 0.30

Table 1: Tables of conditional probabilities for the Bayesian network of figure 2.

P(IE/RDE, MP, RSK)	High Low
High, Good High,	0.90 0.10
High, Good, Low	0.85 0.15
High, Bad, High	0.02 0.98
High, Bad, Low	0.65 0.35
Low, Good, High	0.70 0.30
Low, Good Low	0.55 0.45
Low, Bad, High	0.5 0.5
Low, Bad, Low	0.1 0.9

3.3 Making inference

The ultimate goal is to model the probability distribution of the investment efficiency (IE) for IP portfolio.

Once a Bayesian network is constructed, it can be used to make inferences about the variables in the model. The conditionals given in Bayesian network representation specify the prior joint distribution of the variables. If the values of some are observed (or learnt), then such observations can be represented by tables where 1 is assigned for observed values and 0 for unobserved values. Then the product of all tables (conditionals and observations) gives the posterior joins distribution of the variables. Thus the joint distribution of variables changes each time new information is learnt about item.

Often the interest is on some target variables. In this case, inference is made by computing the marginal of the posterior joint distributions for the variables of interest. Consider the situation described by the Bayesian network in figure1. The interest is in the true state of the R&D investment efficiency. Given the prior model (as per the probability tables given): table 1(a) – (b)), the marginal distribution is computed (giving probability values for "High" and "Low"). Now suppose it is learnt that (i.e new observation) risk (RSK) is "Low" and management performance (MP) is "Good". The posterior marginal distribution of R&D investment efficiency changes.

4 CONCLUSION

The main goal of this paper is to propose a Bayesian network as a tool for modeling the forecast of the distribution of R&D investment efficiency towards the strategic management of IP.

The improvement of investment efficiency by reviewing R&D and the reorganization of a business portfolio have become urgent issue to manage. Many companies are realizing that if they are going to spend any money on IP, it better be IP that has value to the business. So companies are now developing strategic plans as to (a) where they want inventions (b) what

resultant IP they want to create and (c) the efficiency of their R&D investment. Company executives now face a variety of opportunities that require sophisticated analysis and decisions. These requirements are better met by advanced decision support tools.

With the capability to model the forecast of the efficiency of R&D investment given different scenarios of the combination of resources (R&D personnel, management performance, R&D budget), risk and market demand, companies are able to make effective decisions regarding their IP portfolio. This enables companies to evaluate the business benefit of any IP before its creation. This means for example they can strengthen the few patents they will file by focusing on "inventing around their own IP" before filing it. An advanced analytic tool, such as the Bayesian network, helps to work the process of inventing around" to systematic and robust.

There are many off-the-shelf software systems that allow Bayesian networks to be constructed graphically by end-users, for example Bayesialab (www.bayesi.com), Netica (www.norsys.com) and Hugin(www.hugin.com). These tools allow the user to enter the graph and specify numerical relationships among the variables. The software in use calculates the inference based on these inputs. The inference results are shown graphically as probability distribution for the network. The analysis would help executives make better strategic decision regarding their IP portfolio.

5. REFERENCES

[1] Dino Isa, Pter Blanchfield, Zh.yuan 2009"Intellectual Property Management System for the super-capacitor Pilot plant "workshop on advances on advances intelligent computing.

[2] Benintend, S. 2003 "Intellectual property valuation one important step in a successful asset management system "Payton: University of Dayton School of Law, (2003) PP 12, 14; 16-20

[3] John Cronin 2010, "The cause for Developing IP an "Executable" IP strategy in 2010" IP Capital Group, Inc, PP 1.

[4] Masayuki Miyake, Yuji Mune, and Keiichi Himeno, Dec 2004 "Strategic Intellectual Property Management: Technology Appraisal by using the "Technology Heetmop", NRI papers No. 83.

[5] Riza Demiser, Roreld R. Mom, Catherine Shenoy, June, 2005, "Bayesian Networks. A Decision Tool to improve portfolio Risk Analysis", PP 9-11 .

[6] Alpert, M., and H. Raiffa, 1982 "A Progress Report on the Training of Probability Assession. Judgment under uncertainty: Heuristics and Biases. ed. D. Kahneman, P. Slovic, A. Tversky. New York: Cambridge University Press.

[7] Hagstrom, R.G. 1999 "The Warren Buffett Portfolio. Wiley and Sons, New York.

[8] Pearl .J., 1988 "probability reasoning in Intelligent systems: Networks of plausible Inference. Morgan Kaufmann Publishers, 2nd edition on.

[9] Mark Taylor, Charles fox "Inventory Management with Dynamic Bayesian Network Software systems" Adaptive Behavior Research Corp, Sheffield, UK.

[10] Arbib M.A, editor 2003 "The handbook of brain theory and neural networks, chapter Bayesian networks. Press.

[11] K.P. Murphy 2002 "Dynamic Bayesian networks: Representation, Inference and learning "Technical report, University of California, Berkeley.

[12] Kemmeser, B.S. Mishra, and PP. Shenoy 2001 "Bayesian Casual Maps as Decision Aids in venture capital decision making: Methods and Applications" University of kensas, working Paper.

[13] Gillete, P and R.P. Sriverstava 2000. "Attribute Sampling: A Belief Function Approach to Statically Audit Evidence". Auditing: A journal of practice and theory vol.19, no 1. (Spring) 145 – 155.

[14] Bielza, C., S. Rios-Insua, and M. Goniez 1999 "Influence Diagrams for Neorata Jaundice

management" In lecturer notes in Artificial Intelligence 1620, (ed. Werror Horn, Yural Shahar, G. Lindberge, Steen Andresen, J. Wyatt) springing-Verley Berlin Heride/berg 1999, 138-142.

[15] Horvitz, E., J. Breese, D. Heckerman, D. Hovel, and K. Rommelse 1998. "The Lumiere Project: Bayesian user modeling for Inferring the Goals and Needs of software Users" proceeding of the fourteenth conference on uncertainty in Artifical intelligence, July.

[16] Catherine Slonoy, Pracash P. Sloney 1998. "Bayesian network models of portfolio risc and return", School of business working paper No.279.

[17] Slonoy 2001. " A Description of Security Analysis" Decision making process." University of Kansas, working paper.

Sampling of User Behavior Using Online Social Network

J.Amarnath
Department of Computer Science and Engineering
BVSR Engineering College
Chimakurthy, India

Abstract: The popularity of online networks provides an opportunity to study the characteristics of online social network graphs is important, both to improve current systems and to design new application of online social networks. Although personalized search has been proposed for many years and many personalization strategies have been investigated, it is still unclear whether personalization is consistently effective on different queries for different users, and under different search contexts. In this paper, we study performance of information collection in a dynamic social network. By analyzing the results, we reveal that personalized search has significant improvement over common web search.

The mixing time of thee sampling process strongly depends on the characteristics of the graph.

Keywords: Sampling methods, Social network services, Random walks, Multi graph, Graph sampling.

1. INTRODUCTION

The Internet has spawned different types of information sharing systems, including the web. Unlike the web, which is largely organized around content, online social networks are organized around users. Participating users join a network, publish their profile and any content, and create links to any other users with whom they associate. The resulting social network provides a basis for maintaining social relationships, for finding users with similar interests, and for locating content and knowledge that has been contributed or endorsed by other users.

An in-depth understanding of the graph structure of online social networks is necessary to evaluate current systems, to design future online social network based systems, and to understand the impact of online social networks on the internet. For example, understanding the structure of online social networks might lead to the can detect trusted or influential users, much like the study of the web graph led to the discovery of algorithms for finding authoritative sources in the web. Moreover, resent search, and to defend against Sybil attacks However, these systems have not yet been evaluated on real social networks at scale, and little is known to date on how to synthesize realistic social network graphs.

We observe a high degree of reciprocity in directed user links, leading to a strong correlation between user indegree and outdegree. This differs from content graphs like the graph formed by web hyperlinks, where the popular pages (authorities) and the pages with many references (hubs) are distinct. We find that online social networks contain a large, strongly connected core of high-degree nodes, surrounded by many small clusters of low-degree nodes. This suggests that high-degree nodes in the core are critical for the connectivity and the flow of information in these networks.

Online social networks have existed since the beginning of the internet. For instance, the graph formed by email users who exchange messages with each other form an online social network. However, it has been difficult to study this network at large scale due to its distributed nature.

There are many ways a graph can be sampled, e.g., by sampling nodes, edges, paths, or other substructures. Depending on our measurement goal, the elements with different properties many have different importance and should be sampled with a different probability. For example, fig. 1(a) depicts the world's population, with residents of china (1.3B people) represented by blue nodes, of the Vatican (800 people)by block nodes, and all other nationalities represented by white nodes. Assume that we want to compare the median income in china and Vatican. Taking a uniform samples will come from counties other than China and Vatican. Even restricting our sample to the union of china and Vatican will not help much, as our sample is unlikely to include any Vatican resident. In contrast, uniformly sampling 50 Chinese and 50 Vaticanese residents would be much more accurate with the same sampling budget.

(a) Population (b) WIS weights_ WIS (c) social graph G (d) _ WIS applied to WRW (e) S_WRW weight
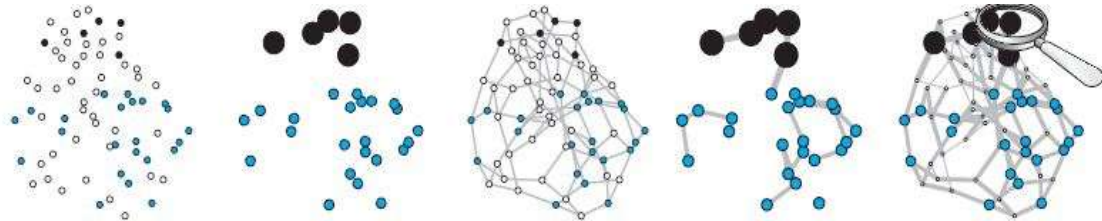
Figure 1. Illustrative example. Our goal is to compare the blue and black subpopulations (e.g., with respect to their median income) in population (a). Optimal independence sampler, WIS (b), over-samples the black nodes, under-samples the blue nodes, and completely skips the white nodes. A naive crawling approach, RW (c), samples many irrelevant white nodes. WRW that enforces WIS-optimal probabilities many result in poor or no convergence (d). S-WRW (e) strikes a balance between the optimality of WIS and fast convergence.

Despite the fact that many algorithms and tools exist for analysis of networks, in general, these mainly focus on analysis of the properties of the network structure and not on the content of the nodes. They also typically not operate on user specific graphs (i.e, users' neighborhoods), but on the whole graph. Instead, for many modern applications, it would be beneficial to design algorithms that operate on a single node. For example, in the case of social search, it would be beneficial to design algorithms that starting from a specific user in the works may consist of thousands of users and their structure may not be static, thus a complete crawl of all social peers is infeasible.

Therefore, efficient methods are required. Based on these observations we focus on improving the performance of information collection from the neighborhood of a user in a social network and make the following contributions:

We introduce sampling-based algorithms that given a user in a social network quickly obtain a near-uniform random sample of nodes in its neighborhood. We employ these algorithms to quickly approximate the number of users in a user's neighborhood that have endorsed an item. We introduce and analyze variants of these basic sampling schemes in which we aim to minimize the total number of nodes in the network visited by exploring correlation across samples

We evaluate our sampling-based algorithms in terms of accuracy and efficiency using real and synthetic data and demonstrate the utility of our approach. We show that our basic sampling schemes can be utilized for a variety of strategies aiming to rank items in a network, assuming that information for each user in the network is available.

These third party aggregators, which we refer to crawlers, represent a significant problem for OSN site operators as well. User data provides OSN operators with a revenue stream (e.g., via targeted advertisements); stopping crawlers is therefore in the OSN operators' business interests. Additionally, OSN operators cannot ensure that data collected by third party is used according to the operator's privacy policy. Yet, OSN operators are likely to be held responsible if crawled data is used in ways that violate the policy, at least in the court of public opinion. For example, Facebook was widely blamed in the popular press when a single crawler gathered public profiles of over 100 million users. Thus, OSN operators need effective mechanisms to thwart large-scale crawling of OSN sites.

A day maintaining a large database is naturally a major problem for just about any organization. Keeping large database includes the issues of operational costs, analytical costs as well as a threat of going to a wrong conclusion by considering the inconsistent data. Databases today can range in sizes straight into the terabytes of a little more than 1,000,000,000,000 bytes hard drive data. Within these masses of data lies hidden information of strategic importance i.e., data rich information poor situation. When there can be so many trees, tips on how to draw meaningful conclusions regarding the forest? The latest response is data mining, that's being utilized both to extend revenues and also to reduce costs. Data mining can refer to extracting knowledge from a lot hard drive data recovery. It is actually the automated extraction of hidden knowledge from large volumes of raw data.

2. BACKGROUND AND MOTIVATION

We begin with a brief overview of online social networks. We then describe a simple experiment we conducted to estimate how often the links between users are used to locate content in a social networking site like Flickr. Finally, we discuss the importance of understanding the structure of online social networks.

Online social networks have existed since the beginning of the Internet. For instance, the graph formed by email users who exchange messages with each other forms an online social network. However, it has been difficult to study this network at large scale due to its distributed nature. Popular online social networking sites like Flickr, You-Tube, Orkut, and LiveJournal rely on an explicit user graph to organize, locate, and share content as well s contacts.

In many of these sites, links between users are public and can be crawled automatically to capture and study a large fraction of the connected user graph. These sites present an opportunity to measure and study online social networks at a large scale.

3. LITERATURE SURVEY

[1] M. Richardson and R.W. White have investigates by predicting the variety of outcomes during the question life cycle. Here they used Synchronous social Q&A system which provides the searches with an additional mechanism beyond the general-purpose search engines through which to get questions answered. They also used the data from an instant messaging for automatically identifies candidate's answers by ranking all users by representation of their interest and route the asker and answerer. Using this we found that they only gathered the interaction data from the real users.

[2]. Horowitz and S.D. Kamvar are presented Aardvark which is used to find the right person to satisfy the user information need. They also find the vast majority of questions get answered promptly and satisfactorily so the users also surprisingly active for both in asking and answering. The questions include the background details but it is not obvious how to translate these information needs in to keyword searches.

[3] M.R. Morris, J. Teevan, and K. Panovich have presented the detailed data from a survey of Facebook and Twitter users on the topic of asking and answering questions via social status messages and those responses are very are valuable. This approach would be most effective if the speed of receiving social answer could be optimized. Survey participants appreciated the context inherent in interactions with their social network. This finding suggests that enhancing the effectiveness of personalized search algorithms is an important area of further study for improving the usability of search engine.

[4] J. Raacke and j. Bonds-Raacke have studied to identify the dimensions of use for friend networking sites that including the information dimension, the friendship dimension, and the connection dimension. The reasons in the information dimension were related to using friend networking sites to gather and share information. The reason in the friendship dimensions was conceptually related to gather and share information. The reason in the friendship dimensions were conceptually related to sustaining friendship. Then the reason in the connection dimensions that were related to making the connection with others and this dimension included to make new friends, to find a significant other, and to feel connected in general.

[5] M.R. Morris, J. Teevan, and K. Panovich have conducted a study to compare social and non-social search for complex, self-motivated information seeking tasks. In this paper they focused on the aspects of social search where the searchers ask a question to the group of people they personally by means of a social search where the searches ask a question to the group of people they personally by means of a social network status message update. They compared this experienced to the searching for the same information with a web search engine. They found an average of 1.4 responses waiting, and a maximum of 5 responses and 5 participants received no responses compared to the social search.

[6] D. Carmel, N. Zwerdling, I.Guy, S., and Ofek-Koifman, N. Har'el, I.Ronen,E. Uziel, S. Yogev, and S. Chernov have investigated the personalized social search based on the user's social relations which provide the effectiveness of Familiarity and overall based network. Their assumption behind this personalization approach is that the preferences to other people, who are expected to have "similar" interests as the searcher and assist the search results that might subjectively, satisfied the searcher needs.

4. SOCIAL CONTENT WORKLOADS

OSN have changed the way content is shared on the Internet. In this section we study the characteristics of OSN content using real traces gathered from popular OSN sites, and focus on the differences between OSN and the traditional web content.

4.1 Datasets

We implemented a web crawler for flicker.com and youtube.com, which are popular sites that allow people to share content with their friends. Our crawler gathered detailed information about the uploads and downloads of publicly available content from these sites.

Flicker: We randomly chose 11,715 users from the list of 2.5 million users gathered by [4]. We crawled the pages of these users daily for 19 consecutive days. In total, these users had uploaded 1,324,080 publicly accessible photos. For each photo, we recorded the number of daily views received, as well as metadata, like photo size, tags, and favorite markings.

YouTube: We randomly chose 77,575 users from the list of YouTube users gathered by [3]. We collected information about the videos uploaded by these users. In total, these users had uploaded 1,251,492 publicity accessible videos. We collected the number of daily views for all of these videos over a period of166 days using the "StatisticsAndData" feature in YouTube.

Ideally we would have liked to include data from an OSN site like Facebook, but obtaining data from such sites is hard because most of the shared content is private. In contrast, allthe data we gathered from Flicker and YouTube is publicly accessible. Furthermore, these two sites provide mechanisms for searching and featuring popular content. Hence, our analysis of content consumption patterns is likely to overestimate the popularity that content would have reached had it been shared on an OSN site like Facebook. On the other hand, the content production patterns are likely to be similar to the ones in Facebook.
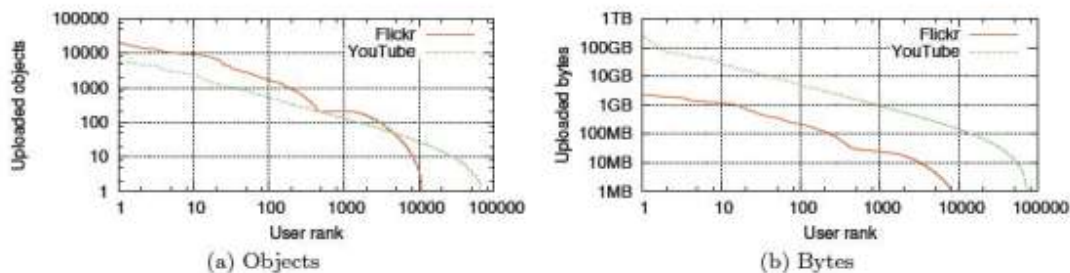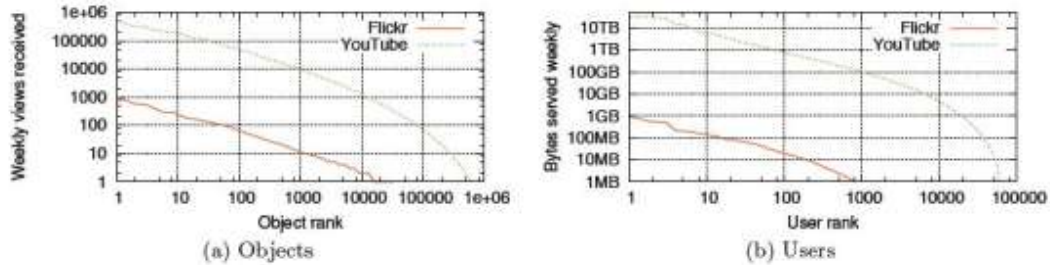

Figure 2. Content production patterns: (a) users ranked by number of uploaded objects and (b) by the total size of uploaded content.

Figure 3. Content consumption patterns: (a) objects ranked by the number of weekly requests and (b) the total amount of content served by Flicker and YouTube on behalf of content up loaders.

4.2 Stratified Crawling

Due to lack of a sampling frame, the independence sampling (include its stratified version) is typically infeasible in online networks, making crawling the only practical alternative. In this section, we show how to perform a weighted random walk (WRW) which approximates the stratified sampling. The general problem can be stated as follows:

Given a category-related measurement objective, an error metric and a sampling budget $|S|=n$, set the edge weights in graph G such that WRW on this graph achieves a minimal estimation error. Although we are able to solve this problem analytically for some specific and fully known topologies, it is not obvious how to address it in general, especially under a limited knowledge of G. Instead, in this paper, we propose S-WRW, a heuristic to set the edge weights. S-WRW starts from a solution that is optimal under WIS, and takes into account practical issues that arise in graph crawling. Once the edge weights in G are set, we simply perform a WRW and we collect node samples.

5. PROPOSED SYSTEM

Perfuse introduces abstractions for filtering source data into visualizable content, providing both scalability and representational flexibility, and using compassable actions to perform batch processing of this content, for example data transformation, layout. Or color assignment. Programmers craft visualizations by stringing together actions in to executable chains that can then be run to manipulate visual data and perform animation.

Interactive views are then created from this visual data through a highly configurable rendering system, to which pre-built controls can be added to specify interactive behaviors. This separation of concerns provides a degree of flexibility unmatched by existing infovis toolkits, supporting multiple views, semantic zooming, data and visual transformations, and application extension and customization. Perfuse further demonstrates that these generalized abstractions can be provided without unduly sacrificing performance.

Best Network Spilt

1.   Find edge between of all edges in network.
2.   Find vertex between of vertices, from edge between, using Eq. (1).
3.   Find candidate set of vertices: those whose vertex between is greater than the maximum edge between.

4.   If candidate set is non-empty, find pair between of candidate vertices, and then find split between of candidate vertices, using Eq.(1).
5.   Remove edge with maximum edge between or split vertex with maximum split between (if greater).
6.   ReFind edge between for all remaining edges in same components(S) as removed edge or split vertex.
7.   Repeat from step 2 until no edges remain.

$$c_B(v) = \frac{1}{2} \sum_{e \in \Gamma(v)} c_B(e) - (n-1)$$

$$\tag{1}$$

The split between of a vertex v is the number of shortest paths passing between any member of n1 and any member of n2 via v, where n1 and n2 are disjoint sets containing all neighbors of v.
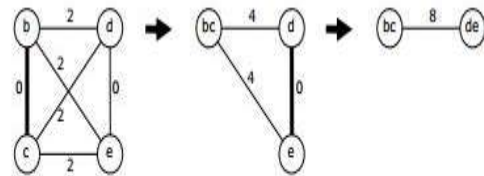
Finding the best split of vertex a

Sampling via Crawling

In contrast to independence sampling, the crawling techniques are possible in many online networks, and are therefore the main focus of this paper.

Simple Random Walk (RW) [30] selects the next-hope node v uniformly at random among the neighbors of the current node u. In a connected and a periodic graph, the probability of being at the particular node v converges to the stationary distribution

$$\pi^{RW}(v) = \frac{\deg(v)}{2 \cdot |E|}.$$

Metropolis-Hastings random walk (MHRW) is an application of the metropolis-Hastings Algorithm [31] that modifies the transition probabilities to converge to a desired stationary distribution. For example, we can achieve the uniform stationary distribution
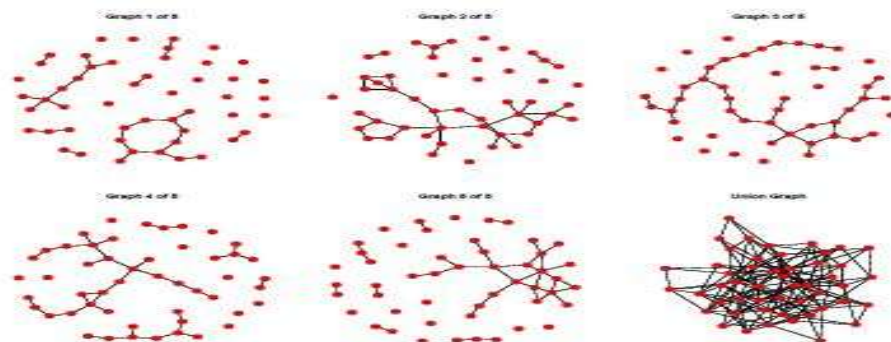
$$\pi^{MHRW}(v) = \frac{1}{N}$$

6. SAMPLING METHODOLOGY

Terminology and Definitions

We consider different sets of edges E= {E1,.....,EQ} on a common set of users V. each Ei captures a symmetric relation between users, such as friendship or group co-membership. (V,Ei) thus defines an undirected graph Gi on V. We make no assumptions of connectivity or other special properties of each Gi. Fig below shows an example of Q=3 different relations and relations graphs Gi defined on the same 5 nodes.

Consider set of graphs Gi=(V,Ei), i=1,…,Q, defined on a common node set V. E can be used to construct several types of combined structures on V. We will employ the following two such structures:

Definition 1: The union (simple) graph G' = (V,E') of G1,…,GQ is defined as the graph on V, whose edges are given by the set E' = UQ i= 1 Ei.

Definition 2: The union multigraph G= (V,E) of G1,….GQ is defined as the multigraph on V, whose edges are given by the multiset E = UQ i=1 Ei.

Note that the union multigraph G can contain multiple edges between a pair of nodes, while the union graphs G' contains only one edge. Every multigraph G can be reduced to the union graph G' by merging together multiple edges between two nodes into one. Our focus, in this paper, is on the union multigraph, also referred to as simply the multigraph; because it allows us to more efficiently implement sampling on multiple relations. However, we also the union graph as a helpful conceptual tool.

6.1 Some False Starts

We seek to draw a sample of the nodes in V, so that the draws are independent and the sampling probability of each node is known up to a constant of proportionality. There are several ways to achieve this goal using multiple graphs. We discuss some of them below.

1)   Naïve Multiple Graph Sampling: A naïve way is to run many random walks, one per each individual graph Gi, and to combine the collected samples. However, if a particular Gi is disconnected a walk

ALGORITHM 1: MULTIGRAPH SAMPLING ALGORITHM
Require: v0 €V, simple graphs Gi, i=1….Q
1: Initialize v ← v0.
2: **while** NOT CONVERGED **do**
3: Select graph Gi with probability di(v)
d(v)
4: Select uniformly at random a neighbor v'of v in Gi
5: v ← v'
6: **end while**
7: **return** all sampled nodes v and their degrees d(v).

ALGORITHM 2: BEST ANSWERER IDENTIFICATION ALGORITHM
The pseudo code of the best answerer identification executed by node i.
Input: IDi, IDj, Q(i,j) (j £Fi)
Output: top-K best answerers
For each friend j in friend list Fi **do**
Update Q(i,j)
end for
If create a question or receive a question it cannot answer **then**
If TTL > 0 **then**
For each friend j in friend list Fi **do**
Calculate s(qi,j) using IDqi and IDj
Calculate BA(I,j) using Q(i,j) and s(q,j)
Add BA(I,j) to a list List

end for
Quick Sort partition around the Kth largest element in List
Find the top-K friends having the highest BA(I,j)
TTL-=1
Send the question to the identified K friends
end if
end if
if does not receive answers for its created question during the time corresponding to TTL
then
Resort to the centralized server for the answers
end if

This algorithm shows the pseudo code of the process for the best answerer metric calculation and best answerer selection conducted by node i. If node I does not receive answers for its created question during the time corresponding to TTL, it resorts to the centralized server for the answers, where all users conduct Q&A activities as in online Q&A sites. Lines 3-5 are used to periodically update answer quality of each of its friends. Lines 7-17 calculate each friend's best answerer metric and generated a list including all metric values.

Lines 13-17 identify the top-K friends with the highest best answerer metric values and send the question to them. Answer

quality Q(i:j) is pre-processed and only interest similarity S(qi:j) needs to be calculated at run time. The S(qi:j) calculation has a time complexity of O(|Fi|). As the number of keywords in a question is generally very small, the calculation of S(qi:j) should take a short time and costs little computation resources of the mobile devices. This top-K friend selection algorithm has a time complexity of O(|Fi|).

6.2 Sampling of Large Data Spaces

Sampling of information signals and images has been extensively studied in signal processing and information theory literature [Candes and Wakin 2008: Kellogg 2006: Romberg 2008]. With the advent of large networked systems, there have been several studies on sampling large networks such as the webgraph [Rusmevichietong et al.2001] and social networks [Frank 1978: Gjoka et al. 2010: Granovetter 1976: Leskovec and Faloutsos 2006]. A large majority of these studies have focused on sampling of nodes from the networks. Different node sampling strategies, have been studied to ascertain whether the subgraphs obtained by these straregies can be used to recover the properties of the complete network, including topological properties such as size and degree distributions [Gjoka et al. 2010: Katzir et al. 2011: Leskovec and Faloutsos 2006] as well as dynamic processes such as information diffusion [Choudhury et al. 2010] over the networks. Specially for the Twitter OSN, [Krisnamurthy et al. 2008] made the first attempt to compare two datasets obtained by different sampling techniques- one through snowball crawling of the Twitter network starting from a small seed set of users, and the other through continuously succession the Twitter publicN timeline and extracting the users who posted the tweets. Again, Choudhury et al. [2010] examined several node (user) sampling strategies on Twitter users (e.g., random sampling, sampling users based on their location or their level of activity) and compared the information diffusion patterns observed from the samples. Whereas all these studies deal with sampling of users, we, in this work, focus on sampling data/content streams generated is OSNs, for which the methodologies discussed in the aforementioned studies are not directly applicable.

7. MEASUREMENT METHODOLOGY

We now describe the data presented in this paper and the methodology we used to collect it. We were not able to obtain data directly from the respective site operators. Most sites are hesitant to provide even anonymized data, and signing non-disclosure agreements to obtain data from multiple competing sites may not be feasible or desirable. Instead, we chose to crawl the user graphs by accessing the public web interference provided by the sites. This methodology gives us access to large data sets from multiple sites.

Since the focus of this paper is to investigate the structure of online social networks, we focus on the large weakly connected component (WCC) of the corresponding graphs in the rest of this paper. As we show later in this section, the large WCC is structurally the most "interesting" part of the network. The nodes not included in the WCC tend to be either part of very small, isolated clusters or are not connected to other users at all.

7.1 Challenges in Crawling Large Graphs

Crawling large, complex graphs presents unique challenges. In this section, we describe our general approach before discussing the details of how we crawled each network.

7.1.1 Crawling the Entire Connected Component

The primary challenge in crawling large graphs is covering the entire connected component. At each step, one can generally only obtain the set of links into or out of a specified node. In the case of online social networks, crawling the graph efficiently is important since the graphs are large and highly dynamic. Common algorithms for crawling graphs include breath-first search (BFS) and depth-first search. Often, crawling an entire connected component is not feasible, and one must resort to using samples of the graph. Crawling only a subset of graph by ending a BFS early is known to produce a biased sample of nodes. In particular, partial BFS crawls are likely to overestimate node degree and underestimate the level of symmetry. In social network graph, collecting samples via the snowball method has been shown to underestimate the power-law coefficient, but to more closely match other metrics, including the overall clustering coefficient.

Some previous studies of social networks have used small graph samples. Example studies have used samples of 0.3% of Orkut users [4], less than 1% of LiveJournal communities [8], and 0.08% of MySpace users [4]. In this paper, we obtain and study much larger samples of the user graphs.

	Flickr	LiveJournal	Orkut	YouTube
Number of users	1,846,198	5,284,457	3,072,441	1,157,827
Estimated fraction of user population crawled	26.9%	95.4%	11.3%	unknown
Dates of crawl	Jan 9, 2007	Dec 9 - 11, 2006	Oct 3 - Nov 11, 2006	Jan 15, 2007
Number of friend links	22,613,981	77,402,652	223,534,301	4,945,382
Average number of friends per user	12.24	16.97	106.1	4.29
Fraction of links symmetric	62.0%	73.5%	100.0%	79.1%
Number of user groups	103,648	7,489,073	8,730,859	30,087
Average number of groups memberships per user	4.62	21.25	106.44	0.25

Table 1. High Level Statistics of our social networking site crawls.

7.1.2 Using Only Forward Links

Crawling directed graphs, as opposed to undirected graphs, presents additional challenges. In particular, many graphs can only be crawled by following links in the forward directed. Using only forward links does not necessarily crawl an entire WCC; instead, it explores the connected component reachable from the set of seed users. This limitation is typical for studies that crawl online networks, including measurement studies of the Web [12].
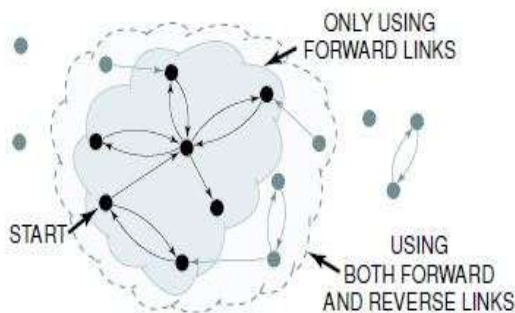
Figure 4. Users reached by crawling different link types. If only forward links are used, we can reach only the inner cloud (shaded cloud): using both forward and reverse links crawls the entire WCC (dashed cloud).

Figure above shows an example of a directed graph crawl. The users reached by following only forward links are shown in the shaded cloud, and those reached using both forward and reverse links allows us to crawl the entire WCC, while using only forward links results in a subset of the WCC.

8. CONCLUSION

While most honest users view the profiles of a modest number of users who tend to be nearby in the social network, even a strong, strategic crawler must view profiles of users who are further away. In our context where data stored at each node are rapidly changing this method is not directly applicable. Our algorithm assumes that information for each user in a network, such as web history logs, is available. Access to personal information infringes on user privacy and, as such, privacy concerns could serve as major stumbling block toward acceptance of our algorithms.

Systems that utilize our algorithms should adhere to the social translucence approach to designing social systems that entail a balance of visibility, awareness of others, and accountability. We also evaluate our anomaly detection technique using extensive ground-truth data of anomalous behavior exhibited by fake, compromised and colluding users. Our approach achieves a detection rate of over 66% (covering more than 94% of misbehavior) with less than 0.3% false positives. However, as of today, most research and applications are centered only on random sampling, and we conclude by calling for equal focus on expert sampling of social network data.

9. REFERANCES

[1] M. Gjoka, M. Kurant, C.T. Butts, and A. Markopoulou, "Walking in Facebook: A Case Study of Unbiased Sampling of Osns," Proc. INFOCOM, 2010.

[2] L.Katzir, E. Liberty, and O. Somekh, "Estimatinh Sizes of Social Networks via Biased Sampling," Proc. 20th Int'1 Conf. World Wide Web (WWW), 2011..

[3] A.S. Maiya and T.Y. Berger-Wolf, "Sampling Community Structure, " Proc. 19th Int' 1 Conf. World Wide Web (WWW), 2010.

[4] J.-T. Sun, H.-J. Zeng, H. Liu, Y. Lu, and Z. Chen, "Cubesvd: A Novel Approach to Personalized Web Search," Proc. 14th Int'1 Conf. World Wide Web (WWW), 2005.

[5] J. Teevan, S. T. Dumais, and E. Horvitz, "Personalizing Search via Automated Analysis of Interests and Activities," Proc. 28th Ann. Int'1 ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2005.

[6] E. Agichtein, E Brill, R. Wetzker, C. Zimmermann, C. Bauckhage, and S. Albayrak, "I Tag, You Tag: Translating Tags for Advanced User Models," Proc. ACM Third Int'1 Conf. Web Search and Data Mining (WSDM), 2010.

[7] S. Chaudhuri, G. Das, M. Datar, R. Motwani, and V.R. Narasayya, "Overcoming Limitations of Sampling for Aggregation Queries," Proc. 17th Int'1 Conf. Data Eng. (ICDE), 2001.

[8] T. Erickson and W.A. Kellogg, "Social Translucence: An Approach to Designing Systems That Support Social Processes," ACM Trans. Computer-Human Interaction, vol. 7, no. 1, pp. 59-83, 2000.

[9] VISWANATH, B., POST, A., GUMMADI, K. P., AND MISLOVE, A. An Analysis of Social Network-based Sybil Defenses. In Proc. of SIGCOMM (2010).

[10] WANG, G., KONOLIGE, T., WILSON, C., WANG, X., ZHENG, H., AND ZHAO, B. Y. You Are How You Click: Clickstream Analysis for Sybil Detection. In Proc. of USENIX Security (2013).

[11] XIE, Y., YU, F., ACHAN, K., PANIGRAHY, R., HULTEN, G., AND OSIPKOV, I. Spamming Botnets: Signatures and Characteristics. In Proc. of SIGCOMM (2008).

AUTHORS PROFILE

P. Eswaraiah working as assistant professor in C.S.E department, BVSR Engineering College, Chimakurthy, A.P, India

J.Amarnath is currently studying M.Tech in BVSR Engineering College, Chimakurthy, A.P, India.

WAP, HTTP and HTML5 Web Socket Architecture Analysis in Contemporary Mobile App Development

Eke B. O.
Department of Computer Science, University of Port Harcourt
Port Harcourt, Nigeria

Onuodu F. E.
Department of Computer Science, University of Port Harcourt
Port Harcourt, Nigeria

Abstract: Accessing current and accurate information anywhere and at anytime is becoming a growing interest nowadays. Wireless Application Protocol (WAP) is an application protocol that creates an opportunity to access information of any interest from WAP servers using mobile phones. WAP is an enabling technology based on the Internet client server architecture model, for developing client application for handheld devices or other wireless terminal which usually have less powerful CPU's, less memory, very restricted power consumption, smaller and variant displays, phone keypads etc. This paper analyses the features of WAP in relation to the well established HyperText Transfer Protocol (HTTP) technology, the web socket API innovations introduced in HTML5, the recent improvements in mobile devices processing capacity by connecting to cloud services and how application can be developed on them using modern tools. The features that are more adapted to client development of micro-devices are used for the technology application test.

Keywords: WAP, HTTP, HTML5, Web Socket API, Mobile App, Client, Servers, Cloud

1. INTRODUCTION

Recently, protocol technologies that enable handheld wireless devices to retrieve information have increased, presenting a more constrain on the computing environment compared to desktop computers. Some of the major protocols used in the wireless devices development include Wireless Application Protocol (WAP) and the HyperText Transfer Protocol (HTTP). The recent introduction of HTML5 and its associated Web Socket API has improved developers experience in building mobile applications and even connecting seamlessly to the cloud. services. WAP is an open protocol for wireless messaging, it provides the same technology to all vendors regardless of the network system. The WAP standard described a protocol suite allowing the interoperability of WAP equipment and software with different network technologies, such as GSM and IS-95 (also known as CDMA). The Hypertext Transfer Protocol (HTTP) on the other hand is an <u>application protocol</u> for distributed, collaborative, <u>hypermedia</u> information systems.

The standard development of HTTP was coordinated by the Internet Engineering Task Force (IETF) and the World Wide Web Consortium (W3C), culminating in the publication of a series of Requests for Comments (RFCs), most notably RFC2616 [1], which defines HTTP/1.1, the version of HTTP in common use.

HTTP functions as a request-response protocol in the client-server computing model. The client submits an HTTP request message to the server. The server, which provides resources such as HTML files and other content, or performs other functions on behalf of the client, returns a response message to the client. The response contains completion status information about the request and may also contain requested content in its message body.

In this paper the features of these two protocols that are more adapted to client development of micro-devices client environments that are reusable and user-friendly are presented.

1.1 WAP Technology

The WAP platform is an open specification that addresses wireless network characteristics by adapting existing network technologies (and introducing new ones where appropriate) to the special requirements of handheld wireless devices. WAP is made up of WAP Session Protocol (WSP), WAP Transaction Protocol (WTP), Wireless Transport Layer Security (WTLS),WAP Datagram Protocol and the Bearer that support WAP architecture and the internet layer on the other hand connected by the WAP Gateway [2]. WAP intends to standardize the way wireless devices (mobile phones, PDA, and so forth) access Internet data and services. WAP's reuse of existing Internet protocols eases the development of WAP services for Wireless Markup Language (WML) and WMLScript. Wireless handheld devices preset a more constrained computing environment and platforms, compared to desktop computers which most of the Internet technology was developed for. The handheld devices tend to have less powerful CPU's, less memory, very restricted power consumption, smaller and variant displays, phone keypads etc.
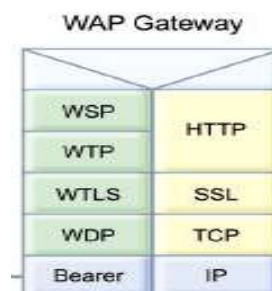

Figure 1: WAP technology GateWay

Furthermore, the wireless networks present additional constraints as communication infrastructures. They have less bandwidth, more latency and less connection stability and unpredictable availability. WAP intends to overcome these difficulties by being interoperable, have scaleable quality of service, efficient in the mobile network resources, reliable and secure.

WAP allows carriers to strengthen their service offerings by providing subscribers with the information they want and need while on the move. Infrastructure vendors will deliver the supporting network equipment. Application developers and content providers delivering the value added services are contributing to the WAP specification. Enabling information access from handheld devices requires a deep understanding of both technical and market issues that are unique to the wireless environment. The WAP specification was developed by the industry's best minds to address these issues[2]. Because, WAP is a standardized way for delivering Internet data over wireless networks and capable of addressing the unique characteristics of mobile terminals and wireless networks.

2. THE WAP ARCHITECTURE

The WAP standard defines two essential elements: an end-to-end application protocol and an application environment based on a browser. The application protocol is a communication protocol stack that is embedded in each WAP-enabled wireless device (also known as the user agent). The server side implements the other end of the protocol, which is capable of communicating with any WAP client. The server side is known as a WAP gateway and routes requests from the client to an HTTP (Hyper Text Transfer Protocol) (or Web) server. The WAP gateway can be located either in an Operator premises as illustrated in Figure 1. or in WAP application provider premises with the web server as shown in Figure 2.
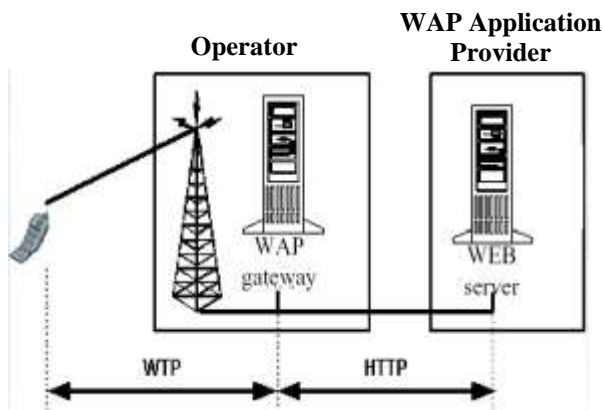

Figure 1. Gateway equipment in the operator premises

Irrespective of the WAP architecture used in an implementation, the core issues is the shielding of the micro-devices with less powerful CPU's, less memory, very restricted power consumption, smaller and variant displays etc from the challenging Web client processing and handling capacity.

In the WAP network the client communicates with the WAP gateway in the wireless network. The WAP gateway translates WAP requests to WWW requests, so the WAP client is able to submit requests to the Web server. Also, the WAP gateway translates Web responses into WAP responses or a format understood by the WAP client [3].

The wireless application environment provides WAP micro browser for interaction between WAP (web applications) and wireless devices. This browser relies on WAP markup

languages such as WML (Wireless Markup Language), WML Script and XHTML MP (Extensible Hypertext Markup Language Mobile Profile.
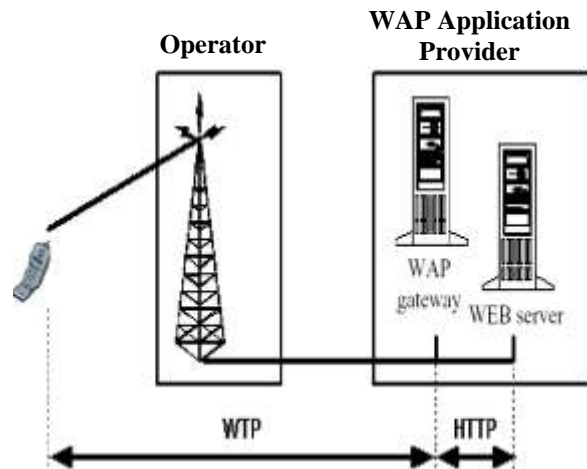

Figure 2. WAP Gateway in the WAP application provider premises

2.1 The WAP Programming Model

The WAP programming model is similar to the Web programming model with matching extensions, but it accommodates the characteristics of the wireless environment. The WAP programming model is based heavily on the Web programming model. But how does the WAP gateway work with HTML (Hyper Text Markup Language)? In some cases, the data services or content located on the Web server is HTML-based.

Some WAP gateways could be made to convert HTML pages into a format that can be displayed on wireless devices [3]. Because HTML wasn't really designed for small screens, the WAP protocol defines its own markup language WML, WML Script, and XHTML MP is the languages that are specifically designed to develop WAP applications for Mobile devices. These languages adhere to the XML standard and are designed to enable powerful applications within the constraints of handheld devices. In most cases, the actual application or other content located on the Web server will be native WAP contents created with WML (XHTML MP) or generated dynamically using WML Script, Java Servlets or JSP (Java Server Page), or other server side programming languages.

WML is an XML-based markup language that was designed especially to present WAP content on a wireless terminal. WML can preserve the content of variables between different WML pages. The basic unit of WML is the card that specifies a single interaction between the user and the user agent. Multiple cards are grouped together in decks, which is the top most element of a WML file. When the user agent receives a deck, it activates only the first card in the deck. There are no functions to check the validity of user input or to generate messages and dialog boxes locally using WML. Therefore, to overcome this limitation, WML Script was developed. WML Script, which is based on ECMA Script (the standard for java script), is a language that can be used to provide programmed functionality to WAP applications. It was defined to enable

the execution of scripts on WAP devices. The goal of using WML Script is to reduce the number of turn around between the client and the server. It is part of the WAP specification, and it can be used to add script support to the client. Its difference from ECMA Script is that it is compiled into byte code before it is sent to the client. The main reason for this is to cope up with the narrowband communication channels and to keep client memory requirements to a minimum. XHTML is a markup language used to create richer web content on an ever-increasing range of platforms including mobile handsets. It is similar with HTML in its tag definition and syntax, but it adds modularity and enforces strict adherence to language rules. It brings a clear structure to web pages, which is especially important for the small screens and limited power of mobile devices. This WAP programming Model is illustrated in figure 3. The XHTML MP is a mobile adaptation of XHTML by excluding those features not appropriate for devices with small screens. It is a strict subset of XHTML that includes additional elements and attributes that are useful in mobile browsers with additional presentation elements and support for internal style sheets.
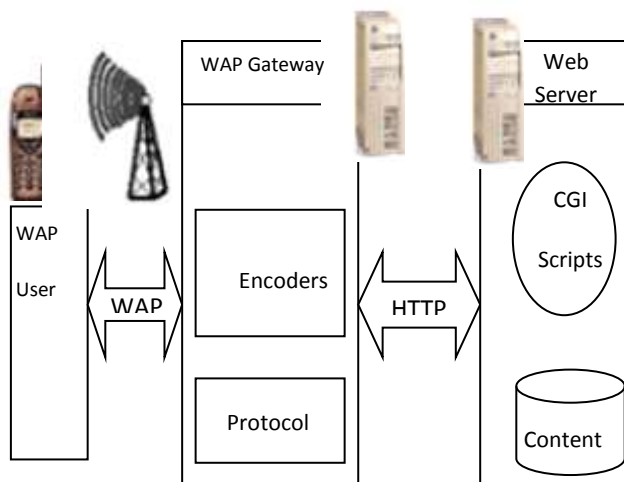

Figure 3. The WAP Programming Model

Mobile browsing technology is evolving from WAP 1.x to WAP 2.0, by introducing different enhancements for mobile content development. Especially WAP 2.0 provides support for protocols such as IP, TCP and HTTP. This provides interoperable optimizations suitable to the wireless environment and to the environment that permits wireless devices to utilize existing Internet technologies. WAP 2.0 also provides different application environment, which enables delivery of information and interactive services to wireless devices.

WAP standard defines the future of wireless browsing technology based on the WML, XHTML MP and WAP CSS (WAP Cascading Style Sheet). Both WML and XHTML MP are a reformulation of the XML. XML is a language for marking up structures in text documents and supports the UTF-8 (8 bit Unicode Transformation Format) coding standard. The UTF-8 coding standard supports several languages character set. So WML and XHTML MP can be used to create WAP pages that are encoded as UTF-8. Browsing from wireless terminals supporting UTF-8 encoding becomes possible.

2.3 HTTP Architecture

The innovations that Berners-Lee added to the Internet to create the World Wide Web had two fundamental dimensions: connectivity and interface [4]. He invented a new protocol for the computers to speak as they exchanged hypermedia documents.

HTTP stands for **Hypertext Transfer Protocol**. It is a TCP/IP based communication protocol which is used to deliver virtually all files and other data, collectively called resources, on the World Wide Web. These resources could be HTML files, image files, query results, or anything else.

A browser works as an HTTP client because it sends requests to an HTTP server which is called Web server. The Web Server then sends responses back to the client. The standard and default port for HTTP servers to listen on is 80 but it can be changed to any other port like 8080 etc.

There are three important things about HTTP, which should be noted:

- **HTTP is connectionless:** After a request is made, the client disconnects from the server and waits for a response. The server must re-establish the connection after it processes the request.

- **HTTP is media independent:** Any type of data can be sent by HTTP as long as both the client and server know how to handle the data content. How content is handled is determined by the MIME specification.

- **HTTP is stateless:** This is a direct result of HTTP's being connectionless. The server and client are aware of each other only during a request. Afterwards, each forgets the other. For this reason neither the client nor the browser can retain information between different requests across the web pages.

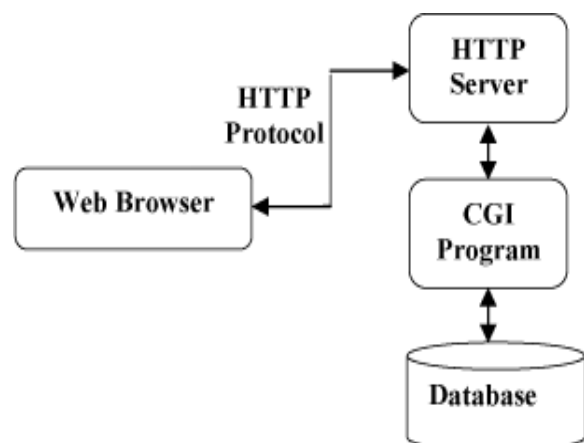Figure 4 below shows where HTTP Protocol fits in communication:

Figure 4. The HTTP Protocol in Communication Model

The set of common methods for HTTP/1.0 is defined below. Although this set can be expanded.

The model makes it possible for a client to reach services on a large number of origin servers; each addressed by a unique Uniform Resource Locator (URL). The content stored on the

servers is of various formats, but HTML is the predominant. HTML provides the content developer with a means to describe the appearance of a service in a flat document structure. If more advanced features like procedural logic are needed, then scripting languages such as JavaScript or VB Script may be utilised.

Figure 5 below shows how a WWW client request a resource stored on a web server. On the Internet, standard

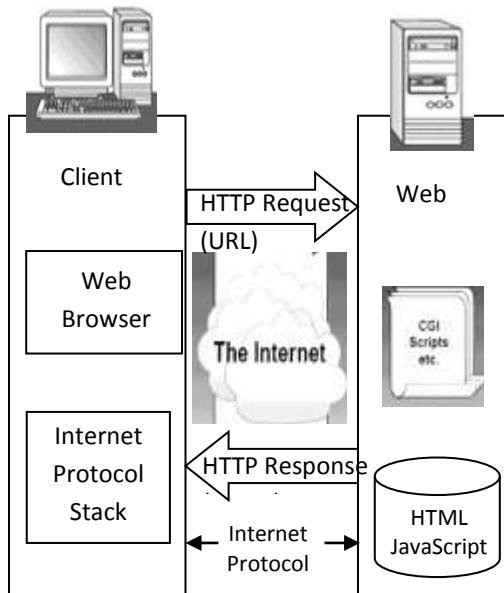communication protocols, like HTTP and Transmission Control Protocol/Internet Protocol (TCP/IP) are used.

Figure 5. A HTTP Client-Server Request-Response Architecture

3. ANALYSIS OF CONTEMPORARY TECHNOLOGIES

It is clear that what we are considering in this work is web mobile development technologies supported by WAP and HTTP. These protocols have a way of supporting web development on mobile devices and even a native-web application development via devices internal client scripting languages. These use the operating systems browser as a runtime environment and are runnable on all mobile operating systems without any installation process [5].

The programming languages to create web applications on WAP include WML,WMLS but on HTTP, HTML5, CSS and JavaScript are used. Those allow platform independent development. The application can be opened by accessing a specific website.. A specifically developed and adjusted website mocking a native application in design and functionality [6]. With particular techniques, these web apps can also be used without an Internet connection, as long as the user has saved them on the device once and not cleared the browser's cache. In iOS, this is achieved via the function "add to home screen". A third way of developing mobile applications is to combine native and web development. Connecting a container application written in platform specific code with a web app containing the application logic

and the user interface leads to a program referred to as "hybrid". In the most extreme case, the native app contains only a single browser window in full screen, running the web app [7]. The corresponding element is called UIWebView in iOS and WebView in Android and Windows.

As the different development approaches for mobile applications have been shown, the question remains which of these – depending on the applied criterion – offers advantages and disadvantages. The criteria "Content vs. Experience", "Performance", "APIs", "Distribution" and "Development Cost" are to be considered. Those are only a subset of possible criteria and represent those with a clear difference in results.

3.1 Analysis of HTML5, HTTP and Mobile Device Development

Before the development of HTML 4 in 1997, World Wide Web Consortium (W3C) believed that XML is the future of the web even with the good features of HTML4, this view never changed. The strict definitional guideline for development of XML and need to abide by the rules of coding made it look exiting. The development of XHTML lend credence to this view but with the release of HTML5 to simplify developing for the web many new elements were added, along with the addition of several new JavaScript API's. Previous to HTML 5, playing media on a browser required a plug-in or an application installation [6]. For example playing a Flash game requires Adobe shockwave player but now thanks to HTML5 and its new multimedia elements audio, video and canvas that could be a thing of the past. Although HTML5 is not without faults most see it as a huge leap forward for web development. Its release plays a vital role how web games are developed. Recently there has been a massive expansion of casual 'pick up and play' games which are frequently played on social networking sites, smart phones, tablets and web sites this has led to a major shift in the game development industry.

The Canvas element provides a way for developers to draw and manipulate 2D images using HTML and JavaScript to implement movement of a canvas element the image must be deleted and redrawn continuously by the web OS. With the canvas element 2D games can be easily implemented and with the use of vector based images they can be manipulated without losing any quality. Websites such as FaceBook have already begun providing an abundance of games showing the effectiveness of HTML5 canvas. The WebGL API is another exciting and innovative feature that allows the customisation of 3D objects by allowing JavaScript to communication with the users GPU. WebGL provides an API that allows 3D graphics to be used with canvas. This could potentially make web browsers a valid gaming platform, which would revolutionise game development.

WebSocket technology provides full-duplex bi-directional communication channels over a single TCP socket in both web browsers and web servers [8]. The WebSocket API provides features in JavaScript that allow the implementation of real-time interactions on applications. For example web based games will be able to provide instant interactions with game objects (such as the canvas images) and provide in game chat functionalities. Combining Canvas and WebSockets allows for the development of a multiplayer web browser

based games framework [9]. The effectiveness of these new HTML5 features provide information on how these technologies can and are being used.

Normal web communication is achieved using HTTP, the problem with this is that it only allows transfer to occur in one direction at a time. HTTP communications also requires constant web page requests each time new data is required. This results in slow communication between client and server. Current technologies such as Ajax and Comet attempt to speed up this transfer. These technologies attempt to simulate a full-duplex connection, but they are merely a hack and still truly only provide one way communication. Comet uses methods such as long-polling or streaming, these methods however still involve using HTTP requests which causes latency issues. This is where WebSockets come in; HTTP was not designed to support real time communication but WebSockets upgrade this HTTP protocol to a WebSocket protocol. This enables true full-duplex, bidirectional connections, what this signifies is the ability to create real time multiplayer games on web applications. WebSockets portray a massive advancement for real time applications on the web. Although WebSockets do not render Ajax completely obsolete they do however supersede its solution for real time functionality. Originally WebSockets were implemented into all web browsers providing developers with the means to begin creating interactive real time games, however due to security vulnerabilities WebSockets were disabled in a number of browsers including Mozilla Firefox 4 [10].

This security issue has been corrected in the latest version of the WebSocket protocol, which means we can still expect to see multiplayer web based games coming becoming available in the latest web browser OS versions. The potential for WebSockets when used alongside other HTML5 technologies such as Canvas and WebGL is incredible, a number of games have already been created displaying different creative features built with WebSockets. Figure 3 shows an online multiplayer game "Rawkets" which looks similar to asteroids only offering multiplayer free for all game play. The game is in fact quite enjoyable and shows the potential of games, signifying that real time multiplayer Games could begin to expand all over social networking sites in the near future. Another game which adds to this theory is the online social chatting game (Figure 4) "Rumpetroll", which transforms players into a tadpole floating around space. The real time chat feature from WebSockets can clearly be seen here as the messages sent between players reach each other almost instantaneously. Using WebSockets in these games shows the availability of real time collision detection and chat features, with almost no latency being displayed WebSockets really is the "*Quantum Leap*" [11] that will bring real-time multiplayer games to the web and mobile devices. The Pusher API [10] enables developers to easily add WebSocket functionality to web games and mobile games, they also offer a number of libraries to be used within your game that explain how to effectively use Pushers client libraries.

With all the new elements of HTML5 and the new JavaScript features there comes an abundance of third party software to optimise features and lessen the learning curve required to delve into these innovative elements.

3.2 HTML5 and the Cloud

It is clear that HTML5 has cross-platform capability and its support is available in any latest mobile device's browser. So, application made in HTML5 can be run on any devices. Ability to use same application and data from different devices without tedious installations. Here applications and data are stored on cloud and not bound to particular devices. So, we can use it on anywhere, anytime and on any device.HTML5 is lightweight than other alternative like Flash, etc. It has not required extra plug-ins because it has built in support for video, audio, canvas, etc.

If WebSocket connection mechanism is used with deflate compression mechanism in HTML5, network consumption will become efficient which supports HTML5-based cloud phone software platform concept. In that, HTML5 worked as software platform for mobile devices in which all end user functionality of devices is downloaded and cached dynamically from web including all applications [10].

The offline application capability and web storage of HTML5 enabled mobile web browser to bring an offline user interface to users. Web storage is a feature that is intended to overcome the limitation of HTTP cookies. Despite it's mainly purpose is for state management mechanism, cookies can be used for storing information at client side, at limited size. Using Web storage, client side will have enough space for storing more data. In order to making it capable to operate while offline, Offline web application feature is used. Offline web application allows set of HTTP objects is able to be accessed from browser without having to connect to Internet.

If mobile application is mainly used to display and interact with online content or services, it is better to avoid the native choice. However, if mobile application is mainly used offline, a native app will offer a better user experience.

WebGL[9] APIs have only recently become available in major browsers. WebGL will provide support for highperformance, direct manipulation 3D graphics content that can run without installation in major web browsers. We believe that WebGL will allow even high-performance gaming applications to run efficiently in a standard web browser.

3.3 HTML5 Mobile Application Cloud Architecture

In mobile application cloud architecture HTML5 on the mobile browser connects to the internet which in turn connects to a cloud server via a firewall. The backend data and the cloud services are connected to the cloud server which then provides the services to the mobile devices. This operation is supported by the mobile device OS and the HTML5 components that supports the cloud services.

Common characteristics of mobile web development frameworks supported by this architectures include [12]:

Cross-platform: Support for multiple mobile device Platforms allows to deliver app to a wide range of users.

Lightweight: Due to current bandwidth limitations, a stronger emphasis on lowering file weight is placed into mobile web development frameworks.

Optimized for touch screen devices: Fingers as input devices instead of mouse cursors provide an extra set of challenges in user interface design. Mobile web development frameworks

provide standard user interface elements and event-handling specifically for mobile device platforms.
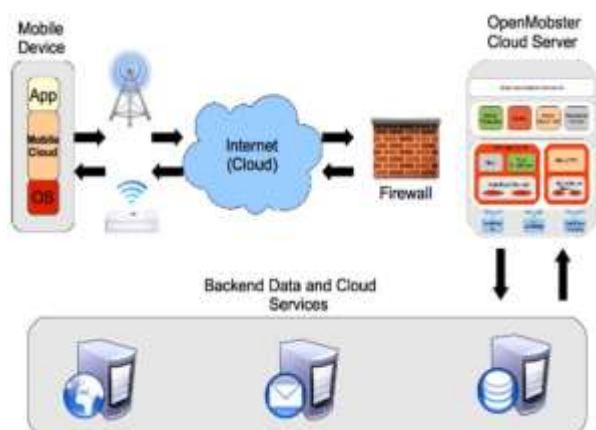

Figure.6. The open mobster architecture [12]

Uses HTML5 and CSS3 standards: Most mainstream mobile devices have web browsers that support HTML5 and CSS3, and so mobile web development frameworks take advantage of new features available in these upcoming W3C specifications for a better user experience.

3.3.1. Use of HTML5 Mobile Cloud Computing

The HTML5 breakthrough is mainly represented by the extreme simplification of the web content creation and by the generality of browsers, alleviated from the needs of complex plug-ins.

These new features are ensured by completely new syntactic elements (like <video>, <audio>, <canvas>), by hooks toward other standards (*e.g.* the possibility of using Web Socket [8]) or by extending existing elements for recent interaction modes (*e.g.* from on Click to on Touch functionalities).

The <canvas>[9] element is the enabler for real-time drawing of complex graphical content (paths, boxes, circles, characters and images) and gives the possibility for dynamic update and creation of the web content on the fly, by using JavaScript [8].

Beyond AJAX (Asynchronous JavaScript And XML), in order to ensure content delivery, HTML5 uses Web Socket, a new mechanism for establishing a connection between the server and the client.

4. DESIGN AND TECHNICAL CHALLENGES OF WAP AND HTTP

There are design and technical challenges that these technologies face which must be remedied in the new technologies. When new technologies such as HTML5 makes entry into mobile application development, there is a need to correct and improve the challenges. These design and technical challenges are found in HTTP and WAP.

4.1 HTTP Design Challenges

The major challenge of HTTP is in its very nature which is its Request-Response design as illustrated in figure 5. In the time HTTP was developed it was design to give a single response for every single request. This means that for 1 million response there is a corresponding 1 million request. This

approach of having HTTP Response for each HTTP Request, is a huge drawback when having dynamic real-time updates sent from the server to the client (browser), requiring a lot of extra-traffic for each packet. This is too much burden on the network as well as on the server. Mobile lightweight nature low network capability will not easily adapt to that challenge. This made designers to develop Asynchronous JavaScript And XML (AJAX) in the preceding years when XML was seen as the answer even by W3C.

However, AJAX was not a complete solution since it has certain drawbacks inherent in HTTP. On the one hand, AJAX (Asynchronous JavaScript And XML) is used for creating asynchronous request from the client to the server and for receiving server side responses in the same manner. Hence, any server request results in creating an XMLHttpRequest. Once the server receives this request, it parses it and sends the corresponding response to the browser as HTTP Response. This improves the dynamic real-time updates sent from the server to the client but does not eliminate the HTTP circle.

On the other hand, HTML5 Web Socket API supports bidirectional, full-duplex communication over a single socket. A Web Socket based communication is established by sending from the client to the server, a simple HTTP Request for upgrading its connection type to Web Socket. If the server positively answers with the Web Socket upgrade message, the subsequent messages are exchanged by using the Web Socket API. The compression mechanisms vary with the connection type: The HTTP provides both gzip and deflate supports [8], while in the Web Socket case, deflate is the only available mechanism [8]. Considering the user interaction, HTML5 reconsiders and extends the previous elements so as to deal with the interaction modes.

Hence, mobile application developed with HTML5 framework will give better performance over web based mobile cloud application requiring high bandwidth. Mobile application will also run across platform efficiently and it will not have any constraint of any proprietary system. Resource starved application which have required more resources like memory, processing will easily run on mobile devices by using appropriate cloud services.

4.2 WAP Design Challenges

We have clearly specified that the languages supporting WAP technology is WML and WMLS(a subset of ECMAScript). One of the major set back of the technology is the language-WML. WML cut users off from the conventional HTML Web, leaving only native WAP content and Web-to-WAP proxi-content available to WAP users. The technology was provided custom-designed content by reducing complexity interface as the citizens of many nations are not connected to the web at the inception of the technology and have to use government funded and controlled portals to WAP and similar non-complex services.

There was equally an under-specification of terminal requirements which meant that compliant devices would not necessarily interoperate properly. This resulted in great variability in the actual behavior of phones, principally because WAP-service implementers and mobile-phone manufacturers may have not obtained a copy of the standards or the correct hardware and the standard software modules. As an example, some phone models would not accept a page more than 1 Kb in size; others would downright crash. The user interface of devices was also underspecified: as an

example, accesskeys (e.g., the ability to press '4' to access directly the fourth link in a list) were variously implemented depending on phone models (sometimes with the accesskey number automatically displayed by the browser next to the link, sometimes without it, and sometimes accesskeys were not implemented at all).

Mover WML have Constrained user interface capabilities making terminals with small black-and-white screens and few buttons, like the early WAP terminals, face difficulties in presenting a lot of information to their user, which compounded the other problems: one would have had to be extra careful in designing the user interface on such a resource-constrained device which was the real concept of WAP.

Developers were also not offered good authoring tools that could allow content providers to easily publish content that would interoperate flawlessly with many models, adapting the pages presented to the User-Agent type. This could have provided solution to the interface constrain WML problems. However, the development kits which existed did not provide such a general capability. Developing for the web was easy: with a text editor and a web browser, anybody could get started, thanks also to the forgiving nature of most desktop browser rendering engines. By contrast, the stringent requirements of the WML specifications, the variability in terminals, and the demands of testing on various wireless terminals, along with the lack of widely available desktop authoring and emulation tools, considerably lengthened the time required to complete most projects. The many mobile devices supporting XHTML, and programs such as Adobe Go Live and Dreamweaver offering improved web-authoring tools, it is becoming easier to create content, accessible by many new devices. This created massive exodus from WML.

Lack of user agent profiling tools and database of device capabilities made it difficult to know which devices users are using to get the correct content.

Often when a technology neglects content providers then it seem to be heading to extinction. Some wireless carriers had assumed a "build it and they will come" strategy, meaning that they would just provide the transport of data as well as the terminals, and then wait for content providers to publish their services on the Internet and make their investment in WAP useful. However, content providers received little help or incentive to go through the complicated route of development. Others, notably in Japan, had a more thorough dialogue with their content-provider community, which was then replicated in modern, more successful WAP services such as i-mode in Japan or the Gallery service in France.

The challenge of lack of openness in the technology also scared away developers. Many wireless carriers sold their WAP services as "open", in that they allowed users to reach any service expressed in WML and published on the Internet. However, they also made sure that the first page that clients accessed was their own "wireless portal", which they controlled very closely. Some carriers also turned off editing or accessing the address bar in the device's browser. To facilitate users wanting to go off deck, an address bar on a form on a page linked off the hard coded home page was provided. It makes it easier for carriers to implement filtering of off deck WML sites by URLs or to disable the address bar in the future if the carrier decides to switch all users to a walled garden model. The vendors, phone manufacturers and W3C must have learned the importance of developers in the sustenance of any given technology.

5. PERFORMANCE ANALYSIS OF WAP, HTTP AND HTMLS

The performance of the technologies on varying phones and mobile devices will be analyzed in other to find the adaptability of the technologies to mobile developers and how widely applications in the market places have deployed the technologies. The period that will be under consideration is 2005-2015 a period of ten years and the data that will be used in the analysis is a secondary data. The analysis will use WML adaptability and development for evaluating WAP technology. HTML version 3 and below for simple HTTP request-responses and HMTL5 for web socket and other recent improvements [13].

In figure 7 the plot of the performance of this technologies estimated based on the number of deployments, resulting in mobile application development over a period of ten years gathered from W3C and Mobile vendors shows an amazing result. In 2004 HTML5 development started [13] hence in 2005 performance was zero since the technology is still at its infancy. WAP was up and running as indicated in figure 7. It is important to note that by 1997 HTML4 had been released making HTML3 which uses pure HTTP request-response to move down-wards. From the chart WAP was on the rise from 2005 to 2007 and by 2008 HTML5 deployment was already at par with HTML3 but still far below WAP deployment [14]
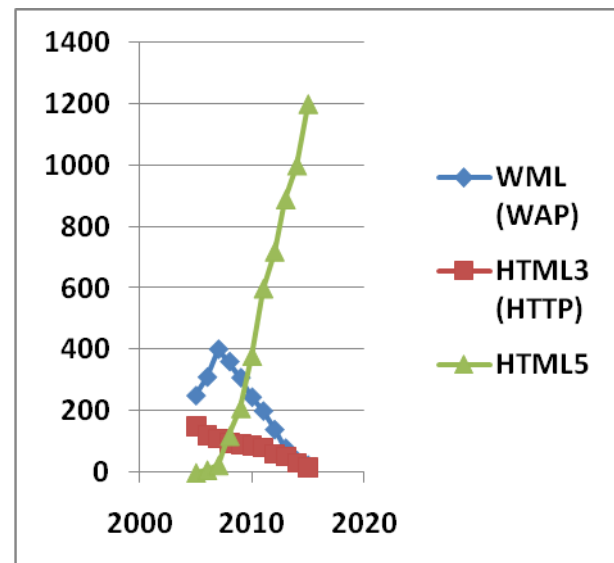

Figure.7. Plot of performance of the Technologies over the a period of ten years.

Due to the flexibility of HTML5 it continued a steady rise on mobile application development while WAP continued to decline as the years goes by and improvement continues in HTML5 mobile capability. HTML3 and WAP seem to continuously remained on the decline and many modern smart phones and PDAs do not have WAP specification for them at all. One may ask if that is the end of the road for the WAP technology or if Japan and Asia nations that still support WAP will do a technological reversal either in WAP, Dynamic WAP or any future WAP technology upgrade. Only the future can tell where the WAP technology is headed but the continuous improvement in HTML5 is making matter worse

since it seems to by integrating many other scripting technologies which were separate from the HTML developers have known.

5.1 WAP Performance Challenges

The challenge is that WAP (WML) still rely heavily on WMLS for dynamism and WMLS is only a subset of ECMAScript (JavaScript) which HTML5 developers are already trying to integrate into the main stream of HTML5 technology [15]. The one application for all mobile devices which HTML5 have achieved and is further improving upon make matter worse when compared to the variations in specification for different mobile devices which is used in WAP. It is even worse that many new devices do not have WAP specifications making difficult if not impossible for developers to adapt in those devices.

Moreover the time spent in developing the same application for different devices can be used in improving the single application that will run in different devices. The time may also be used in developing different applications which will make more fortune for the developers.

6. CONCLUSION

Mobile application developed with HTML5 technology will give better performance over simple HTTP web based mobile application and WAP based Mobile application. The support of HTTP5 in enabling Mobile application to run on cross platform efficiently means it will not have any constraint of any proprietary system. It also imply that resource constrained application which have required more resources like memory, processing will be run on mobile devices by using appropriate cloud services. The WAP technology seem to be at the verge of been abandoned by developers making the technology to be on its way to extinction if its supporters do not act very fast. The investment in its development and expansion will coolly go down the drain just as many investments in different obsolete IT technologies. Simple HTTP technology supported in HTML1 to HTML3 and even HTML4 may still remain for beginner developers and classic applications that lay on the internet for mobile devices which usage has continued to decline. W3C and mobile browser developers may not scrap the support of this versions immediately but there is a clear possibility in the nearest future that many mobile browsers may not continue to support them.

HTTP as an internet technology will continue to support HTML5 and even the newer technologies but they may no longer be the base for the architectural development of the mobile development technologies. In conclusion the research have analyzed selected mobile technologies and have presented their architecture with the aim of analyzing their strength and weaknesses and recommend actions to both technology vendors and mobile developers.

6.1 Recommendation

In the light of the analysis and the result presented in figure 7 developers may have no option than to jump into the HTML5 band-wagon in their effort in developing web mobile application that will be efficient and developed once for majority of mobile devices. The growth of Android Operating System and its support for HTML5 support browser seem to add benefit of usage across different mobile devices powered by Android. Apple and Microsoft supported mobile devices are also not exempted. Developers investment in WAP development even if it is little my end up as an academic exercise if the vendors are not offering free and efficient tools and support. The lower versions of HTML have no future in mobile development but may still be very relevant in desktop development since many classic web site are not likely to go away in the nearest future and may still need to be supported by tomorrow browsers.

6.2 Future Work

In future, HTML5 will be very useful in the application field of mobile gaming, mobile healthcare, mobile commerce, mobile learning. Complex mobile application requiring high processor capability will be executable on mobile devices by leveraging cloud service for mobile application. We may need to find out whether the end of the road has come for WAP. Research can also be carried out on how to create a cloud computing application which uses Software as a Service for executing HTML5 application using Mobile Web Development Framework such as Phone Gap, Sencha, etc.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] RFC2616 (1999) "Hypertext Transfer Protocol – HTTP/1.1", R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, T. Berners-Lee, June 1999. URL: http://www.rfc-editor.org/rfc/rfc2616.txt

[2] WAP-AS(2000) Wireless Application Protocol Architecture Specification,WAP Architecture WAP-210-WAPArch Proposed version 17-October-2000, Wireless Application Protocol Forum Ltd. (http://www.wapforum.org/what/copyright.htm).

[3] APION (2001) WAP Wireless Application Protocol, Apion Telecommunication Software Systems, The International Engineering Consortium http://www.iec.org

[4] IDC (2013). IDC Raises Tablet Forecast for 2012 and Beyond As iOS Picks Up Steam, Android Gains Traction, and Windows Finally Enters the Market. http://www.idc.com/getdoc.jsp?containerId=prUS23833612#.UV77q78pG3h. Accessed on 28.03.2015.

[5] Patrick M. (2012) Native versus HTML5 – where's mobile programming heading to? FHWS SCIENCE JOURNAL, Vol. 1, No. 1, 2013, 23-33, Germany

[6] Willnecker, F., Ismailovic, D. Und Maison, W. (2012). Architekturen Mobiler Multiplattform-Apps. Verclas, S. and Linnhoff-Popie, C. (ed.): Smart Mobile Apps - Mit Business Apps ins Zeitalter mobiler Geschäftsprozesse. Heidelberg, London, New York et. al. 2012. 403-418.

[7] Franke, F. and Ippen, J. (2012). Apps mit HTML5 and CSS3 für iPad, iPhone und Android. Bonn. 2012.

[8] Rama R. G., Mihai M., Bojan J. and Françoise P. (2012) HTML5 as an application virtualization tool 2012 IEEE 16th International Symposium

[9] Yang J., Zhang J. (2010) Towards HTML 5 and Interactive 3D Graphics 978-1-4244-8035-7/10 © 2010 IEEE

[10] Pavel S. (2012) Mobile development tools and cross-platform solutions 978-1-4577-1868-7/12©2012 IEEE

[11] Han Qi, Abdullah Gani (2014) "Research on Mobile Cloud Computing: Review, Trend and Perspectives".

[12] Nimit S Modi ,Proff.Yask Patel(2013) Web Interface using HTML5 for Interaction between Mobile Device & Cloud-Services, International Journal of Engineering Trends and Technology (IJETT) - Volume4Issue5- May 2013

[13] Kevin C., Aaron B., Gavin F. (2012) HTML5 and the Mobile Web, International Journal of Innovation in the Digital Economy, 3(2), 40-56, April-June 2012

[14] Hoang T. Dinh, Chonho Lee, Dusit Niyato, and Ping Wang (2015) A Survey of Mobile Cloud Computing: Architecture, Applications, and Approaches http://onlinelibrary.wiley.com/doi/10.1002/wcm.1203/abstract

[15] Corral, L.. (2011). Evolution of Mobile Software Development from Platform-Specific to Web-Based Multiplatform Paradigm. Proceedings of the 10th SIGPLAN symposium on New ideas, new paradigms, and reflections on programming and software. New York. 2011. 181-183.

9. ABOUT THE AUTHORS

Eke Bartholomew PhD, MCPN, MACM, FIPMD is a Software Engineering / Computer Science Lecturer at the University of Port Harcourt and Mobile Application Developer in Oyol Computer Consult Inc. His research interest is in SE Methodologies and Mobile IT deployment.

Dr. Onuodu, Friday E. is a Lecturer at the University of Port Harcourt. His research interest is in Data Mining and Data Extraction using both Mobile Devices and Desktops. He also has interest in IT deployment analysis. He has many publication in learned journals.

Expression of Query in XML object-oriented database

Roghayeh Madani
Department of Computer Engineering, Ardabil
Science and Research Branch, Islamic Azad
University,
Ardabil, Iran
Department of Computer Engineering, Ardabil
Branch, Islamic Azad University,
Ardabil, Iran

Ali Harounabadi
Department of Computer Engineering, Central
Branch, Islamic Azad University,
Tehran, Iran

Abstract: Upon invent of object-oriented database, the concept of behavior in database was propounded. Before, relational database only provided a logical modeling of data and paid no attention to the operations applied on data in the system. In this paper, a method is presented for query of object-oriented database. This method has appropriate results when the user explains restrictions in a combinational matter (disjunctive and conjunctive) and assumes a weight for each one of restrictions based on their importance. Later, the obtained results are sorted based on their belonging rate to the response set. In continue, queries are explained using XML labels. The purpose is simplifying queries and objects resulted from queries to be very close to the user need and meet his expectation.

Keywords: Object Oriented, disjunctive queries, conjunctive queries, object-oriented database, XML.

1. INTRODUCTION

Upon paying attention to the data existing around ourselves, we understand that plenty of these data may not be stored to the same existing form. For storage of this type of data as well as combined and complex data existing around ourselves, we have no option but to use the new type of database that can record complex and combined data, therefore we approach to object-oriented database.

Contrary to the relational database (classic) that availability samples in certain perceptual level are exhibited by table records, in object-oriented database, objects are observed; in other word, queries in this type of database are similar to classic database, but responses are different.

Database queries are analyzed in this paper as simple and combined modes. In combined mode, disjunctive and conjunctive queries were used. In continue, queries are explained by XML labels. In second part of paper, background and in third part, Related Works and in fourth part, Proposed Method was presented. The queries are explained by disjunctive and conjunctive predicate, in addition can be stored by XML labels. Fifth part was allocated to Case Study and sixth part to Conclusion.

2. BACKGROUND

The background of this study includes object-orientation concept, object-oriented database and XML.

2.1 Concept of object-orientation

Object-orientation emphasizes on the objects, in the object-oriented programming, the objects are propounded abstractly. Object-oriented programming is a modern style therein components may be created and used in different programs. Readability of programs written in this method is high and their testing, fault detection and amendment is easy. Object-orientation is formed based on several principles, as follows:

- Inheritance: In object-oriented programming, the concept of inheritance is used frequently. There are classes that inherited their attributes from main classes referred to as super class. Reusability of code is one of main advantages of inheritance.
- Encapsulation: it is referred to as placing implementation in a capsule so that external user is not aware of implementation procedure and only knows that this capsule performs a specified work.
- Polymorphism: Assume that you inherit the characteristics of father and his works, but do it in another manner. It is just the same thing that is called polymorphism.
- Abstraction: abstracting! It is referred to an abstract class that methods are implemented therein.
- Interface: it shows the relationship between objects.

2.2 Object-oriented database

The weaknesses of relational model caused the experts to think of make new models that one of their best types is object-oriented model. Maximum object-oriented power is arising out of its ability in real world modeling phenomena. Object-oriented databanks model combines the power of object-oriented languages and data storage and retrieval systems and converted to very powerful systems. Definition of concepts such as class, inheritance and polymorphism etc. has empowered it extraordinarily. In the object-oriented databank model, the data is stored only as shape and includes the relationship between data, attributes, restrictions and time and place problems. In the object-oriented model, each phenomenon is related to a class and output of queries may be a part of an object or combination of a few objects.

2.3 XML (extensible markup language)

Extensible Markup Language (XML) is a language for showing data that was standardized by consortium web W3C [1]. Within recent years, XML has been taken into consideration as a standard language and structure for exchange and exhibition of data on various applied program particularly in web-based applications. Following web integration and continuous production of a great volume of electronic data, XML has been converted to a standard for electronic exchange of information documents. The XML can be used in the following:

- Exchange the data between disparate applications, businesses and databases.

- Enable a model to display the same data on various devices with different purposes and readers.

- A suitable format for storing data.

3. RELATED WORKS

Upon upraising the demand and need to new methods for working with the data and relationships between these data and applied programs focusing on enormous volume of data, it was proved that object-oriented databases are better than relational database and other databases.

Fong [2] proposed a reengineering engineering methodology for transfer of an EER model to object modeling technique (OMT) for production of object-oriented database plan. In this paper, a series of mapping rules from EER model to general object-oriented model were provided.

In [3], a language was presented that one of its characteristics is defining classes that differ from other classes in a few aspects. It is possible through separating the concept of "type" and "class" and allows queries compilers to detect the type errors even at the presence of objects contrary to the definition of classes.

In [4], an independent query language with high class range was introduced for visual and alphabetic database management that is called PICQuery.

In [5], XML in the most specific state is assumed as database and XML documentations are self-defined, means that structure and name of types are specified at the use time. In addition, data may be presented in two tree and graph forms.

In [6], a software was offered for direct use of Query languages on XML. Whereas for exchanging the data between XML and database, it is required to establish a correspondence between documented XML schema and database schema, relevant software for data exchange is founded in high level of this correspondence.

In [7], use of XSLT was presented that allows the user before data exchange to convert the data to respective format.

Two XQuery and XPath languages are accepted more extensively for whom working with XML. In XPath language, to select each data, a XML document of Path Expression corresponding thereto is used [8] and XQuery language is mostly similar to SQL and more familiar for programmers [9].

4. THE PROPOSED METHOD

Our proposed method is mapping queries by XML, designing an object-oriented database and applying on the queries. In this paper, the queries can be stored with XML labels. So, labels in XML may be presented for support of object-oriented database.

4.1 Definition of class structure in XML document

To define class structure in XML, respective class must be shown in XML schema. For this purpose, class exhibition must be mapped to DTD schema.
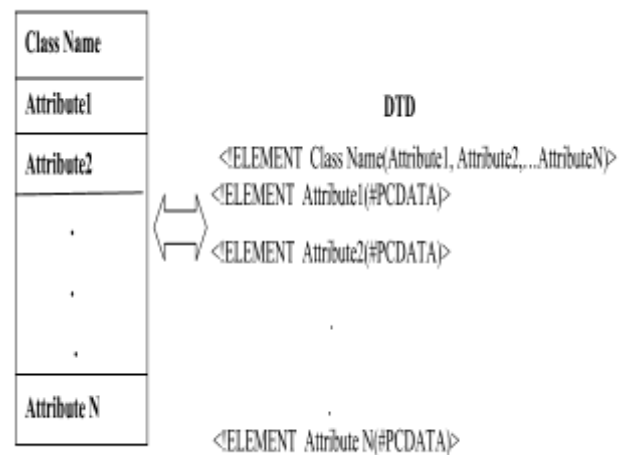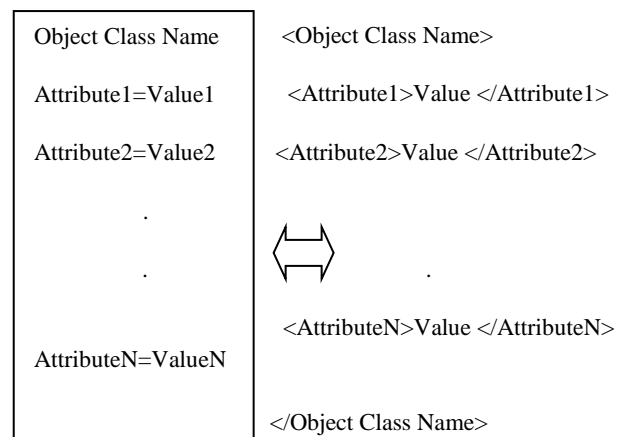

Figure 1. Class mapping by DTD

4.2 Exhibition of samples of class in XML document

Whereas class was formed based on XML schema, at this stage, the defined samples of class may be easily shown in XML document as below:

4.3 Query procedure

Query processing in the object-oriented database refers to a process therein objects of classes are selected that meet the

designated condition. In the method proposed by Zung Ma, it was not possible for the user to prioritize his assumed attributes and only membership rate of each object in the relative class (μ) was analyzed. In the proposed method of this paper, upon allocating weight to queried attributes, setting the priority of each attribute in the query is assigned to the user. Therefore, syntax rule of a SQL query based on object-oriented database will be as follows:

SELECT <attribute list> FROM <class₁,…, classₘ> WHERE <query condition >

In above relation, attribute list includes the attributes that are going to be appeared in the output and related to the objects that meet the user condition. The classes that are written in front of FROM include classes that query operation in performed therein. Query condition is a condition that user expected to be met. Queries related to object-oriented database may be written in XML format that a sample thereof is shown in figure 2:
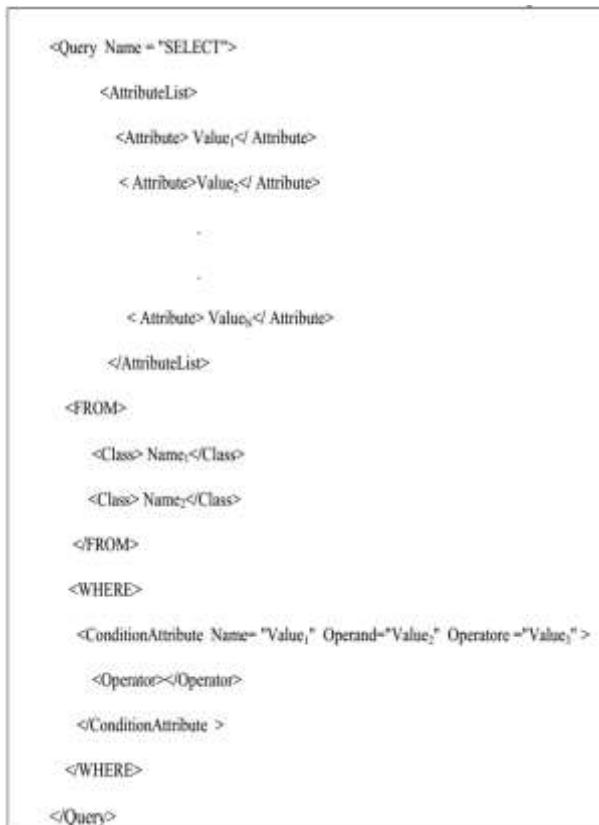

Figure 2. Schema of Select query using XML

5. Case study

In this part, the book information registration system is used for evaluation of proposed method. In this system, the books information is registered so that different books can be searched based on different fields. In addition, considering proposed method, insert , delete, select and edit operation of XML object-oriented database is carried out.

The classes related to book information registration system include two classes in the name of book and Author with the below-mentioned particulars:

```
Class Book {
Int ID;
String Title;
String ISBN;
String Publisher;
Date Date;
double Price;
Float Weight;
String Subject;
Author Author; //Object as Data in Object Oriented Database
}
Class Author{
String FirstName;
String LastName;
}
```

According to the defined classes and proposed method, a mapping to DTD model must be applied to obtain XML document related to respective system, easily.

```
<!ELEMENT Book(ID, Title, ISBN, Publisher, date, Price,
Weight, Subject, author)>
<!ATTLIST Book ID CDATA #REQUIRED>
<!ELEMENT Title (#PCDATA)>
<!ELEMENT ISBN (#PCDATA)>
<!ELEMENT Publisher (#PCDATA)>
<!ELEMENT Date (#PCDATA)>
<!ELEMENT Price (#PCDATA)>
<!ELEMENT Weight (#PCDATA)>
<!ELEMENT Subject (#PCDATA)>

<!ELEMENT Author(FirstName, LastName)>
<!ELEMENT FirstName (#PCDATA)>
<!ELEMENT LastName (#PCDATA)>
```

6. Conclusion

In this paper, a method was presented for showing samples of class in the XML database and so the concept of XML object-oriented database was defined. Furthermore, different queries related to XML object-oriented database was defined and showed. Consequently, the concepts of object-oriented database and its relevant queries can be shown by XML labels.

7. REFERENCES

[1] B. Tim, P. Jean, C. M. Sperberg, M. Eve," ExtensibleMarkup Language (XML)1.0", World Wide Web Consortium (W3C) , (2006), http://www.w3c.org/TR/REC-xml. Fourth Edition, 29 September. (Visited on 2008-02-09).

[2] J. Fong, " Mapping Extended Entity Relationship Model to Object Modelling Technique", SIGMOD, (1995), Record 24.PP.18-22.

[3] A. Borgida, " Modelling Class hierarchies With Contradictions", Proc of ACM SIGMOD International Conference on data management Chicago,US,Vol.17,no.3,(1998), pp.434-443.

[4] A. F. Cardenas, I. T. Leong, R. K. Taira, R. Barker, C. M. Breant, " The Knowleadge Based Object Oriented PICQuery+ Language", IEEE Transaction on Knowledge and Data Engineering, Vol.5, no.4, (1993), PP.644-657.

[5] R. Bourret, " XML Database Products", WWW.rpbourret.com/xml/XML And Databases.htm, (2003).

[6] J. E. Funderburk, " XTABLES:Bridging Relational Technology and XML", IBM Systems Journal, Vol.41, (2002).

[7] J. shanmugasundaram, " A General Technique for Querying XML Documents Using a Relational Database System", Special ISSUE: Special Section on advanced XML data Processing, (2001).

[8] A. Berglund, S. Boag, D. Chamberlin, M. F. Fernández, M. Kay, J. Robie, J. Siméon, " XML Path language (XPath) 2.0", World Wide Web Consortium (W3C), (2007), http://www.w3.org/TR/xpath20 ,. (Visited on 2008-03-17).

[9] S. Boag, D. Chamberlin, M. F. Fernández, D. Florescu, J. Robie, J. Siméon, " XQuery 1.0: An XML Query Language", World Wide Web Consortium (W3C), (2007), http://www.w3.org/TR/xquery, January 2007. (Visited on 2008-03-17).

Improved Strategy for Distributed Processing and Network Application Development

Eke B. O.
Department of Computer Science,
University of Port Harcourt
Port Harcourt, Nigeria

Onuodu F. E.
Department of Computer Science,
University of Port Harcourt
Port Harcourt, Nigeria

Abstract: The complexity of software development abstraction and the new development in multi-core computers have shifted the burden of distributed software performance from network and chip designers to software architectures and developers. We need to look at software development strategies that will integrate parallelization of code, concurrency factors, multithreading, distributed resources allocation and distributed processing. In this paper, a new software development strategy that integrates these factors is further experimented on parallelism. The strategy is multidimensional aligns distributed conceptualization along a path. This development strategy mandates application developers to reason along usability, simplicity, resource distribution, parallelization of code where necessary, processing time and cost factors realignment as well as security and concurrency issues in a balanced path from the originating point of the network application to its retirement.

Keywords: Parallelization, EE-Path, Distribution, Usability, Concurrency

1. INTRODUCTION

The software strategy referred in this work proffers solution to distributed software development by using the abstraction of user requirements and design-time distribution of processes across multi-core computer powers in multidimensional visualization, network development, parallelism and alignment of conceptualization along a path known as the EE-Path [1]. The technique uses ideas in computational geometry in trying to resolve a given network, parallelism and distributed software engineering problem. It is common to see software specified from the view point of the owners and from the ideas of similar existing application. It can also be seen from the point of cost and benefit as well as processing time and computer resources in a combined or peered manner.

Multi-core computers have shifted the burden of software performance from chip designers to software architects and developers. In order to gain the full benefits of this new hardware, we need to parallelize our code [2]. Parallelization, therefore, need a design strategy that can guide the software development process in a distributed system from the inception to the deployment of the software.

The goal of this paper is to use a development strategy (EE-Path) that aligns parallelization, usability, distribution, user requirement abstraction along a balance path during software development. Since we must overcome software complexity paradox to achieve the level of simplicity demanded by users we must think not just along the specified requirements of the user as classical strategy demand but also on the unspecified requirements and machine commitment which belong to the other dimensions in the EE-path. The weakness in the other strategies is their inability to distribute software design and development load across processes and processors as well as the unspecified requirements into the software building plan. They often ignore or allow programmers to take distribution and parallelism responsibility. Problems often arise where programmers depend on the software plan in developing the system.

2. PARALLELIZATION

Parallelism is a form of computation in which many calculations are carried out simultaneously, operating on the principle that large problems can often be divided into smaller ones, which are then solved concurrently, or 'in parallel'. Parallelism is all about decomposing a single task into smaller ones to enable concurrent execution [2]. Usually, processor would execute instructions sequentially, which meant that the vast majority of software was typically written for serial computation. While we were able to improve the speed of our processors by increasing the frequency and transistor count, it was only when computer scientists realized that they had reached the processor frequency limitation that they started to explore new methods for improving processor performance [3]. They explored the use of the germanium in place of silicon, co-locating many low frequency and power consuming cores together, adding specialized cores, 3D transistors, and others. In this era of multi-core processors exploiting large-scale parallel hardware will be essential for improving application performance and its capabilities in terms of executing speed.

Multithreading can be on a single-processor machine, but parallelism can only occur on a multi-processor machine. **M**ultiple running threads can be referred to as being concurrent but not parallel. Concurrency is often used in servers that operate multiple threads to process requests. However, parallelism is about decomposing a single task into smaller ones to enable execution on multiple processors in a collaborative manner to complete one task. Distributed systems are a form of parallel computing; however, in distributed computing, a program is split up into parts that run simultaneously on multiple computers communicating and sharing data over a network. By their very nature, distributed systems must deal with heterogeneous environments, network links of varying latencies, and unpredictable failures in the network and the computers.

3. THE SOFTWARE DEVELOPMENT STRATEGY

The EE-Path software development technique aims at resolving the software development complexity resulting from improper or lack of provision for the Unknown network, parallelism and user requirements at the time of the software specification. An architectural pattern understanding is complex in terms of the three quality attributes: modifiability, performance, and availability. Software architects on the other hand think in terms of architectural patterns [4]. However, what the architecture needs is a global characterization of architectural patterns in terms of the factors that affect quality attribute behaviour so that a software design can be understood in terms of those quality attributes. Software engineers must not shay away from complexities of seemingly intractable parallelism concerns in specifying software requirement analysis and design.

The quality attributes of architectural patterns are the design primitives of the software and they are system independent. In designing software architecture for a product line, the long life and the flexibility of the software must be of paramount importance. The full set of requirements of the system is sparsely known. When the actual products are created there still remain the Unknowns or better still the unknowable in the product line. New users may emerge, newer needs may arise and working environments may change such as operating system, databases, server changes and machine speed improvements, multi-core processor changes. These changes will drive the entire system to reflect the realities of the trend, creating the need for rapid response to such changes. Our software development strategy provides a solution to this need by making architectural provision for the Unknown and also a room for the Unknown in the entire life cycle of the software. In areas where parallelism is previously envisaged dummy checks can be deployed to recover from drawbacks such as system slowdown, thread race conditions and unforeseen dependencies.

The Unknown implies all the unspecified requirements of the software at inception. It also includes all the unforeseen user need that could give rise to the deployment of parallelism such as video and heavy image inclusions in software. It seems that irrespective of the software development method used, it is the user or software client that specifies what the software is to do. Irrespective of the way it was specified or the way the information is collected the target of the software will depend largely on what the users or software clients actually want whether they know what they want or not. It is also true that in most cases the users do not know how to specify the details of what they want even when they are well consulted. Some of their specifications are capable of compromising speed, multi-core processor efficiency and concurrency. Clients may not be software gurus and may not specify the software requirement to the extent that all requirements are covered. Even where all requirements are covered, external environmental factors such as operating system changes, network expansion or upgrades and database upgrades, introduction of new data for processing, new formulas as well as security loop holes may make the software vulnerable, and the need to update the software based on the new requirements may arises. These unforeseen requirements we generally refer to as the **Unknown**.

The capturing of the Unknown involves software abstraction embedded in the conceptual architecture of the system. The conceptual architecture is one of four different architectures identified by Hofmeister, Nord and Soni [5]. It describes the system(s) being designed in terms of the major design elements and the relationships among them. The EE-Path strategy determines the balance of the known architectural drivers, the known environmental factors, the known parallelism conditions, the known user simplicity factors as well as all other hidden factors-(Unknown). The software is then built along this path at least conceptually. A model of the EE-Path strategy is illustrated in figure 1.

The architectural drivers are the combination of business, quality and functional requirements that "shape" the architecture. The known architecture drivers are represented in the y-axis while the unknown architectural drivers are represented in its shadow as Architectural drivers 2. Similarly, other well known parallelism requirement specification are represented in the z-axis while their unknown is also represented using its envisaged shadow as Others RS 2. In analysis, design and construction the EE-Path takes all the axis into consideration as providing the necessary balance it requires to remain on its path of move as the software tends to retirement. The EE-Path will terminate at a point when the software peters out, but the issue of when this will take place also throws up another unknown.
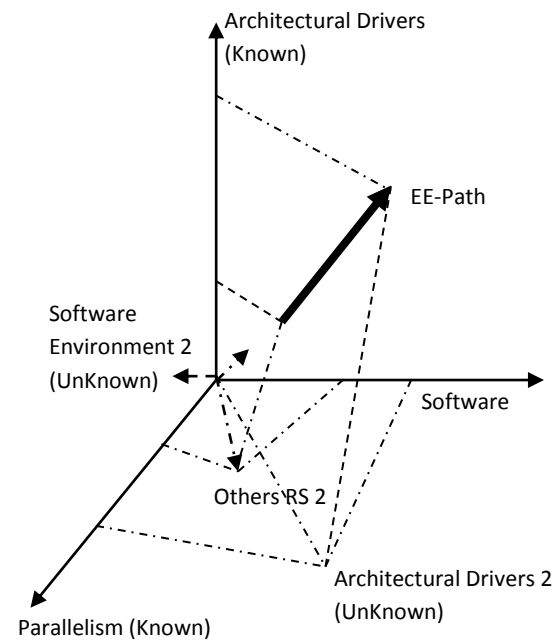
Figure 1: The EE-Path Software Strategy Model

Newer upgrades are more likely to surface with additions of parallelism requirements that were hitherto unknown at the earlier versions when the software first hit the market. In integrating parallelism, two types of data parallelism are considered:

- Explicitly Data Parallelism
- Implicitly Data Parallelism

In Explicitly Data Parallelism one just plans a loop that executes in parallel. This can be done by adding OpenMP

pragmas around the loop code, or using a parallel algorithm from **Intel® Threading Building Blocks (TBB)**, from other source or by developing one.

In Implicitly Data Parallelism one just call some method that manipulates the data and the infrastructure (i.e. a compiler, a framework, or the runtime) that is responsible for parallelizing the work. For instance, the .NET platform provides LINQ (Language Integrated Query) that allows the use of the extension methods, and lambda expressions to manipulate the data like dynamic languages. The following example demonstrates implicit data manipulation and parallelism:

C# implicit data manipulation using LINQ
string[] students = { "Bartho","Yuntho", "Barry","Friday"};

var student = students.Where(p => p.startsWith("B"));

C# parallel implicit data manipulation using LINQ (Note the AsParallel method)

string[] students = { "Bartho", "Yuntho", "Barry","Friday"};

var student = students.AsParallel().Where(p => p.startsWith("B"));

In language and compiler-based parallelism, the compiler understands some special keywords to parallelize part of the code; for example, in OpenMP you can write the following to parallelize a loop in C++ [6]:

```
#pragma omp parallelfor
for ( int j = 0; j< max; j++)
{
     Num[j] = 1.0;
}
```

Language and compiler-based parallelism is easy to use because the majority of the work falls on the compiler. In library-based parallelism, the programmer should call the exposed parallel APIs. For example, if one want to parallelize a for loop in .NET 4 (C#, or VB) a call on For method from the System.Thread.Parallel class will suffice in C#:

```
 Parallel.For(0,max,j =>
{
    Num[j] = 1.0;
});
```

This method accepts two integers (from, to) and delegates to the loop body.

4. EE-PATH IN THE APPLICATION LIFE CYCLE

The EE-Path is a very flexible strategy. It is suited for complex; highly interactive applications, where very high integration is required providing good utilization of underlying hardware within the network and within the multi-core machine. The strategy promotes reusability of application components and possibly performance since design components planned as unknown is reused when the requirement gets clear. Software requirements have functional both abstract and concrete, quality and business constraints. The abstract requirements are used to generate the software design while the concrete requirements are used to validate

the decisions made as a result of the abstract requirements [7]. The EE-Path remains a guiding path which the requirements need to follow during the specification. The path is not introducing any requirement but it provides a structure and a reference point in the specification of the software requirements. The use case has the functionality in the system that gives a user a result of value and captures the functional requirements [7]. The use case therefore needs to be projected along the EE-Path to be able to reflect both the known and the unknown requirements.

Klain [9] believes that the choice of architectural style is based on the architectural drivers for the design elements that fit the need at hand. We however believe that architectural style should not just be based on the need at hand but also on envisaged need and the unknown future needs. These unknown needs should be represented using any appropriate representation in the architecture. Design consideration also must take into account the specified unknown so that the unknown can be well specified at least at the abstract component design level where commitment is yet to be made to actual software components. The unknown is therefore well represented in the modular design and aggregated in the object-oriented class abstraction even if the abstraction is at worst a dummy. The class abstraction has an inert effect at making sure some force is exerted to keep the software development effort on the EE-Path. In the path, the logical, process, implementation and deployment views are realigned with the parallelism views even when the software is targeted at a standalone machine. The standalone can be multi-core and can equally migrate easily to multi-user when new requirements surface. This path alignment boosts the modifiability of the software even when it is already deployed.

5. DISCUSSION OF EE-PATH BENEFITS AND PARALLELISM

The EE-Path strategy increases the usability of software since aggregation is encouraged by patterning one or more actions on more than one object, even when the object is unknown. It also makes the system, rather than the user, responsible for iteration. Furthermore, it is very easy to recover from failure since the unknown is taking into consideration right from the architectural stage of the software. Recovery could easily be based on the unknown functionality of the environment, such as OS failure and machine failures and even unknown dependency conditions in parallelized system implementation.

In order to take advantage of the EE-Path in software development specification for multi-core machines, programs must be parallelized. Multiple paths of execution have to work together to complete the tasks the program has to perform, and that needs to happen concurrently, wherever possible and in an integrated manner with other requirements. Only then is it possible to speed up the program . Amdahl's law expresses this as [10]:
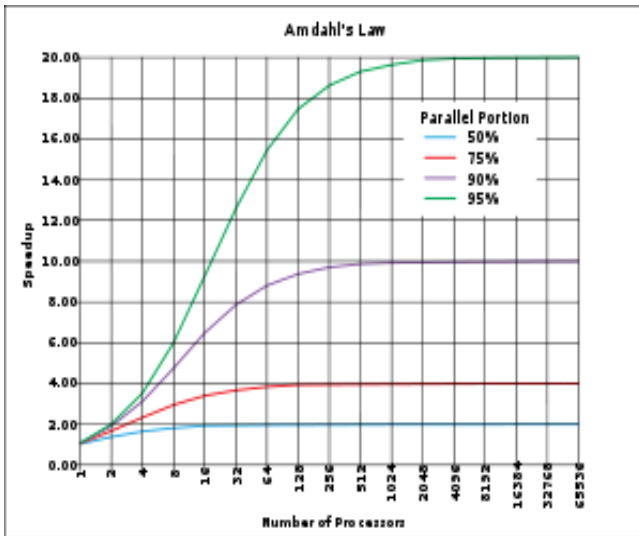
$$S = \frac{1}{1 - P}$$

Figure 2 : Illustrating Speedup and Number of processors

where *S* is the speed-up of the program (as a factor of its original sequential runtime), and *P* is the fraction that is parallelizable.

Determining when and where in the software to inject parallelism is a challenge and if wrongly decided could have retrogressive consequences hence most of decision could be provided as unknown at certain stage of the system. The good thing in the EE-Path is that strong provision is made for its implementation at worst as a dummy implementation. This will help developers to plan ahead even when it is not feasible to implement it at the earlier releases of the software. One probably do not need to parallelize if the application is really simple and the code is running fast enough already. But we know that a simple application today may turn to a complex application with time and a fast code could slow down when new users use it in a network or when newer features are added hence the need to plan for the unknown via the EE-Path. Network applications use shared data, hence dependency issues can be planned using the EE-Path to take other factors into considerations to avoid pitfalls of delays as a result of dependency and thread locks. Decisions when made at the planning stage guides developers in the choice of parallel frameworks and APIs to use during the application implementation. These will help in leveraging the power of all the extra cores on developers and users machines. The EE-Path strategy encourages developers to leverage their knowledge and also to develop systems in relatively unfamiliar parallelized contexts as offered by distributed application environments. The unknown is not fixed but it remains the unknown as long as it has not been unraveled and since no human can have full and final insight of any matter at any given time, progressive development is encouraged by our strategy.

6. OUR CONTRIBUTIONS.

In this paper, we incorporate a new software strategy which is able to implement a multidimensional requirement visualization of three or more lines of simultaneous requirement alignment. It inculcates the unspecified requirement that we see as forming the core of modern system design and allow parallelism requirement analysis and design to varying level of implementation. When the requirement is not needed at the moment we postulate it can be allowed to be implemented as an abstract class in the system without any derivation or with dummy derivation. Some of the requirement issues to be considered include security, concurrency, interoperability, reusability and the Unknown. The Unknown class can be specified with all possible abstraction that can be modified in the future when the need for the Unknown requirement arises. This design technique takes care of the Unknown making the system to be extendible without the need to redesign the system. This design technique takes care of the light-speed changes in requirements resulting in the development of newer versions of software within very short period of time ranging from few days to few months. It breaks parallelism conditions in the software requirement to determine where it can be implemented to maximize speed. It also articulates pitfalls to avoid deployment of parallelism to those areas where parallelism could lead to processing slow-down or incorrect generation of result. There are many parallelism frameworks, and debugging tools aimed at simplifying the task of parallel programming, such as:

Intel Parallel Studio, Microsoft CCR and DSS, MS PPL - Microsoft Parallel Pattern Library (was released in 2009 Q4), MS .NET 4 - Microsoft .NET Framework 4 (will released in 2009 Q4), Java 7 (will release in 2009), PRL - Parallel Runtime Library (Beta 1 released in June 2009) [2]. Software engineers need to integrate this entire requirement in system development early in the system life-cycle while making provision for the unknown.

7. CONCLUSION

The EE-Path software development strategy provides a means of guiding developers and software architects in qualitative measures of marginal building blocks in choosing and developing architectural styles and in conceptualization of the system at hand from the inception to the conclusion. Based on the evaluation of software complexity and other models it can be seen that if the EE-path is followed, a better preparation for the unknown is made. Furthermore, it can be seen that for the parallelization of network application the EE-Path model offers variables for consideration and integration of other factors and requirements in the development of software in a hitherto different network platforms. These provide improved standardization of development even at the architectural level of software development. It can therefore be concluded that developing network software using the EE-Path concept results in building a software today with provision made for **change** which itself appears to be a constant in the world of software engineering.

8. REFERENCES

[1] Eke B. O. and Nwachukwu E. O. (2011), Software Engineering Process: Yaam Deployment in E-Bookshop Use Case Scenario, Journal of Theoretical and Applied Information Technology, Vol 30 No. 2 August, 2011, JATIT and LLS, www.jatit.org E-ISSN:1817-3195, ISSN:1992-8645, Islamabad,Pakistan, http://www.jatit.org/volumes/Vol30No2/2Vol30No2.pdf

[2] Fadeel, H. (2009) Getting Started with Parallel Programming, http://library.dzone.com/category/ Accessed, June 2015.

[3] James Reinders (2009) Getting Started With Intel Threading Building Blocks, http://library.dzone.com Accessed July 2009

[4] Katzman, R.; Barbacci, M.; Carriere S. J.; and Woods, S. J. (2014) Experience with Performing Architectural Styles, Software Architecture Tradeoff Analysis. 54-63. Proceedings of ICSE99. Los Angeles, CA, May 1999.

[5] Hofmeister, C.; Nord, R.; and Soni, P. (2000) *Applied Software Architecture.* Reading MA: Addison Wesley.

[6] James P. C.; Jack W. D.; C++ Program Design An Introduction to Programming and Object-Oriented Design WCB McGraw-Hill USA

[7] Bass L.; Klein M.; Bachman F. (2000) Quality Attribute Design Primitives (CMU/SEI-2000IN-2000-017). Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University.

[8] Jacobson, I.(1992); Object-Oriented Software Engineering, Addison-Wesley, USA

[9] Klein M.; Kazman R.; Barbacci M.; Carriere S.;and Lipson, H.(1999) Attribute-Based Architectural Styles, Software Architecture. 225-243. Proceedings of the First Working IFIP Conference on Software Architecture (WICSA1), San Antonio, TX, February 1999.

[10] Wikipedia, Parallel computing, www.en.wikipedia.org/wiki/parallel_computing accessed July 2014

7. ACKNOWLEDGMENTS

8. ABOUT THE AUTHORS

Eke Bartholomew PhD, MCPN, MACM, FIPMD is a Software Engineering / Computer Science Lecturer at the University of Port Harcourt and Mobile Application Developer in Oyol Computer Consult Inc. His research interest is in SE Methodologies and Mobile IT deployment.

Dr. Onuodu, Friday E. is a Lecturer at the University of Port Harcourt. His research interest is in Data Mining and Data Extraction using both Mobile Devices and Desktops. He also has interest in IT deployment analysis. He has many publication in learned journals.

A Survey of Existing Mechanisms in Energy-Aware Routing In MANETs

Nazila Majdkhyavi
Departement of computer,
Meshkin branch,
Iclamic Azad University, Meshkin, Iran

Raziyeh Hassanpour
Departement of computer,
Urmia branch
Iclamic Azad University,Urmia, Iran

Abstract: A mobile ad hoc network (MANET) is a distributed and Self-organized network. In MANET, network topology frequently changes because of high mobility nodes. Mobility of nodes and battery energy depletion are two major factors that cause loss of the discovered routes. battery power depletion causes the nodes to die and loss of the obtained paths and thus affects the network connectivity. Therefore, a routing protocol for energy efficiency should consider all the aspects to manage the energy consumption in the network. so introducing an energy aware routing protocol, is one of the most important issues in MANET. This paper reviews some energy aware routing protocols. The main purpose energy aware protocols are efficiently use of energy, reducing energy consumption and increasing the network lifetime.

Keywords: MANET, routing, energy, transmission power, lifetime

1-INTRODUCTION

MANET is a collection of independent mobile nodes. In such a network, nodes can move freely and randomly. Network topology frequently changes and is unpredictable. This type of networks are limited in terms of resources such as power, memory and processing power [1].

Routing data packets in the network according to specific features of MANET, requires special mechanisms. In MANET, due to mobility of nodes and depletion of battery of nodes, the obtained paths are destroyed and data forwarding faces with failure and delay. In the event of any of these two events, the obtained routes will not be valid anymore and therefore the route discovery phase should be done again.

Re-runnig path discovery, imposes additional costs to the network, delays data forwarding and consumes network resources such as power and bandwidth. In the most of the existing routing protocols, there is a high probability to frequently select a node for data forwarding. This leads to high depletion rate of the node and cause the node to die and thus affects the network connectivity.

A routing protocol in ad hoc networks must be efficient in terms of energy consumption. This protocol must consider all aspects and factors of ad hoc networks to save energy. This is due to the nature of wireless networks which are battery constraint. Dynamic source routing (DSR) [2] and Ad hoc on demand distance vector (AODV) [3] routing protocols have been developed as the basic and most popular routing protocols by the Network Community for MANETs. However, these protocols have not been developed for energy efficient routing. Feeney [4] showed that these protocols are not optimized for energy efficient

routing. Energy efficient routing have entirely different criteria than traditional metrics such as end to end delay and throughput and so on. So many researchers based on these protocols, have been proposed their own energy aware protocols.

Routing protocols in MANET are classified to reactive, proactive and hybrid ones. In this paper we focus on reactive routing protocols. The most famous of reactive protocols are DSR and AODV protocols. Most of the energy aware protocols consider these protocols as the underlying protocol to provide a new energy aware protocol. For example, the EDDSR protocol [8] is based on DSR and EA_AODV [17] and EEAODR [14] protocols are based on AODV. These protocols modified discovery/reply/maintenance phases of the traditional routing protocols to provide energy efficiency features.

The remainder of the paper is organized as follows. section 2 introduces some basic concepts about energy issues in MANET. section 3 introduces famous reactive protocols in MANET. Section 4 presents an overview of energy aware routing protocols in MANET and section 5 presents the conclusion of this paper.

2-BASIC DEFINITIONS FOR ENERGY AWARE ROUTING

In ad hoc networks, nodes are powered by batteries. Energy, communication and computational capacity at each node are limited. Nodes that lose all their energy, can not recharge their batteries and thus are removed from the network. For the below reasons, energy consumption management must be done:

−Limited energy storage: the nodes have limited battery power.

−**Battery replacement problems**: in situations such as floods, earthquakes and battlefield, recharging or battery replacement is very difficult.

−**lack of a central coordination**: Due to the distributed nature of these networks and the lack of a central coordinator and multihop routing, some nodes act as relaying nodes and if the traffic load is heavy, it can consume a lot of energy of the relaying nodes.

-**Battery resource constraints**: node weight may increase by increasing the weight of the battery. By decreasing the weight of the battery, battery life may be reduced. So battery management techniques should be considered in these cases.

-**choosing the optimal transmission power**: with increasing the transmission power, the amount of battery consumption is also increases. Optimal transmission power reduces the interference between the nodes and therefore increases simultaneous transmissions.

High lifetime of the nodes can be achieved by increasing the capacity of the battery of nodes. Increasing the battery capacity of the nodes can be achieved with battery or power management.

2-1 Energy Model

Each node in the MANET consumes energy in four states:

- send
- Receive
- Idle
- Sleep

Suresh Singh, Mike Woo, and CS Raghavendra [5] described five energy aware criteria that affect the energy aware routing:

-**minimizing the energy consumed per packet**: This measures minimizes the average energy consumption per packet.

-**Maximizing network partitioning time**: The purpose of this measure is to increase the network lifetime.

-**Minimizing difference in power levels of nodes**: This criteria ensures that all the nodes in the network stay alive and work together.

-**Minimizing cost per packet**: This metric minimizes total cost of sending a packet.

-**Minimizing the maximum cost of the node**: This metric minimizes the maximum cost of the node.

2-2 Energy Aware Routing

Energy aware routing in MANET is of great importance. The

main objective of energy aware routing protocols are reducing energy consumption in packet sending between source and destination, avoiding routing packets through nodes with low energy levels, optimizing the routing information in the network and increasing the network lifetime.

3- REACTIVE ROUTING PROTOCOLS

In reactive routing protocols for reducing overhead, the path between two nodes is established only when is needed. The most popular protocols in this area are DSR and AODV protocols. Most of the proposed energy aware protocols, have used DSR and AODV protocols as the underlying protocol. These protocols, with make some modification in route discovery/selection and in some cases route maintenance phase of the above mentioned protocols, have been added energy aware measures to these protocols.

3-1 DSR Protocol

In the DSR protocol, the source node generates a route request (RREQ) packet and inserts the source and destination address in the packet and broadcast it to all its neighbors. Each node receiving a RREQ packet, if did not know any path to the destination, inserts its own address into the path list of the packet and broadcast it. Upon recieving this packet on the destination, destination node creates a route reply (RREP) packet and sends it back to the source. When a node could not send a data packet to the next node on the path to the destination, generates a RERR packet and returns it to the source. The source node receiving RERR packet noticed a disconnection between the two nodes and route discovery phase is resumed.

3-2 AODV Protocol

In the AODV routing protocol, the routing operation is done hop by hop using routing tables. When a source wants to send data to a destination, if there is no route to the destination in its routing table, the RREQ message is broadcasted to all neighbors. Intermediate nodes can reply the RREQ packet if they have any path to the destination. To identify the fressness of the path provided by intermediate nodes, a sequence number are contained in the RREQ packets. When RREQ packet reaches the destination, destination node, sends a RREP packet back to the source using the reverse route. When traveling RREP packet on the reverse path, each intermediate node on the path establishes a forward path to the node from which the RREP packet is received.

4 ENERGY AWARE ROUTING PROTOCOLS

Energy aware routing in ad hoc networks is done in several ways. Some approaches try to minimize energy consumption during the route discovery/reply/maintenance. In this section some energy aware routing protocols for MANETs is presented.

4.1 Energy Aware Routing Protocols Based on DSR

The DSR Protocol that is a source routing protocol, is one of the most popular protocols in MANETs. However, this protocol is not energy efficient. So many researchers with changes on this protocol phases and to improve its performance in terms of energy consumption, have added energy measures to the protocol. Some of these protocols are described below.

4-1-1 Energy aware routing based on the energy drain rate

In energy aware routing mechanisms based on drain rate (MDR) [6], a new cost function is presented to predict the lifetime of nodes. In this method, each node in the network includes its current cost in the received RREQ packet. During a certain interval T after receiving the first RREQ, the destination node, collects all the subsequent RREQs. When timer expires, the destination node chooses the routing path using Min-Max algorithm. Each node calculates its drain rate (DR_i) every T seconds according to DR_{old} and DR_{sample} (old and new drain rate) values as relation (1):

$$DR_i = \alpha \times DR_{old} + (1 - \alpha) \times DR_{sample} \qquad (1)$$

In this protocol, the cost function $C_i = RBP_i/DR_i$ specifies the lifetime of node i in which RBP_i is the remaining battery power of node i. MDR mechanism chooses route with the highest lifetime among all the possible routes.

4.1.2 Local Energy Aware Routing Protocol

In the local energy aware routing protocol (LEAR-DSR) [7] which is based on DSR, if the remaining battery power of a node is higher than a certain threshold, the node will participate in the route discovery. Otherwise, the node drops RREQ packet and generates a new message, called drop route request (DRREQ) to inform other node about the droped RREQ. Thus, the RREQ packet will reach destination only when it has passed through nodes with higher energy than the specified threshold.

The path inserted in the RREP packets is the shortest path among all the paths that have sufficient energy. Each node has already removed a RREQ packet or has received a DRREQ packet, reduces amount of its battery threshold using a specific adjustment value and only if its residual energy is higher than the new threshold, will broadcast the RREQ packet.

4.1.3 A Novel DSR-Based Energy Efficient Routing Protocol

In the DSR-based energy efficient routing protocol (EDDSR) [8], each node based on its current level of battery power, calculates its willingness to participate in routing packets. EDDSR mechanism tries to discourage nodes that have low lifetime from participating in the route discovery and as a result

increase node and network lifetime.

In this protocol, each node in the network periodically computes its remaining battery power (RBP) and if it has enough battery power, will participate in the network operations. When the remaining battery power of a node is below a certain threshold, the node delays broadcasting of the received RREQ. This operation is done by calculating a time period which is inversely proportional to its expected lifetime i.e. $C_i = RBP_i/DR_i$. In the route maintenance phase of EDDSR, when the energy level of the node reaches below a critical threshold, node sends a RERR packet to the source.

4-1-4 Energy Aware Multi-Path Source Routing Protocol

Energy aware multi-path source routing protocol (MEA-DSR) [9], is based on DSR protocol. In this protocol two path from source to the destination node are found and are stored in the routing table. The primary route selection in MEA-DSR protocol is done using residual energy of nodes on the path and the total transmission power required to transmit data. About selection of second route, the path with high disjointness ratio from the first path has high priority. A new field called min_bat_lev which stores the minimum residual energy of nodes on the path is added to the route table and RREQ packets.

In the MEA-DSR, only the destination node can responds to the RREQ packet. When an intermediate node which is located in the neighborhood of the source node, recieves the RREQ packet, inserts its remaining energy in the min_bat_lev field. Otherwise, each intermediate node compares its residual energy with the min_bat_lev value contained in the packet. If it is smaller, replaces min_bat_lev value by its remaining energy. After receiving first RREQ, the destination node waits for more RREQs. After that, the destination node, choose a primary route. After selecting the primaary route, the destination node immediately sends a route reply packet to the source node.

4-1-5 Energy Aware Multi-Path Routing Protocol

Energy aware multi-path routing protocol (EMRP) [10] which is derived from DSR protocol, applies some changes in the phases of DSR protocol. In the route reply phase of EMRP, each node on the path, inserts its current status such as the distance between the node i with the next hop, number of retransmission attempts related to the last successful transmission, current length of queue and its current energy in the RREP packet. Firstly, the source node calculates weight of all possible paths using relation (2):

$$W = \sum_{i=1}^{n}(\alpha \times W_{energy}^{i} + \beta \times W_{queue}^{i}) \qquad (2)$$

In relation (2), W_{energy}^{i} is the energy weight of node i and W_{queue}^{i} is the queuing weight of node i. the source node after calculating the weights of all paths, sortes the routes in descending order based on W and selects the best N sets of paths as primary path for sending real-time data and the N next sets of

paths as backup paths.

4-1-6 Power Aware Source Routing Protocol

In the Power Aware Source Routing Protocol (PSR) [11], during the route discovery procedure, the path π at time t is choosen so that the cost function (3) is minimized:

$$C(\pi, t) = \sum_{i \in \pi} C_i(t) \quad \text{where} \quad C_i(t) = \rho_i \cdot \left(\frac{F_i}{R_i(t)}\right)^\alpha$$
(3)

In relations (3), ρ_i is the Transmission power, F_i is the capacity of battery, R_i is the remaining battery of node i at time t and α is a weighting factor. In the PSR protocol, all nodes except the destination, calculate their link costs (C_{ij}) and insert it in RREQ packet. When an intermediate node receives a RREQ packet, it sets a timer T_r and holds the cost in the packet header as Min-Cost. If another RREQ packets reach, their costs are compared with the saved Min-Cost value. If new packets cost is less than saved Min-Cost, Min-Cost is changed to the new value and new RREQ packet are sent. Otherwise, the new RREQ is droped. In PSR, after receiving the first RREQ, destination node waits for a threshold T_r. When this timer expires, the destination node chooses the path with the least cost and responds to the source.

4-1-7 Energy Aware and Reliable Routing Protocol

In the Energy aware and reliable routing protocol (EARR) [12], each node depends on its residual energy and traffic requirements of higher layer decides whether to send RREQ packets or not. Only when its residual energy is higher than the required energy, will send RREQ packets. In this protocol, only when the intermediate nodes of the route have enough energy, the destination will receive the RREQ packet.

So only the valid paths are selected as candidate paths and this will avoid reconstruction of path because of lack of enough battery power and thus the life of the network will be increased. In this protocol, when intermediate nodes decide to send RREQ, insert their current energy in the packet. In the route cache of EARR, remaining energy also is stored. In this protocol if multiple candidate paths exist, the path with maximum battery capacity is selected for data transmission.

4-1-8 Reliable Energy and Signal Strength Aware Routing Protocol

The main objective of Reliable energy and signal strength aware routing protocol (SEA-DSR) [13] is to improve the reliability of the discovered path with taking into account both signal strength and remaining energy of nodes. This protocol reduces link failure and increases network lifetime by distributing traffic load. The benefits of this protocol are reduction of path failure, packet loss, route discovery and routing control overhead.

SEA-DSR protocol, adds an additional field called trust count (RELCOUNT) to the RREQ packet header of DSR protocol. After receiving first RREQ packet, the destination sets a timer and stores all RREQ in its route cache. After the timer, chooses the path with high reliability factor and sends a RREP packet for that. Reliability factor for the path is calculated using relation (4).

$$\text{Reliability Factor} = \frac{\text{Reliability Count}}{\text{No.of Hops}}$$
(4)

4-2 Energy Aware Protocols Based on AODV Protocol

AODV routing protocol is a reactive routing protocol which only holds the next hop for each destination in the routing table. In this section an overview of the multiple energy aware routing protocols based on AODV are performed. The protocols for improving the efficiency of AODV, added energy measures to the protocol.

4-2-1 Energy Efficient Routing Protocol

In the energy efficient routing protocol (EEAODR) [14], an optimization function based on time, the amount of energy of nodes and number of hops factors decides which path is the best path. In this protocol, if the path has one node with low energy level, the optimizer function will not choose it. if time is low, the route will be shorter and less energy will be consumed. but some nodes may die too quickly. Therefore optimizer function should do a trade off between time and network/node lifetime.

In this protocol when the destination receives first RREQ, waits for a while and collects all the subsequent RREQs. After this time, the destination uses the optimizer function to choose the best path and inserts it in the RREP packet. The destination also chooses some backup routes to avoid wasting energy and time for re-calculation of paths.

4-2-2 Distributed Energy Efficient Routing Protocol

Distributed energy efficient routing protocol (DEEAR) [15], obtains the necessary information for energy aware routing only using RREQ packet and requires only battery power level of the network. In DEEAR, intermediate nodes control rebroadcast time of RREQ packets and nodes with relatively high battery power broadcast the RREQ packets sooner. DEEAR protocol does a trade off between min-hop path and the path with balanced energy consumption.

To estimate the average power, average battery energy of nodes on the path (i.e. P) (source and intermediate nodes) and number of hops (i.e. N) fields are added to the RREQ packet. To select the optimal route, the source, adds the amount of its battery to the P field and broadcast the RREQ. Average remaining battery, P new, of the path is calculate as:

$$P_{new} = [(P_{old}*N)+B_i]/(N+1)$$

(5)

In the relation (5), B_i is remaining battery power of node i and P_{old} is the amount of P field in the RREQ packet. With receiving a RREQ packet, node i calculates average remaining energy (i.e. E_{new}) based on the prior average remaining energy of network (i.e. E_{old}) and P_{old} using relation (6):

$$E_{new}=[(1-\alpha*E_{old}) + (\alpha*P_{old})] \qquad (6)$$

4-2-3 Energy Based QoS Routing Protocol

In energy-based QoS Routing Protocol (EQRP) [16], the problem of QoS routing is considered. According to this protocol, nodes based on their energy, decide how long delay the RREQ packet transmission. delay calculation is defined as:

$$W_e = 0, E_r > E_{max} \qquad , \qquad W_e = \frac{E_i}{E_r}$$
(7)

In the relations (7), E_r is current residual energy of node, E_i is initial energy of node I and E_{max} is the maximum energy for each node. The key idea is that EQRP choose the nodes on the path based on their residual energy. Due to the energy depletion of nodes, the goal of this protocol is to reduce route re-construction.

4-2-4 Integrated Energy Aware Routing Protocol

In the integrated energy aware routing protocol (EA_AODV) [17], when each intermediate node receives the RREP packet, estimates its remaining battery power by using the energy available at the node and the required power to transmit a packet at the node and appened it in the RREP packet. Each node on the path, estimates its battery and only if it is less than the value contained in RREP packet, it will be replaced. If multiple paths reach the source, the path which its minimum battery energy is maximum is selected.

Available energy level and power level are taken into account when making routing decisions. In this protocol, source node chooses the path based on load balancing method to balance the battery depletion and applies transmission power control during data sending.

4-2-5 Energy Aware Routing Protocol

In the Energy Aware Routing Protocol (AODVEA) [18], routing is based on the minimum residual energy. Node with the minimum residual energy is marked and the path that has the largest minimum residual energy is selected. For this purpose, a new field called residual energy (Min-RE) is added to RREQ and RREP control packets.

This field, shows the node with the minimum residual energy. In this protocol, when the source receives the first valid

RREP packet, starts the communication. Once the source node receives the next RREP packet starts the main algorithm. Sends RREQ packets to all its neighbors. Gets various routes to destination. Compares parameters of routes with respect to the remaining energy level and the minimum number of hops and then chooses the right path to the destination.

4-2-6 Dynamic Energy Aware Routing Protocol

Dynamic energy aware routing protocol (EA-DYMO) [19], works based on the DYMO protocol. DYMO is a successor of AODV protocol. DYMO is an energy and load aware routing protocol that tries to provide a mechanism to select the optimal route based on the residual energy and traffic load of each node. In this protocol, route selection factor is calculated as the ratio of average energy to the average traffic load. The path which has maximum PS value is chosen for sending data. These two factors in relation (8) are combined to calculate the route selection factor:

$$PS_{sdi} = \frac{EF_{sdi}}{LF_{sdi}} \qquad (8)$$

In the relation (8), EF_{sdi} and LF_{sdi} are the average power and average traffic load of path i from source S to destination D.

4-2-7 Reliable and Energy Efficient Routing Protocol

Reliable and energy efficient routing protocol (E2R2P) [20], uses a new concept called backbone nodes, which are nodes with optimal signal strength. This project focuses on the concept of battery power and signal strength or required energy for the routing process. Backbone nodes helps in the construction phase and finding new routes. Each node has a routing table entry for the number of backbone nodes connected to it, its battery status and signal strength.

when any of the nodes on the communication path is damaged or move out of communication range, the backbone nodes can take care of routing process. Backbone nodes broadcast their signal strength and power status to neighboring nodes. In this protocol, selection process is based on power and signal strength factors. Although the path may be longer, but the chosen path is very stable and reliable to deliver the data packets.

4-2-8 Reliable and Energy Aware Routing Protocol

Reliable and energy aware routing protocol (TEA-AODV) [21], uses battery power and reliability value of each node for a reliable and trusted routing. Trust value of each node is determined based on various parameters such as length of the association, ratio of the number of packets sent successfully by neighbors to the total packets sent to that neighbor and the average time spent to response to a RREQ.

In the route discovery phase, the source node broadcast RREQ packet containing the values of trust and reliability of source. Upon receiving RREQ packet, each intermediate node measures its reliability and adds it to RREQ packet. When two or more RREQs reaches the destination, destination finds the most reliable path using average reliability. The average reliability is obtained using relation (9):

Avgerage reliability=(Cumulative reliability/Number of hops) (9)

4-2-9 Local Energy Aware Routing Protocol

In the local energy aware routing protocol (LEAR-AODV) [22], each node based on its remaining battery power E_r, decides whether transmit the received RREQ packet or not. If the battery is more than a certain threshold, RREQ packet is sent, otherwise the packet is droped. If all nodes on the path have enough energy, RREQ packet will reach its destination. In the route maintenance phase if the battery is below a threshold, a RERR packet is sent to the sender node that this is done locally in the node sending the RERR packet. In this protocol, when a node removes a RREQ packet, it will broadcast a new message called ADJUST_Thr. Next nodes towards the destination find out a RREQ has been removed and adjust their threshold value.

4-2-10 Lifetime Prediction Routing Protocol

In the Lifetime prediction routing protocol (LPR-AODV) [22], the path with high lifetime is chosen. This protocol uses battery lifetime prediction. Each node estimates its battery lifetime based on the past activity. In the route discovery phase, all nodes except the destination and source calculate their estimated lifetime. An additional field called Min-lifetime is added to each route request packet which shows the minimum lifetime.

In the route selection phase, the source node chooses the path with maximum lifetime. In this protocol when the estimated lifetime reaches below a threshold, node sends a RERR packet to the source. This decision is only based on node residual energy and its depletion rate.

4-2-11 Load Aware Routing Protocol

Load Aware Routing Protocol (ELB-MRP) [23], uses collision window size and queue size to calculate load at the node and its single hop neighboring nodes. Each node except source and destination, collects information about the collision window size (ACW), power factor (EF) and queue factor (QF) for itself and its single hop neighbors. The cost for node i is calculated using relation (10):

$$C(i) = ACW * EF * QF$$
(10)

To discover Paths, intermediate nodes add their traffic and energy information to the RREQ packets. Hello packets also collect information about ACW, EF and QF values. Using the neighbor information obtained from hello packets, intermediate nodes calculates the cumulative cost using relation (11) and adds it to the RREQ packet. If a node i, has neighbors j and k, the cumulative cost of CC(i) is obtained using the relation (11):

$$CC(i) = C(i) + (C(j) + C(k)) \qquad (11)$$

When the first RREQ reaches the destination, destination waits for more RREQs. Then destination, selects two paths with least cost as primary and backup path.

5 CONCLUSIONS

In this paper, an overview of recently proposed routing protocols which conidered energy metric of nodes to improve routing operation is introduced. no routing protocol can not provide the best performance in all the different scenarios and the performance of a network depends on the condition and parameters of the network. parameters such as node mobility and QoS features are the most important factors in ad hoc networks. In this paper we focused on the characteristics of the nodes energy and reviewed the protocols that considered energy measure as the main parameter.

6 REFERENCES

1] Subir Kumar Sarkar, T G Basavaraju and C Puttamadappa, "Ad hoc mobile wireless networks, principles, protocols and applications".

[2] D.B. Johnson, D.A. Maltz, "Dynamic Source Routing in Ad Hoc Networks", Mobile Computing, T. Imielinski and H. Korth, Eds., Kulwer, 1996, pp. 152-81 protocol for mobile ad hoc networks (RDMAR)", CCSR ,UK.

[3] C.E. Perkins, E.M. Royer, "Ad-Hoc On Demand Distance Vector Routing", Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Application New Orleans, LA, February 1999, pp. 90-100.

[4] Laura Marie Feeney. An energy consumption model for performance analysis of routing protocols for mobile ad hoc networks. Mobile Networks and Applications, 6(3):239–249, 2001.

[5] S. Singh, M.Woo, and C. S. Raghavendra, "Power-aware with Routing in Mobile Ad Hoc Networks," Mobicom 1998, Dallas, TX, 1998.

[6] D.Kim, J.J.Garcia-Luna-Aceves, K.Obraczka, J.Carlos Cano, P.Manzoni, "Power Aware Routing Mechanism Based on the Energy Drain Rate Metric in Mobile Ad Hoc Networks," 11th IEEE International Conference on Computer Communications and Networks (IC3N), 2002.

[7] K.Woo, C.Yu, H.Y.Youn, and B. Lee,"Non-Blocking, Localized RoutingAlgorithm for Balanced Energy Consumption in Ad Hoc Networks," International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication System, MASCOTS, 2001.

[8] Garcia, J.-E.; Kallel, A.; Kyamakya, K.; Jobmann, K.; Cano, J.-C.; Manzoni, P, "A novel DSR-based energy-efficient routing algorithm for mobile ad-hoc networks", 58th IEEEVehicular Technology Conference, 2003.

[9] S Chettibi, M Benmohamed, "A Multipath Energy-Aware On demand Source Routing Protocol for Mobile Ad-Hoc Networks", arxiv.org, 2009.

[10] Meng Li, Lin Zhang, Victor O. K. Li, Xiuming Shan, Yong Ren, "An Energy-Aware Multipath Routing Protocol for Mobile Ad Hoc Networks", ACM Sigcomm Asia, 2005.

[11] Morteza Maleki, Karthik Dantu, and Massoud Pedram, "Power-aware Source Routing Protocol for Mobile Ad Hoc Networks", ACM, August 12-14, 2002.

[12] Fang Xie, Lei Du, Yong Bai, and Lan Chen, "Energy Aware Reliable Routing Protocolfor Mobile Ad Hoc Networks", IEEE Communications Society subject matter experts for publication in the WCNC proceedings, 2007.

[13] P.Srinivasan & K.Kamalakkannan, "Signal Strength And Energy Aware Reliable Route Discovery In MANET", International Journal of Communication Network Security ISSN: 2231 – 1882, Volume-1, Issue-4, 2012.

[14] Sanjay Kumar Dhurandher, Sudip Misra, "An Energy-Efficient On-Demand Routing Algorithm for Mobile Ad-Hoc Networks", 15th IEEE International Conference on Electronics, Circuits and Systems, 2008.

[15] radhika d.joshi, priti p.rege, "Distributed Energy Efficient Routing in Ad Hoc Networks", IEEE, Fourth International Conference onWireless Communication and Sensor Networks, 2008.

[16] Jin LIAN, "An Energy-based QoS Routing Protocol in mobile Ad Hoc Network", IEEE, Pacific-Asia Conference on Circuits, Communications and System, 2009.

[17] M. Tamilarasi, T.G. Palanivelu, Integrated Energy-Aware Mechanism for MANETs using On-demand Routing, International Journal of Computer and Information Engineering 2:3 2008.

[18] Radhika D.Joshi, Priti p.Rege, "Energy Aware Routing in Ad Hoc Networks", 6th WSEAS International Conference on CIRCUITS, SYSTEMS,ELECTRONICS,CONTROL & SIGNAL PROCESSING, Cairo, Egypt, Dec 29-31, 2007.

[19] Sumit Kumar Singh, Shiva Prakash, Kapil Kumar, "Energy Aware Dynamic On-deman (EA-DYMO)", International Journal of Computer Applications (0975 – 8887), Volume 25– No.11, July 2011

[20] G.Madhukar Rao, M.Nadeem Baig, Md. Fareed Baba, k.kanthi kumar," Energy Efficient Reliable Routing Protocol for Mobile Ad hoc Networks", IEEE, 2011.

[21] M. Pushpalatha, Revathi Venkataraman, and T. Ramarao, "Trust Based Energy Aware Reliable Reactive Protocol in Mobile Ad Hoc Networks", World Academy of Science, Engineering and Technology 32 2009.

[22] S.-M. Senouci and G. Pujolle, "Energy Efficient Routing in Wireless Ad Hoc Networks", IEEE International Conference on Communications, 2004.

[23] Sujatha P Terdal, Dr. V.D.Mytri, Dr. A.Damodaram, "A Load Aware Routing Mechanism for Improving Energy Efficiency in Mobile Ad Hoc Networks", International Journal of Computer Applications, November 2010.

A Survey on Decision Support Systems in Social Media

M. Thangaraj
Madurai Kamaraj University
Madurai, India

R. Indra
Alagappa University
Government Arts College For   Women
Sivagangai, India

Abstract: Web 3.0 is the upcoming phase in web evolution. Web 3.0 will be about "feeding you the information that you want, when you want it" i.e. personalization of the web. In web 3.0 the basic principle is linking, integrating and analyzing data from various data sources into new information streams by means of semantic technology. So, we can say that Web 3.0 comprises of two platforms semantic technologies and social computing environment. Recommender system is a subclass of decision support system. Recommendations in social web are used to personalize the web [20]. Social Tagging System is one type of social media. In this paper we present the survey of various recommendations in Social Tagging Systems (STSs) like tag, item, user and unified recommendations along with semantic web and also discussed about major thrust areas of research in each category.

Keywords: Social Tagging System, Recommendation, Semantic web**,** Tag Recommendation, Item Recommendation, User Recommendation, Unified Recommendation.

1. INTRODUCTION

Social tagging systems allow users to annotate web resources using tags. While not restricted to controlled vocabulary, tags are freeform keywords that convey meaning and interpretation from the user about the resource being annotated. It facilitates navigation and improves searching without dependence on pre-configured categories. It provides one way of organizing resources resulting in Folksonomy. Folksonomy is different from taxonomy which is hierarchical and exclusive whereas Folksonomy is nonhierarchical and inclusive. According to what kind of resources are supported, there are different systems like Flickr, citeUlike, Connotea, last.fm, Bibsonomy, BIBTEX and Technorati. Advantages of Folksonomies are their flexibility, rapid adaptability, free-for-all collaborative customisation and their serendipity. Formally a folksonomy is a tuple $F := (U, T, R, Y)$ where

- U, T and R are finite sets, whose elements are called users, tags and resources respectively and
- Y is a ternary relation between them, i.e. $Y \in U \times T \times R$, whose elements are called tag assignments

Recommender systems can alleviate information overload and combat noise by personalizing the user's view of the system. Such a system provides information with three decisive factors of "customized", "interested" and "useful" for any individual user by analyzing his/her preferences and the content of the items.   These factors distinguish the recommender systems from traditional information retrieval systems and search engines. Recommender algorithms should favor items bookmarked by more users. But recommender algorithms without tags do not take into account the number of raters and neighborhood calculation may not be the most efficient because due to the large volume of items and low overlap between user bookmarks, two users who are very

similar in their interests may still have too few common items bookmarked. In this context, tags can provide a more reliable approach to find similar users and this can be used to get better recommendation.

First based on what is recommended we have categorized Recommendations in STSs into three types as tag, item and user recommendations. All three recommendations combined together referred to as unified recommendation. Then in each type of recommendation based on the methods used for recommendations we have classified as shown in Figure 1. This paper is organized as follows. Tag, item, user and unified recommendations are specified successively from Section 2 through Section 5 and Section 6 deals with semantic web. Advantages and limitations of these recommendations are given as findings in Section 7 and we conclude in Section 8.
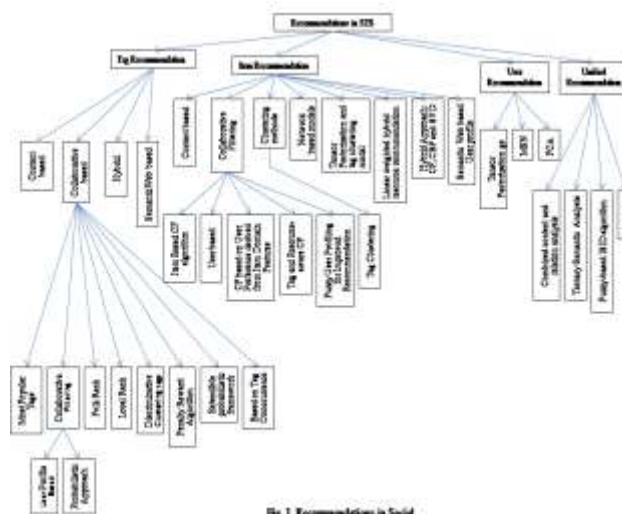

Fig. 1. Recommendations in Social Tagging Systems

2. TAG RECOMMENDATION

Tags are recommended at the time when a user wants to annotate a resource. As most tags in the STSs are uncontrolled, redundant and ambiguous the tag recommenders are useful in simplifying the tagging process for users to find good tags and consolidate the tag vocabulary across users as well. There are two types of tag recommendation such as prediction or Collective tag recommendation which does not assume a query user for recommendation and Personalized tag recommendation recommends tags for a query user. Tag recommendation techniques can be classified into four categories: content based, collaborative, hybrid approaches and semantic approaches.

2.1 Content based

It focuses on the suggestion of keywords (tags) extracted from resource contents and meta-data. It exploits the technology of the automatically textual extracting keywords. One method used in content-based tag recommendation is Discriminative clustering approach [22]. In this approach two clustering models of the posts are created: one based on the tags assigned to the posts and second based on the content terms of the posts. From the clustering model ranked lists of tags and terms for each cluster is generated. The final recommendation is done by using both lists, together with the user's tagging history if available. Prediction results of, tag based clustering model is more accurate than term based clustering model.

Another method used in content based tag recommendation is three-step tag recommendation system [18]. In this system basic tags are extracted from the resource title. In the next step, the set of potential recommendations is extended by related tags proposed by a lexicon based on co-occurrences of tags within resource's posts. Finally, tags are filtered by the user's personomy – a set of tags previously used by the user.

Yet another method compiles a set of resource specific tags, which includes tags related to the title and tags previously used to describe the same resource (resource profile). These tags are checked against user profile tags – a rich, but imprecise source of information about user interests. The result is a set of tags related both to the resource and user [17].

RSDC'08 Tag Recommendation [19] comes under content based recommendation. In this method document (web pages and publications) model is constructed using the textual content associated with bookmarks, user model is constructed based on their tagging and based on these models tags are suggested for new bookmarks. A combination of statistical and semantic features are used to build document and user models.

2.2 Collaborative based

It exploits the relations between users, resources and tags of the folksonomy graph to select the set of recommended tags. Following methods come under this approach.

2.2.1 Most Popular Tags [14]

Tags are recommended based on tag counts. Some variants of this approach are as follows.

For any user u and any resource r, recommending the most popular tags of the folksonomy is the most simplistic approach.

$$\tilde{T}(u,r) := argmax_{t \in T}^{n}(|Y_t|)$$

Tags that are most specific to the resource globally will be recommended when using the most popular tags by resource.

$$\tilde{T}(u,r) := argmax_{t \in T}^{n}(|Y_{t,r}|)$$

Since users might have specific interests for which they already tagged several resources, using the most popular tags by user is another option.

$$\tilde{T}(u,r) := argmax_{t \in T}^{n}(|Y_{t,u}|)$$

Another approach is to recommend a mix of the most popular tags of the user with the most popular tags of the resource. The simplest way to mix the tags is to add their counts and then sort them by their count

$$\tilde{T}(u,r) := argmax_{t \in T}^{n}(|Y_{t,r}| + |Y_{t,u}|)$$

2.2.2 Collaborative Filtering (CF) [14]

Because of the relational nature of folksonomies, traditional CF cannot be applied directly. Reduce the ternary relation Y into two 2-dimensional projections i.e. user's resources and user's tags. Either can be used to find user's neighborhood as follows.

$$N_u^k := argmax_{v \in U \setminus \{u\}}^{k} sim(\vec{x}_u, \vec{x}_v)$$

For determining, for a given user u, a given resource r, and some n ∈ N, the set $\tilde{T}(u,r)$ of n recommended tags we use

$$\tilde{T}(u,r) := argmax_{t \in T}^{n} \sum_{v \in N_u^k} sim(\vec{x}_u, \vec{x}_v) \, \delta(v, t, r)$$

where $\delta(v,t,r) := 1$ if $(v,t,r) \in Y$ and 0 else.

Some variations of CF approaches are as follows.

User profile based tag recommendation[32]

Tag based user profile : profile (u) = {(w₁,t₁), (w₂,t₂)..(wₙ,tₙ)} where $t_i \in T$, w_i is the weight of the t_i, represents the importance of this tag to the user. Items are selected using balanced strategy, item tag matrix is constructed, for user u the preference relation between t_i and t_j for item k is calculated and ranking is done using voting, a vector is formed by the ranking tags, the profile of user u for item k is presented, ITW – Itemtagweight matrix is constructed, user similarity using Pearson correlation coefficient is calculated and most frequently used tags from the k similar user are recommended.

Probabilistic Approach [13]

For personalized tag recommendation, a probabilistic framework that is based on personomy translation which translates from the resource tags to personomy tags is used. It is used for translation from similar users for expanding the candidate tags for recommendation. To compute the relevance score for a candidate tag the likelihood of the tag is estimated.

The overall likelihood of a candidate tag is the weighted average of the likelihoods estimated from users. Here weight is the similarity between the neighbor and the query user. Personomy translation method can be used for estimating the likelihood. Users are profiled by a set of translation probabilities one for each t_r and similarity between users can be measured by using distributional divergence metric.

2.2.3 Graph-based Approach - Folksonomy-Adapted PageRank – FolkRank[14]

Folksonomy is converted into an undirected Graph with V set of nodes consisting of the disjoint union of the sets of tags, users and resources. Each triple (u,t,r) in Y gives rise to the three undirected edges {u, t}, {u, r} and {t, r} in E. The rank of the vertices of the graph is computed with the weight spreading computation

$$\vec{w}_{t+1} \leftarrow dA^T\vec{w}_t + (1-d)\,\vec{p}$$

where $\vec{w}$ is the weight vector with one entry for each node in V. A is the row-stochastic version of the adjacency matrix (a_{ij} := $\frac{1}{\deg(i)}$ if {i,j} ∈ E and 0 else) of the graph G, $\vec{p}$ is the random surfer vector – which is the preference vector and d ∈ [0,1] is determining the strength of the influence of $\vec{p}$. The rank of each node is its value in the limit $\vec{w} := \lim_{t\in\infty} \vec{w}_t$ of the iteration process. For a global ranking, one will choose $\vec{p} = 1$, i.e., the vector composed by 1's. In order to generate recommendations, however $\vec{p}$ can be tuned by giving a higher weight to the user node and to the resource node for which one currently wants to generate a recommendation (i.e. $\vec{p}[u] = 1+ |U|$ and $\vec{p}[r] = 1 + |R|$). The recommendation $\tilde{T}(u,r)$ is then the set of the top n nodes in the ranking, restricted to tag nodes

2.2.4 LocalRank [21]

In Folk rank scalability and update are the main issues. To avoid these issues new tag recommendation LocalRank has been proposed in which the rank weights are calculated only based on the local "neighborhood" of a given user and resource. Instead of considering all elements in the folksonomy, local rank focuses on the relevant ones only. The rank computation in LocalRank takes into account, how often certain tag was used by a user and how often a tag was attached to a resource. Rank computation and weight propagation in LocalRank is done similar to FolkRank but without iteration.

$Y_u \subseteq Y$ is the set of all (u, t, r)-assignments of Y where u is the given user.

$Y_r \subseteq Y$ is the set of all (u, t, r)-assignments of Y where r is the given resource.

T_u is the set of all tags appearing in the (u, t, r)-assignments of Y_u.

T_r is the set of all tags appearing in the (u, t, r)-assignments of Y_r.

The overall set of relevant tags to be ranked by the algorithm is $T_u \cup T_r$

The rank of each t ∈ T_u is calculated as

$$rank(t) = |Y_{u,t}| \; x \; \frac{1/N}{|Y_u|}$$

The rank of each t ∈ T_r is calculated as

$$rank(t) = |Y_{r,t}| \; x \; \frac{1/N}{|Y_r|}$$

Tags that appear in both sets (t ∈ $T_u \cap T_r$) are on principle more important than the others and should receive a higher weight. Therefore we sum up the individual rank weights obtained from the two calculations.

LocalRank propagates the weight of the given user and the resource nodes to all their adjacent tags. Therefore, it computes rankings for user and resource relevant tags and returns a list of tags and their ranks. The recommendation of tags can then be done by picking the top n elements with the highest rank values.

2.2.5 Discriminating Clustering [11]

Tag recommendation based on discriminative clustering recommends the top tags of the most similar clusters for a post. A discriminative method is used to cluster the historical data of posts based on the posts' tags. This method maximizes the sum of the discrimination information provided by posts and outputs a weighted list of discriminating tags for each cluster. Given a new post, and based on the post's contents, the top 5 tags of the most relevant cluster for the post are recommended.

2.2.6 Extendible Probabilistic Framework [25]

It aggregates and exploits the knowledge that exists at four different contextual layer ie. Personal Context, Social Contact Context, Social Group Context and Collective Context. For each context weighted network of tags has to be derived, with nodes representing unique tags t_i and edges occurring when two tags have been used to annotate the same photo. Weights are defined by the number of times this happens in our data set. For all the tags of all the photos in the collection, we calculate the occurrence tally $o(t_i)$ and the co-occurrence tally $c(t_i,t_j)$. The probability of a tag occurring and the conditional probability of two tags co-occurring are formulated as:

$$p(t_i) \;=\; \frac{o(t_i)}{\sum_{t\in T} o(t)}$$

$$p(t_i,t_j) \;=\; \frac{c(t_i, t_j)}{o(t_j)}$$

To produce a set of recommendations for a given set of input query tags, each query tag is first used to generate a intermediate set of recommendations and these sets are then

combined. The set of recommendations for a given query tag for a given context is the complete set of tags that co-occur with that tag in that context's network. Probability of an intermediate suggestion given a query set of tags is calculated. Each resultant probability is then used to produce an ordered list of tags in descending order of probability. The top N tags are then the final recommendations as given by that context's network of tags for a given query tag set. The four individual ranks produced from the tag networks of four different contexts are then combined using Borda Count method.

2.2.7 Penalty-Reward Algorithm[37]

First, it favors tags that are used by a large number of people (with good reputation). Then it aims to minimize the overlap of concepts among the suggested tags to allow for high coverage of multiple facets. It also honors the high correlation among tags for example if two tags are used together by most users for a given object, they will co-occur in the suggested tags. All these things are done using Penalty-reward algorithm. It rewards good tags (i.e. high coverage of multiple facets, high popularity, uniformity and least effort) and penalizes redundant information yielding best output.

2.2.8 Tag Recommendation based on Tag Co-occurrence [29]

Users not only tag the visual contents of the photo, but also provide a broader context in which the photo was taken, such as, location, time, and actions. Given a photo with user-defined tags, an ordered list of m candidate tags is derived for each of the user-defined tags, based on tag co-occurrence. The lists of candidate tags are then used as input for tag aggregation and ranking, which ultimately produces the ranked list of n recommended tags. We define the co-occurrence between two tags to be the number of photos [in our collection] where both tags are used in the same annotation. It is common to normalize the co-occurrence count with the overall frequency of the tags. There are essentially two different normalization methods: symmetric and asymmetric.

Symmetric measures.: According to the Jaccard coefficient we can normalize the co-occurrence of two tags t_i and t_j by calculating:

$$J(t_i, t_j) := \frac{|t_i \cap t_j|}{|t_i \cup t_j|}$$

Alternatively, tag co-occurrence can be normalized using the frequency of one of the tags. For instance, using the equation:

$$P(t_j | t_i) := \frac{t_i \cap t_j}{t_i}$$

it captures how often the tag t_i co-occurs with tag t_j normalized by the total frequency of tag t_i. Two aggregation strategies exist. One strategy is based on voting, and does not take the co-occurrence values of the candidate tags into account, while another strategy ie. the summing strategy uses the co-occurrence values to produce the final ranking. In both

cases, we apply the strategy to the top m co-occurring tags in the list.

2.3 Hybrid

The hybrid approaches combine two or more approaches and outperforms well in precision and recall. But have higher computational complexity.

2.4 Semantic Web based [1]

Title Recommender Model

It extracts words i.e. adjectives, nouns and Non-WordNet words from the resource's attribute (title or URL) and suggests them as tags.

Tag to tag recommendation

It may be useful to recommend tags based on other recommended tags. Given a tag, other tags can be produced from its related terms like synonyms and hypernyms. This type of meta-recommenders improves the quality and quantity of the recommendation, when the main recommender fails to provide a sufficient number of tags.

3. ITEM RECOMMENDATION

Recommender systems apply knowledge discovery techniques to the problem of making personalized recommendations for information, products or services during a live interaction. Item recommender systems helps users to find the items that they would like to purchase at E-commerce sites by producing predicted likeliness score or a list of top-N recommended items for a given user, using data analysis techniques. Recommendations can be based on demographics of the users, overall top selling items or past buying habit of users as a future predictor of future items.[27] Various methods available for item recommendation are Content based, Collaborative based, Hybrid, Network based, Clustering based and Semantic based. In hybrid approaches two or more above mentioned methods are combined.

3.1 Content based Recommendation [27]

The Content Based Filtering (CBF) approach creates a profile for each user or product to characterize its nature. For example, a movie profile could include attributes regarding its genre, the participating actors, its box office popularity, and so forth. User profiles might include demographic information or answers provided on a suitable questionnaire. The profiles allow programs to associate users with matching products. A known successful realization of content filtering is the Music Genome Project, which is used for the Internet radio service Pandora.com.

The system can build a profile based on the attributes present in the items that user has rated highly. The interest a user will have in an unrated item can then be deduced by calculating its similarity to their profile based on the attributes assigned to the item.

3.2 Collaborative Filtering (CF) [27]

The main aim of this algorithm is to suggest new items or to predict the utility of a certain item for a particular user based on the user's previous likings and the opinions of other like-minded users. In this algorithm

$$U = \{u_1, u_2, \ldots u_m\} \rightarrow \text{List of m users}$$

$$I = \{i_1, i_2 \ldots i_n\} \rightarrow \text{List of n items}$$

Each user u_i has a list of items I_{ui} which the user has expressed his opinions about.

There are two types of collaborative filtering algorithms namely Memory-based and Model-based algorithms. Memory based approaches which include user-based and item-based algorithms, employ statistical techniques (correlation or vector similarity) to find a set of users, known as neighbors that have a history of agreeing with the target user (i.e. they either rate different items similarly or they tend to buy similar set of items). Once a neighborhood of users is formed, these systems use different algorithms to combine the preferences of neighbors to produce a prediction or top-N recommendation for the active user. It is also known as nearest-neighbor. Correlation can be extended using default voting and vector similarity can be extended using Inverse user frequency. Later algorithm, provide item recommendation by first developing a model of user ratings. It computes the expected value of a user prediction, given his/her ratings on other items using probabilistic approach. The model building process is by different machine learning algorithms such as Bayesian network, Clustering and rule-based approaches. Matrix factorization approaches, such as SVD and NMF have been proved useful in model-based CF, which predict unobserved user-matrix.

3.2.1 Item based Collaborative Filtering Algorithm [27]

Item-based approach looks into the set of items the target user has rated and computes how similar they are to the target item and then selects k most similar items. At the same time their corresponding similarities are also computed. Once the most similar items are found, the prediction is then computed by taking a weighted average of the target user's ratings on these similar items.   The basic idea in similarity computation between two items is to first isolate the users who rated both of these items and then to apply similarity computation technique to determine the similarity. There are different similarity computation techniques like Cosine-based Similarity, Correlation-based similarity etc. Once we isolate the set of most similar items based on the similarity measures, the next step is to look into the target users' ratings and use a technique such as weighted sum or regression to obtain predictions.

Obviously, the users' opinions and interests can be driven implicitly or explicitly. Examples of explicit data include the following:

_ Rating score to the items based on a defined scale

_ User's interests and preferences

_ Information of any questionnaire being filled out by a user

Discovering implicit information is a difficult task because they are usually hidden. Depending on the domain of the application, there are different methods for extracting implicit information from available data. Examples of implicit data include the followings:

_ User's behaviors and activities

_ User's social and relational behaviors in a group

_ Items being visited by users

_ Expended observation time for the items

_Web usage data mining regarding user's navigation

3.2.2 User-based Collaborative Filtering [5]

Classic Collaborative Filtering(CCF) uses Pearson Correlation to calculate similarity between users and a classic adjusted ratings formula to rank the recommendations. Neighbor-weighted Collaborative Filtering, takes into account the number of raters in the ranking formula of the recommendations. Another approach explores an innovative way to form the user neighborhood based on a modified version of the Okapi BM25 model over users' tags. In the said three approaches, users are collected and for each user, the neighborhood of users who posted her same articles and the neighborhood of users who share the same tags are included in this collection.

3.2.3 CF based on User Preference Derived from Item Domain Features [33]

In Social tagging system, user-created tags are utilized to depict user preferences for personalized recommendation but it is difficult to identify users with similar interests due to the difference between users' descriptive habits and the diversity of language expression. So item domain features are utilized to construct user preference models and combined with CF for personalized recommendation. Because of the diversity of domain characteristics, traditional personalized recommendations do not adapt well to all domains. Hence, it is required to combine domain characteristics and personalized recommendation. It could make recommendations to users who have not selected any common items with others. Here first item domain features are used to model user preference matrix then the user preferences vector is derived from user preference matrix and user preference models are combined with CF to provide personalized recommendations.

3.2.4 Tag and Resource-Aware Collaborative Filtering Algorithms for Resource Recommendation [7]

Sparsity is a problem which occurs in a Social Recommender system, when the number of tags and resources to profile a user are inadequate, to provide quality recommendations. To address this problem Tripartite Nearest Neighbor Algorithm (TRNNA) which combines the similarity based on the Tag Vector Model, the Resource Vector Model and the Resource Vector of tags with weights is used.

3.2.5 Fuzzy User Profiling for Improved Recommendation [4]

Recommender systems, generally have, at their disposal, information regarding genres/categories that a movie/ book belongs to. However, the degree of membership of the objects in these categories is typically unavailable. Such information, if available, would provide a better description of items and consequently lead to quality recommendations. Fuzzy item profile is formed by first identifying the set of tags representative of each genre and then measuring the genre content of each movie. Information about actors and directors are also included in this profile. This item profile is combined with user preference data, which is also represented as a fuzzy set membership value to automatically derive a fuzzy user preference profile. These profiles are then used in multiple ways for both CF and CB systems to derive recommendation. Fuzzy user profile formed here contains user preference information, content information and tag information.

3.3 Clustering Methods [27]

Clustering techniques work by identifying groups of users who appear to have similar preferences. Once the clusters are created, predictions for an individual can be made by averaging the opinions of the other users in that cluster. Each user may be represented with partial participation in several clusters in some clustering techniques. The prediction is then an average across the clusters, weighted by degree of participation.

3.3.1 Tag Clustering [28]

Clustering Algorithm is able to aggregate tags into topic domains.   Hierarchical clustering is proposed to generate a taxonomy from a folksonomy. Tag clusters are presumed to be representative of the resource content. Thus, a folksonomy of Web resources is used to move the Internet closer to the Semantic Web. Tag clustering can support tag recommendation, reducing annotation to a mouse click rather than a text entry. Well-chosen tags make the recovery process simple and offer some control over the tag-space diminishing tag redundancy and ambiguity to some degree.

Each user, u, is modeled as a vector over the set of tags, where each weight, $w(t_i)$, in each dimension corresponds to the importance of a particular tag, t_i.

$$\vec{u} = (w(t_1), w(t_2)...w(t_{|T|}))$$

Resources can also be modeled as a vector over the set of tags. In calculating the vector weights, a variety of measures can be used. The tag frequency, tf, for a tag, t, and a resource, r, is the number of times the resource has been annotated with the tag. We define tf as:

$$tf(t,r) = |\{a = (u, r, t_i) \in A : u \in U\}|$$

Likewise, the well known term frequency * inverse document frequency can be modified for folksonomies. The tf*idf multiplies the aforementioned frequency by the relative distinctiveness of the tag. The distinctiveness is measured by the log of the total number of resources, N, divided by the number of resources to which the tag was applied, n_t . We define tf*idf as:

$$tf*idf(t,r) = tf(t,r) * \log(N/n_t)$$

With either term weighting approach, a similarity measure between a query, q, represented as a vector over the set of tags, and a resource, r, also modeled as a vector over the set of tags, can be calculated. The user interacts with the system by selecting a query tag and expects to receive resource recommendations. A query is a unit vector consisting of a single tag and Cosine Similarity can be calculated as follows:

$$\cos(q,r) = \frac{tf(q,r)}{\sqrt{\sum_{t\in T} tf(t,r)^2}}$$

To recommend resources, we can calculate the similarity of the selected tag to each resource and recommend the top n. Let it be R' subset of resources R.

The Personalized recommendation process proceeds in two stages. First, given a user's click on a tag, the standard non-personalized recommendation algorithm is applied to produce a set of recommended resources R'. This set is then personalized by taking the user profile and tag clusters into account and re-ranking the results accordingly.

3.4 Network-based Models [35]

A tag-based network can be viewed as a tripartite graph which consists of three integrated bipartite graph or a hypergraph. Therefore, network-based methods are widely used to describe the tag-based graph. In a typical bipartite graph, there are two mutually connected communities, which contrastively have no link within each community. Probability Spreading(random walk), Heat Spreading, tag-aware diffusion-based method, user-centric diffusion-based similarity are the network based methods for item recommendation.

3.5 Tensor Factorization and Tag Clustering Model [26]

Tensor Factorization and tag clustering model consist of three steps. The first step involves of tag propagation by exploiting content, so as to face the issues of sparsity, "cold start" and "learning tag relevance". It is based on a relevance feedback

mechanism, in order to perform tag (less noisy tags) propagation between similar items only if they belong to the same concept. The second step of the TFC model is tag clustering in order to reveal topics and identify the taste of users in these topics. Here the sparsity problem is solved. After producing tag clusters, an innovative tf-idf weighting scheme is followed to calculate users' interests. The third step is using High Order Singular Value Decomposition (HOSVD) which is used to reveal the latent association among users, topics and images. Finally for making recommendation the elements of the reconstructed tensor is used.

3.6 Linear Weighted Hybrid Resource Recommendation[15]

It composed of KNN user-based collaborative filtering algorithm, tag-specific user-based collaborative filtering algorithm, item-based collaborative filtering, tag-specific item-based collaborative filtering, tag model similarity and popularity model. It provides a flexible, general and effective approach to capitalize on strong relationships across different dimensions of the data and to incorporate the most effective components into a single recommendation framework.

3.7 Hybrid Approach : CF, CBF and SVD[2]

This approach is used in queveo.tv, a system for highly customized TV content recommendation. The gray-sheep problem, cold-start problem and first-rater problem in CF recommender system can be eliminated using CBF. The problem with the CBF is its tendency to overspecialize item selection because it only recommend items similar to those the user has previously liked. So both CF and CBF are combined. Again there will be a problem of sparsity and scalability. So SVD is used to reduce the dimensionality of the recommender system database.

3.8 Semantic Web based User profile [30]

Amazon.com provides a recommendation service that is based on collaborative filtering: if a user buys an item that has been bought by a number of other users in combination with some other items, then those other items will be recommended by Amazon.com to the user. These recommendations are entirely based on what goes on inside the system ignorant of any external knowledge about the items or the users themselves. To improve such recommendation techniques, we think it might be useful to incorporate data from as many sources as possible to build richer profiles that model many facets of interest that might be difficult and impractical to capture by a single system or service. Users information can be obtained by analyzing their shared profile in MySpace, bookmarks in del.icio.us, photos in Flickr, references in Connotea and any other popular Web 2.0 applications.

This approach is used in movie recommendation. Data representation is done by importing both IMDB database and Netflix rating data into a relational database. String matching is then used to correlate the movie titles in the Netflix with

their counterparts in the IMDB data set. To provide a homogeneous view over both data sources, an ontology is used in conjunction with the D2RQ mapping technology, supplying a SPARQL end-point which can be queried to find extensive amounts of information on movies.

For making recommendation, first tag clouds are formed based on the keywords and ratings assigned to the movies by the users. Then three methods are used to make recommendation. 1. Average-based rating 2. Simple Tag-Cloud comparison 3. Weighted tag cloud comparison.

4. USER RECOMMENDATION

This recommendation service helps users to discover new interesting multimedia resources, encourages interaction between users with similar interests and improve users' satisfaction (which means more advertising revenue for web sites). There are three techniques for making user recommendation namely Tensor Factorization, Formal Concept Analysis (FCA) and Multidimensional Social Network (MSN).

4.1 Tensor Factorization in Social Networks [36]

First, new model is proposed with tensor factorization to capture the potential association among user, user's interests and friends. Second a novel approach is proposed to recommend new friends with similar interests for users. This method considers both link structure and user's tagging content. User's tags indicate user's interests. User makes friends with others based on similar interests. Given a user u, the user recommender system is set to predict a personalized ranking list of Top-N users whom u wants to make friends with.

4.2 Mining and Representing User Interests using FCA [10]

Tags are used to depict people's interest in online resources, and tags in most social Web sites have become an available feature. Since tags are used to organize individuals' ideas and thoughts as well as to encourage their social interactions, tagging activities on social web sites can be considered a new way of collective authorship. Once a set of tags is assigned to a resource, a network structure can be constructed given a number of users and the tags that they assign to a set of shared resources. An interest group based on tagging data is extracted from tagging behaviors; sets of tags can be used to build social networks and promote their use by other people in online communities. Formal Concept Analysis (FCA) is a mathematical theory used for conceptual data analysis and unsupervised machine learning. FCA models the world of data through the use of objects and attributes. The relations between objects and attributes in a data set form the formal context. A formal concept refers to the relationship between a set of formal objects and a set of attributes. Based on FCA, given a set of users U and considering the tags that they have in common, the interest group of U is the set of users who are using these tags. The intent of a set of users U is the set of

tags which are used by every user in U. The extent of a set of tags T is the set of users using every tag in T. Thus an interest group would be a set of users that use a significantly similar collection of tags to identify their resources. Folksonomy represents a formal context. List of concepts can be extracted from a formal context using the algorithm by Ganter and Kuznetsov and the significance score of each concept is also calculated.

4.3 Social Recommender System using MSN [24]

All online sharing systems gather data that reflects users' collective behavior and their shared activities. We can use this data to extract different kinds of relationships which are grouped into layers and forms the basic components of the multidimensional social network (MSN). The layers created are based on the two types of relations between humans such as direct social links between individuals and object based semantic links between individuals. This MSN is used in recommender system to suggest one human being to another so as to expand the human community. It mainly makes use of relationships come from indirect connections via Multimedia Objects (MOs) rather than from direct links. The system and personal weights that are assigned independently to each layer make the recommendation process personalized. Also the system is adaptive due to personal weights that are adaptively recalculated when the user utilizes the recommendations.

5. UNIFIED RECOMMENDATIONS

It provides all three types of recommendation i.e. user, tag and item. There are four ways of providing all recommendations namely combining content and relation analysis, Ternary Semantic Analysis, Internet Social Interest Discovery algorithm (ISID) and FolkRank.

5.1 Combining Content and Relation Analysis for Recommendation [34]

One recommendation system for social tagging system combines content and relation analysis in a single model. By modeling the generative process of social tagging systems in a latent Dirichlet allocation approach, we can build a fully generative process of social tagging. By leveraging it to estimate the relation between users, tags and resources we can achieve tag, item and user recommendation tasks.

5.2 FolkRank [12]

FolkRank takes into account the folksonomy structure for ranking search requests in folksonomy based systems. It is used for two purposes : determining an overall ranking and specific topic-related rankings. FolkRank is already discussed in Tag Recommendation. The preference vector p is used to determine the topic. We can define a topic by assigning a high value to either one or more tags and/or one or more user and/or one or more resources. FolkRank provides one topic-specific ranking for each given preference vector. FolkRank yields a set of related users and resources for a given tag and vice versa. Thus FolkRank can be used to generate

recommendations within folksonomy systems. These recommendations can be presented to the user at different point in the usage of a folksonomy system.

5.3 Fuzzy-based Internet Social Interest Discovery Algorithm (ISID) [16]

Fuzzy-based ISID algorithm consists of the component called Syntactic variation which avoids the syntactic variations of the posts. And also provides functions such as finding topics of interests, resource clustering and topics of interest indexing. In the function finding topics of interest, for a given set of bookmark post find all topics of interest. Each topic of interests is a set of tags with the number of their co-occurrences exceeding a given threshold. In the function Clustering, for each topics of interests, find the URLs and the users such that those users have labeled each of the URLs with all the tags in the topic. In the function Indexing, import the topics of interests and their user and URL clusters into an indexing system for application queries.

5.4 Ternary Semantic Analysis [23]

The three types of entities (user, item, tag) that exist in STS are modeled using 3-order tensor. Dimensionality reduction technique Higher Order Singular Value Decomposition (HOSVD) is applied in 3-order tensor to reveal latent semantic associations between users, items and tags. Smoothing technique based on Kernel-SVD is also applied to address the sparseness of data. HOSVD uses the taxonomy and outputs the reconstructed tensor Â. Â measures the associations among the users, items and tags. Each element of Â can be represented by a quadruplet {u, i, t, p}, where p measures the likeliness that user u will tag item i with tag t. Therefore items can be recommended to u according to their weights associated with {u, t} pair. A similar approach is followed for user and tag recommendation

6. SEMANTIC WEB

In general, tag ontologies in STSs can contribute in the following three areas:

• *Knowledge Representation Sophistication*: A tag ontology can robustly represent entities and relationships that shape tagging activities. It could make the knowledge structure of tagging data explicit and facilitate the Linked Data (Berners-Lee, 2006) of tagging data on the Web.

• *Facilitation of Knowledge Exchange*: Ontologies enable knowledge exchange among different users and applications by providing reusable constructs. Thus, a tag ontology can be shared and used for separate tagging activities on different platforms.

• *Machine-processable*: Ontologies and Semantic Web technologies in general (knowledge representation, processing and reasoning) expose human knowledge to machines in order to perform automatic data linking and integration of tagging data.

Various ontologies in STSs

An Ontology for tagging is not just a way to define meanings of certain tags, but it can also robustly represent the relationships among the entities that shape tagging activities, explicitly stating the knowledge structure of tagging data. Social Semantic Cloud of Tags (SCOT) aims to describe folksonomic characteristics and to offer social interoperability of semantic tag data across heterogeneous sources. This model can express the structure of, features of and relationships between tags and users, allows the exchange of semantic tag metadata for reuse in social applications, and enables interoperation among data sources, services or agents in a tag space. Typical social tagging systems do not provide explicit links between the involved entities, nor do they expose their data in a standard form. The design of the tag ontology was an attempt to provide a common conceptualization of 'what tagging means' by providing a standardized way to collect, interpret and use tagging data. One of the advantages of this ontology is that isolated tagging data can be easily made mobile and can be integrated across applications.

User information is represented using Semantically Interlinked Online Communities(SIOC). The Tag Cloud class in SCOT aggregates all tagging instances with their relevant information. At this level, tagging entities are represented with their collective feature underlying their relationships. The SCOT ontology can be utilized by SPARQL, the query language for Semantic Web data, to get minimal information to compute the significance of tagging data.

Newman's model describes the relationships between an agent, an arbitrary resource and one or more tags. MOAT(Meaning of a Tag) is intended for semantic annotation of content by providing meanings for free-text tagging. It provides the meaning class to represent customized, user-provided 'meanings' for tags. This class provides the meaning of tags to be unambiguous. The Nepomuk Annotation Ontology (NAO) is provided for annotating resources on the Social Semantic Desktop.

All earlier tag ontologies including Gruber and Newman's models do not provide a way of fully representing the meaning of a tag and the relationships between tags, since they focus on expressing individual tagging instances. On the other hand, SCOT offers various properties for representing tag semantics and collective characteristics of tagging entities.

As a subclass of tag:Tag from Newman's model, the class scot:Tag describes a natural-language concept, which is used to annotate a resource. The purpose of this class is to describe the semantics and collectiveness of tags that are aggregated from individual tagging activities.

The tag class has some properties for eliminating tags' ambiguity:

- scot:spellingVariant refers to variations in the way in which a word is spelt;

- scot:delimited is used to describe a multiple-word tag name where each word is separated by a certain character;
- scot:synonym defines synonymous terms

These properties can reduce tag ambiguity resulting from the use of different conventions and even recommend more common patterns for tag names. Furthermore, in order to represent tag frequencies, SCOT introduces two properties: scot:ownAFrequency and scot:ownRFrequency. The former is intended to describe the absolute value of popularity for a specific tag. The purpose of the latter is to represent the relative value in order to identify the significance of a tag proportional to all the tags.

A single tag can have both frequency formats. The popularity of a tag plays a key role in distinguishing its significance in folksonomies.

Collecting and sharing tag metadata [9]

Int.ere.st is a prototype of a tag-sharing platform conceived for reusing tagged resources across heterogeneous platforms. A major goal of int.ere.st is to create a Semantic Web-based tagging application capable of solving the common problems of tags and tagging systems. From a technical point of view, int.ere.st is built on a variety of technologies: Apache, PHP, and MySQL. These frameworks are used to implement most functions in typical social websites. The need to provide interoperability between tagging information has led to the use of Semantic Web technologies – RDF and SPARQL. The majority of social websites now provide APIs based on popular mechanisms (e.g. REST, SOAP, and XML RPC). These APIs provide community users and applications with easy and intuitive access to data from the sites. As the amount of aggregated sources may increase exponentially, it is difficult to generate and update semantic tag metadata synchronously. To solve this problem, we use the D2R Server, a tool that maps relational databases to RDF and that is accessible through SPARQL.

Converting tags into senses [31]

Tag Disambiguation algorithm allows easily to semantify the tags of the users of a tagging service : it automatically finds out for each tag the related concept of Wikipedia in order to describe web resources through senses. There are some services that analyze the tags of a specific user to detect tag usage inconsistencies like slightly different keywords : eg. Bookmark Cleaner and Del.icio.us tag cleaner. The systems and the procedures that use some sort of semantic information to better organize tags and understand their meaning can be divided into two groups. The first one comprises all the methods that introduce some sort of structure to the sets of tags taking into account only the information retrievable from tagging services, i.e. the collections of users, tags and tagged resources. They mainly try to group together similar tags on the basis of their relations with users and resources. In this way they identify sets of strictly related tags or understand the

sense of ambiguous ones. The second group of semantic based approaches exploits external semantic resources to structure sets of tags. Some of them try to define the right meaning of each tag retrieving the semantic relations that occur between related tags so as to visualize tags on the basis of their sense and relevance. In order to achieve that, data extracted from different ontologies available over the web are collected and merged. The sense-based tagging is also a way to connect the social data collaboratively created through tagging and the Semantic Web.

Tagpedia is based on the model of term-concept networks; for each meaning of Wikipedia, Tagpedia groups together all the different words used to refer to it.   Tagpedia is built to support the characterization of web contents through sense-based tagging and thus it easily disambiguate the meaning of a tag, is accessible over the web at the URL http://www.tagpedia.org/. It can also be queried by means of a dedicated Web API and it can be collaboratively edited. Tag Disambiguation Algorithm(TDA) relies upon Tagpedia. TDA collects the tags of a user from a tagging service and for each of them finds out the relative sense by linking it to the corresponding page of Wikipedia.   In particular it identifies for each tag a list of candidate senses, referred to also as concepts or meanings and assigns them a number, called sense-rank SR; the higher the rank of a meaning, the better that meaning defines the sense intended by the user for that tag. First advantage of the adoption of sense-based tagging is represented by the possibility to group together user tags that refer to the same concept.

The TDA manages automatically to convert each tag into the intended concept of Wikipedia and thus into the related URI of DBpedia. In this way we are able to generate for each user of a sense-based tagging service, a set of RDF triples describing his tagging profile; they include, for instance, a triple for each sense associated to a tagged resource. TDA is used to clean user tags grouping them by sense and also to classify the tagged resources on the basis of Wikipedia categories, YAGO classes and Wordnet synsets By characterizing web resources through Wikipedia concepts we can also connect the social data produced by tagging systems to the datasets of the Linked Data community.

Semantic Web Recommender Systems [3]

There are two novel approaches for recommender system in open decentralized scenarios, namely trust networks along with trust propagation mechanisms, and taxonomy-driven profile generation and filtering.

7. FINDINGS

Tag recommendation :

Content based approach is better to solve the cold start problem but it is difficult to apply in the field of multimedia data where as collaborative approach can be used here but has the problem of Sparsity, because users are not able to evaluate the items and they are not wishing to rank purchased items or

viewed items, Reduced coverage due to sparsity, Acccuracy of the recommendation may be poor because of little ratings, Scalability and synonymy. In Collaborative approaches, user profile based tag recommendation provides better performance than Collaborative filtering, Probabilistic approach provide an ample improvement when adopting translation from neighbors, CF and most popular tags are Cheaper to compute than FolkRank and requires no iteration and both CF and most popular tags have similar costs but some advantage on the side of the mix of most popular tags. Compared to FolkRank, Local rank is easy to implement, simple, quicker, process large data sets and produce accurate result. Whereas extendible probabilistic framework has benefits for users who do not use English while interacting with Flickr, Penalty-reward algorithm rewards good tags (i.e. high coverage of multiple facets, high popularity, uniformity and least effort) and penalizes redundant information yielding best output, tag recommendation based on tag co-occurrence can be incrementally updated when new annotations become available and thus gracefully handle the evolution of the vocabulary.   Hybrid approaches outperforms well in precision and recall, but have higher computational complexity. Semantic web based approaches improves the quality and quantity of the recommendation, when the main recommender fails to provide a sufficient number of tags but it's drawback is it cannot be used independently.

Item Recommendation :

Content-based strategies require gathering external information that might not be available or easy to collect. CBF may not be suitable for recommending products such as music, art, movie, audio, photograph, video, etc. which are sold in e-commerce sites since these products may not be easily analyzed for relevant attribute information.   Both collaborative and content based recommendations have the problem when the new user or new items are added to the system because both depend on the user ratings.   Hybrid recommender systems are used to alleviate these problems. Ontologies can also be used to represent user profiles. The benefits of this approach are more intuitive profile visualization and the discovery of interests through inferencing mechanisms. CF is domain free and it is used in application domains where it is difficult to profile using Content based filtering. Model based approaches generally perform better than memory-based approaches in terms of top-k recommendations metric.   Second, models with tags perform better for dense data and are more efficient than the corresponding models without tags. Third, incorporating tags in the recommendation algorithms can help to obtain more accurate recommendation in the top 2% ranks.

In clustering methods the accuracy may be less compared to nearest neighbor algorithm but it may be used as the preprocessing step to reduce the candidate set. Clusters therefore provide an effect means to bridge the gap between users and resources. It works better on the denser dataset.

Network based methods and tensor based methods can overcome the sparsity of large-scale data but they only focus on the network structure without considering the relations among tags.

Linear-weighted hybrid resource recommendation provide a flexible, general and effective approach to capitalize on strong relationships across different dimensions of the data, and to incorporate the most effective components into a single recommendation framework.

The hybrid approaches works well because the algorithms complement each other.

Folksonomy-generated movie tag-clouds can be used to construct better user profiles that reflects a user's level of interest and provides a basis for prediction.

User Recommendation

User recommendation with Tensor Factorization outperforms well compared to Collaborative Filtering, Google Follower Finder and recommendation based on the KL-divergence between user interests.

In FCA, the disadvantage is the size of the concept lattice and the users and the tags in a lattice structure have a high degree of overlap among concepts. Therefore it is difficult to recommend the concepts when some queries are performed.

In STSs using MSN, as all the process is performed online, efficiency problem arises. So some tasks can be performed offline.

Unified Recommendation

When the relation between users, tags and resources becomes much sparser than usual, combining content and relation analysis for recommendation can extract knowledge required by the recommendation tasks from content information and reveal relation between different objects. By this characteristic, combining content and relation analysis for recommendation overcomes sparsity problem.

In Folk Rank, ranking is based on tags only, without regarding any inherent features of the resources at hand. This allows to apply FolkRank to search for pictures and other multimedia content, as well as for other items that are difficult to search in a content-based fashion.

Fuzzy-based ISID algorithm improves the performance of algorithms for interest discovery, through the clustering of syntactic variation in the data sources of social systems.

Ternary semantic analysis improves recommendations by capturing users multimodal perception of item, tag and user.

Semantic Web

SCOT, combined Gruber's conceptual model and Newman's vocabularies, is the ontology that must be suitable to represent collaborative tagging activities and it provides the most appropriate representations for the Folksonomy model. In addition linking between SCOT and MOAT is useful way to complement to define a meaning of tag. [8]

8. CONCLUSION

In this paper we have explored many possible recommendations in Social Tagging Systems. Research can be extended in any of the given recommendation technique by combining with other recommendation or semantic web. Since it explored two main building blocks of web 3.0 such as social web and semantic web, it will be more useful for future enhancement.

9. ACKNOWLEDGMENTS

10. REFERENCES

[1]   Alepidou, Zinovia I., Konstantinos N. Vavliakis, and Pericles A. Mitkas. "A semantic tag recommendation framework for collaborative tagging systems."*Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on.* IEEE, 2011.

[2]   Ana Belen Barragans-Martinez, Marta Rey-Lopez and Enriqe Costa-Montenegro, Fernando A. Mikic-Fonte, Juan C. Burguillo and Ana Peleteiro. Exploiting social Tagging in a Web 2.0 Recommender System. Information Overload November/December2010 IEEE Computer Society.

[3]   Cai-Nicolas Ziegler. Semantic Web Recommender Systems. EDBT 2004 Workshops, LNCS 3268, pp.78-79, 2004.

[4]   Deepa Anand, Bonson Sebastian Mampilli. Folksonomy-based fuzzy user profiling for improved recommendations. Expert Systems with Applications 41 (2014) 2424-2436

[5]   Denis Parra, Peter Brusilovsky. Collaborative Filtering for Social Tagging Systems: An Experiment with CiteULike. RecSys '09, October 23-25,2009, New York, USA.

[6]   Dominik Benz, Andreas Hotho, Stefan Stutzer, Gerd Stumme. Semantics made by you and me: Self-emerging ontologies can capture the diversity of shared knowledge. Web Science Conf. 2010, April 26-27, 2010, Raleigh, NC, USA.

[7]   Gayatri Gadepalli, ElkeRundensteiner, David Brown, Kajal Claypool. Tag and Resource-aware Collaborative Filtering Algorithms for Resource Recommendatoin. 2010 Seventh international conference on Information Technology.

[8]   Hak Lae Kim, Simon Scerri, John G. Breslin, Stefan Decker. The State of the Art in Tag Ontologies: A Semantic Model for Tagging and Folksonomies. *2008 Proc. Int'l Conf. on Dublin Core and Metadata Applications.*

[9] Hak-Lae Kim and Stepan Decker, John G. Breslin. Representing and sharing folksonomies with semantics. Journal of Information Science OnlineFirst, Published on September 24, 2009 pp. 1-16.

[10] Hak-Lae Kim, John G. Breslin, Stefan Decker, and Hong-Gee Kim. Mining and Representing User Interests: The Case of Tagging Practices. IEEE Transactions On Systems, Man and Cybernetics-Part A: Systems and Humans. Vol-41.No.4, July 2011.

[11] Hassan, Malik Tahir, et al. "Self-Optimizing a Clustering-based tag recommendation in social bookmarking systems." *ECML PKDD Discovery Challenge 2009 (DC09)* (2009): 85.

[12] Hotho, Andreas, et al. *Information retrieval in folksonomies: Search and ranking*. Springer Berlin Heidelberg, 2006.

[13] Hu, Meiqun, Ee-Peng Lim, and Jing Jiang. "A probabilistic approach to personalized tag recommendation." *Social Computing (SocialCom), 2010 IEEE Second International Conference on*. IEEE, 2010.

[14] Jäschke, Robert, et al. "Tag recommendations in social bookmarking systems." *Ai Communications* 21.4 (2008): 231-247.

[15] Jonathan Gemmell, Thomas Schimoler, Bamshad Mobasher, Robin Burke. Resource recommendation in social annotation systems: A linear-weighted hybrid approach. Journal of Computer and System Sciences 78 (2012) 1160-1174.

[16] Jose JavierAstrain, Alberto Cordoba, Francisco Echarte, Jesus Villadangos. An Algorithm for the Improvement of Tag-based Social Interest Discovery. SEMAPRO 2010 : The Fourth International Conference on Advances in Semantic Processing.

[17] M Lipczak, Y Hu, Y Kollet Tag sources for recommendation in collaborative tagging systems … - ECML PKDD discovery …, 2009 - kde.cs.uni-kassel.de

[18] M Lipczak. Tag recommendation for folksonomies oriented towards individual users - ECML PKDD discovery challenge, 2008 – Citeseer

[19] M Tatu, M Srikanth, T D'Silva Rsdc'08: Tag recommendations using bookmark content - ECML PKDD discovery …, 2008 - kde.cs.uni-kassel.de

[20] Malik Muhammad Imran Pattal, LI Yuan, ZENG Jianqiu. Web 3.0: A real personal Web. 2009 IEEE Third International Conference on Next Generation Mobile Application, Services and Technologies.

[21] Marius Kubatz, Fatih Gedikli and Dietmar Jannach. Local Rank – Neighborhood-based, fast computation of tag recommendation. E-Commerce and Web Technologies, 2011 - Springer

[22] MT Hassan, A Karim, S Manandhar… Discriminative clustering for content-based tag recommendation in social bookmarking systems- ECML PKDD …, 2009 - kde.cs.uni-kassel.de

[23] Panagiotis Symeonidis, Alexandros Nanopoulos and Yannis Manolopoulos. A Unified Framework for Providing Recommendation in Social Tagging systems Based on Terenary Semantic Analysis. IEEE Transactions on Knowledge and Data Engineering. Vol. 22. No. 2. February 2010.

[24] PrzemyslawKazienko, Katarzyna Musial and Tomasz Kajdanowiez. Multidimensional Social Network in the Social Recommender System. IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans. Vol. 41. No. 4. July 2011.

[25] Rae, Adam, Börkur Sigurbjörnsson, and Roelof van Zwol. "Improving tag recommendation using social networks Proceeding RIAO'10 Adaptivity, Personalization and Fusion of Heterogeneous Information." *Paris France*(2010).

[26] Rafailidis, Dimitrios, and Petros Daras. "The TFC model: Tensor factorization and tag clustering for item recommendation in social tagging systems."*Systems, Man, and Cybernetics: Systems, IEEE Transactions on* 43.3 (2013): 673-688.

[27] Sarwar, Badrul, et al. "Item-based collaborative filtering recommendation algorithms." *Proceedings of the 10th international conference on World Wide Web*. ACM, 2001.

[28] Shepitsen, Andriy, et al. "Personalized recommendation in social tagging systems using hierarchical clustering." *Proceedings of the 2008 ACM conference on Recommender systems*. ACM, 2008.

[29] Sigurbjörnsson, Börkur, and Roelof Van Zwol. "Flickr tag recommendation based on collective knowledge." *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008.

[30] Szomszor, Martin, et al. "Folksonomies, the semantic web, and movie recommendation." (2007)..

[31] Tesconi, Maurizio, et al. "Semantify del. icio. us: Automatically turn your tags into senses." *The 7th International Semantic Web Conference*. 2008.

[32] Xu, Dihua, et al. "A Collaborative Tag Recommendation Based on User Profile." *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012 4th International Conference on*. Vol. 1. IEEE, 2012.

[33] Zhang, Jing, et al. "Collaborative filtering recommendation algorithm based on user preference derived from item domain features." *Physica A: Statistical Mechanics and its Applications* 396 (2014): 66-76.

[34] Zhang, Yin, et al. "Combining content and relation analysis for recommendation in social tagging systems." *Physica A: Statistical Mechanics and its Applications* 391.22 (2012): 5759-5768.

[35] Zhang, Zi-Ke, Tao Zhou, and Yi-Cheng Zhang. "Tag-aware recommender systems: a state-of-the-art survey." *Journal of computer science and technology* 26.5 (2011): 767-777.

[36]   Zhenlei Yan, Jie Zhou. User Recommendation with
        Tensor Factorization in Social Networks. ICASSP
        2012.
[37]   Zhichen Xu, Yun Fu, Jianchang Mao and Difu Su.
        Towards the Semantic Web : Collaborative Tag
        suggestions. **Collaborative web tagging** workshop
        …, 2006.

A Model for Predicting Students' Academic Performance using a Hybrid of K-means and Decision tree Algorithms

Thaddeus Matundura Ogwoka
School of Open, Distance and eLearning (SoDeL)
Jomo Kenyatta University of Agriculture and Technology (JKUAT)
Nairobi, Kenya

Wilson Cheruiyot
School of Computer Science and Information Technology
Jomo Kenyatta University of Agriculture and Technology (JKUAT)
Nairobi, Kenya

George Okeyo
School of Computer Science and Information Technology
Jomo Kenyatta University of Agriculture and Technology (JKUAT)
Nairobi, Kenya

Abstract: Higher learning institutions nowadays operate in a more complex and competitive due to a high demand from prospective students and an emerging increase of universities both public and private. Management of Universities face challenges and concerns of predicting students' academic performance in to put mechanisms in place prior enough for their improvement. This research aims at employing Decision tree and K-means data mining algorithms to model an approach to predict the performance of students in advance so as to devise mechanisms of alleviating student dropout rates and improve on performance. In Kenya for example, there has been witnessed an increase student enrolling in universities since the Government started free primary education. Therefore the Government expects an increased workforce of professionals from these institutions without compromising quality so as to achieve its millennium development and vision 2030. Backlog of students not finishing their studies in stipulated time due to poor performance is another issue that can be addressed from the results of this research since predicting student performance in advance will enable University management to devise ways of assisting weak students and even make more decisions on how to select students for particular courses. Previous studies have been done Educational Data Mining mostly focusing on factors affecting students' performance and also used different algorithms in predicting students' performance. In all these researches, accuracy of prediction is key and what researchers look forward to try and improve.

Keywords: Data Mining; Decision tree; K-means; Educational Data Mining

1. INTRODUCTION

To predict how students may perform during their learning process is a complex task despite continuous increase of data in the databases relating to students academics in institutions of higher learning. According to (Marquez et al., 2013), the academic management systems are not designed properly to support educational managers to investigate which students are at risk of dropping out of university. Data Mining is the process of discovering interesting patterns and knowledge from large amounts of data (Han and Kamber, 2003). From educational Data Mining (EDM) website , "Educational Data Mining is an emerging discipline , concerned with developing methods for exploring the unique types of data that come from educational settings, and use using them to better understand students and the setting in which they learn" . To analyze students' learning process is complex, but thanks to EDM in which its methods and approaches can be used to predict students' performance like the model being proposed in this research paper. Hence university managers will have options of to come with strategies of improving student academic performance (Borka and Rajeswari, 2013). This research proposes the use of Decision tree classification and K-means clustering algorithms to develop a model for predicting the academic performance of students in higher level institutions like universities. Prediction is a method of carrying out Educational Data Mining (EDM) using clustering algorithms like K-means and classification algorithms like decision trees to predict student performance (Ramesh et al., 2013). In university, the performance a student at the end of every semester determines whether a student is to progress to the nest academic year which leads to the completion of his/her studies. Passing in these semester examinations is crucial since it will determine whether a student is to get to final year

which later realizes a student graduating and ushered into the economic development of a country. This is one of the reasons institutions of higher learning are established for (Patel et al., 2013). In this research paper, WEKA knowledge analysis software tool is used to for the analysis the algorithms used and the model performance.

2. RELATED WORK

Great work has been done and is always being done by this area of Educational Data Mining. From (Shovon and Haque, 2012)'s research, where they used k-means algorithm to predict the student learning activities by clustering them into: "Good", "Medium", and "Low" based on their GPA. They used 50 students as the training samples and concluded that their prediction accuracy was low and needed improvement in future. Our proposed model has realized improved accuracy by using 173 students as the algorithm training samples. (Yedav and Pal, 2012) using decision trees' ID3, CART, and C4.5 classifiers, conducted a study to predict student academic performance and realized an accuracy of 62.22%, 62.27%, and 67.77% respectively. In our research, using decision tree's J48, we realized a prediction accuracy of 98.8439% on the student training instances. (Kalpesh and Pal, 2013) in their model of predicting students' performance using decision tree's ID3 and C4.5 on 173 training datasets, achieved a prediction accuracy of 75.145% at 47.6 milliseconds execution time.

2.1 Data Clustering

Data clustering is unsupervised statistical analysis technique, which is used to segment large data into homogeneous groups called clusters, in order to discover hidden patterns and relationships to help in quick decision making (Shovon and

Haque, 2012). K-means is simple algorithm that partitions "n" observations into k clusters in which each member belongs to a cluster of nearest mean (Mustafa et al., 2010).

Figure 1. Basic K-means algorithm (Mustafa et al., 2010)

2.2 Decision tree classification algorithm

Decision tree is a data mining technique which can be applied in prediction tasks, it is a tree-like structure in which the root and each internal node are labelled with a question (Shovon and Haque, 2012). It is a classifier with the following structure: **Decision node** specifies a test on a single attribute, **Leaf node** indicates the value of the target attribute, **Arc/edge** split of an attribute, and **Path** is a junction test to make final decision.

3. PROPOSED MODEL APPROACH

In The proposed model is to take into account the educational data mining requirements as cited by other authors (Baker and Yasef, 2009) of keeping continuous record of student progress. The goal of this study is to predict students' academic performance. The approach will contain among others: Data tier which is the collection of student semester records, Application layer that will contain data extraction, loading, and knowledge repository. Then finally the presentation layer for displaying results analyzed to the user.

3.1 Data processing and analysis

In data mining, before running tests on the collected data instances, it is necessary to clean and prepare the data for use. In our research, sample data from the Technical University of Mombasa-Kenya student management system was cleaned to look at the relevance of the data attributes to be able to remove any redundancy, or irrelevant features, and analyzed using WEKA software tool. In our study, we have considered dataset of undergraduate students pursuing Bachelor of Science in Technology (BTIT), Bachelor science in Information Technology (BSIT) both government sponsored and self-sponsored, part-time and full-time students at the department of computer science and information technology (CSIT) in the Technical university of Mombasa which was our case study. According to (Suchita and Rajeswari, 2013), on the basis of the data collected some variable attributes are considered to predict student academic performance: Attendance%, Assignment%, Unit tests% and University result%. But for the purpose of this research, only useful fields from the single combined table were selected for our study. Some of the selected variables from the database and recommended attributes for academic prediction are shown below.

Table 1. Selected variables from student records

	VARIABLE	DESCRIPTION	POSSIBLE VALUE
1			
2	Dept.	Student's department	{CSIT}
3	AdmNo	Student's admission number	{Char}
4	Gender	Student's sex	{M,F}
5	Course	Student's course	{BSIT,BTIT}
6	EntryMode	Student's entry mode	{JAB, Self-sponsored(SS)}
7	CourseMode	Student's mode of studying	{Full-time(FT), Part-time(PT)}
8	semesterMarks	Student's total semester marks	{Numeric}
9	CAT	Whether Student sat for Continous Assessment Tests	{Y, N}
10	Attendance	Student's Average Semester Attendance	{Good, Poor}
11	MeanMark	Student's Average Semester Mark	{Numeric}

3.2 Analysis of data

Here we are evaluating the Decision tree algorithm using J48 classifier and K-means algorithm using Simple-Kmeans option from WEKA. From our framework architecture, it consisted of Data Tier where data from students is collected and interesting variables selected. Application tier where data cleaning and mining using algorithms is done producing a prediction model. We used First semester results to train the algorithm (this was the training Dataset), the second semester final results was predicted by the algorithms (this was our Test dataset). After successful learning, the algorithms were tested for prediction of second semester results in which the final grade column was left blank for the algorithm to predict the students' performance (predicted results).

Figure 2. Part of test dataset used for prediction

Part of the training data used for training the algorithms (first semester student records-cleaned) is shown below:

Figure 3. Part of training data set

This are part of the student datasets which is used to train the algorithms in our model. It represents the first semester student results. The test dataset of records represents second semester students' records in which the model is supposed to predict how the student will perform. And use the model to

cluster students in groups according to their predicted result relationships.

4. RESULTS & DISCUSIONS

Here we evaluated the experiments conducted using the algorithms (Decision tree using J48 and K-means) the analysis and results was done in WEKA. The decision tree algorithm was used to do the prediction and show in tree-like structure the results which will help in making decisions from the predictions made. The K-means algorithms run on the same dataset on the same WEKA tool, was used to group the predicted students into several groups of our choice in relation to the attributes of the records as shown below. Several classification metrics are used to evaluate the results (Manahaes et al., 2013):

> **Accuracy :** the measure of correctly classified instances
> **True positive (TP):** the proportion of positive cases (P) correctly classified as such and
> **True Negative (TN):** the proportion of negative cases (F) correctly classified as such.

After loading the training dataset into WEKA, the preprocessing yielded the following results as shown in Figure 4 below.

Figure 4. preprocessed training datasset results

From preprocessing results of training dataset, 143 students are identified as passed-P (in blue color as per the bar graph) and 30 students as failed-F (in red color as per the bar graph) in the fig 1above. This is true as per the original dataset, we had a total of 173 student samples. After preprocessing, by using Decision tree's J48 classifier, the algorithm was trained using the same dataset. The results are as shown in fig 5 below.
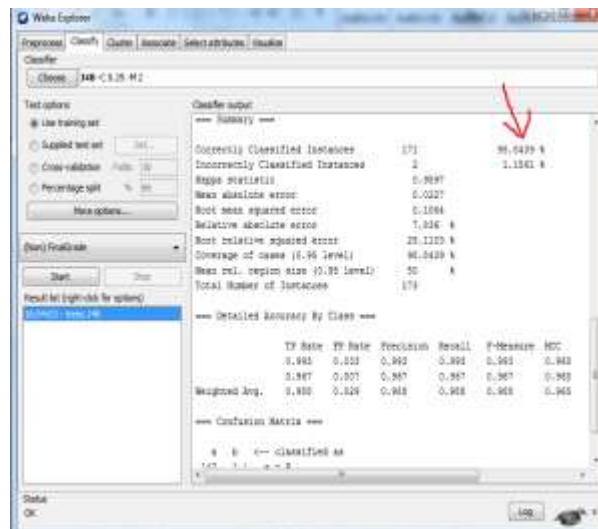
From the training result, 171 intances were classified correctly with an **accuracy of 98.8439%,** while 2 instances were incorectly classified with an **error** of **1.1561%.** The kappa statistic which takes account of similarities between classsses was 0.95978% which is better. :A kappa value greater than zero indicates the classifier is doing better than chance" (Manhaes et al, 2011). **TP Rate:** Rate of true positives (these indicated students who were correctly clasified in relation to final grade as the class)-the model realised 0.988 out of 1. **FP Rate:** Rate of false positives (these indicated students who were wrongly classified as belonging to a given final grade as aclass)-the model generated 0.029 out of 1. **Precision**: these represents a proportion of students that are truly of aclass(given final grade) divided by the total students classified as that class-the model generated 0.988 out of 1. **Recall**: these represents the proportion of students from the experiment who were classified as a given class (final grade) diveided by the actual total in that class (eqivalent to TP Rate)-the model generated 0.988 out of 1. **F-measure**: this is acombined measure of Predcision and Recall calculated as 2*Precision*Recall/(Precision+Recall) and it was 0.988. From all these acuracy measures, they all approached 1, hence showing our algorithm was learning well from the training instances.

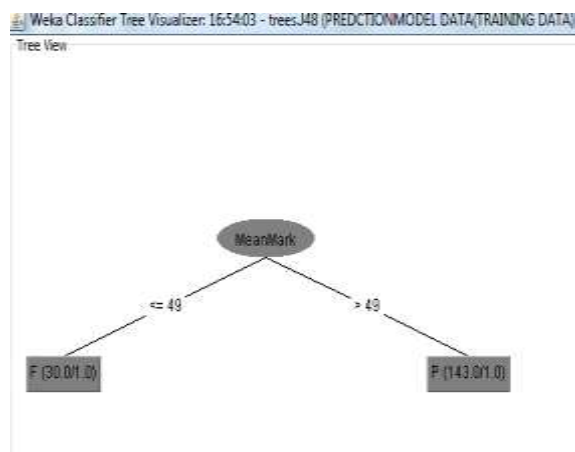The figure 6 below shows decision tree view generated by the model

Figure 6 decision tree view of the model

The view above shows in "tree"-structure one of the views that can be used in decision making. From the above results, 30 students whose mean marks were <=49 failed (F) while those with >49 mean marks in their first semester results passed (P).

4.1 Prediction results

After learning using first semester students results, we tested the algorithm with same students but on different results (second semester records) and the Class attribute "FinalGrade" was left blank for the algorithm to predict as was shown in figure 2. Using decision tree algorithm, on the test data, the following output were realized as predicted results as shown in figure 7 below.

```
@attribute AdmNo {BSIT/0018/2012,BSIT/029J/2012,B
@attribute Gender {M,F}
@attribute Course {BSIT,BTIT}
@attribute EntryModde {JAB,SSP}
@attribute CourseMode {FT,PT}
@attribute SemesterMarks numeric
@attribute CAT {Y,N}
@attribute Attendance {Good,Poor}
@attribute MeanMark numeric                   predicted
@attribute 'prediction margin' numeric
@attribute 'predicted FinalGrade' {P,F}
@attribute FinalGrade {P,F}

@data
BSIT/0018/2012,M,BSIT,JAB,FT,468,Y,Poor,59,1,P,?
BSIT/029J/2012,M,BSIT,JAB,FT,508,Y,Good,64,1,P,?
BSIT/048J/2012,M,BSIT,JAB,FT,441,N,Poor,55,1,P,?
BSIT/050J/2012,M,BSIT,JAB,FT,460,Y,Good,58,1,P,?
BSIT/052J/2012,M,BSIT,JAB,FT,516,N,Poor,65,1,P,?
BSIT/054J/2012,M,BSIT,JAB,FT,455,Y,Good,57,1,P,?
```

Figure 7 part of predicted results from the model

As shown from figure 7 above, the model was able to predict the second semester's final grades (as shown by arrow), and as shown from the decision tree view of the predicted results in figure 8 below, 30 students who failed (F) had their end of semester two mean mark <=49 while 143 students passed with a mean mark of >49.
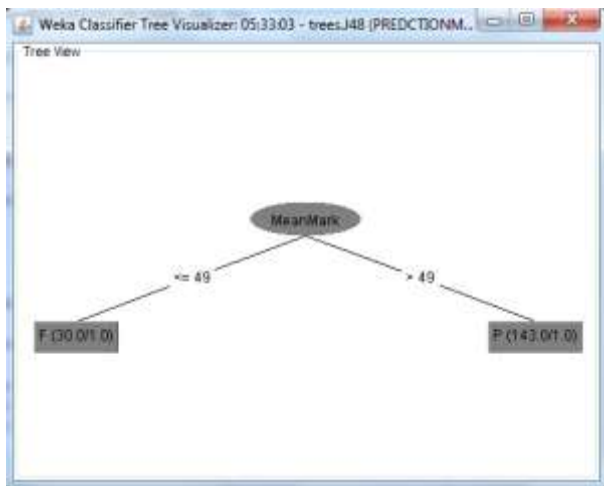

Figure 8 predicted results decision tree view

4.2 K-means clustering predicted results

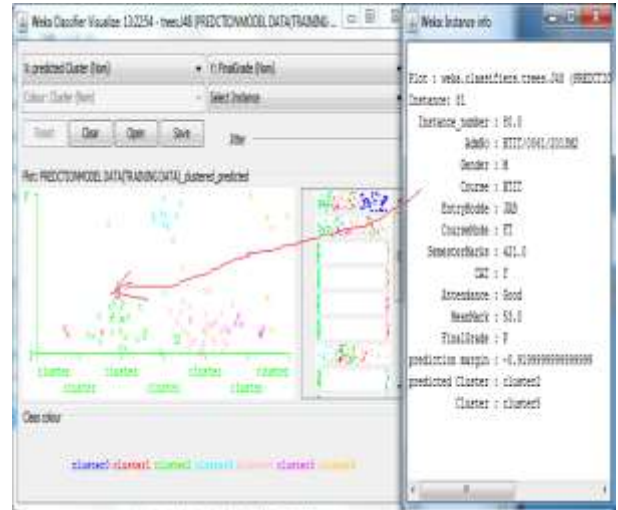From the same model, using WEKA, the following results were output.

Figure 9 predicted results clusters

"Clustering is the process of grouping a set of elements in such a way that the elements in the same group or cluster are more similar to each other than to those in other groups or cluster" (Kalpesh et al., 2013). As shown from the above fig 4.8, a student of admission number "BTIT/0061/2013M2" who is a male, entry mode by JAB, learning full-time, belonging to cluster0 passed with 53 mean mark, his attendance was good and had sat for the continuous assessment test.

4.3 Results summary

According to (Kalpesh et al., 2013) who predicted students' academic performance using ID3 and C4.5 algorithms, obtained the following accuracy as compared to what we obtained using J48 and K-Means algorithms as shown in table 2 and table 3 below.

Table 2 Algorithm accuracy. Source: (Ogwoka et al., 2015)

Algorithm	Total students	Correctly predicted students	Accuracy (%)	Execution time in milliseconds
J45	173	171	98.8439	20

Table 3 Algorithm accuracy. Source: (Kalpesh et al.; 2013)

Algorithm	Total students	Correctly predicted students	Accuracy (%)	Execution time in milliseconds
ID3	173	130	75.145	47.6
C4.5		130	75.145	39.1

5. SUMMARY OF FINDINGS

The main objective of our research was to apply decision tree and k-means algorithms to create a model for predicting students' performance. To achieve this, our model was realized through stepwise specific objectives:

1. We evaluated decision tree and k-means algorithms in terms of their operations using WEKA free software tool and other written literature.
2. We successfully applied decision tree and k-means algorithms and created a model of predicting students' academic performance where we analyzed 173 undergraduate students of Technical University of Mombasa's Computing and information technology department using first semester results to predict second semester results.
3. We tested the model of predicting students' academic performance and realized an accuracy of 98.8439% at an execution time of 20 milliseconds.

From the most previous researchers we looked into, their biggest challenge was to have an increased accuracy. All the their models as was seen in the previous literature, their accuracies were less than 90% with an execution time as big as 47.6 milliseconds (Kalpesh et al., 2013). Our results of 0.05 milliseconds execution time and accuracy of 98.8439% has reduced this gap.

6. CONCLUSIONS

A model for predicting students' academic performance using Decision tree and k-means algorithms has an improved accuracy and easily be implemented in institutions of higher to do prediction of students' performance and also mine interesting features pertaining academics of students.

7. RECOMMENDATIONS AND FUTURE WORK

From the results and findings of the experiments done in this study, the researcher recommends the adoption of student performance prediction models as Education Data Mining is an emerging data Mining discipline. In our research, WEKA does not update automatically on test dataset predicted as is the case on training dataset, hence to view the results you have to save in a file. In future, we will explore if WEKA has improved on this feature to use in our model or research more on more other open source data mining and analysis tools on this recommendation.

8. REFERENCES

[1] Marquec-vera, C. Cano, A. Romero, C., and Ventura, S. Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. Applied Intelligence, 2013, 1:1-16.

[2] Han, J., Kamber, M., and Pei, J. (2013). Data Mining Concepts and Techniques, third edition. Morgan Kaufmann.

[3] JEDM-Journal of Educational Data Mining, (JEDM. ISSN2157-2100). Meaning of Educational Data Mining. Retrieved from www.educationaldatamining.org/JEDM/index.php/JEDM.

[4] Burka, S., and Rajeswari, K. Predicting Student Academic Performance using Data Mining, 2013, 2(7), 213-219.

[5] Ramesh, P., Parkavi, P., and Ramar, K Predicting Student Performance. A statistical and Data Mining approach, 2013, 63(8), 0975-8887.

[6] Patel, M., Abdul, K., and Pappal, P. Educational Data Mining and its Role in Education Field, 2013, 5(2), 2458-2461.

[7] Ajay, P., and Saurabh, P., Data Mining Techniques in EDM for Predicting the Performance of Students, 2013, 2(6), 2279-0764.

[8] Shovon, M., and Haque, M., An approach of Improving Student Academic Performance by using K-means clustering Algorithm and Decision tree, 2012, 3:8

[9] Ahmed, A., and Elaraby, I., Data Mining: A prediction of Student performance using classification method, 2014, 2:43-47.

[10] Mustafa, T., Ayesha, S., and Khan, M. 2010. Data Mining Model for Higher Education System. European Journal of Scientific, 1450-216X.

Model-Driven Architecture for Cloud Applications Development, A survey

Amar Ibrahim E.Sharaf Eldein
College of Graduate Studies,
Computer Science and Information Technology
Sudan University for science and technology
Khartoum, Sudan

Hany H. Ammar
Lane Department of Computer Science and
Electrical Engineering,
College of Engineering and Mineral Resources
West Virginia University
Morgantown, USA

Abstract: Model Driven Architecture and Cloud computing are among the most important paradigms in software service engineering now a days. As cloud computing continues to gain more activities, more issues and challenges for many systems with its dynamic usage are introduced. Model Driven Architecture (MDA) approach for development and maintenance becomes an evident choice for ensuring software solutions that are robust, flexible and agile for developing applications.
This paper aims to survey and analyze the research issues and challenges that have been emerging in cloud computing applications with a focus on using Model Driven architecture (MDA) development. We discuss the open research issues and highlight future research problems.

Keywords: MDA; cloud computing; application development; Design Driven Software; SaaS;

1. INTRODUCTION

Architecture models are helpful for modeling complex software systems structure and dynamics[1].The software architecture model facilitates cloud providers to smoothly interact with the environment with interfaces that are embedded in all types of objects. The development environment could be useful for supporting applications, since they enable analysis, and experimentation in an easy and seamless way[2].Models of software requirements, structure and behavior at different levels of abstraction help all stakeholders decide how the system mission should be accomplished and maintained[3].Models consist of three design elements: processing, data and connecting elements [4].

Background of the Study

Models are used to predict systems properties and to gain better understanding of the systems, since changes in some parts of it will affect the rest of a system. In addition models help to communicate system characteristics to various stakeholders. The models are developed as blueprints to implement the physical system, and to understand its behavior[16].

1.1 Model Driven Architecture (MDA)

In this section we define the important terms used in this paper.

1.1.1 MDA Definitions

Definition 1: The MDA is an Initiative proposed by the Object Management Group (OMG)as an open, vendor neutral, approach to software development which is characterized by the use of models as the primary artifacts for understanding, design, construction, deployment, operation, maintenance and modification of a system [17].

Definition 2: Is a software development approach where the models are used as prime artifacts throughout the process of software development. These models are defined at different levels of abstraction to represent various aspects of the system [18].

Definition 3: Is the Attribute Driven Design (ADD) method is an approach to defining a software architecture in which the design process is based on the software quality attribute requirements,

follows a recursive process that decomposes a system or system element by applying architectural tactics and patterns that satisfy its driving requirements [18].

1.1.2 MDA Models

The models in MDA are abstracted at three different levels:-

1. **CIM(Computation Independent Model):** is a software independent business domain model that bridges the gap between business experts and system experts.
2. **PIM(Platform Independent Model):** specifies the functionality of the system independent of the technology that would be used for its implementation.
3. **PSM(Platform Specific Model):** specifies the system in terms of implementation constructs that are specific to the implementation technology. A single PIM can be transformed into one or more PSMs, each PSM being specific to the technology platform on which the system would finally be implement [18].
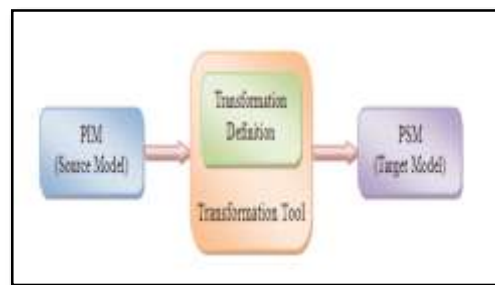

Figure 1.Transformation from PIM to PSM models [19].

The key to the success of MDA lies in automated or semi automated model to model and model to code transformations. The transformation tool executes a transformation definition that is specified for the purpose of transforming higher level, platform independent models into lower level platform specific models and finally into executable code[19].

1.2 Cloud Computing

Cloud computing appears to be a high technology; it has inherited legacy technology as well as new ideas on distributed systems. The computing power nowadays is needed for massive processing [5].The cloud architecture consists of three abstract layers: infrastructure, platform, and application.

Infrastructure is the lowest layer and is a means of providing processing, storage, networks and other fundamental computing resources as standardized services over the network.

Cloud computing regards infrastructure, platform, and software as services, which are made available as order based services in a pay as you go model to users [1].

The wide spectrum of available Clouds, such as those offered by Microsoft, Google, Amazon, HP, AT&T, and IBM, provides a vibrant technical environment, where small and medium enterprises (SMEs) can create innovative solutions and evolve their existing services offers[2].

1.2.1 Cloud Definitions

Definition 1:*"A Cloud is a type of parallel and distributed system consisting of a collection of inter connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resource(s) based on service level agreements established through negotiation between the service provider and consumers."* The user data and software reside on the Internet, possibly in datacenters and clusters located in different countries, which reduces the role of personal computer to a "dumb terminal" to access cloud via Internet. Computing resources are owned and managed by a cloud service provider (CSP). Using virtualization techniques, these virtualized resources, such as hardware, platforms, or services, are dynamically allocated to scale on demand according to customers' needs. If a CSP fails to offer the demand, the CSP may outsource to other CSPs [6].

Definition 2: According to the IEEE computer society cloud computing is:"A paradigm in which information permanently stored in servers on internet and cached temporarily on clients that include desktops, entertainment centers, table computers, notebooks, wall computers handhelds, etc." so cloud computing provide every facility as a service . It provides infrastructure as a service, software as a service and platform as a service[7]. In other words is a source for the dynamic provisioning of computing services, typically supported by state of the art data centers containing groups of networked Virtual Machines[8].

1.3 Objective

The main objective of our survey is to identify and analyze the current published research on model driven architecture for cloud applications development. We aim at identifying current architecture solutions for achieving quality of services and utilizing features of cloud computing for building reliable applications.

1.4 Organization

The paper is structured as follows. Section 2 presents the research method and specify the research questions and contributions. Insection3, we present and discuss related work. In addition we disuses the results of comparing different MDA methodologies and identify open issues of research. Finally in Section4, we present our conclusions.

2. RESEARCH METHOD

2.1 Research Questions and Contributions

The goal of our survey is to answer research questions as below: -

(1) RQ1. What are the research issues and challenges that are evolving in SaaS research?

(2) RQ2. Why do we need model driven architectures (MDA) in cloud computing?

(3) RQ3: what are the important research problems related to using MDA in developing cloud computing applications?

2.2 Search Steps

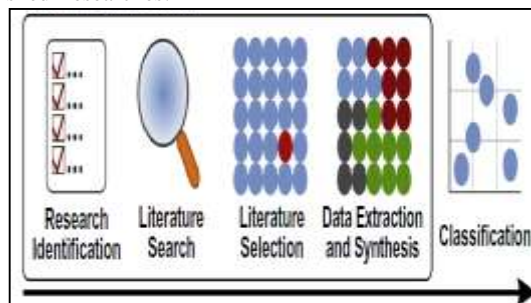To survey on area we follow the steps on figure below for published Researches:-

Figure 2. Steps of Research adapted from [47].

Search Processes are:-

2.2.1 Search Identification

The survey carries out a mapping study to find peer reviewed literature in order to classify and define topics related to software architectures development in cloud computing. From initially 95 papers, we selected 51 papers related to Model Driven Architecture, then analyzed and categorized of the papers using a taxonomy of8 categories.

The search process covered journal articles and conference papers available in electronic databases, namely, IEEE Explorer and Elsevier. These databases were selected because they are available for access and known for including high quality publications in this area. The databases include a representative sample of the literature produced in the subject matter as pertinent to this research. We were interested in recent articles limited to articles published between 2005 and 2014.

Table 1. Papers Publication Trend over Years

#	Year	Cloud Issues	MDA &Cloud	MDA &Cloud Dev
1	2014	8	5	5
2	2013	10	11	7
3	2012	10	1	6
4	2011	6	7	9
5	2010	1	1	1
6	2009	-	10	-
7	2008	-	2	-
8	2007	-	1	-
9	2006	-	2	-
10	2005	-	1	-
Total		**35**	**41**	**28**

2.2.2 Inclusion and Exclusion Criteria

This paragraph answers RQ1: issues and challenges for Software as Services. Major researchers have been mainly focusing on issues related to security and privacy, infrastructure, data management and Interoperability across different service providers. Some issues and challenges that practitioners consider important are under studied such as software related to evolving technologies.

A Number of challenges can be identified for software systems developers and operators such as:-

- **Migrating Legacy Systems to the cloud**: Such as modern the architecture to be more service oriented, data access layer, dealing with non functional requirements such as quality and using agile methods in the migration process.
- **Software Evolution and Inconsistency** in multi-tiered systems.
- **Software Testing** such as measuring the reliability of testing frameworks on the cloud.
- **Standardizing Application:** Cloud models and technologies are in their initial stage and are characterized by many critical issues which pose specific challenges from a software architecture perspective. SaaS model can be grouped into the following 10 factors: costs, security, availability, usability, implementation, ubiquity, flexibility, compatibility, analytics and best practices [13, 14].
- **Mobile Cloud:** With the increasing usage of mobile computing, exploiting its full potential is difficult due to its inherent problems such as resource scarcity, frequent disconnections, and mobility [9].Mobile devices are still resource poor. Battery life, weight, insufficient memory and heat dissipation limit the computational resources, and make mobile devices much more constrained [11].
- **Software Development:** The cloud platforms diversity and complexity pose challenges to the communication and coordination between software engineering and cloud providers during every stage in the software development process.
- **Security and Privacy:** Most papers aim to highlight the major challenges of security, privacy and trust issues in the existing cloud computing environments and help users recognize the tangible and intangible threats associated with their uses [12].
- Modeling and Simulation of Cloud Environments.
- **Business** to Cloud Services issues.
- **Management** issues.

Many of the above challenges are out of scope of our work, we focus on Model Driven Architecture for cloud applications development in next section.

3. DISCUSSION AND RELATED WORK

In this section we survey current work and determined different ways to compare achievements in development methodologies for cloud applications.

The Model Driven approach to the development of practical large scale applications is not an easy task **[40]**.For example, one of the difficult problems is how to obtain models at the early phases of development to address the concerns of different stakeholders.

3.1. Related Work

In this section we present a taxonomy of eight categories to in order to classify the research work accordingly.

3.1.1 Data Collection (Literature Selection)

The benefits of using the MDD in cloud computing are discussed as follows:-

- To "*reduce the primary software artifact's sensitivity to the inevitable changes that affect a software system*" this result in a reduction of cost, the main selling argument of MDD approaches, Others see the main goal in a *n increase of abstraction which makes the design of complex systems easier.* Since the abstraction from programming level issues makes the involvement of domain experts easier, by providing several different views on the same system [19].
- Improve the manageability of services without abandoning portability by create highly scalable applications and services using the providers' proprietary, also leading paradigm for efficiently managing services and applications in a highly automated manner.
- Cloud Computing technology evolution, incurs additional expenditure on part of the cloud service providers, as the applications in the cloud need to be reengineered with newer technologies , MDA approach is an asset which facilitates creation of good designs that easily cope with multiple implementation technologies and extended software lifetimes.
- Towards dynamic resource provisioning and define a cloud provisioning system based on mapping of functional and non functional tenancy requirements with appropriate resources [20].
- Improving the quality of cloud software services making them more robust, flexible and agile.
- Models Reflect separation of concerns by separating business functionality from implementation technology. While traditional software design and development processes create applications for deployment to a specific technology platform, MDA introduces higher levels of abstraction, enabling organizations to create models that are independent of any particular technology platform. The strength of MDA lies in the fact that it is based on widely used industry standards for visualizing, storing and exchanging software designs and models [16].
- Contemporary approach to software engineering that used to achieve greater portability of software in cloud.
- Migration applications across cloud provider for specific solutions.

Table 2. Gives taxonomy focus on eight category uses in cloud support by MDA

#	Category	Purpose	papers	Papers No
1	Interoperability	• The ability of computer systems to access, and exchange resources with one or more other performers and to use resources to accomplish its performed activities according to expected criteria.	[21];[44];[46];	3
2	Deployment	• Help developers to be able design their software systems for multiple clouds and for operators to deploy and re-deploy these systems on various clouds.	[22];[23],[43]	3
3	Development	• Development of software to mitigate unfavorable effects of technology changes.	[16];[24]; [25];[26];	4

#	Category	Purpose	papers	Papers No
4	Evolution	• Used as a process of developing software as a service initially on the basis of some requirements, also to model requirements from iterations to be evolving. • Used to improve dynamic cloud service in heuristic manner with healthiness validated.	[27];[42]; [46];	3
5	Quality attribute	• Multi tenancy helps to determine the number of resource provisioning to meet Service Level Objectives.	[28];[29];[40]	3
		• To run and mange multi-cloud systems, allows cloud solution that optimize the performance, availability and cost of the applications.	[46]	1
		• Achieves Reliability and scalability	[41]	1
		• Achieves Resource Scalability &Provisioning	[16];[41];	2
6	SOA	• Deliver services to either users or other services; created to satisfy business goals, using web services to handle communications.	[34];[35];[36]	3
7	Migration	• Provide holistic view to inform decisions when migrating to clouds. • Benefits organizations to select efficient transition architectures to increase productivity and reduce complexity.	[37];[38];[39]	3
8	Evaluation	• To analyze the impact of cloud adoption to identify potential risks and verify that the quality requirements have been addressed in the design, also to determine the robustness of systems.	[6]	1

3.1.2 Data Filtering &Analysis (Extraction & Classification)

Recently, many researchers presented in different studies approaches to use MDD to develop cloud applications.

Distributed data environment cannot be easily moved to any platform provider and may even cause problems to be used by a specific service (SaaS). It is almost impossible to move a service / environment between providers on the same level.

DaniloArdagnae.tal[6]implemented Model Driven Approach for the design and execution of applications on multiple Clouds (MODACLOUDS) that aims to support system developers and operators in exploiting multiple Clouds and in migrating systems from Cloud to Cloud as needed.

The work presented a framework and an Integrated Development Environment (IDE) used for developing and deploying applications in multi Clouds. In addition, a Decision Support System (DSS) is proposed to enable risk analysis for the selection of Cloud providers and for the evaluation of the Cloud adoption impact on internal business processes. Furthermore, the work offers a runtime environment for observing the system under execution and for enabling a feedback loop with the design environment that allows system developers to affect performance change and to redeploy applications on different Clouds on the long term.

However, there are many challenges such as Vendor Lock-in **[42]**to a specific vendor deployment.

The main concerns of cloud providers are how to deal with quality issues, the need for auto scaling mechanism for interoperability between different cloud platforms**[42],** and for developers to be able to design their software systems for multiple Clouds and for operators to be able to deploy and redeploy systems on various clouds. The portability of applications, data between Clouds, moving applications and data from Cloud to non-Cloud environments are not addressed. In addition Risk Management in the DSS uses only primitive tools and decision to support selecting and binding to a specific target cloud.

FrancescoMoscatoe.tal[40] Providers offer new different services to order customers' needs, Cloud Environments needs to define value added services (VAS) since users required increase complexity of services. The methodology uses Model driven Engineering and Model Transformation Techniques to analyze services , focused in using ontology to build modeling profiles that help to analyze complexity of systems, by developing open source platform that enables applications to negotiate cloud services requested by users via interface and targeted platform for developing multi cloud applications. One of the main goals is that of obtaining transparent, simple access to heterogeneous Cloud computing resources and to avoid locked in proprietary solutions. However, using ontology achieves interoperability but requires hard challenge.

Hugo & Manu Sood[41]paper explores the interaction between service oriented engineering and model driven engineering. The work shows how these methodologies can benefit from each other. The paper introduces modeling as a Service (MaaS) to provide modeling that allows the deployment and on demand execution for services on the cloud. There are many applications of Maas such as distributed modeling tools to allow the specification, sharing of software models, Definition of modeling, availability of model transformation engines, improving scalability, model execution

and evolution, solving tool interoperability problems, and distributed global model management.

However MDE models useful in the development of new SaaS applications could be used as part of service oriented architecture (SOA).There is no general agreement on the right set of models, languages, model transformations and software processes for the model driven development of SaaS systems. Furthermore, legacy system need to evolve and be adapted to be executed as service.

Mohammad Hamdaqa et al.[43]Presents a model driven approach for building cloud application solutions .The proposed approach presents a Reference model (Meta model) that facilitates cloud applications development from the design to implementation without depending on specific PaaS or IaaS components. This approach can be used by developers to better understand cloud applications independently of any specific cloud development environment. Moreover, the approach can improve flexibility by forcing developers to select a cloud vendor before porting the legacy application to the cloud.

Frey and Hassel bring,[44]presents a framework to facilitate the migration of legacy software to the cloud. The steps begin from existing legacy systems, extract the actual architecture then use a Meta-model to generate the target model to system migration. A reference model starts from the cloud platform to extract elements and vocabulary to create the cloud Meta-model.

The model needs to refine the syntax and create a platform independent modeling language for cloud applications.

There are several projects that aim at addressing challenges by providing solutions for provisioning, deployment, monitoring and adoption of cloud systems such as Modeling QoS constraints. In addition time consuming services are identified as a challenge for adaptive systems. Moreover handling failures during adaptation because multi cloud systems involves complex actions may consists of sub actions.

In addition, there is a lack of a systematic engineering process and tools supported by reusable architectural artifacts**[45]**.

The authors **Nicolas Ferry et al.[46]** proposed Cloud Modeling language (CloudML) explain that model driven techniques and methods facilitating the specification of provisioning and deployment concerns of multi cloud systems , this will enables the continues evolution of system between design time and run time activities. Argue model driven is suitable for developing complex systems.

Table 3.Related work using MDA for cloud applications development.

REF	Author/s	Approach	Advantages	Limitations	Domain
[6]	**DaniloArdagnae.tal** **2012**	**MODACLOUDS, model Driven Approach** for the design and execution of applications on multiple Clouds. Allows early definition and assessment of quality at design time.	• Supporting system developers and operators in exploiting multiple Clouds. • Migrating applications from Cloud to Cloud (performance). • Inform evolution process to design time.	• Vendor lock in on cloud customer to decide on adoption model. • Risk management used primitive tools. • Quality Assurance need mechanisms to be able deploy and redeploy systems.	Business Application
[40]	**Francesco Moscatoe.tal** **2012**	**MOSAIC Ontology methodology and Framework,** aims at creating, promoting, open source (API) and platform for developing multi cloud oriented applications. **Frameworks** enhance modeling profile for verification QoS of cloud services.	• Simple access to heterogeneous resources. • Design interface for users and implemented existing services. • Enable intelligent service discovery. • (QoS) given for users to avoid locked-in and for providers to build on demand services.	• Do not provide approaches to model and verify dependability during all phases of the life cycle. • Difficult to achieved interoperability.	Multi Agent System
[41]	**Hugo e.tal** **2010**	**Modeling as a Service (MaaS)** to provide modeling and model driven engineering services from the cloud.	• Using MDE for the development of SaaS applications. • Using SaaS to deploy modeling services in the cloud.	• No general agreement on the right set of models, languages, model transformations and software processes for the model driven development of SaaS systems.	SaaS application (SOA)

REF	Author/s	Approach	Advantages	Limitations	Domain
[43]	**Mohammad Hamdaqa et al** **2011**	Defining **A Reference Model (Meta model)** for developing cloud applications environment.	• Present Meta model shows main cloud vocabulary, design elements, configuration rules and semantic interpretation. • Facilitates cloud application development from the design to implementation.	• The lack of standardization and terminologies challenges portability and migration between different cloud environments.	Cloud Application
[46]	**Nicolas Ferry et al** **2013**	**Cloud Modeling language(CloudML)** aims at facilitating the provisioning, deployment, monitoring and adaptation of multi cloud systems.	• Enables the evolution of system between design time and run time activities. • Enables developers to work at higher level of abstraction of cloud concerns rather than implementation details.	• Model under development and many challenges identified such as. • Time consuming development activities for adaptation. • Techniques and methods to prevent failure. • Data movements from region to another without legal consequences.	Multi Cloud Systems

3.2 Discussion and Result

3.2.1 Result Analysis

In order to assess the benefits of MDA we provide a survey analysis on crosscutting concerns within the MDA context. We analyze MDA with respect to cloud applications development scenarios concerns**.** We first discuss the literature review, and outline literature selection based on cloud taxonomy issues stated in table (2).Furthermore, we compare different MDA used for cloud applications development to determine research challenges and limitations, as shown in table 3.

3.2.2 Open Issues

We discuss some issues and challenges that need further research in developing cloud applications based on MDA as follows:

- There is no generic cloud software architecture for designing and building cloud applications **[43]**, applications in cloud need to be reengineered with newer technologies **[20]** to distinguish the cloud development paradigm from the existing ones.
- Developers argue for the need for model driven techniques and supporting tools that facilitate the specification **[46]** of provisioning, deployment, monitoring and adaptation concerns at design time and their enactment at run time. The need to bind configuration management in order to minimize shortcomings**[19]**.
- Need to build cloud applications that offer cloud providers for auto scaling mechanism for interoperability between clouds **[6].**
- The lack of standardization and common terminologies that challenges portability, also need to migrate application's components between cloud providers as needed [**6**].
- Quality of applications pose a need for developers to be able to design their applications and for operators to operate, monitor and assure performance change of cloud applications to be able to deploy and redeploy on multi Cloud environments.

The above described effort in the area of applications on clouds indicates that the topic still requires research on new programming abstractions, developing and presents best solutions. We believe that the work described in this paper be a part of an interesting approach in this concern.

4. CONCLUSION

This paper surveyed research work on using model driven approaches for cloud applications development.

We reviewed literatures of cloud computing with emphasis on Model Driven architecture. We covered the major methods, and summarized their features in Table 3. We also discussed several open research issues.

The result shows that most of methods focus on the main concept of model Driven Design to provide arguments that justify the effort of using these approaches that promise to reduce the overhead of developing, configuring, deploying and maintaining cloud applications.

We believe our work provide a better understanding of principles and challenges of developing applications in clouds to help developers, architects and researchers to use, support, evaluate and predict different methods and techniques.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Chengzhang Peng, Zejun Jiang, Building a Cloud Storage Service System, Elsevier, Procedia Environmental Sciences 10 (2011) 691 – 696.

[2] Katia Vilae.tal Model-driven adaptation of question answering systems for ambient intelligence by integrating restricted-domain knowledge/ Elsevier,Volume 4, 2011, Pages 1650–1659.

[3] Tom Mens, Pieter Van Gorp, A Taxonomy of Model Transformation, ELSEVIER, Electronic Notes in Theoretical Computer Science 152 (2006) 125–142.

[4] BasGraaf,Model-Driven Evolution of Software Architectures, thesis, 2007.

[5] Bhaskar Prasad Rimal, Eunmi Choi, Ian Lumb, taxonomy, issues for cloud computing Ecosystems springer, 2010.

[6] DaniloArdagna&e.tal ,MODACLOUDS: A Model-Driven Approach for the Design and Execution of Applications on Multiple Clouds, IEEE, 2012 .

[7] National Institute of Standard and US Department of Commerce Technology, "The NIST Definition of Cloud Computing,"2012.http://csrc.nist.gov/publications/nist pubs/800-145/SP800-145.pdf ,1=(Emerging Issues & Challenges in Cloud Computing— A Hybrid Approach).

[8] Fa-Chang e.tal , The Impact of Cloud Computing Technology on Legal Infrastructure within Internet— Focusing on the Protection of Information Privacy, International Workshop on Information and Electronics Engineering, 2012.

[9] Niroshinie Fernando &e.tal, Mobile cloud computing: A survey, 24th DAAAM International Symposium on Intelligent Manufacturing and Automation, 2014.

[10] AleksandarBahtovski&e.tal, Cloudlet Challenges, 24th DAAAM International Symposium on Intelligent Manufacturing and Automation, 2014.

[11] Xiao Ma, Yong Cui, Ivan Stojmenovic, Energy Efficiency on Location Based Applications in Mobile Cloud Computing: A Survey The 9th International Conference on Mobile Web Information Systems, Procedia Computer Science 10 (2012) 577 – 584.

[12] Dawei Sun e.tal, Surveying and Analyzing Security, Privacy and Trust Issues in Cloud Computing Environments, Elsevier, 2011.

[13] Cloud Computing the New Frontier of Internet Computing, Published by the IEEE Computer Society, 2010.

[14] Björn Johansson, Pedro Ruivo, Exploring Factors for Adopting ERP as SaaS, Elsevier, 2013.

[15] Karakostas&Zorgios, Model-Driven Service Engineering, IGI Global, 2008.

[16] Ritu Sharma and Manu Sood, ENHANCING CLOUD SAAS DEVELOPMENT WITH MODEL DRIVEN ARCHITECTURE, International Journal on Cloud Computing: Services and Architecture(IJCCSA),Vol.1, No.3, November 2011.

[17] Attribute-Driven Design (ADD), Version 2.0, RobWojciketal, November 2006.

[18] AnnekeKleppe, Jos Warmer, WimBast: MDA Explained: The Model Driven Architecture: Practice and Promise. Pearson Education, 2003.

[19] R. Gitzel et al, Using established Web Engineering knowledge in model-driven approaches, ELSEVIER, 2007.

[20] AtulGohad and e.tal, Model driven provisioning in multi-tenant clouds, Service Research and Innovation Institute Global Conference, 2012.

[21] Towards Model Driven Tool Interoperability: Bridging Eclipse and Microsoft Modeling Tool.

[22] Integrating Configuration Management with Model driven Cloud Management based on TOSCA Towards Model Driven Tool Interoperability.

[23] Hugo Bruneli`ere, Jordi Cabot and Fr´ed´ericJouault, Combining Model-Driven Engineering and Cloud Computing, Modeling, Design, and Analysis for the Service Cloud - MDA4ServiceCloud'10: Workshop's 4th edition (co-located with the 6th European Conference on Modeling Foundations and Applications - ECMFA ,2010.

[24] Abderrahman Matoussi R´egineLaleau,A Survey of Non-Functional Requirements in Software Development Process,2008.

[25] Junichi Suzuki e.tal, Model-Driven Integration for a Service Placement Optimizer in a Sustainable Cloud of Clouds,2012.

[26] Ardagna, D. e.tal,A Model-Driven Approach for the Design and Execution of Applications on Multiple Clouds,IEEE,2012.

[27] Dr Kevin Chalmers ,Evolution Feature Oriented Model Driven Product Line Engineering Approach for Synergistic and Dynamic Service Evolution in Clouds,2013.

[28] AA Falasi, MA Serhani, R Dssouli, A Model for Multi-levels SLA Monitoring in Federated Cloud Environment,2013.

[29] FunmiladeFaniyi et al., A Dynamic Data-Driven Simulation Approach for Preventing Service Level Agreement Violations in Cloud Federation, International Conference on Computational Science ,Elsevier , 2012.

[30] Anders Henrikssonet& Uwe Aßman, Improving Software Quality in Safety-Critical Applications by Model-Driven Verification, Electronic Notes in Theoretical Computer Science 133, Elsevier, 2005.

[31] Petra Marešová&KateřinaPůžová, Application of the Cost Benefit Analysis method in Cloud Computing in the Czech Republic, WCBEM, Elsevier, 2014.

[32] Danny Thebeau II e. tal, Improving cyber resiliency of cloud application services by applying Software Behavior Encryption (SBE), Conference on Systems Engineering Research, Elsevier, 2014.

[33] R. Gitzel et al, Using established Web Engineering knowledge in model-driven approaches, ELSEVIER, 2007.

[34] Catalina Martínez-Costa et al, model-driven approach for representing clinical archetypes for Semantic Web environments, ELSEVIER, Journal of Biomedical Informatics, 2009.

[35] Mohsen MohammadiandMuriati Mukhtar, A Review of SOA Modeling Approaches for Enterprise Information Systems, The 4th International Conference on Electrical Engineering and Informatics, 2013.

[36] J. Knodel et al., An Efficient Migration to Model-driven Development (MDD), Elsevier , 2005.

[37] Mark Harman et al, Cloud engineering is Search Based Software Engineering too, The Journal of Systems and Software 86 (2013) 2225– 2241, Elsevier , 2013.

[38] Imran Khan et al, Migrating To The Cloud: Lessons And Limitations Of 'Traditional' IS Success Models,

Conference on Systems Engineering Research (CSER'13),Elsevier , 2013.

[39] AtifFarid Mohammad e.tal, Software Evolution as SaaS: Evolution of Intelligent Design in Cloud, Elsevier, Procedia Computer Science 19 , 486 – 493, 2013.

[40] FRANCESCO MOSCATO_AND B. DI MARTINO AND R. AVERSA, ENABLING MODEL DRIVEN ENGINEERING OF CLOUD SERVICES BY USING MOSAIC ONTOLOGY, Scalable Computing: Practice and Experience, Volume 13, Number 1, pp. 29–44. http://www.scpe.org, 2012.

[41] Hugo Bruneli`ere e tal., Combining Model-Driven Engineering and Cloud Computing published in "Modeling, Design, and Analysis for the Service Cloud - MDA4ServiceCloud'10:Workshop's 4th edition (co-located with the 6th European Conference on Modelling Foundations and Applications - ECMFA, 2010.

[42] Tsaie.tal. ,Service Oriented Cloud Computing Architecture, inITNG10,7th International conference on Information Technology: New Generations, pages 684-689, 2010.

[43] HamdaqaM., Livogiannis T. and Tahvildari L., A Reference Model for Developing Cloud Applications, 1st International Conference on Cloud Computing and Service Science, pages 98-103, 2011.

[44] Frey and Hasselbring, Model Based Migration of Legacy Software Systems into the cloud: The Cloud MIG approach ,12th Workshop Software Reengineering ,pages 1-2, 2010.

[45] Liang-Jie Zhang ; T.J. Watson Res, Architecture Driven Variation analysis for Design Cloud Applications, IEEE, 2009.

[46] Nicolas Ferry, Alessandro Rossini, Franck Chauvel, Brice Morin, Arnor Solberg, Towards model-driven provisioning, deployment, monitoring, and adaptation of multi-cloud systems, Sixth International Conference on Cloud Computing, IEEE, 2013.

[47] Maria Leitner, Stefanie Rinderle-Ma, A systematic review on security in Process-Aware Information Systems – Constitution, challenges, and future directions, Elsevier, Information and Software Technology 56 (2014) 273–293.

7. AUTHRORS BIOGRAPHIEES

Amar Ibrahim E.Sharaf B. Sc and M.Sc both in Computer science, Sudan. He is IT manager, College of Computer Studies-Khartoum at National Ribat University, also work as a lecturer and supervision of under graduate projects since 2008. He was worked as Software Engineer from 2004 to 2006 on software Requirement Specification, analyzing various companies systems. During 2007 to 2013 work as par timer in Islamic Research Studies center (CSICW) responsible for IT solutions and technical support. Currently, he is PHD student at Sudan University of Science and Technology , College of Computer Science and Information Technology (SUST), my research interests in Software Architecture and Cloud Computing.

Hany H. Ammar BSEE, BS Physics, MSEE, and PhD EE, is a Professor of Computer Engineering in the Lane Computer Science and Electrical Engineering department at West Virginia University. He has published over 170 articles in prestigious international journals and conference proceedings. He is currently the Editor in Chief of the Communications of the Arab Computer Society On-Line Magazine. He is serving and has served as the Lead Principal Investigator in the projects funded by the Qatar National Research Fund under the National Priorities Research Program. In 2010 he was awarded a Fulbright Specialist Scholar Award in Information Technology funded by the US State Department - Bureau of Education and Cultural Affairs. He has been the Principal Investigator on a number of research projects on Software Risk Assessment and Software Architecture Metrics funded by NASA and NSF, and projects on Automated Identification Systems funded by NIJ and NSF. He has been teaching in the areas of Software Engineering and Computer Architecture since 1987. In 2004, he co-authored a book entitled Pattern-Oriented Analysis and Design: Composing Patterns to Design Software Systems, Addison-Wesley. In 2006, he co-authored a book entitled Software Engineering: Technical, Organizational and Economic Aspects, an Arabic Textbook.

Impacts of Object Oriented Programming on Web Application Development

Onu F. U.	Osisikankwu P. U.	Madubuike C. E.	James G.
Computer Science Department, Ebonyi State University, Nigeria	Department of Computing and Engineering, University of Sunderland, UK	Computer Science Department, Akanu Ibiam Federal Polytechnic, Uwanna, Nigeria	Computer Science Department, Obong University Akwa-Ibom, Nigeria

Abstract: Development of web application nowadays can hardly survive without object oriented approach except for the purpose of just information display. The complexity of application development and the need for content organization has raised the need for web application developers to embrace object oriented programming approach. This paper exposes the impact of object oriented programming on web application development. The exposition was done through a detailed study and analysis of information from secondary sources. The internet was usefully employed to access journal articles for both national and international sources. Our study enables web developers and designers to understand web application features, tools and methodologies for developing web application. It also keeps researchers and scholars abreast of the boost which OOP has brought into Web Applications development.

Keywords: Object-Oriented Paradigm; Web; Web 2.0; RIAs; URL.

1. INTRODUCTION

According to [17], Web Applications are web sites which are dynamic in nature and uses server side programming to allow a good interaction between the user form at the front end, and the database at the back-end. Such applications include but are not limited to: Online Banking, Online Reservations, e-commerce or Shopping Applications, Online Training, Online Polls, Blogs, Online Forums, Social media sites, Location based services, and Content Management Systems. Web applications are mostly preferred to their off-line counterparts because of their ability to provide multiuser interface; centralized maintenance and updates; thin client integration; distributed processing; multilingual programming; and cross platform interoperability using either the Internet, Intranet or Extranet technologies. Web applications have become very popular in the software market because of this popularity; most application developers are converting their legacy applications to the Web platform. Hence, today, virtually all kinds of applications are supported on the Web. These applications range from small scale services to large scale enterprise software.

Early web implementations were only based on low level tools that do not support high level abstraction for sharing and reuse. This lack of appropriate abstractions made it difficult to create frameworks that supported the reuse of different existing components. However, with the evolution of web application in the late 90s, Web (software) engineers realized the need for modularization of concerns to accommodate big and complex applications [1]. Hence, technologies like Document Object Model (DOM) and Extensible Markup Language (XML) which supports object oriented principles (such as modularization, encapsulation and abstraction) were developed. This bridged the gap between higher and lower level implementation [12]. Today different frameworks and tools have been developed with maximum support of object oriented concept for web application development.

Our objective in this paper is to expose the impact of object oriented programming paradigm on web application development. Our study will also enable web developers and designers to understand web application features and popular languages, tools, and methodologies for developing web application.

The remaining sections of this paper are arranged as follows: In section 2 and 3, we reviewed computer classes and computer application; the features and popularity of web applications; web application development languages and tools, and methodologies for developing web application. However, our discussion in this section will be mainly focusing on the functionality and history/evolution of the tools. In section 4, we discussed the impacts of object oriented programming on web application. Finally, we concluded our paper with the contribution of object oriented programming on web application development.

2. LITERATURE REVIEW

Many of the studies in the area of web application development have mainly focused on the evolution of web application and comparison of web application development languages. Jazayeri wrote on trends and status quo of web application [16], Ronacher presented security related issues in web application [21], Voslro and Kourie wrote on concepts and web framework [25]. Purer highlighted some differences, advantages and drawbacks of PHP, Python and Ruby [20]. He compared the languages based on history, evolution, popularity, syntax, semantics, features, security and performance in web application environments. Cholakov analyzed PHP and summarized some drawbacks [6]. Gellersen and Gaedke in their article [12], overviewed object oriented web applications and identified object-oriented model for web applications, they found that XML technology contributes in enabling high level abstractions for design level modeling in a markup language. Mattsson identified the strengths and weaknesses of object oriented frameworks [18]. Finifter and Wagner explored the relation between web application development tools and security [10]. Chatzigeorgiou et al, evaluated object oriented design with link analysis [5]. Paikens and Arnicans explored the use of design patterns in PHP-based web application

frameworks [19]. French presents a new methodology for developing web applications and web development life cycle [11]. Copeland et al, in their article titled "Which web development tool is right for you" discussed and compared various tools for web application development [7].

However, not too many studies have been conducted in the area of impact of object oriented programming on web application development. This research aims at discussing the impact of object oriented programming on web application.

2.1 Computers and Computer Applications

Although, early computer applications runs on mainframe computers, and application development were mainly on a low level languages accessed through a low technology terminal device. However, the popularity of computer applications started when the Personal Computers (PC) were invented. This technological advancement made computers more affordable and accessible. Also software developers started developing standalone applications that run locally on the PC. PC networking technology was subsequently discovered and client-server applications were created, which allowed PCs to remotely connect to servers and share information. Create a link between PC networking and the internet. With the constant decrease in prices of microcontroller; high speed of internet connectivity and increase user participation on Web platforms, the Web became a dominant platform for applications development.

2.2 Classes of Computer Applications

There are several classifications of computer applications: based on uses (Graphics applications, accounting applications, etc); based on the developer (proprietary or non-proprietary). In-lieu of our discussion, we shall classify computer applications based the platform they run. These include:-

- **Desktop Application:** These are applications which run on the desktop computers. They are developed to be installed and run on a particular system hardware and operating system. One major advantage of desktop applications is that they are standalone applications; hence they do not need network connection to operate [4].

- **Web Applications:** these are client-server software applications that run on Web browser. Unlike desktop applications that depend mainly on the operating system or hardware platform, a Web application depends on the web browser and server. Any computer device with web browser and internet connection can run web application irrespective of the operating system and hardware platform [3].

- **Embedded Applications:** these are applications that are written inside Read-Only Memories (ROM) of computing controlled appliances. These types of applications differ significantly from desktop and web application mainly in environmental conditions such as real time, performance expectations, maintenance difficulty, changing platforms, just to mention [9].

- **Mobile Applications:** these are applications that run on mobile devices such as PDA, GSM and Android phones, etc. The early mobile phones were mainly capable of making and receiving calls, sending and receiving Short Message Service (SMS), without sophisticated applications. With Wireless Application Protocol (WAP) standard, new mobile applications that run on WAP browser started to emerge [24]. Just like standalone desktop applications, mobile applications can be standalone and as well web based or even embedded.

2.3 Web Application, Their Features and Popularity

Though early stage of the Web evolution, Web 1.0, which existed between 1990 and 2000 [8] enjoyed some level of growth due to its multiuser interface; single point maintenance and updates; distributed and hyperlinked documents etc. The level of popularity and user activity was still low owing to the fact that most of the websites developed using web1.0 was static and operated in brochure architecture with only professional web designers producing the content for users to access.

The Web platform we enjoy today, Web 2.0 has grown tremendously with resurgence of popularity and interest from millions of companies and billions of users across the world. Web 2.0 operates in architecture of participation were companies only provide the platform and users generate the content. Most sites on the Internet today like wikis, blogs and social media sites all present user generated content bringing the shift from few powerful professionals (programmers) to many empowered users [15]. In-lieu of this development, so many tools have been introduced to enable more user participation such as the Rich Internet Applications (RIA) and AJAX (Asynchronous JavaScript and XML) technologies. These technologies are used to develop web applications, which look and behave like desktop applications.

At the root of this Web 2.0 evolution, which is triggered by technologies like AJAX, Document Object Model (DOM), RIA, frameworks etc, is enshrined the concept of Object-Oriented Programmed (OOP). Hence, the knowledge of these tools and OOP cannot be ignored.

2.4 Web Application Tools and Languages

The World Wide Web operates in client-server architecture with the user making request using his/her Web browser at the client-side and the Web server, on the server-side, responding to user request. The entire Web 2.0 technologies can simply be categorized into AJAX technologies, RIA technologies and Frameworks. Hence, we shall discuss Web application tools and languages under the following: Client-side technologies, Server-side technologies and Web application Frameworks

2.4.1. Client-Side Technologies

These are Web technologies, tools and languages that run on the browser. Some of these tools include but are not limited to the following:

- **Web Browser:** The major responsibilities of web browsers are: (1) to generate and send request to web servers. These requests can be from hyperlinks, submitted forms, Uniform Resource Locator (URLs) and Hypertext Markup Language (HTML) pages. The request could also involve auxiliary resources like applets, and images, which accept and interpret responses from web servers to produce visual representation; (2) to render results in web browser window or tools from third party based on the type of content. [23]. Some of the most popular web browsers are Internet Explorer, Chrome, Firefox, Safari and Opera, etc.

- **RIAs client technologies**: Rich Internet Applications (RIA) are Web applications that behave and look like desktop applications. RIAs are the major attraction of the Web 2.0. RIAs client technologies are RIA technologies that run on the browser. This include Microsoft Silverlight, Linux Moonlight, Adobe Flash, Adobe Flex, etc

- **AJAX technologies**: Asynchronous JavaScript and XML (AJAX): this is one of the key technologies of Web 2.0 and RIAs. AJAX allows different sections/regions of a Web page to be refreshed separately while the user is still

on the same page. In RIA, AJAX technologies provide the fast response of the application. Common AJAX technologies are XHTML (Extensible Hypertext Markup Language), CSS (Cascaded Style Sheet), JavaScript, DOM, XML (Extensible Markup Language) [22]. AJAX implements the Object Oriented Model. For example using the DOM, each element of a Web page can be treated as an object which can be created, modified or removed independently.

2.4.2. Server-Side Technologies

These are technologies, tools and languages that run on the Web server. Some of tools include:

- **Web Server:** these are software applications that store the hyperlinked documents and other Web resources for users to access. IIS (Microsoft Internet Information Services) and Apache (developed by Apache Software Foundation) are the most popular Web servers today.
- **Database Management Systems (DBMS):** most Websites of today's Internet use DBMS to store their data. These databases are managed with DBMSs which are installed separately or as a suite. Common DBMS on the Internet are Structure Query Language (SQL) based. These include: SQL Server, MySQL, and Java DB.
- **RIAs Server technologies:** these are RIA languages and frameworks that run on Web Server. They include Perl, PHP, Ruby, Ruby on Rails, JavaServer Faces, Active Server Pages (ASP & ASP.NET) and Web Services.

2.4.3 Web Application Frameworks

Web application frameworks are sets of precompiled libraries, components and software tools arranged in such a way to separate content from presentation. This is done by dedicating the task of presentation formatting to web designers, while developers will be responsible for program logic and access the content [17]. This practice creates a consistent architecture among designers and developers for developing and accessing request framework elements that can be embedded within the presented page. Such requests including support for authentication, data access and transformation, session and state management. Frameworks provide the rich interfaces for RIAs giving them similar appearance as their desktop counterparts.

The table 1 shows some of the frameworks and tool support for web application development including IDEs.

Table 1. Web Application Development Tools, Frameworks and IDEs.

Tools	Frameworks	IDE
PHP	Zend, CakePHP, Symfony Project, Seagull, WACT, Prado, Zoop, PHP on TRAX, eZ Components, CodeIgniter etc.	Micromedia Dreamweaver, NetBeans, JetBrains, WebMatrix3 etc
JSP	Struct 2, JavaServer Faces (JSF), Spring MVC, Wicket, Stripes, Tapestry, RIFE, Seam, Google Web Toolkit (GWT), OpenXava etc.	Eclipse, JavaBean, JBuilder, Micromedia Dreamweaver, Sun Java Studio Creator, Borland JBuilder, Stylus Studio, Oracle JDeveloper etc.
ASP	ASP .NET MVC Framework, ASP.NET Dynamic Data, Mono Rail, OpenRasta, Component Based Scalable Logic Architeture(CBSLA), Base One Foundation Component Library(BOFCL) etc.	Visual Basic .NET, Cloudio, WebMAtrix3, MonoDevelop, JetBrains, WeBuilder, NetBeans etc.

3. WEB APPLICATION DEVELOPMENT METHODOLOGIES

A web application is a client to server application that generally adopt web browser as its client, the browser sends requests to the server and the server generate responses and send to the browser [23]. There are several approaches to web application development. The choice of development approach depends on several factors such as the nature of the information, scale of the application, frequency of content update, experience of the developer, and many more. In this paper, we briefly look at 3 methodologies.

1. **System Development Life Cycle (SDLC)**: is the traditional process of developing software or web applications by including research to identify and define the application requirements, information analysis, architectural design and specifications blueprint, team involvement, programming, testing and bug fixing, system testing, implementation and maintenance [11].

2. **Agile Software Development:** is a set of methodologies (like scum) that uses few resources to build software faster. Agile methodologies are often used for systems whose user requirements are constantly changing. [17].

3. **Object- Oriented Methodologies**: over the years a lot of software methodologies have been proposed by different authors using object-oriented concept to address contemporary issues on the Web. They include:

   a. WebComposition: is an object-oriented methodology that decomposes Web applications into components. Each component (page, site etc) is defined in terms of its state and behavior. This approach maintains support for software reuse, high level abstraction, run-time management and maintenance throughout the development cycle of the web application [13].

   b. W3Objects: this methodology encapsulates the web elements as objects with well defined interfaces for interaction with other objects. Each object can inherited to build new ones. Using W3objects, the problem of referential integrity and transparent object migration can be well handled [14].

   c. Jessica: is a web engineering system that employs object-oriented abstraction model of HTML++ (a methodology similar to W3objects). Jessica system is made up of two major component: an Object-oriented language component of the web application and a compiler that maps the abstract system definition into highly dynamic Web services [2].

4. DISCUSSION

Object oriented programming is currently the flexible programming pattern that every software developer or programmer employs, this OOP concept applies to both web application development and standalone software development. Many programming languages and frameworks particularly for web applications are now structured along-side OOP techniques. It is clear that all the recent frameworks for web application development maintain a solid structure of object oriented techniques and design patterns. It is therefore necessary to understand what impact object oriented approach

has on web application development. We have considered this under positive impacts and challenges posed by the approach.

4.1 Positive Impacts of OOP on Web Application

The popularity of web applications (web Apps) was the reason for the applications developers to evolve a more robust technique for the web App development. The adoption of OOP techniques in web App development has brought about the following:

1. Reduce the number of code a programmer needs to write: if the application a programmer is developing requires a similar functionality of already developed application, the programmer just need to edit little code to meet the programmers required function than writing the whole code from scratch.
2. Enable reuse of design and code function: in developing a web application a programmer do not need to write the whole code for the application, he/she can easily copy design or code function from other Apps and add to his/her project to suite his/her application needs.
3. Transfer design knowledge and experience to a developer: from the reuse of design developed by others, design knowledge and experience can easily be transferred.
4. Improved Maintainability of web application: since object oriented approach breaks complex programs into separate modules. Functions can easily be removed or updated on web page without affecting other functions.
5. Reduce the cost and time of developing a web application: With availability of function codes (for example one can get a JavaScript code that displays time in an application and integrate into a program) and rich library of object oriented programming, there is a reduced time and cost for web App development.
6. Improve application access restriction from unauthorized person: The concept of encapsulation in object oriented approach enables web application to specify access right to a function or content on web application. In fact, this benefit to web application has actually contributed a lot to the acceptance of web applications in several sensitive areas like e-Banking and e-commerce.
7. Create opportunity for web application development framework: using the object oriented concept, reusable framework elements have been developed. This can be embedded within the pages of a website to provide authentication, data access and transformation, session and state management.
8. It creates opportunity for integration of external web application with other web application such as Intranet and Extranet, by allowing object/module to be accessible to other functions or programs.
9. Proper computation and organization of complex tasks: object oriented approach helps to specify responsibilities and the distribution of these responsibilities among the cooperating objects. This helps to solve complex tasks and create opportunity for multi-tasking.

4.2 Challenges with OOP on Web Application development

Though the OOP technique sounds wonderful for web App development, it does present some challenges to web App developers. The most serious among these challenges include:

1. Requires expert knowledge: developers or programmers who are just starting-out will find it difficult to integrate HTML and Scripting language as it requires a descent level of understanding of various Object Oriented Programming (OOP) techniques and principles.
2. Difficult to find bugs in code especially when embedding scripting language with HTML code. Again, one can introduce unknown bug through code reuse from copying a function from other sources.
3. It creates opportunity for cyber attack through the support for integration of code from third party application. Object oriented pattern accepts integration of object from other source which can permit application attack.

5. CONCLUSION

The popularity of the Web and its impacts as a client-server system has led to a vast collection of web application. That was not so until frameworks and other technologies that supported object-oriented concepts were introduced. Our study has shown the various classes of computer and its application development more especially in the area of web applications. The object-oriented paradigm has been shown to be the most secured and flexible pattern for developing web application by shifting task from functional modules to a more confined level. Despite the few challenges with the integration of object-oriented concept in Web development, our conclusion is that the impact of object oriented approach on web application development (especially in the areas of resource reuse, extensibility, maintenance, scalability and delegation) are implicit: This impact has been tremendously positive and has brought about and a large acceptance of web applications. Hence web App development has become a boom due to OOP.

6. REFERENCES

[1] Athula Ginige, (2014). Web Engineering: Methodologies for Developing Large and Maintainable Web Based Information Systems.

[2] Barta Robbert A., & Schranz Markus W. (1998). "Jessica: An Object-Oriented Hypermedia Publishing Processor: Computer Networks and ISDN Systems. Proceedings of the 7th International WWW Conference. Vol. 30(1-7), pp.281-290.

[3] Borland JBuilder (2002) Web Application Developers Guide; Borland Software Corporation, 100 Enterprise Way, Scotts Valley, CA 95066-3249

[4] Chappell David (2011). What is an Application Platform?; Microsoft Corporation, Chappell & Associates.

[5] Chatzigeorgiou Alexander, Xanthos Spiros & Stephanides George (2004). Evaluating Object Oriented designs with link analysis; Proceeding of the 26th International Conference on Software Engineering, IEEE Computer Society

[6] Cholakov Nikolaj, (2008). On some drawbacks of the php platform, CompSysTech '08: Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing (New York, NY, USA), ACM, pp. II.7.

[7] Copeland Dennis R., Corbo Raymond C., Falkenthal Susan A., Fisher James L., & Sandler Mark N. (2000). Which Web Development Tool Is Right for You? IT Pro IEEE

[8] Deitel, P. J. & Deitel, H. M. (2007), Java How To Program, USA, Pearson Inc., 7th Ed., pp. 421 - 423.

[9] Ebert Christof & Jones Capers (2009): Embedded Software: Facts, Figures, and Future: IEEE Computer Society

[10] Finifter Mathew & Wagner David (2010). Exploring the Relationship between Web Application Development Tools and Security

[11] French M. Aaron (2011). Web Development Life Cycle: New Methodology for Developing Web Applications, Journal of Internet Banking and Commerce, vol.16, no.2

[12] Gellersen, H. & Gaedke, M. (1999). Object-Oriented Web Application Development. IEEE Internet Computing. Accessed from http://computer.org/internet on 3/8/15

[13] Gellerson, H. W.; Wicke, R.; and Gaedke, M. (1997). WebComposition: an object-oriented support system for the Web engineering lifecycle, Computer Networks and ISDN Systems, 29(8–13): 1429–38, April 1997.

[14] Ingham, D.; Little, M.; Caughey, S. and Shrivastava, S. (1995) W3Objects: Bringing Object-Oriented Technology to the Web , Fourth International World Wide Web Conference, Boston, Massachusetts, 1995, USA. URL: http://www.w3.org/pub/Conferences/WWW4/Papers2/14 1/

[15] James Governor, Dion Hinchcliffe & Duane Nickull (2009). Web 2.0 Architectures. 1st ED, O'REILLY Media Inc. ISBN: 978-0-596-51443-3

[16] Jazayeri Mehdi, (2007) Some trends in web application development, FOSE '07: 2007 Future of Software Engineering (Washington, DC, USA), IEEE Computer Society, pp. 199

[17] Kohan, B. (2015). Guide to Web Application Development: Guides, Resources, and Best Practices. Accessed from http://www.comentum.com/guide-to-web-application-development.html on 02/08/15

[18] Mattsson Michael, (1996). Object-Oriented Frameworks: A Survey of Methodological Issues. CODEN: LUTEDX(TECS-3066)/1-131

[19] Paikens Andris and Arnicans Guntis (2008). Use of Design Patterns in PHP-Based Web Application Frameworks, Latvijas Universitātes Raksti:Datorzinātne Un Informācijas Tehnoloģijas

[20] Purer Klaus (2009). PHP vs. Python vs. Ruby - The web scripting language shootout; Vienna University of Technology, Institute of Computer Languages, Compilers and Languages Group, 185.307 Seminar aus Programmiersprachen

[21] Ronacher Armin, (2006) Sicherheit in Webanwendungen, http://dev.pocoo.org/ blackbird/fachbereichsarbeit.pdf

[22] Sareh Aghaei, Mohammad Ali Nematbakhsh & Hadi Khosravi Farsani (2012). Evolution of the World Wide Web: From Web 1.0 to Web 4.0, International Journal of Web & Semantic Technology (IJWesT), Vol.3(1)

[23] Shklar Leon and Rosen Richard (2003) Web Application Achitecture: Principles, Protocols and Practices, John Wiley & Sons Ltd, ISBN 0-471-48656-6

[24] Tejas Vithani & Anand Kumar, (2014). Modeling the Mobile Application Development Lifecycle: IMECS, Hong Kong, ISBN: 978-988-19252-5-1, Vol.1

[25] Voslro Iwan and Kourie Derrick G., (2008). Server-centric web frameworks: An overview, ACM Comput. Surv. 40 (2).

Study on Efficient Algorithm and Differential Private Frequent Item set Mining

M.Banu priya

Department of Computer
Science and Engineering
Jansons Institute of
Technology
Coimbatore, India

Abstract: Frequent item set mining is to find the set of item occur frequently in the database Transactional database are insufficient to analyze the data in current shopping trends and dynamic dataset that update in data set. Discovering frequent item set play an important role in mining association rules, clusters ,web long mining and many other interesting pattern among complex data Efficient algorithm for analyze frequent item set based on the memory utilization and performance at the run time .Differential private FIM to find high data utility and high degree of privacy in the database

Keywords: Frequent item set (FIM), cluster association rule, Differential Private frequent item set

1. INTRODUCTION

Frequent item set play an essential role in many data mining task that try to find out interesting pattern from databases such as association rule, correlation ,sequences, classifiers and clusters .association rule helpful for analyzing customer behavior in retail trade, banking system etc.Frequent item set tries to find item set that occur in transaction more frequently than given threshold.FIM treat all the item having the same unit profit.Differential privacy offer strong privacy of released data without making assumption about an attacker background knowledge . Sequential pattern mining is define to finding statistically relevant pattern . The customer buying first a mobile phone,data cable and memory card if it occur frequently in a shopping history database is a sequential pattern. Association rule can be defined as{X,Y}=>{Z} the customer buys X,Y product and they like to buy Z..

2 RELATED WORK

2.1 Association rule

An association rules provide information in the form "if-then" statements. Association rules are computed from the data and unlike the if-then rules of logic but the association rule are probabilistic in nature .if 90% of transaction that purchase mouse and pen drive ,then also purchase headset.

Antecedent: mouse and pen drive

Consequent: headset

Confidence factor: 90%

As an addition to the antecedent(the "if" part)and the consequent (the "then" part) an association rule has two number that express the degree of uncertainty about the rule .Association analysis the antecedent and consequent are set of item.

of the item set are maintained in a tree-based structure called utility pattern tree. The construction of the UP-tree can be performed

	ASSOCIATION RULE	APERORI	FREQUENT PATTERN
Advantages	Predictability ,support	Easily parallelized	Only 2 pass over data set No candidate generation
Disadvantage	- Obtaining non interesting rules - Huge number of discovered rules - Low algorithm performance	- Bottle neck problem - Requires many database scan	- It is tree based structure when the database size increase when the node disconnect difficult to find the item set - It does not fit into the memory

FIGURE 2: ALGORITHM MERITS AND DEMERITS

2.2 Apriori Algorithm

This algorithm uses the association rule to find out frequent item set in transactional database. It uses bottom up approach for finding frequent item set and breath first search method .Hash tree to count the item set effectively

2.3 FP growth Algorithm

Frequent pattern algorithm is fragment growth method .it store all the item set in tree structure .FP growth algorithm use both horizontal and vertical database to store data set in main memory.FP growth algorithm is based on the depth first search method A frequent item pattern tree is a tree structure .

2.4 UP growth Algorithm:

Utility pattern growth algorithm for mining high utility item set. The information

With the two scan original database Step involved while construct UP-tree During First scan of the Database

STEP1: TU (Transaction Unit) of each transaction is computed
STEP2: TWU (Transaction Weight Unit) of each single item is also calculated
STEP 3: Discarding global unpromising items.
STEP 4: Unpromising item are removed from the transaction
STEP 5: The remaining promising item in the transaction are sorted in the Descending order TWU
Second scan
STEP 6: Transaction is inserted into UP (Utility Pattern)

3 DIFFERENTIAL PRIVACY

Differential privacy aim to publishing the frequent item set from the original database without providing the information about individual details. Differential privacy

ensures that the removal or addition of a single database item

Differential Privacy main concept is sensitivity. The amount of injected noise is carefully calibrated to the sensitivity .the Laplace Mechanism add noise drawn roughly from the Laplace distribution. To support multiple differentially private computation sequential composition guarantees overall privacy

4 EFFICIENT METHOD FINDING FREQUENT ITEM SET AND DIFFERENTIAL PRIVACY

Cheng –Wei Wu[3] :paper proposed lossless and compact representation named high utility item set(frequent item set).To mine the representation three algorithm are proposed AprioriHC(Apriority-based approach for mining high utility closed item set),AprioriHC-D (AprioriHC algorithm with discarding unpromising and isolated item) and CHUD(Closed High utility Item set).AprioriHC discard the global unpromising item and Isolated Item Discarding Strategy for finding itemset.AprioriHC perform breadth-first search in horizontal database and CHUD perform depth first search in vertical database .Derive High Utility Item set (DAHU) for efficiently recovering all the High utility item set from the CHUD

Vincent Tseng[4]:Proposed two efficient algorithm UP-growth and UP-growth+ for mining high utility item set from transactional database .In this algorithm construct UP-tree for maintaining information about high utility Item set .Potentially High Utility Item(PHUI) set have been generated from the UP-tree with only two scan of the original database

Freddy Chong Tat Chua [5]: Proposed a social correlation framework that incorporates a probabilistic social correlation into a latent space approach. Two generative models are used they are Sequential Generative model describe the social correlation matrix and Unified Generative Model. In this method efficient parameter estimation solution based on the expectation maximization. In this paper is focused on item adoption predication based on the social links Differential private transaction splitting methods

Sensu[6]: proposed Private FP-growth (PFP-growth) Algorithm. It consist of two phase they are preprocessing phase and mining phase .In the preprocessing phase improve the utility–privacy Tradeoff and smart splitting method to transform the database. In the mining phase a run time estimation method is used to estimate the actual support of item set Downward closure property used in dynamic reduction method to dynamically reduce the amount of noise added to provide privacy during the mining process

Cynthia Dwork:[7]:proposed privacy mechanism add noise to the transferring data item set. In this Gaussian noise are used instead of Laplacian.A finer analysis of realistic attack and to provide differential privacy are used sharpen the result and difficult to analyze the techniques used for transforming the data sets.

5 CONCLUSIONS

Frequent item set is very important to find out from the large data set. Online transaction has increases need to find out which item has frequently access.Privacy mechanism discussed adds an amount of noise in the data set. Summary of the in-depth analysis of few algorithm s is done which made a significant contribution to the search

of improving the efficiency of frequent item set mining algorithm

6 FUTURE WORK

In this paper, we done comparative study and analysis of efficient algorithm techniques, but these techniques have pros and cons, therefore there is necessity to develop such technique to overcome the entire disadvantage to find frequent item and to provide privacy accessing data from the database

7 ACKNOWLEDGEMENTS

8 REFERENCES

[1] R.Agrawal, T.Imelinski and A. Swami "Mining association rule between set of items in large databases" , in proceedings of the ACM SIGMOD Int'l Conf.on Management of, data pp 207-216 , 1993.

[2] Rakesh Agrawal and Ramakrishnan Srikant , "Fast Algorithms for Mining Association rules" ,in Proceedings 20th Int.conf.very large Data Bases ,pp,487-499 ,1994

[3]Cheug-wei wu, Philippe Fournier –viger, Philip S.Yu "Efficient Algorithms For Mining The Concise and Lossless Representation of Closed+ High Utility Item sets" Unpublished

[4] Bai-En, Shie, Cheng-Wei Wu, Philip S.Yu and Vincent S.Tseng "Efficient Algorithm for Mining High Utility Itemset from Transactional Database", IEEE, Vol 25 No8 ,2013

[5] Freddy ChongTat Chua, Hady W.Lauw,Ee-peng Lim "Generative Models for Item Adoption using Social Correlation", unpublished

[6]Sen Su,Shengzhi Xu,Xiang Chang,Zhengi Li and Fangchun Yang "Differentially private Frequent Itemset Mining via Transaction Splitting",IEEE 2014

[7]Denning,D.E."Secure Statistical Databases Sample Queries ", ACM, 1998