Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2002 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2002

E-MARKET INFRASTRUCTURE PLANNING AND INTERNET GROWTH

Subhashish Samaddar
Georgia State University

Satish Nargundkar
Georgia State University

Samir Chatterjee
Claremont Graduate University

E-Market Infrastructure Planning and Internet Growth

Subhashish Samaddar
J. Mack Robinson College of Business
Management Department
Georgia State University
s-samaddar@gsu.edu

Satish Nargundkar
J. Mack Robinson College of Business
Management Department
Georgia State University
s-samaddar@gsu.edu

Samir Chatterjee
School of Information Science
Claremont Graduate University
samir.chatterjee@cgu.edu

Abstract

Prior research studies have shown that the Internet was growing at an exponential pace during its early stage of growth with no predictable upper bound or saturation limit. In this paper we use popular network growth prediction models to track the Internet diffusion. The data that was analyzed were obtained from the Internet Software Consortium that uses a variety of sophisticated techniques to learn about autonomous systems, domain names and host count on the Net. We report that the exponential model is not an accurate model for anticipating the growth of the Internet at present. We also show that a finite saturation limit of the Internet host count worldwide appears now to be in sight. Furthermore we provide critical insights into inflection point on Internet host growth and discuss the extraneous factors that could lead to a more optimistic growth count. We comment on what the findings here mean for planners of e-market systems.

Keywords: Internet growth, host count, exponential models, Gompertz model, logistic model, IPV6

Introduction

During the last decade there has been a lot of optimism regarding the Internet. In a 1994 article, Goodman et al (1994) observed, "If the Internet were a stock it would be considered a market phenomenon, with sustained double-digit growth and no apparent end in sight to the upward spiral." Indeed, studies have shown that the Internet was growing at an exponential pace during its early stage of growth with no predictable upper bound or saturation limit. Consequently it was concluded that diffusion growth models based on the concept of external influence were explaining the diffusion of the Internet. Put differently, diffusion growth models based on the contagion phenomenon, or otherwise known as internal influence models, were not necessarily explaining and capturing the diffusion of the Internet (Rai, Ravichandran, Samaddar 1998). In this paper we use popular theoretical network growth prediction models to test and track the Internet diffusion. We report that the exponential model or external influence is not necessarily an accurate model for anticipating the growth of the Internet at present. We also show that a finite saturation limit of the Internet host count worldwide appears now to be in sight. Based, on our findings in this paper, we would like to put forward the notion that the Internet diffusion is more propelled by internal influence factors at present as opposed to the external factors as found in the earlier studies mentioned above. Furthermore we provide critical insights into inflection point on Internet host growth and discuss the extraneous factors that could lead to a more optimistic growth count.

The global diffusion of the Internet is of interest to many including infrastructure planners and policy makers (Press, 1997). It is also of critical importance to various other interest groups or constituencies such as the Internet Engineering Task Force (IETF) that overseas the technical research and development of the Internet. Other constituencies range from Internet infrastructure and service providers, to e-commerce and other commercial users of the Internet, to technological innovators. One important interest of these groups is in obtaining a working knowledge of the Internet growth process (Press et al 1998). For example, explaining the growth process by virtue of appropriate modeling is critical for policy formulation, capacity planning, introducing new networks of hardware and software, and planning and deploying IP based telecommunications. All e-commerce and other business planners and marketing firms can also benefit by orienting their strategic plans to accommodate and exploit the knowledge of the Internet diffusion process.

However important, tracking and modeling the Internet's global diffusion is at best a daunting task today. It is well known that the Internet is growing rapidly, but measuring that growth with a degree of precision is difficult (Press, 1997). What started as a 4-link network in 1969 now has a 109 million plus hosts worldwide–as advertised in the DNS directory as of January 2001. Understanding Internet growth involves assessing alternative models for the growth process. Internal influence models are one class of popular models that can help in such assessments.

Internal Influence Models

These models capture the diffusion of an innovation as a rate of adoption over time and express the rate in form of some suitable S-shaped curve. These models assume variable growth rate where the growth rate first increases and then, after an inflection point, decreases over time to reach finite subscriber saturation level. The diffusion rate is defined as the speed at which members of a social system adopt the innovation. Two key assumptions here are: (1) that the existing number of adopters positively drives the rate of growth and (2) the difference between the potential number of adopters at the saturation level and the number of existing adopters also influences the rate of growth. Two basic diffusion theories are used to explain the logic behind these models. The first, *diffusion of innovation theory*, studies the diffusion as a process by which an innovation is communicated internally over time among social members. The distribution of adopters is expected to be a bell-shaped curve. The total number of adopters overtime is expected to follow an S-shaped curve. That is, the total numbers of adopters is expected to grow slowly at first due to uncertainty about the innovation in the early phase. This results in a relatively flat curve in the beginning of the process. If the innovation succeeds, positive feedback fuels the innovations process and the adoption rate catches momentum and increases rapidly causing a steep curve which levels off later due to saturation. The second, *utility of network theory*, suggests that potential return from adopting a network depends on the number of existing users. This dependence is especially strong for computer and telephone networks where the value of a network increases, as does its number of users. This is also sometime referred to as *network externality*. Once a critical mass of membership is reached it motivates further adoption of the innovation. The total numbers of adopters plotted over time is expected to be an S-shaped curve with the take-off point representing the critical mass.

Internal influence models and consequent S-curves have been used in many disparate fields such as management, sociology, marketing, communication networks and medicine to model diffusion throughout a population of adopters or subscribers (Gurbaxani 1990, Rai, Ravichandra, Samaddar 1998).

Popular Forms of S-curve

Two fundamental forms of S-curves – Gompertz and Logistic – have been used widely in network growth studies. Both allow for growth rate that changes over time and with an eventual slowing down to a finite or bounded saturation level. An approximation of the S-curve can also be achieved by an Exponential curve (Gurbaxani 1990). Exponential curves assume a constant ratio of growth rate that generally characterizes the early stage of an innovation. Such curves assume an unlimited saturation point, which may be representative of the early stage of a highly popular innovation but will, in time, turn out to be an over aggressive estimation of the future growth as the innovation matures to a finite saturation level. Indeed, based on Internet host count data until January 1994, an exponential model, with unlimited saturation level, outperformed both Gompertz and Logistic models in characterizing the Internet growth (Rai, Ravichandran, Samaddar 1998). However authors of the study then anticipated that "It is likely that due to the absence of an upperbound, the Exponential model will eventually overestimate the growth of the Internet." Our latest analysis shows that that eventuality has started to occur based on the Internet host data available since then.

Assumptions: The functional forms of the three models are presented here. For all the models below, Y_T and y_T are the cumulative number of existing adopters (of a given innovation) and the number of adopters joining at the period T, respectively. K, A and B are constants.
Gompertz Model: In this model the rate of diffusion is a function of existing adopters and the difference between the logarithms of the number of adopters at the saturation level and the existing number of adopters. The mathematical form below expresses this relation: The rate of diffusion, dy/dt = $f\{y^*(\log y_{saturation} - \log y_{existing})\}$. This relation leads to the following integral form: $Y_T = KA^{B^T}$ For 0<A<1 and 0<B<1, Y_T is an increasing S-curve which reaches the upper bound or the saturation point of K (total number of adopters of the innovation) as time T approaches its theoretical limit of infinity. This curve reaches its inflection point (i.e., the point in the S-curve where the diffusion growth reaches its maximum rate and then switches from an increasing rate to a decreasing one) at Y_T=K/e where e is the Euler's constant of 2.7027. That is the inflection point is when Y_T reaches 37% of its saturation level.
Logistic Model: This model has a similar structure as above except it does not use the logarithmic form of the number of adopters. Thus the rate of diffusion is expressed as: The rate of diffusion, dy/dt = $f\{y^*(y_{saturation} - y_{existing})\}$. This relation leads to the following integral form: $Y_t = 1/(K + AB^T)$ For A>0 and 0<B<1, Y_T is an increasing S-curve which reaches the upper bound or the saturation point of 1/K as time T approaches its theoretical limit of infinity. This curve reaches its inflection point at Y_T=K/2. That is the inflection point is when Y_T reaches 50% of its saturation level.
Exponential Model Unlike the above two, this model is characterized by a constant ratio of growth and takes the integral form of $Y_t = A + e^{(BT)}$ For B>0, Y_T is an ever increasing growth function that reaches infinity as T approaches its theoretical limit of infinity.

Sidebar 1. Mathematical Formulae for the Models Used in the Analysis

The Data

We collected the total number of Internet hosts worldwide from August 1981-March 2001 (see Table 1), from the reports published by Internet Software Consortium at www.isc.org (Internet Software Consortium, WWW reference).

We define a "host" to mean a machine that is assigned an IP address. This would imply user desktops, servers and routers. Estimating the exact number of hosts on the Internet is a daunting task. However, the data set obtained from ISC uses a very sophisticated technique using DNS registrations and pings. Ping is a popular program that is commonly used to communicate with a machine to find out if it is alive on the net. It is also important to note that the current version of the Internet Protocol (version 4) uses a 32 bit for its addresses. Hence the maximum number of hosts that could ever exist using the present version is 2^{32} = 4.3 billion. The actual number is however slightly less since the IP addresses are classified into three primary classes with certain bits representing network ID and other representing the host ID.

Table 1. Internet Host Growth Data by Quarter (August 1981-March 2001)
Source: Internet Software Consortium at www.isc.org

Period	Quarter	Number Of Hosts		Period	Quarter	Number Of Hosts
1	Q4 81	218		**41**	Q4	617,000
2	Q1 82	225		**42**	Q1 92	727,000
3	Q2	233		**43**	Q2	890,000
4	Q3	279		**44**	Q3	992,000
5	Q4	344		**45**	Q4	1,136,000
6	Q1 83	409		**46**	Q1 93	1,313,000
7	Q2	475		**47**	Q2	1,486,000
8	Q3	540		**48**	Q3	1,776,000
9	Q4	628		**49**	Q4	2,056,000
10	Q1 84	727		**50**	Q1 94	2,217,000
11	Q2	826		**51**	Q2	2,757,948
12	Q3	925		**52**	Q3	3,212,000
13	Q4	1,024		**53**	Q4	3,864,000
14	Q1 85	1,258		**54**	Q1 95	4,852,000
15	Q2	1,493		**55**	Q2	5,747,000
16	Q3	1,727		**56**	Q3	6,642,000
17	Q4	1,961		**57**	Q4	8,057,000
18	Q1 86	2,221		**58**	Q1 96	9,472,000
19	Q2	2,926		**59**	Q2	11,176,500
20	Q3	3,853		**60**	Q3	12,881,000
21	Q4	4,780		**61**	Q4	14,513,500
22	Q1 87	8,641		**62**	Q1 97	16,146,000
23	Q2	13,968		**63**	Q2	17,843,000
24	Q3	19,295		**64**	Q3	19,540,000
25	Q4	24,622		**65**	Q4	24,605,000
26	Q1 88	28,863		**66**	Q1 98	29,670,000
27	Q2	30,932		**67**	Q2	33,204,500
28	Q3	33,000		**68**	Q3	36,739,000
29	Q4	56,000		**69**	Q4	39,984,500
30	Q1 89	80,000		**70**	Q1 99	43,230,000
31	Q2	105,000		**71**	Q2	49,724,000
32	Q3	130,000		**72**	Q3	56,218,000
33	Q4	159,000		**73**	Q4	64,308,046
34	Q1 90	197,500		**74**	Q1 00	72,398,092
35	Q2	236,000		**75**	Q2	82,722,939
36	Q3	274,500		**76**	Q3	93,047,785
37	Q4	313,000		**77**	Q4	101,311,107
38	Q1 91	376,000		**78**	Q1 01	109,574,429
39	Q2	455,500				
40	Q3	535,000				

The Results

Which model best characterizes the Internet growth now? We fitted non-linear regression to estimate parameters for each of the three models by using the data from August 1981-December 1999. Further to the each model's statistical fit, their predictive validity should be considered before selecting the best model. Mead has observed: "The ability of a growth curve to forecast future development is a crucial requirement, and thus it is desirable if this ability can be evaluated on some available data" (Meade 1984). We tested the predictive ability of each of the three models on actual growth data from January 1999- January 2001 which was not used for model estimation.[1]

The Gompertz model explains 99.77% of the variability in the data. (see Table 2 and Figure 1). The model predicts an approximate saturation level of 2.6 billion Internet hosts (as of January 2001, the Internet has reached 0.11 billion hosts). The model suggests an inflection point in March 2012. This indicates that the rate of Internet growth is increasing at present and will start to decrease sometime during the beginning of 2012.

The Logistic model explains 99.81% of the variability in the data (see Table 2 and Figure 2). The model predicts an approximate saturation level of 0.1 billion hosts which is less than the current network size of 0.11 billion hosts. The model suggests an inflection point in the middle of 1999. That is the model suggests that Internet has been growing at a decreasing rate since the middle of 1999. The Logistic model seems to be underestimating the Internet growth. The Exponential model can explain only 88.89% of the variability. (see Table 2 and Figure 3).

Table 2. Estimated Model Results

Model	**Parameter Estimates (Using data over 70 quarters)**	**R-Squared**	**Saturation Limit (SL) in number of hosts**	**Inflection Point**
Gompertz $Y_t = KA^{B^T}$	K = 2,599,716,414.69 A = 0.0000000000014 B = 0.9731512677871	0.9977	2,599,716,414	March 2012 or 961,895,173 hosts (=0.37*SL)
Logistic $Y_t = 1/(K + AB^T)$	K = 0.0000000099 A = 0.0022200255 B = 0.841610354	0.9981	101,010,101	June 1999 or 50,505,050 hosts (=0.5*SL)
Exponential $Y_t = A + e^{(BT)}$	A = 1,741,272.6181 B = 0.2542314948	0.8889	No Limit	None

[1]Out of the 78 quarters of host count data we had, we used the first 70 for estimation and the later 8 quarters data were used as 'hold-out' data to check predictive ability of the models.

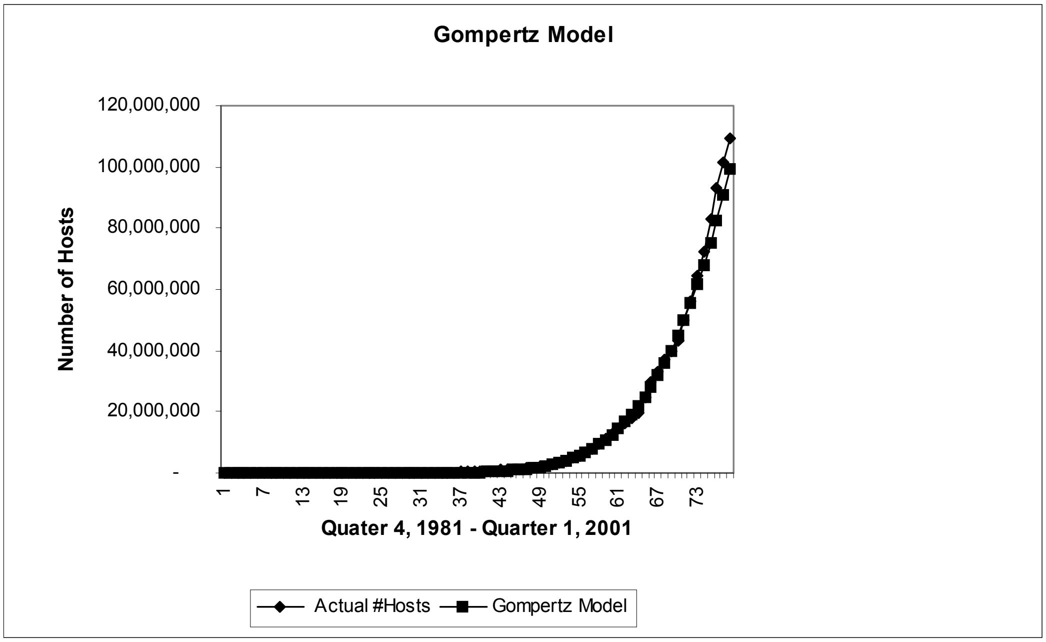

Figure 1. Gompertz Model

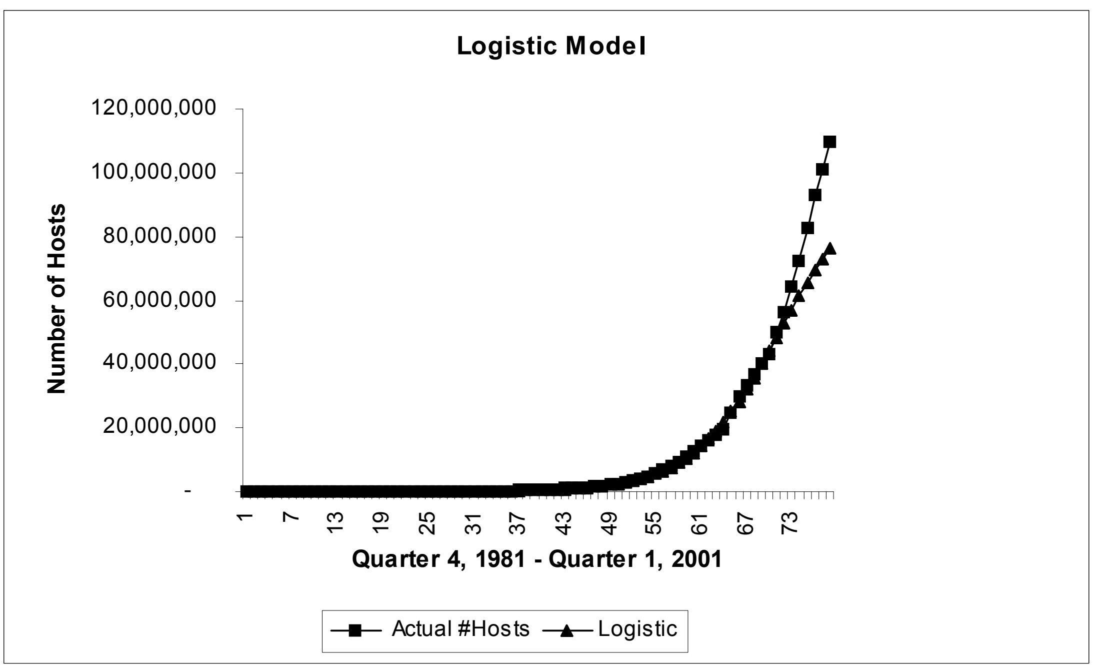

Figure 2. Logistic Model

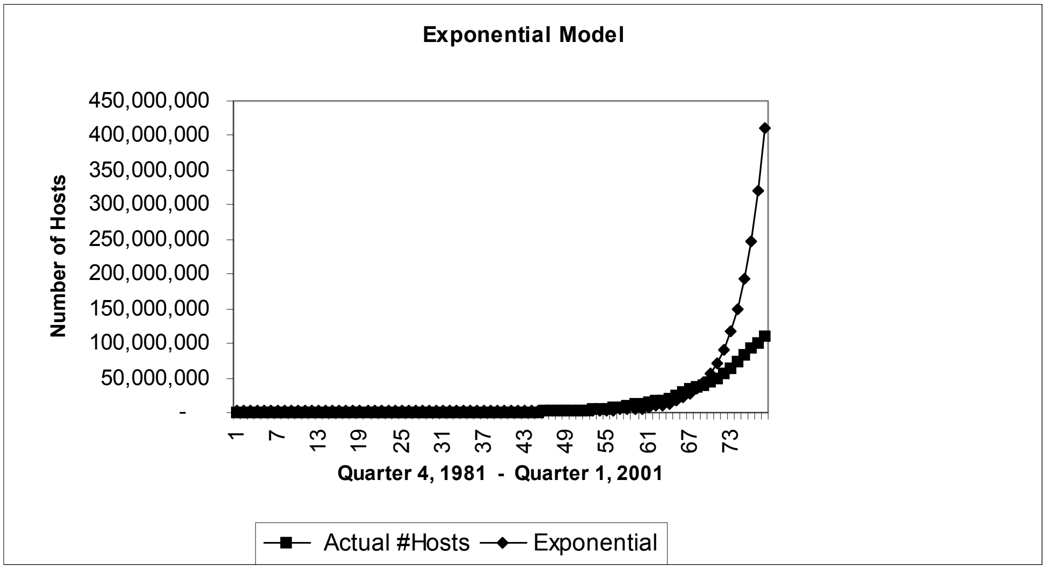

Figure 3. Exponential Model

We projected the growth of the Internet for a decade – January 1996 to January 2005, i.e., from the middle of the last decade to the middle of this decade – with each model (see Figure 4). In order to check predictive ability of all three models, we compared their predictions with the actual Internet growth from January 1999 to January 2001. The Gompertz model performs better than the Logistic and the Exponential. The Exponential model substantially over predicts future growth and may not be a suitable predictor of the Internet diffusion any more. Consequently we rejected the Exponential model. The Logistic model under predicts the growth of the Internet. Increases in the actual Internet host growth suggest that the model's inflection point in the middle of 1999 and the model's prediction of a saturation point of 0.1 billion hosts are unrealistic. The Gompertz model provides the closest fit to the data but seems to underpredict growth up to January 2001 very slightly, and these deviations from actual growth are much less than the deviations of the other two models.

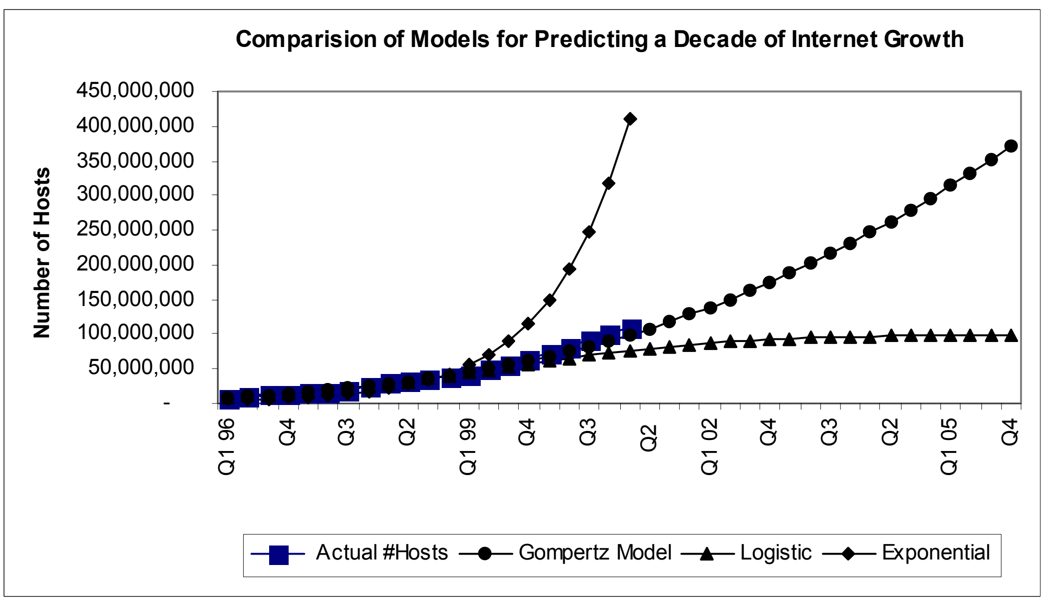

Figure 4. Comparison of Models

That the Gompertz model - a bounded S-curve – best characterizes the current Internet growth is a new and interesting finding for three reasons. First, it is different from the earlier findings that the Internet was growing at an exponential pace (Rai, Ravichandran, Samaddar 1998). Second, earlier during mid nineties, both anecdotal expectations (Goodman et al. 1994) and empirical studies (Rai, Ravichandran, Samaddar 1998) suggested an unlimited saturation point for the growth of Internet. Our present analysis seems to reject, for the first time, the notion of unlimited final size of the Internet. In stead, it suggests that the final size of the Internet will be somewhere around 2.6 billion hosts. Finally, earlier reports did not support any inflection point or slowing down in the rate of growth of Internet hosts. The model supported in our current analysis suggests that although the Internet is currently growing at an increasing rate, it will switch to grow at a decreasing rate starting around the first quarter of 2012. These results are compared in Table 3.

Table 3. Comparison of Results

	Earlier Results (1998) Source: Rai, Ravichandran, Samaddar 1998.	**New Results (2001)**
Best Fitting Model	*Exponential*	*Gompertz*
Estimated Saturation Limit on the number of Internet Hosts	*No Limit*	*2,599,716,414*
Estimated Inflection Point in Time	*No Inflection Point*	*Quarter 1, 2012*
Estimated Inflection Point expressed as the number of Internet hosts	*Not applicable*	*961,895,173 hosts*

Discussion and Conclusion

The internal influence models, in general, have been popular in studying systemic diffusion process; it is somewhat unclear whether these models completely represent the Internet diffusion process. There seems to be at least two limitations due to the models' assumptions (Rai, Ravichandran, Samaddar 1998). First, an underlying assumption in internal influence models is that the composition of the social system is considered unchanging, i.e., member homogeneity is assumed over time. The demographics of potential adopters, however, are changing in terms of age, race, gender, and economic status. Second, external factors are considered irrelevant by internal influence models. However, that may not completely be the case.

From an IETF perspective, there is enough optimism already that the growth will continue and there will be need for more IP addresses. This is evident from the recent standardization effort of IPV6 protocol that uses a 128-bit IP address space, which is significantly a much bigger space to choose from. What current trends support such an explosion of hosts? Two important technological breakthroughs namely, mobility and nano-systems; can make this happen. As miniaturization matures, small and powerful systems will be embedded around us (pervasive computing (Ark and Selker 1999)) and these systems will all talk to each other using TCP/IP. Secondly, to support mobility, we are already seeing hand-held devices (cell phones, pocket PCs, PDAs, wearable computers) flourishing at a rapid pace that all require IP addresses to be connected to the Internet. It is further believed that as the cost of such small yet powerful systems come down we may see a further explosion in the usage and implementation.

Various authors have studied the Internet growth phenomenon from the vantage point of the influence caused by factors external to the system of the adopting population (Goodman et al. 1994). Examples of external factors can include new breakthroughs in the Internet technology, governmental initiatives directed at promoting/demoting the growth, new and novel developments, such as web hosting, from outsourcing vendors fostering technical ease for the users to adopt the innovation, and so on. Unlike internal influence models, studies of external factors help explain the growth phenomenon as a consequence of some important external events. These models are informed by analyzing real external events after those events have occurred and are very useful in gaining insights to such events and their expected contribution to the growth of the Internet. External influence models can offer rich qualitative description of the factors and resulting patterns of growth. Such models can also be useful in identifying segments of growth that are not adequately explainable by the help of an internal influence model. For example, factors such as networking investments (Press et al 1998), governmental sponsorships, availability of access tools and browsers (Rai, Ravichandran, Samaddar 1998), the developmental initiatives in building national backbones, grass root nets in countries where IT sophistications are limited to poor telephone lines, the presence of commercial carriers and resellers can influence the patterns of global diffusion of the Internet (Goodman et al. 1994). Consequently, external influence models can possibly supply additional insights to the external *causes* behind the growth of the Internet. The results presented here have hopefully shed more light into the incredible Internet growth phenomenon.

References

Ark, W. S. and Selker, T., "A look at human interaction with pervasive computers" *IBM Systems Journal*, 38:4, pp. 504-507, 1999.

Coffman, K. G. and A. M. Odlyzko "Growth of the Internet" in Optical Fiber Telecommunications IV, I. P. Kaminow and T. Li, eds. Academic Press, 2001.

Coffman, K. G. and A. M. Odlyzko "Internet growth: Is there a "Moore's Law" for data traffic?" in Handbook of Massive Data Sets, J. Abello, P. M. Pardalos, and M. G. C. Resende, eds., Kluwer, 2001. To appear.

Goodman, S.E., Press, L.I., Ruth, S.R. and Rutkowski, A.M., "The global diffusion of the Internet patterns and problems" *Communications of the ACM* 37, 8 (Aug. 1994), 27-31.

Gurbaxani, V. "Diffusion of computing networks: The case of Bitnet" C*ommunications of the ACM* 33, 12 (Dec. 1990), 65-75.

Internet Software Consortium, www.isc.org.

Meade, N., "The use of growth curves in forecasting market development: A Review and Appraisal" *J. of Forecasting* 3, 4 (Oct.-Dec. 1984), 429-451.

Odlyzko, A. M. "Internet growth: Myth and reality, use and abuse" *iMP: Information Impacts Magazine*, November 2000.

Press, L., Burkhart, G., Foster, W., Goodman, S., Wolcott, P. and Woodard, J., "An Internet diffusion framework" *Communications of the ACM* 41:1010, pp. 21-26 (Oct. 1998).

Press, L., "Tracking the global diffusion on the Internet" *Communications of the ACM* 40:11, pp. 11-17, (Nov. 1997).

Rai, A., Ravichandran, T. and Samaddar, S., "How to anticipate the Internet's global diffusion" *Communications of the ACM* 41:10 pp. 97-106 (Oct. 1998).