Breaking Into Social Nervous System

Architecture for Reality Mining

Deepak Kr Vashisth
Department of Information Technology, MD University, Rhotak ,India
vashisth1@live.com

***Abstract*— Gathering and analyzing of real time communicational traces in a society empowers us to formulate behavioral structure of its members. Reality Mining the core concept to support this enables us to collect digital breadcrumbs left by people while they perform their daily activities. Collection of these signals through sociometric badges and then formulating them for a visual view is shown in the further section of this paper. The model proposed in this paper is based on multi-level information gathering and filtration system. In this model society is divided in groups on the basis of their intra-group and inter-group interactions. It determines the sequestered groups and the quickest information distributing group. This filtration is processed on the server and all the data transactions are accomplished with secure protocols. For collection of communicational traces we argue use of mobile devices as sensors, which process data to further server. Further incorporation of influential model and centrality approaches enable us to detect most influential person in the sub-group. Implementation of web based multi-level architecture allows easy extension, wider area coverage, storing and processing large log records and easy integration with preexisting communication network.**

Keywords— Sociomertic badges; multilevel-architecture; intermediate server;centrality; clique; styling; SAML.

I. INTRODUCTION

Nourishing a strong, innocuous, and effectual society is a challenge that takes us back to our Independence when the newly emerging power required dealing with the social and environmental problems. The therapy then was to contrive consolidated network that delivers hygienic water & food, eradicate waste, provided energy, facilitate transportation, and offers access to centralized healthcare, police, and educational services. Since then all these problems are regularly addressed by every government and same solutions are consecutively iterated without much amendments. These century-old solutions are progressively obsolete. Today's social structures are not designed as integrated systems and do not take advantage of new digital feedback technologies that would allow them to be dynamic and responsive. We need to fundamentally restructure our methodology. Rather than acknowledging basic social requirements like water, food, waste, transport, education, energy, and so on as isolated factors we must consider them holistically. Instead of concentrating only on access and distribution systems, we require dynamic, networked, self-regulating systems that take into account complex social interfaces. In short, to ensure a sustainable future society, we must use embryonic technologies to create a nervous system for society that upholds the constancy of public health systems. We need to reinvent societal systems to generate a framework that analyze the situation, integrate the observation with dynamic demand and response prototypes and then use the resulting prediction techniques to implement the measures.

Here comes our most widely used sensor network i.e. cellular-phones in focus. According to Telecom Subscription Data as of 31 Oct 2013 total numbers of mobile phone subscribers have reached 875.48 million. So there is no other technology that is in constant touch with such huge subscriber strength. Every working person has its share in this technology. Hence with the use of mobile base system a huge amount and equipment requirement is reduced. It is already used in India to trace terrorists' activities and proved use full, as government claims that it has caught huge bulk of terrorists with their telecommunication traces with the help of social sensors called sociometric badges [2][3]. These sociometric badges provide a reliable source to examine unusual changes in a human behavior facing particular circumstances. On analyzing data of these sociometric badges we can determine behavioral stereotypes of considered group and with these we can predict act of the group on a particular incident.

This concept was first introduced by Jacob Moreno in "Who Shall Survive?" where he defined the process of converting a social communicational interaction into an image of graph. Graph of connected nodes corresponding to each person in the group, the connection shows the interaction between these nodes [5][4]. This method was based on the ad hoc which was not well-defined procedure. To overcome this others used computational procedures so that the points which are in continuous interaction or member of same group are placed in close proximity [4]. They used two famous approaches (i.) MDS (multidimensional scaling) proposed by Kruskal and Wish [6][4] (ii.) SVD (singular value decomposition) proposed by Weller and Romney [7][4]. MDS are matrix of proximity and distances which are used to construct graphs, in SVD the data matrix is transformed into basic structures by representing the considered points into

2014 IEEE Canada International Humanitarian Technology Conference - (IHTC)
978-1-4799-3996-1/14/$31.00 ©2014 IEEE

new vector points. These were the earlier approaches used by researchers to collect and to represent the behavioral stereotypes.

The model proposed in this paper is developed to generate better, understandable demographics that displays societal nervous system while accommodating basic "Influence Model" properties as determined by MIT's Human Dynamic Lab. This model collects communicational signals through mobile device and social networking sites. By analyzing these communicational traces we forecast the behavioral stereotypes and most valuable node in a social group or clique [4][8]. This valuable node is the base on information transfer in a clique. We divides group of people into small sub-groups (clique) until some kind of disjoint factor is present in them. By analyzing the clique we can omit individual analysis, so the overall workload is reduced.

II. RELATED WORK

The term Reality Mining refers to field of recognizing complex social signals using statistical analysis and machine learning algorithms [9][10]. These social signals provides us extensive information about once life in correlation with the surrounding people , and are efficiently sensed and gathered by cellular-phones, which offers wide range of mechanism to collect such data. Earlier Bluetooth enable phones were used to recognize social patterns in daily user activity [9]. Badges were used to provide data to build statistical graph to display the daily communicational practices adopted by the subscribers. Earlier the analysis of the group behavior was manual where it required lots of time consuming stuffs like observing the group, recording their movements. To enhance this concept Dong and Pentland presented an automated process for better analysis and greater efficiency using the concept of reality mining by mobile devices [11].

Reality mining had been used in managing traffic congestion, where GPS technology in cars are used as sensors and a proper feedback from a server forecast the prediction of congestion hours and days before it happen. Similarly mobile sensors had been used in India by the Intelligence Bureau to trace terrorists' activities. These are also used by the disaster-management agencies to gather information on affected portion where hazard occurred.

It had also been used in advertising industries to create first publicity blitz which increases the chances of initial selling. If we talk about marketing reality-mining is very efficiently used by Professor Alex "Sandy" Pentland and Tracy Heibeck [12] in their article "Making Meeting Work" where they used device "meeting mediator" which collects the voice and motion signals of the members to provide interaction based data [13]. Such sociometric badges are widely used to capture social interaction and sophisticated non-linguistic social signals [9].

Previously proposed Client-Server architecture is divine but *group specific*, for large population it is necessary to divide in highest proximity group [2]. These groups illuminate behavioral stereotypes of its members. Other limitation like *Server failure*-as all the data is transferred to a single server, failure of which can cause immense increase in log records within sensors , and *memory limitation* - sensors have small and limited memory, large amount of log records can even lead to reduction in its functioning ability. To overcome these three levels architecture is proposed in this paper.

Data visualization should be more human understandable to enhance the efficient formation of conclusions. This visual support is achieved by incorporating D3 statistical documentation.

III. PROPOSED MODEL

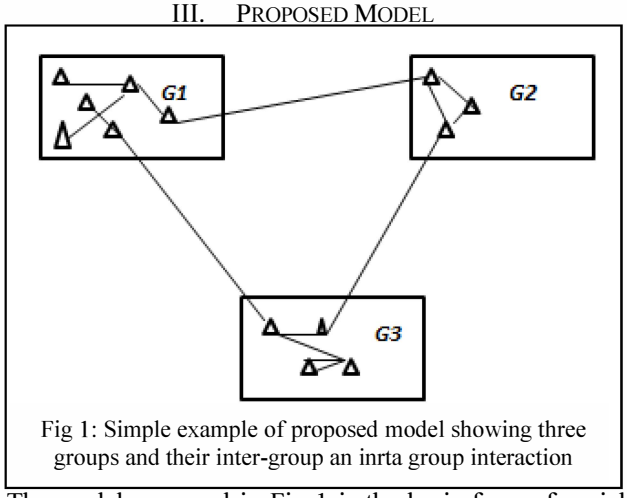

Fig 1: Simple example of proposed model showing three groups and their inter-group an inrta group interaction

The model proposed in Fig 1 is the basic form of social linkage as seen by Jacob Moreno [5] with amendments. Moreno model doesn't provide well-defined procedure so utilized MDS concept and clique approach. As defined by Luce and Perry "clique" cohesive subgroup, provides us the capability to analyze network data with respect to social groups formed on the basis of communicational traces.

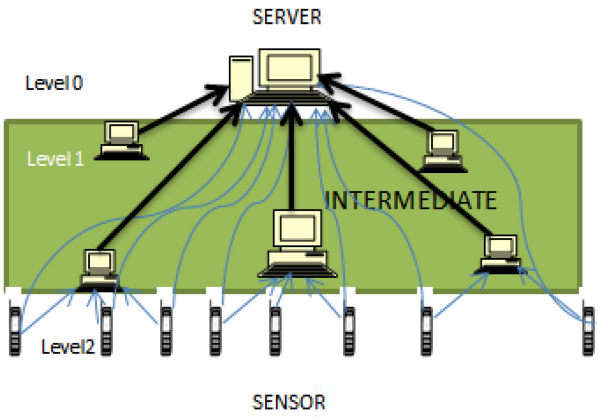

Fig 2 Multi-level architecture, Server; intermediate and sensor.

Architecture is based on multi-level; client-intermediate-server architecture. That provides sufficient solution to above mention problems. In proposed model as shown in fig.2 three level approaches is defined.

A. Level 2- Sensor level

At this level client-end sensors are implemented that provides all the communicational traces to the dedicated server. Most of the cellphones offer number of position based sensors like GPS, network location provider, Wi-Fi and Bluetooth. Among these GPS and data from network location provider can be used for locating the place and other position based sensors like Wi-Fi and Bluetooth for detecting face-to-face interaction. Motion based sensors like Gravity, Linear Acceleration, Orientation Sensors, Gyroscope, Rotation Vector and activity sensor. To gather the above information efficiently I use "Fun framework" developed in MIT Media Lab [14]. Funf provides us easy application-building system to collect mobile based communication traces. This is used at client-end and transfers data to two servers i) dedicated server ii) intermediate servers via multipart POST protocol using RemoteFileArchive interface which has HttpArchive that stores server URL .The Bluetooth based face-to-face interaction tracing is set to 20 sec so the system checks every new interaction in this time period. The data is stored in several Sqlite database files to protect them from any kind of corruption losses. The data uploading is advised to schedule at frequent time intervals. Schedule pipeline contains action like name , value/object , interval - define schedule period , duration - time of run in connection and strict - disable flexible scheduling.

B. Level 1- Intermediate level

This level comprises of intermediate servers i.e. work-pc, home-desktop, special-servers (only for data mining) which are used to back-up the logs incase dedicated server fails. Primary level filtration of data is also done at this level. This is a SOA based interface which communicates with Level2-sensors and Level0-server through secure SSL/TLS channels triggered by HTTPS. Data from sensors are maintained and archived in JSON data-interchange format, data is further stored in encrypted format under CSV files. CSV files contains data in - id, device, event and timespan headers .For further security communication between the services are done through Identity Management Protocol [16][17] using SAML- Security Assertion Markup Language .

C. Level 0- Server level

A dedicated server which analysis all the data received from beneath levels is implemented. This server filters out all the data and provides us proper result for further manual analysis. The filtration is done through two filters clique-filter which collects and filters out data based on a particular clique. Every clique-filtered data is stored separately in different CSV files. Second time-filter that chose records from certain time periods and decides which record is outdated. This is necessary as proper clique formation is a combined result of various transformations and readjustments [19][2]. Hence logged out data is to be filtered before utilization.

This is SOA based server , that use same secure SSL/TLS channels of HTTPS protocols and Identity Management Protocol for inter service communication. Here decryption key of the data is present to formulate and redefine it in different forms.

Data visualization using D3 - Data-Driven Document [18] works well on JSON format data. Different visualizations are made to provide on basis of different algorithms.

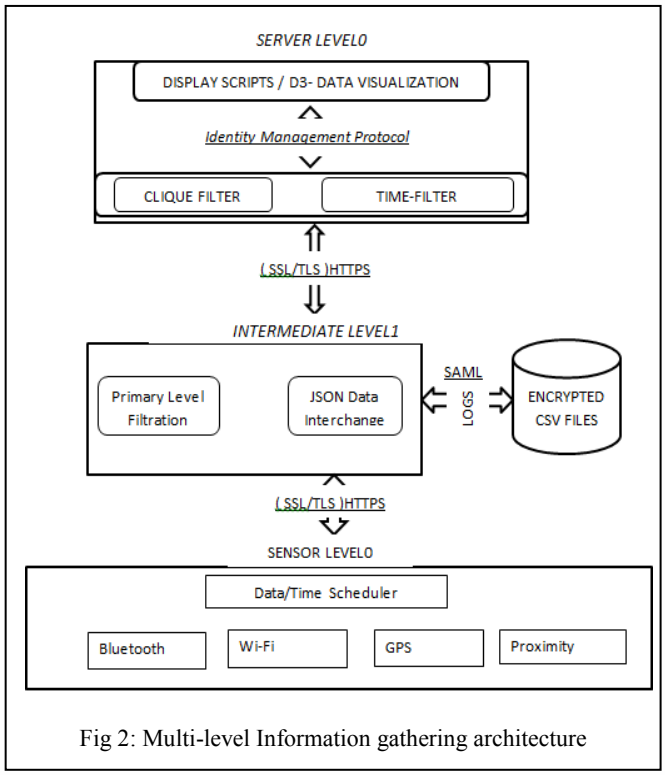

Fig 2: Multi-level Information gathering architecture

IV. APPLICATION

The graphical representation shown in Fig4 represents communicational connection of 77 people taken under consideration. These people are divided into different clique each clique member is displayed in same but isolated color dots. While the logical model holds a log record for each real connection that was registered, the visualization shows, at most, one edge for each direction between any two nodes. This graphical model also allows implementation of filters for the different type of record to be exhibited. This allows the formation of social network models, which displays different visual for different types of communication like different view for e-mail communication data or Wi-Fi or face-to-face (Bluetooth proximity) data. These way users are enabled to

quickly view the different communication patterns that have occurred on different types of communication channels.

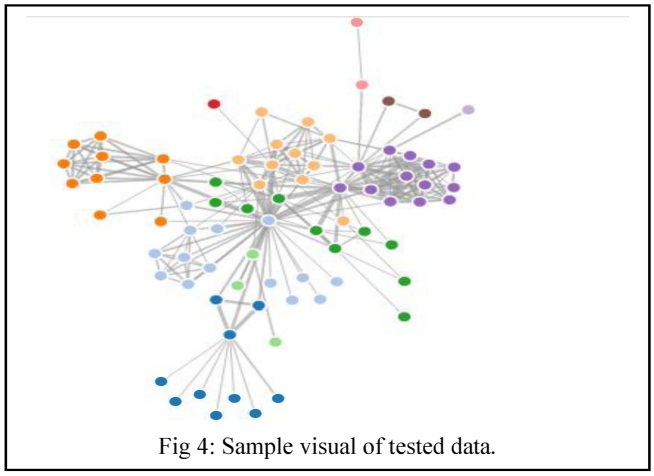

Fig 4: Sample visual of tested data.

This model is basically formed to work more efficiently with the Influence Model [21]. It shows beret visual as compared to the original MIT model.

Further it is used to provide the most influential node in the model using Evertt and Borgatti's [15] group centrality approach. This individual is the best source of information spreading in a clique.

V. WHY THIS MULTI-LEVEL ARCHITECTURE

For most of the applications of Reality Mining earlier proposed simple client-server model doesn't counts much, as an alternative this model provides a much reliable system with less faults. Multi-level approach allows the sensors to backup data at intermediate nodes.

VI. FUTURE WORK

Model can be used for other fields of data-mining and reality mining. It is best for testing and analyzing vaccinations. It provides quick result and predicts impact of vaccination on the considered clique.

It can be used in collecting behavioral stereotype data for large population. It can be used in advertisement, marketing and other commercial activities for improving standard of lifestyle.

When incorporated with universal identification like "Aadhar" can improve grass-root implementation of government policies. It keeps continuous check on how the budget is distributed? Is the distribution is correct? Is the scheme actually beneficial for the needy? gives feedback of all these questions for amendment.

As its other application its incorporation in government scheme distribution system can increasing efficiency and credibility

VII. CONCLUSION

This model uses the existing reality mining techniques with amendment to formulate a better system that eliminates the manual knowledge base technique. It provides better solution for data-loss flaw in the current sensors. This multi-level approach defines clique and most influential person in the clique. Data delivered by cell-phone networks, sensor networks, and other digital arrangements are empowering us with a third-eye view of ourselves. It provides opportunity to accurately record the activities of large numbers of people as they go about their daily lives. It would be very beneficial if we can use this new in-depth understanding of human behavior to increase the efficiency and responsiveness of industries and governments. We can achieve a world in which everything is arranged for once convenience and we would acquire things before we require them.

REFERENCES

[1] Alex Pendland and Trac Heibeck. Honest Signals, How They Shape Our World. The MIT Press, 2008.

[2] Alex Pentland, Massachusetts Institute of Technology (MIT),Reality Mining of Mobile Communications: Toward a New Deal on Data

[3] Matthias Steinbauer and Gabriele Kotsis,Building an Information System for Reality Mining , Based on Communication Traces, IEEE 2012 15th International Conference on Network-Based Information Systems.

[4] Linton C. Freeman, Visualizing Social Group, University of California, Irvine 2143 SSPA, School of Social Sciences, Irvine, CA 92697-5100.

[5] Moreno, J.L., Who Shall Survive?, Beacon, N.Y.: Beacon House Inc.

[6] Kruskal, J.B., And Wish, Multidimensional Scaling, Beverly Hills, CA: Sage.

[7] Weller, S.C., And Romney, A.K., Metric Scaling: Correspondence Analysis, Beverly Hills, CA:Sage.

[8] Luce, R.D., And Perry, "A method of matrix analysis of group structure," Psychometrika.

[9] Nathan Eagle and Alex Pentland. Reality mining: Sensing complex social systems. Personal and Ubiquitous Computing, 10(4):255–268, 2006.

[10] Alessandro Vinciarelli, Maja Pantic, Herv´e Bourlard, and Alex Pentland. Social signal processing: State-of-the-art and future perspectives of an emerging domain. In ACM Multimedia, pages 1061–1070, 2008.

[11] Wen Dong and Alex (Sandy) Pentland. Quantifying group problem solvingwith stochastic analysis. In International Conference on Multimodal Interfaces, 2010.

[12] Alex "Sandy" Pentland and Tracy Heibeck, How do social media influence people's buying decisions? Psychology Today Published on January 19, 2011

[13] Alex "Sandy" Pentland and Tracy Heibeck, Making meetings work ... and why chatty meetings are good meetings. Psychology Today Published on October 1, 2010

[14] Funf-open-sensing-framework, developed at MIT Media Lab 2011.

[15] Everett, M. G., & Borgatti, S. P. 1999. The centrality of groups and classes. Journal of Mathematical Sociology. 23(3): 181-201.Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[16] Birgit Pfitzmann and MichaelWaidner ,"Federated Identity-Management Protocols –Where User Authentication Protocols May Go–", IBM Zurich Research Lab,bpf,wmi}@zurich.ibm.com

[17] Min Huang, Yan Zhao and Lizhe Zhu , " Research for E-Commerce Platform SecurityFramework Based on SOA", 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI).

[18] Michael Bostock, Vadim Ogievetsky, Jeffrey Heer, " D3: Data-Driven Documents", IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), 2011

[19] David Coleman and Stewart Levine. Collaboration 2.0. HappyAbout. info, 2008.

[20] Everett, M. G., & Borgatti, S. P. 1999. The centrality of groups and classes. Journal of Mathematical Sociology. 23(3): 181-201.Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[21] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring Networks of Diffusion and Influence," Proc. 16th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD 10), ACM, 2010, pp. 1019-1028.