Analysis of Human Mobility Modeling Methods

Sen Wang

Florida Institute of Technology

500 West University Blvd

Melbourne, Florida, 32901

(321)704-6590

swang2011@my.fit.edu

ABSTRACT
The study on human mobility becomes more and more important since the popularity of mobile devices rises up rapidly. And the development of mobile equipment also helps scientists study on human mobility. It's very meaningful to make a report on this area. This report presents the state of some areas that relate to human mobility. I will introduce some relevant technologies and some other usability of the current study, such as the research on the civil protection, the new transfer policy in DTN. In the meantime, I will interpret some experiments in these researches.

General Terms
Algorithms, Performance, Design, Experimentation, Theory.

Keywords
Human mobility, civil protection, data Dissemination

1. INTRODUCTION
After watching and analyzing the motion of animals, it is easily to conclude that most of the time, animals are driven by food. But what do people move for? Of course, not only for food. There must be some more advanced factors in the interaction between the nature and human beings. Does any pattern or mode exist? And if they do exist, how we find them? For a long time, these problems could not be resolved without proper techniques. However, with the invention of mobile phones and other mobile-devices, where people have gone and how long people have stayed in a place can be recorded [15][16] by mobile phone towers. Then the data can be constructed into mobility networks. It's a map with many paths among the nodes, which are specific locations After analyzing mobility networks, some laws of human mobility can be revealed and models can be made by these laws, one step closer, more researches are developed based on them.

Scientists have tried to make a model to predict the spread of epidemics through the research of the human mobility pattern. In fact, under the help of Dirk Brockmann of Northwestern University in Evanston, Illinois, who conducted the project that tracked human mobility by the circulation of U.S. dollar bills [6], a prediction of the spread of H1N1 was made by the data from that project, even though the results were not perfect. In addition, the study of human mobility can also be used to benefit disabled people [1], to predict the spread of viruses among the smart phones by Bluetooth or SMS [2][4][7], and to develop the Pocket Data Switching in Delay Tolerant Network. [8][9]

2. LITERATURE REVIEW
This report focuses on the introduction of human mobility models and applications, rather than mobility computing and mobile devices. However, some essential theories and mathematic concepts relates to human mobility research will be included, such as Lévy flight [5][10] , Random Waypoint and Gauss-Markov. In this section, I will list the key papers.

In research conducted by Nils Aschenbruck, Matthias Frank, Peter Martini, Jens' Tolle [3], they analyze two real disasters which occurred in Germany in 1999 and 2001, then they present a realistic approach to realize the mobility for MANETs in disaster areas based on tactical issues of civil protection. Furthermore, they also simulate the model and get the first results, then find out some specific characteristics in the disaster area scenarios. In their research, they find the defects in the Random Waypoint and Gauss-Markov, and then they use some newer approaches to simulate the most realistic situation. I think their research is very practical and realistic, there is another paper also about this topic [12]; I will interpret in detail in the following section.

In another research conducted by Belik, V.V., Geisel, T., Brockmann, D [6]. They introduce two approaches about the spatial epidemics---diffusive dispersal and direct coupling, or effective force of infection; in the meantime, they analyze the models based on the approaches above, which are both related to human mobility, the most important thing being the bidirectional movements between individuals' houses and distant locations. Then they create a novel epidemiological model which contains the bidirectional nature of human movements. They also compare the difference among the three models and find the key point is the parameters in models. Finally, their model is supported by the different networks and provides a disease dynamics modeling framework. I think their research can be seen as a supplement for the civil protection research and it's very helpful to read their paper.

In a paper made by Matthew Stabeler, Davide Cellai, Simon Dobson and Paddy Nixon [13], they analyze the data switching techniques in Delay Tolerant Networks, including some experiments by organizations in some different locations, such as the 9-month experiment on the MIT campus, and other experiments involved with different types of mobile agents, such as, rollerblades and taxi cabs. After that, they find out the merits of the datasets from the real case scenarios, and they conclude that human mobility patterns and proximity networks could be used to design an environmental sensing system. It's very inspired to read their research, especially their idea of experiments design.

In a paper published by Hui Ye, Zhigang Chen,Zuoqun Xia and Ming Zhao [14]. They focus on the policies of the data dissemination in the Delay Tolerant Networks and analyze the efficiency of the different data transfer approaches. They test a couple of different solutions and conclude that these policies are not good enough to transfer data accurately and efficiently. After studying some other research of DTN, they try to design an ideal data dissemination policy to improve efficiency and balance in the network connectivity. In their paper, they define three properties in human mobility models, which are Destination-Matching, Arrival Time of Destination and Network Hops, respectively, and then with the Comprehensive Evaluation mechanism, they identify how to select relay nodes and analyze the data dissemination efficiency later. About their research, I will introduce more in the following sections.

In one research conducted by J.Dowling, W.Boles and A.Maeder [1], they try to use the technology of human mobility pattern to benefit the blind. They introduce the study of Artificial Human Vision and then propose a design of a programmable PDA-based AHV simulator. This application is just used to resolve the problem of the blind mobility. They start from the two main problems of blind mobility, then point out the defects of many Electronic Travel Aids, next assess the blind mobility in artificial or real world environment and define the formulas. In the following sections, they describe their design of the PDA-based AHV simulator, from the program flow to interface design. They also simulate the results of this model with a simple sample. In my view, the work is very valuable. Just like the civil protection, the researchers try to use the current technologies to benefit the disable people, and their work is very practical and meaningful.

3. ANALYSIS OF HUMAN MOBILITY MODELING

3.1 Research on civil protection

In Germany, A train hit a forgotten tool in the railway and was derailed, then crashed into the river Wupper in 1999. 47 people injured and five fatalities. In 2001, one attraction, the roller-coaster, caught fire in an amusement park near Cologne. There were 70 people injured in that accident. Some similar principles and modes exposed in the rescue of both disasters enlightened people to study how to make the rescue more efficient. In [3], the researchers modeled the two accidents and gave us a test of their model. When building the model, they define three assumptions.

Some people that took part in the mobile communication are divided into couples of groups; however, one group just shared one node.

There is no vehicle in the areas. Everybody has to walk, instead of running. On the one hand, walk is more efficient than running below a speed of 2m/s. On the other hand, it's more risky to hurt people when running more than 2 m/s.

There are not any obstacles imported or simulated. Because it's impossible to build a rescue area with lots of obstacles. The firefighter would destroy bigger obstacles, and then little obstacles could be ignored. In addition to this, the radio propagation are not considered, since radio propagation is not totally suppressed by obstacles, which would hinder a straight line communication between two nodes. Then, it's very complex to import into the model.

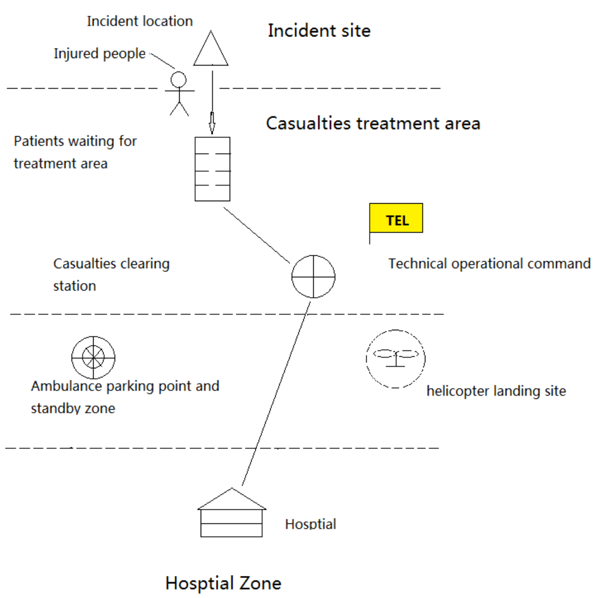

Fig.1 Incident area model

In Fig.1, there are 4 layers total, the top layer stands for the incident site where the accident happens, the layer below stands for casualties' treatment area, and the third layer stands for the ambulance and helicopter transportation area, then the layer in the bottom stands for the hospital. This is the simple simulation of the scene.

The number of units walking in the areas depends on the scenario and tactical considerations; the movement of them is modeled by the Radom Waypoint and Gauss-Markov.

The nodes in Random Waypoint model can move without any restriction. To be more accurate, the destination, direction and speed are all selected randomly and independently.

Gauss-Markov theorem states that in a linear regression model in which the errors have expectation zero and are uncorrelated. This linear also means the lowest possible mean squared error of estimate.

In the movement, some areas will be overlapping, for example, the incident site and patients waiting area, because firefighters or health care people will carry people from the former to the latter. Thus they use a direct line to model the movement between patient waiting area and incident area. At the same time, the speed of node moving on this line varies randomly between 1m/s and 2m/s.

They use 4 kinds of metrics to compare the difference among the models.

1) Relative Mobility
2) Average Node Degree
3) Average Absolute Delta of Mincut
4) Packet Delivery Fraction

In the simulation part of their own model, they choose to use NS-2 with the MANET routing protocol AODV, over an implementation of IEEE802.11b (wireless-LAN), transmission range of 50m-100m. Simulation time is 5 minutes. The load is modeled as real-time point-to-point voice traffic, length of about 40 seconds, using VOIP, with a data rate of 8 Kbps.

The metrics are:

1) Packet Delivery Fraction =

$$\frac{\text{Data packets received}}{\text{Data packet sent}}$$

2) Normalized Routing Load =

$$\frac{\text{Number of routing packets sent}}{\text{Number of routing packets received}}$$

3) Transmission Delay:

Time a packet needs for being transmitted.

And the result shows that their model works as what they expect.

3.2 Research on data transfer in DTN

In [14], researchers focus on how to improve the efficacy of data package transformation among the mobile nodes and how to reduce the consumption. In their paper, they compare four kinds of policies which used to transfer data package in Data Tolerant Network and then they confirm that the HMP, the new policy that they create, has less computation. In the process, they adapt 2 different scenarios for the data transfer, the normal movement state and the concentration movement state.

In the former state, the key point is how to select the proper nodes to transfer data hop by hop. They define a set of Z for the multiple factors, which consists of Destination-Matching (DM), Arrival Time of Destination (ATD) and Network Hops (NH). The mechanism runs like this: sending node N1 looks for some nodes to send data items. If there are some available nodes, it will compare the ATD, then select the one which has a minimum value; if some nodes have similar values, they confine the number of relay nodes when the NH reaches a certain threshold. The NH is used to balance the network cost and network congestion. If there is no one available, it will keep looking for nodes, till find someone; if it exceeds the time threshold. N1 just randomly select a node and send the data item to it. Then, it would follow steps node N1 has done above.

In the concentration movement state, they define a circle Cr with radius R and the data destination is in its center. The nodes in this circle have close relationship in this area and they consider this state as a "small world model". They choose to use flood policy in this state in which R is equal to 100 meters.

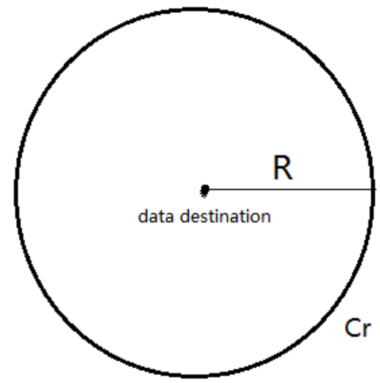

Fig.2 Concentration movement model

In the system model, they define 500 nodes locates randomly with moving speed of 1.4m/s in a rectangular area. The nodes also divided into 2 groups. One group follows mobility pattern, the other one is not.

They consider 2 situations in the simulation of HMP and the other 3 policies: 1) Increase from 1.4m/s to 8m/s, the end to end delay drops in all algorithms, when the speed increase beyond 8m/s, since the Bluetooth can't build the connection promptly, the data delay started to go up. However, in the network cost, the cost of HMP is 12% less than that of flooding policy in the duration of 50s, as well as 40% less in the range of 300-500s.

After study their research, I found the set of factors which impact the data transformation is incomplete, if they can consider more realistic factors in the set, then their simulation would be more persuasive.

3.3 Research of Habit

Habit[11] is an efficient multi-layered approach, which is used in content distribute network, based on the human connectivity network and the interest network, like source-based routing, the sending node would pre-computes the whole paths and the consumption from it to the known interested recipients. Although there are numbers of uninterested intermediaries, the sending node would choose the minimized way of consumption according to its pre-computed result.

When building a Content Dissemination Network, node A need to record the nodes that it regularly meets. If A encounters B frequently enough, B is called a familiar stranger. In addition, A will record the time that it encounters with B, and maintains the regularity information about a maximum number of familiar strangers only, and then A generates a regularity graph. With this mechanism, a Content Dissemination Network consists of many nodes that are similar to A.

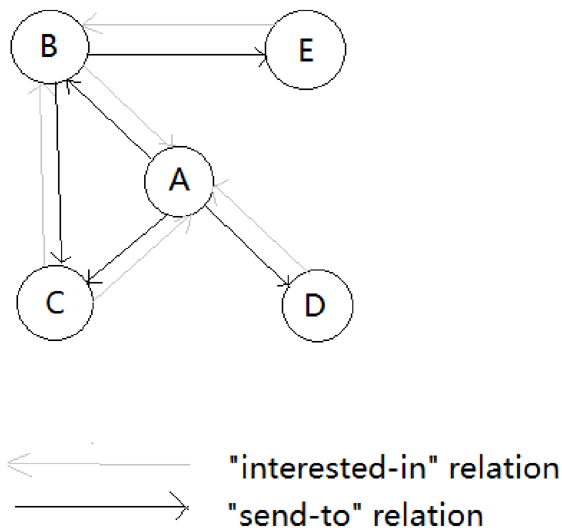

Fig.3 Publishing node A sends messages to neighbor nodes

For example, the publishing node A sends message to node B, C, D frequently, then node B, C, D become the familiar stranger to A, in addition, node B, C, D are interested in A.

Assume node A publishes a message to other nodes, there are 4 steps total:

1) Determine recipients: A will collect information that how many nodes are interested in A.
2) Find cheapest paths: A will compute that how many non-interested nodes have to be through in the path when A sends message to other nodes, and it keeps paths with the interested nodes with minimal non-interested nodes, then abandons others. While exploring the regularity graph to find routes, a delivery probability is also computed for each path, which is the addition of the whole non-zero regularity weight between pairs.

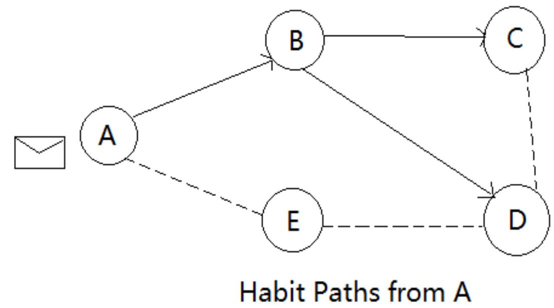

Fig.4 Habit paths from node A to nodes B, C, D

For example, in Fig.4, node A want to send message to D, then there are 3 paths available, 1) A-B-C-D, 2) A-E-D, 3) A-B-D. In these 3 paths, the second one contains one non-interested node E, so this path will be abandoned.

3) Select paths: After step 2, the path with the highest delivery probability will be selected.
4) Optimization: Because the publishing node may not always be aware of all nodes interested in itself, so intermediaries will check if there are nodes which are interested in the publishing node that are not in the regularity path. If there are some nodes like this, the steps above will be run again.

The researchers use OMNeT++ open source discrete event simulator and they use two metrics to measure the effectiveness and efficiency, which are Precision and Recall.

$$\text{Precision} = \frac{\{Relevant\ Msgs\} \cap \{Received\ Msgs\}}{\{Received\ Msgs\}}$$

$$\text{Recall} = \frac{\{Relevant\ Msgs\} \cap \{Received\ Msgs\}}{\{Total\ Relevant\ Msgs\}}$$

In addition, the researchers use other protocols as comparison, which are Epidemic and WaitForDestination. For Epidemic, the publishing node just send message to all nodes regardless of their interest. So it achieves the best performance in effectiveness. For WaitForDestination, the publishing node waits for the requests of other nodes, then responses them and send them the message, so it achieves the best efficiency in experimental.

The results shows that epidemic woks best in Recall, while WaitForDestination works worst, however, when TTL is longer than 7 days, the performance of Habit is nearly equal to Epidemic. For the precision, Habit performs about 70% of WaitForDestination, which has the 100% performance, while Epidemic has the worst performance.

3.4 Research on human mobility patterns

In [17], the researchers collect some GPS trace datasets, and then they define two conceptions to analyze these data, one is geo-location: where users spend a lot of time in a specific location; the other one is geo-community that a place shared by a kind of people, not necessarily at the same time. After they extract from the GPS data, they get two results, one is stay-location: where they spend a piece of time and there are two kinds of stay-locations, one is transit-location, which is not the main-destination that the people pause a little time when they move to the main-destination; the other one is the geo-location, also called main-destination, which is the end of people's movement and in which people stay for a long time, at the same time, a geo-community consists of these people, even if there are no social relationship among them. The second one consists of several probability distributions of quantities involved in the mobility model.

Name	Distribution	Note
Pause times in geo-locations	Tapered Pareto	
Intra-distances	Pareto	Inside a geo-location
Inter-distances	Lognormal distribution	Between geo-locations
Next geo-location choice	Upper-truncated Pareto distribution	

Table.1 Human mobility patterns extracting from GPS traces

The next geo-location choice is used to explain how distance and popularity influence people's choices, and the result show people prefer near geo-locations, don't care about popularity too much.

The researchers separate two phase in human movement, when they extract locations from real GPS traces. A static phase when a person doesn't move and stay in a place for a period; a dynamic phase that he moves towards some locations.

Because the static phase implies people 's interest if a specific site, it's more important than dynamic phase, and then, to make the results more accurate, the researchers use $\|pi\text{-}pi{+}1\| \geq \Delta$, $\Delta = 20$m(the speed of human walk is about 1.1-1.4m/s, time threshold is 30s, so$\Delta = 20$m), to screen the results that people are in the movements, if two points doesn't satisfy $\|pi\text{-}pi{+}1\| \geq \Delta$, pi+1 is abandoned.

Then the stay-location can be defined as a set of GPS coordinates sl={pm,pm+1,…,pn} such that $\forall m < i \leq n$, $\|pm\text{-}pi\| \leq D$ for a fixed D, the geo-location can be defined as a particular stay-location that an individual pause-time exceeds 5 minutes, and a stay-location is either a geo-location or a transit-location.

Then they respectively extract the stay-locations, geo-locations, geo-communities from the datasets.

After get the kinds of results, they consider distance distribution in stay-location and between geo-locations, pause-time distribution in geo-location and how many people choose successive geo-location. They use Maximum Likelihood Estimation (MLE) to fit some common distributions.

A. Intra-distance inside geo-locations, they perform MLE over x-axis range between 10m, then they found the figure match perfectly with Pareto distribution.
B. Inter-distance between geo-locations.When use the results from small datasets (elements<170), they found the distribution is prone to tapered Pareto or exponential, however, when they use a considerable dataset, the distribution tends to be lognormal distribution. In fact, the latter is more rational, because it can be interpreted that people prefer short paths between geo-locations, but take long jumps more frequently than expected.
C. Pause-time distributionSince they eliminate transit-locations when extract the pause-time result, the distribution has a little difference with two

datasets. But the best match distribution is tapered Pareto.

D. Next geo-location distributionThe best match distribution is the upper-truncated Pareto, which implies that people often move towards near geo-locations but sometimes they choose to go towards distant geo-locations.

4. ACKNOWLEDGEMENTS

Thanks to Florida Institute of Technology and Computer science department for providing seminar. Also, thank to Dr.Menezes for giving me the useful suggestions and helping me for correct this report.

5. REFERENCE

[1] Dowling, J., Boles, W., Maeder, A. 2005. Mobility assessment using simulated Artificial Human Vision. Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference. (2005), 32.

[2] Kalpaxis,A.2007.Wireless.Temporal-Spatial Human Mobility Analysis Using Real-Time Three Dimensional Acceleration Data. Computing in the Global Information Technology, 2007. ICCGI 2007. International Multi-Conference. (2007). PhoenixNGT, Inc, New York,NY.

[3] Aschenbruck, N., Frank, M., Martini, P., Tolle, J. 2004. Human Mobility in MANET Disaster Area Simulation - A Realistic Approach. Local Computer Networks, 2004. 29th Annual IEEE International Conference. (2004), 668 – 675. University of Bonn, Germany.

[4] Chaintreau Augustin, Hui Pan, Crowcroft Jon, Diot Christophe, Gass Richard, Scott James. Impact of Human Mobility on Opportunistic Forwarding Algorithms. Mobile Computing, IEEE Transactions. (Mar, 2007), Vol 6, Iss 6,606-620.

[5] Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, Song Chong. 2008. On the Levy-Walk Nature of Human Mobility. INFOCOM 2008. The 27th Conference on Computer Communications, IEEE. (2008), 924-932.

[6] Brockmann, D., Theis, F. 2008. Money Circulation, Trackable Items, and the Emergence of Universal Human Mobility Patterns. Pervasive Computing, IEEE. (2008), Vol 7, Iss 4, 28-35.

[7] Pan Hui, Crowcroft, J. 2008. Predictability of human mobility and its impact on forwarding. Communications and Networking in China, 2008, ChinaCom 2008, Third International Conference. (2008), 543-547. University of Cambridge.

[8] Ming Zhao, Mason, L., Wenye Wang.2008 Empirical study on human mobility for mobile wireless networks. Military Communications Conference, 2008, MILCOM 2008. IEEE. (2008), 1-7. North Carolina State University.

[9] Azevedo, T.S., Bezerra, R.L., Campos, C.A.V., de Moraes, L.F.M. 2009 An Analysis of Human Mobility Using Real Traces. Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE. (2009), 1-6. Federal University of Rio de Janeiro.

[10] Kyunghan Lee, Seongik Hong, Seong Joon Kim, Injong Rhee, Song Chong. 2009 SLAW: A New Mobility Model for Human Walks. INFOCOM 2009, IEEE. (2009) 855-863.

[11] Mashhadi, A.J., Ben Mokhtar, S., Capra, L. 2009 Habit: Leveraging human mobility and social network for efficient content dissemination in Delay Tolerant Networks. World of Wireless, Mobile and Multimedia Networks & Workshops, 2009. WoWMoM 2009. IEEE International Symposium. (2009), 1-6.

[12] Belik, V.V., Geisel, T., Brockmann, D. 2009 The Impact of Human Mobility on Spatial Disease Dynamics. Computational Science and Engineering, 2009. CSE '09. International Conference. (2009), 932-935

[13] Stabeler, M., Cellai, D., Dobson, S., Nixon, P. 2009 Delay tolerant networks and spatially detailed human mobility. Ultra Modern Telecommunications & Workshops, 2009. ICUMT '09. International Conference. (2009), 1-2.

[14] Hui Ye, ZhiGang Chen, ZuoQun Xia, Ming Zhao. 2010 A Data Dissemination Policy by Using Human Mobility Patterns for Delay-Tolerant Networks. Communications and Mobile Computing (CMC), 2010 International Conference. (2010), Vol 1, 432-436.

[15] Shusen Yang, Xinyu Yang, Chao Zhang, Spyrou, E. 2010

Using social network theory for modeling human mobility. Network, IEEE. (2010) 6-13.

[16] Calabrese, F. Di Lorenzo, G. Ratti, C. Human mobility prediction based on individual and collective geographical preferences. Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference. (2010) 312-317.

[17] Zignani, M., Gaito, S. Extracting human mobility patterns from GPS-based traces. Wireless Days (WD), 2010 IFIP. (2010), 1-5.