Regions of Time and Internet: Modelling: An Application of the Space-time Trip (RASTT) model to the USA Internet Market

Robert G. V. Baker

School of Human and Environmental Studies
University of New England, Armidale 2351, Australia
Fax: 61-2-67733030
Email: rbaker1@metz.une.edu.au

ABSTRACT

The mathematical analysis of the Internet and World Wide Web (WWW) is distinctly aspatial at present, with the transaction flows defined specifically by time-dependent indices (such as, the Internet Weather Report). How should the Internet and WWW be viewed as a geographical system where both space and time are fundamental to interaction? The retail aggregate space-time (RASTT) model has been developed previously to study trips to and from shopping malls and this model may provide some insights into the framing of this question. The RASTT model can be developed from a time-dependent random walk from an ensemble of home-based computers sending and receiving transactions through a network of sites. The spatial solution forms very weak gravity interactions and the time-dependent solutions are demand waves circumnavigating the Earth. Recent experimental results from Microsoft Research support these conclusions. These flows have the interesting property of moving either forwards or backwards through regions of time relative to the rotation of the Earth. The model can be developed to show bias in flows to USA sites.

A paper to be presented at the European Regional Sciecne Conferenece, Dortmund Germany. A version of this paper is appearing in "Modelling Geographical Systems: Statistical and Computational Applications" edited by Barry Boots, Atsuyuki Okabe, and Richard Thomas and the journal *Geojournal* both to be published by Kluwer Academic Publishers B.V., Dordrecht, Netherlands.

1. Introduction

The mathematical description of the Internet is a new challenge facing applied modellers. There are now new spatial and temporal accessibilities to consider and new concepts emerging, such as,'e-tailing', where commercial transactions can take place globally and almost instantaneously. This freedom of access into the Internet for consumers means issues of physical location, travel time or market area may be less relevant and the research frontier has to deal with such things as 'virtual distance' and unrestricted shopping opportunities between countries. There even appears to be some sort of time substitution for spatial interaction (particularly from time-poor affluent households). A key theoretical question is whether cyberspace is a product of what Marx described as 'time annihilating space'.

The Internet forms the physical network of connectivity (such as, optical cables and phone wires), where there are nodes or 'routers' that navigate packets of data from one computer to another (Barabasi, 2001). The Internet is therefore spatially specific in that flows occur through physical space. Conversely, in the World Wide Web (WWW) or the Web, links can be easily established arbitrarily as virtual connections between any two computers independent of spatial co-ordinates. The idea of the WWW originated from a hypertext technique where a considerable amount of multimedia is interconnected (Jiang and Ormeling, 2000). It is more content-specific and its properties are analysed by 'maps' that tell how the pages are linked together.

Both the Internet and the Web can be regarded as a network of nodes and links forming a complex graph defining what is known as 'cyberspace'. Cyberspace is a computer generated landscape which integrates these networks into a virtual space. An important question therefore emerges: how does this new landscape affect the flows of information and the importance of distance? What is increasingly becoming apparent is that we have to try to understand the relationship between cyberspace and geographic space and how to develop models that recognise their distinctiveness.

Much of the current Internet research involves the application of graph theory to the study of the Internet and the Web (for example, Barabasi and Albert 1999, Albert *et al.* 2000; Cohen *et al.* 2000). It is interesting that this connectivity and its theoretical descriptions are expressed in terms of time and that there is little recognition of the spatial domain. For example, The Internet Traffic Report (2001) uses a time-based index describing the round trip travel time of major paths on the Internet (also termed 'latency'). The distance factor is replaced by how much time it takes to transfer data. Further, the so-called maps plot connectivity and are essentially aspatial. Within a geographical context this is not satisfactory, because the flows of time-dependent Internet traffic around the world are passing through countries and time zones relative to a 24-hour boundary (Figure 1). This is in contrast to an aspatial view of Web traffic and connectivity using graph theory (Figure 2). The aim of this review is to look at the Internet as a geographical system in space and time and endeavour to set a modelling context for future research.

A second question concerns how retailing fits into a model of the Internet. This is not the aim of this review. However, despite the considerable euphoria and a stock market boom in the late 1990s in technology stock and 'dot com' companies, there are very few retail success stories from marketing on the Internet. Even such an e-tail 'success' story as Amazon.com recorded a $390 million loss in 1999. Why is this the case? There is the possibility that the structure of central places (particularly in terms of cities as points of distribution) are different fundamentally to the evolution of connectivity within the Internet and the Web. Much of the efficiency of cyberspace is seen in time minimisation, but there still could be a place for distance minimisation strategies (and the gravity model) for the distribution of goods and services. The failure to understand this difference and its geographical underpinnings, could be a major barrier to successful marketing and profitability for Internet retailing. This suggests that there is much research needed to understand the operation of the Internet as a geographical system and as a way of receiving and distributing commercial and retail transactions.

2. Modelling Background

2.1 Introduction

The application of random graph theory to define the connectivity of the Internet and the Web is a growing research area at present in the physical sciences. This work will be briefly reviewed. An alternative is to view random processes of connectivity along a time line through differential equations. One such application is the retail aggregate space time trip (RASTT) model (for example, Baker 1994; 2000). It has been applied extensively to study trips to and from point densities (shopping malls) along a time line, where its time-dependent solutions relative to a periodic boundary, suggests some fruitful insights into how the Internet can be modelled. Its underpinnings will also be summarised to set the context for its application to cyberspace.

2.2 Graph network models of the Internet

The analysis of complex networks can be divided into two major classes based on their connectivity distribution $P(m)$ which defines the probability that a node in the networks is connected to m other nodes (Albert *et al.* 2000).

(1) The first type of networks is characterised by a $P(m)$ that peaks at an average $\langle m \rangle$ and decays exponentially for large m. These networks are homogeneous in that each node has approximately the same number of links. Exponential networks (such as the random graph model of Erdos and Renyi, 1960) have a connectivity that follows a Poisson distribution peaked at $\langle m \rangle$ which decays for $m >> \langle m \rangle$.

(2) The second type belongs to inhomogeneous networks (or 'scale-free' networks) where $P(m)$ decays as a power law (or $P(m) \sim m^{-\gamma}$) free of the characteristics of scale. This network has a majority of nodes with only one or two links, but a few large nodes of links guaranteeing that the system is fully connected. An example of this type of network is the World Wide Web and this type of model can be visualised by the Internet tree simulated by Chiswell (1999) (Figure 2).

Barabasi (2001) uses geographical examples to distinguish both types of networks and these are pertinent to the development of an Internet model. An exponential network is a road map that has cities as nodes and expressways as links, because most cities are central places located at the intersection of the motorways. Conversely, an airline route map is a Type 2 network, because although most airports are served by a small number of carriers, they have a few hubs (such as, London) from which links emerge to almost all other US or European airports. The WWW is seen as an example of the latter because a majority of documents have only a few links. It appears that Type 2 networks are also hierarchical. They are also preferential, since they contain nodes that have a high probability of being connected to another node with a large number of links. For example, a new Web page is more likely to be linked to the most popular documents on the Web, since these pages are the ones we know about. Research by Faloutsos *et al.* (1999) have shown that the network behind the Internet also appears to follow the power-law distribution of inhomogeneous networks. This means that the physical wiring of the Internet is also dominated by several highly connected hubs. As Barabasi (2001) states: why do systems as different as the Internet, which is a physical network, and the Web, which is virtual network, develop scale-free networks with a power-law decline in connectivity?

This analysis is distinctly aspatial, but is still imbedded in time-dependent variables for the transfer of information. How should the Internet and WWW be viewed as a geographical system where both space and time are fundamental to interaction? The RASTT model may provide some insights into the framing of this question.

2.3 The retail aggregate space time trip (RASTT) model

The RASTT model defines ‘when’ and ‘where’ consumers enact aggregate shopping behaviour and is underpinned by the so-called ‘supermarket’ equation (Baker 1994, 2000). It is constructed around a differential equation of spatial and temporal operators acting on a population function f (which is standardised to f_o, the equivalent calibration function per hundred shoppers divided by the size of the aggregation unit). A mathematical operator describes what has to be done on a function much like a verb does within a sentence. The ‘supermarket’ equation is different to classical diffusion, because time is differentiated twice and space once. This second order time operator is of immediate interest to theoretical geography, because when this is made equivalent to a first order spatial operator, the solution involves a gravity model of trip distributions in space and a periodic function of time-based demand. This relevant differential equation takes the usual form of:

$$\frac{\partial f_o}{\partial x} = \frac{1}{M}\frac{\partial^2 f_o}{\partial t^2} \tag{1}$$

where x defines the spatial coordinate, t, the trip time and M a transport constant for a calibrated population density of f_o of shoppers. This linear equation (with the transport coefficient M constant) can only apply to one shopping centre, but the operators ($\boldsymbol{d}/\boldsymbol{d}x$ and $\boldsymbol{d}^2/\boldsymbol{d}t^2$) can apply equally for individual or group shopping. Equation (1) is stating there is a trade-off between trip operators through space and time. In other words, ‘where’ a consumer shops is dependent on the shopping cycle (time of the day or day of the week). In the particular solution, the gravity model of trip distance D is the spatial solution between residences (aggregated in concentric one kilometre bands) and the shopping centre (or $x - x_o$, where $x_o = 0$ defines the location of the shopping centre). The underlying behavioural assumption is distance minimisation in trip assignment. The corollary in the time solution of Equation (1) is that such shoppers make regular time-based trips to and from a shopping centre at $x_o = 0$. A population density of shoppers f_o in this model therefore assumes populations regularly select trips that minimise distance (through exponential decay) to and from a shopping centre. A socio-economic group that best approximates this assumption are ‘over 65 years’, whilst for trip purpose, the weekly food and grocery trip fits this model well (Baker 1994; 1996). Mathematically, this statement can be expressed as a particular solution of Equation (1) for one centre as:

$$f_o = A\,exp(-\boldsymbol{b}\;D)\begin{Bmatrix} sin(kt) \\ cos(kt) \end{Bmatrix} \tag{2}$$

where $\boldsymbol{b}$ is the gravity coefficient and k, the interlocational trip frequency (ITF) defines how many trips are made by individuals or groups to the shopping centre. The ITF is introduced arbitrarily as the separation constant to solve the differential equation. This solution is stating that the undertaking of regular or periodic trips (*sin kx*) to a shopping centre is discounted by how far away we are to a centre ($exp - \boldsymbol{b}D$), since there are increasing opportunities to shop elsewhere the further they live from this centre.

Implicit in this approach is that destinations are located along a time line of shopping opportunities. The advantage of this method is that time boundaries can be introduced as part of solving the differential equation and this will affect spatial patterns of the gravity model of trips to and from the centre. The policy corollary is that the shopping hours a mall trades will affect the extent of the surrounding market area (Baker 2000). The RASTT model therefore deals with shopping trip distributions from a particular centre where time boundaries on a time line of destinations can change the spatial distributions. Other methods have specific problems dealing with time. For example, entropy maximising can also derive gravity spatial interaction by optimising the assignment of trip origins to destinations through statistical laws of large numbers (Wilson, 1967; Roy and Lesse, 1981). However, this method has problems dealing with time in maximising the assignment problem to and from a shopping centre. The increase in entropy only occurs between opening and closing times: when the shopping centre is shut, the complexity of the trip assignment problem disentangles and trips approach zero (contrary to the physical analogue which approach infinity). The RASTT model does not suffer from this difficulty because the solution is defined by the hours that the centre is open (0 to T) and all other possibilities are zero.

3. The RASTT Model and Internet Transactions

The operators of the RASTT model ($\delta / \delta x$ and $\delta^2 / \delta t^2$) are not affected by the numbers involved in interaction (applying equally to individuals or populations of billions) and are therefore classified as scale invariant. These operators are defined relative to time boundaries for movement through physical space. Yet are these mathematical operators applicable to the Internet, where there is still a real time boundary (the 24 hour rotation of the Earth) defining the movement of transactions? Consumers can also make virtual rather than real trips to retail sites and the RASTT model therefore offers the scope to explore the movement of demand through virtual space as well as physical space.

The immediate problem in the RASTT model is that relative time functions to the boundary can be either positive or negative. In previous work, the idea of negative time in the context of the process of shopping trips was thought initially to be meaningless and the time boundaries were only applied from 0 to T (and the 0 to –T range discarded; see Baker, 2000). However, the idea of negative values relative to the direction from the boundary for Internet transactions is not as nonsensical as it first appears. Rather than framing the problem over 168 hours per week, we set it for trips or transactions through space over a 24-hour period (the daily cycle). The spatial origin could be located at a computer at an arbitrary location and the consumer can either go forward or backward along a time line relative to this 24-hour boundary. For example, if the individual is located at Sydney (33^o S Lat and 161^o E Long), that person can either go two hours forward in time to a site in Auckland (37^o S Lat and 175^o E Long) or two hours backwards in time to Perth (32^o S Lat and 116^o E Long). The RASTT model can be derived for physical trips to a mall and such trips are only viewed positively along time lines. Conversely, virtual trips on the WWW can be defined as either moving backwards or forwards relative to the 24-hour time boundary. This is a radical statement because it gives a plausible example of how relative time can exist as a corollary of virtual distance and have different properties to physical time. Boulding (1985) states that in the physical sciences, time is assumed to approach infinity in order to focus on spatially specific solutions. Alternatively here, we assume initial spatial locations and produce time-specific solutions, including solutions that can be negative. The study of the Internet as a geographical system therefore provides an opportunity to introduce a new concept and to see if it has any further properties of interest.

There is a possibility of a convergence of virtual distance into a fixed point (the computer screen) at any time. An important question is: can relative time influence the patterns of virtual space? It returns to a concept of a dynamic convergence of locations found in the geographical literature of the late 1960s, where the evolution of spatial reorganisation changes in space-time connectivity, particularly from improvements in transportation and technology (Janelle, 1968, 1969; Forer, 1978; Gatrell, 1983). The Internet is perhaps the next stage in the evolution of this space-time connectivity. Blaut (1961) argued that every empirical concept of space must be reducible by a chain of definitions to a process and Janelle (1969) states that inherent in Blaut's view is the implicit existence of a temporal pattern in each and every spatial pattern. In the RASTT model, this process is summarised as a second order time differential (or operator) that can yield positive and negative time-based solutions. This means that unlike physical time, relative time can lead to reversible time-based processes, a truly remarkable possibility. For example, it means that in the election of a US president, polling booths can be closed in the east, yet the proportion of votes counted and reported on TV can feedback simultaneously to voters on the west coast who are still voting (and can change their votes based on the east coast trends). Reversibility of a result is possible within the boundaries of relative time. It is possible to have two simultaneous sites connected by virtual distance on a computer screen in different time zones. The Internet presents a new horizon to geographical systems because we have to now distinguish between relative time (to a rotating boundary) and physical time (to infinity) and real trips (where people change their spatial co-ordinates) and virtual trips on the Internet (where the location is still at the initial spatial co-ordinate). As this stage we can only make guesses as to how this evolution in space-time connectivity develops, but there are some clues already discernible from the nature of the RASTT model.

What features could be expected from a RASTT model representation of Internet transaction? There are two areas of immediate interest.

3.1 The condition for space time convergence

The condition for space-time convergence in the solution of the supermarket equation (Equation 1) when a 24-hour boundary is applied, yields the same relationship between the gravity coefficient β and the square of the mean interlocational trip frequency *k* divided by the transfer constant *M*, namely:

$$\beta = \frac{k^{2}}{M} \qquad (3)$$

The interlocational trip frequency (ITF) defines the average number of trips or transactions undertaken per day by users and because it is squared it can be applied to virtual trips either forwards or backwards through relative time. The RASTT model suggests that there would still be gravity interaction of physical distance for Internet patronage, but this would be at least one order of magnitude lower than gravity coefficients computed from shopping trips to malls using concentric aggregation. Yet we would expect that it would vary for the type of transaction. For example, weekly food orders would have (with *k* higher) greater β values, indicating the distribution of food would be more localised than for a lower frequency consumption item such as compact disks. We would therefore expect that one feature of the Internet as a geographical system would be ‘very weak’ gravity interactions, but this would still be relative to the type of transaction and the limits of the distribution system.

If β is assumed to be very small and the frequency of patronage the same order of magnitude as shopping trips to malls (one visit to a site per day), the transfer coefficient *M* for Internet usage would have to be very large compared to its value for physical trips to a mall. This is not hard to visualise, with Internet traffic at least one order of magnitude higher than the physical trips to particular shopping malls. The other key question is whether *M* remains a constant, or a variable, making the differential equation non-linear.

3.2 Space-time distributions of internet demand

The type of space-time distributions that could apply to Internet patronage are simulated in Figure 3 for β = 0.0001, T=24 hours, $x_0 = 0$ to $x_0 = 10,000$ km and an arbitrary population density $f_o = 10$ for a sequence of *k* values where *k* = 0.1, 0.2, 0.3, 0.4, 0.5....1.0. The simplest distribution of spatial demand for Internet patronage at a site, receiving both positive and negative flows of transactions, is a gaussian–type distribution between *k* = 0.1 and 0.2 (Figure 3). This is not surprising since a gaussian distribution is an equally valid solution to Equation (1) for a time-based random walk problem. The solution has some advantages in this probabilistic form, because variables can be expressed as average quantities, such as, ‘distance’ and ‘number destinations per visit’. A gaussian distribution can be expressed as probability distribution P(t,x) for a density of web transactions at a site $f_o = f_{os}$. If this site receives n_a transactions per unit distance d, with total transactions $\Phi = n_a d$, the probability distribution is defined as:

$$\frac{f_o}{f_{os}} = P(t,x) = \frac{1}{2(pMx)^{1/2}} exp\left(-t^2 / 4Mx\right) \qquad (4)$$

where *t* is equal to the time for each transaction to travel to the site. The transfer coefficient M can be defined alternatively as :

$$M = \frac{1}{2} nt^2 \qquad (5)$$

The transfer constant is then the number of transactions per unit distance multiplied by the relative time *t* taken to reach the site. Equation (4) is the type of distribution that has been simulated in the k = 0.1 to k = 0.2 range in Figure 3. It is an unbounded gaussian time distribution, where transaction densities can be plotted for $ln\ f$ versus t^2 and the slope of the straight line is $(4Mx)^{-1}$. The average time taken by the transaction is defined by the mean square displacement ($\overline{\Delta} t^2$), namely:

$$\overline{\Delta} t^2 = 2Mx. \qquad (6)$$

The RASTT model can therefore define the possibility of a number of distinctive features of Internet patronage relative to traditional spatial interaction modelling:

(1) the gravity model of spatial interaction would have very small β coefficients compared to a regional shopping mall;

(2) technology allows for a space-time convergence to occur on the computer screen rather than shopping malls and virtual distance allows for the possibility of simultaneous connections both forwards or backwards in relative time;

(3) such connections can have implications for activities in different time zones, such as the US presidential elections or stock market activity; and

(4) transactions to sites should be represented in their simplest form as time-based gaussian distributions.

This type of model (and differential equation) is not found in traditional applications of applied mathematics because of the problem of dealing with positive and negative time-based functions. In the case of the Internet, such difficulty is an advantage because transaction flows can be modelled globally relative to a time boundary. It means that time has to be viewed differently at this particular scale (defined by the rotation) and has different properties to physical time (such as reversibility).

The next step is to look more formally at the derivation of the ‘supermarket’ equation in this Internet context for transaction interaction between a number of web sites.

4. Deriving the RASTT Model for Internet Transactions (after Ghez, 1988)

Consider a network of web sites linked by a time line with an arbitrary origin at W, where these sites are designated through integers $i = 0, \pm 1, \pm 2, \pm 3, \ldots\ldots\ldots$ For example, a household at Sydney could have a choice of other sites at $i = \pm 1$ at Auckland or Perth (Figure 4). Each web site serves a number of households at a particular locality and there are Φ_i households linked to each site *i*. Assume that each of these households can jump to adjacent web sites with a frequency Γ that does not depend on the characteristics of *i*. These households can access sites forward in time or backwards in time. It is assumed the movement forwards or backwards are equally likely. Therefore, movement from site *i* to site

i +1 per unit distance occurs at a rate of ½ Γ Φ_i. Likewise a household or web page at site i +1 can reply to the household or web page at ***i*** at a rate of ½ Γ Φ_{i+1}. The resulting rate of exchange is:

$$\mathrm{E}_{i+1/2} = \frac{1}{2}\Gamma\left(\Phi_i - \Phi_{i+1}\right) \tag{7}$$

and for i-1 into the site i, the flux is

$$\mathrm{E}_{i-1/2} = \frac{1}{2}\Gamma\left(\Phi_{i-1} - \Phi_i\right) \tag{8}$$

The change in web traffic into and out of the i^{th} site at an origin or hub (such as Sydney) is given by a definition of all possible transitions:

$$\frac{d\Phi_i}{dx} = -\frac{1}{2}\Gamma\ \Phi_i + \frac{1}{2}\Gamma\ \Phi_{i+1} - \frac{1}{2}\Gamma\ \Phi_i + \frac{1}{2}\Gamma\ \Phi_{i-1} \tag{9}$$

The space discounting equation (or rate equation) in terms of the distribution of users in and out of the i th site is (by collecting terms)

$$\frac{d\Phi_i}{dx} = \frac{1}{2}\Gamma\ \left(\Phi_{i+1} + \Phi_{i-1} - 2\Phi_i\right) \tag{10}$$

This exchange rate of web or Internet traffic is between nearest-neighbour destinations on the time line around site i and this can be expressed in terms of the exchange rate between sites (using Equation 1 and 2)

$$\frac{d\Phi_i}{dx} = -\left(\mathrm{E}_{i+1/2} - \mathrm{E}_{i-1/2}\right) \tag{11}$$

The change in the web site content is defined by the difference between flows in and flows out of transactions within the connectivity.

Comments

(1) The jump frequency of transactions between sites is constant and it is assumed independent of the site index i and its location in space.

The data signal should not change its frequency within the network and does not depend on the location of the computers. This appears a reasonable assumption and agrees with the aspatial nature of the graph theory approach.

(2) This frequency of movement does not depend on the distribution of households or users in the neighbourhood of the i th site.

The distribution does not have to be homogeneous. This also appears to be a good approximation and parallels the assumptions of graph theory.

(3) The time distance between sites and the type of transfer network does not influence the process, the only thing that is important is the time-based ordering of the points.

The receipt of transaction does not depend on physical location, but on a time-dependent ordering of site hits. Once again this is a reasonable assumption for the Internet or Web.

Equations (7 to 11) define a time line between sites where the distance between points and the hierarchical network of sites is not relevant, rather, what counts is the ordering of hits to the site. Equation (11) states a conservation law where the transactions in and out defines the content of a web site. The time distance p between web sites is assumed to be equal between the origin (such as Sydney) and the i^{th} site and has the co-ordinate of $t_i = i\,p$ on the time line of transaction flows. The transaction density $\boldsymbol{f}_o\,(x,t)$ is assumed to interpolate the previous function at site i, with the co-ordinate $i\ \boldsymbol{f}_o(x)$ by the following assumption:

$$f_o(t,x) = \Phi_i(x) \tag{12}$$

at destinations located at $\boldsymbol{t = t_i}$. but is arbitrary elsewhere. The assignment of this transaction density function around this web site at t_i can be expanded by a Taylor series:

$$f_o(t_{i\pm1}) = f_o(t_i) \pm p\frac{\partial f_o}{\partial t}\Big|_{t_i} + \tfrac{1}{2}p^2\frac{\partial^2 f_o}{\partial t^2}\Big|_{t_1} + \text{terms of order } \boldsymbol{p^3} \tag{13}$$

and using the condition in Equation (12), the expansion becomes when substituted into Equation (10) (noting that we are interested in both forward and backward motion relative to the 24-hour cycle):

$$\frac{\partial f_o}{\partial x} = \frac{1}{2}\Gamma p^2\frac{\partial^2 f_0}{\partial t^2} + \text{ terms of order } \boldsymbol{p^4} \tag{14}$$

Comment

The condition for this approximation is that p ´k <<1, or in other words, the time distance between sites is very much smaller than the smallest significant wavelength and the A amplitude of the demand wave (A sin kt) must be insignificant outside ½k_{max}½

With the definition of latency between sites expressed in milliseconds, such a condition appears reasonable for the Internet of Web traffic.

The continuous exchange function between sites with transaction densities can be written as:

$$\mathrm{E}(x,t) = -\frac{1}{2}\Gamma p\frac{\partial f_o}{\partial t} \tag{15}$$

and the conservation law can be rewritten from Equation (10) as:

$$\frac{\partial f_o}{\partial x} = -p\frac{\partial E}{\partial t} + \text{terms of order } \boldsymbol{p^4} \tag{16}$$

Now the average transaction density $\overset{\&\&}{f}_o$ is equal to $\overset{\&\&}{f}_o = \boldsymbol{f_o / p}$ where $\boldsymbol{p}$ is equal to the average transaction time between sites and Γ equal to the average number of web sites visited per unit trip distance. The transport constant M to the centre is defined in the previous context as:

$$M = \Gamma p^2 \tag{17}$$

The rate of exchange, the conservation law and supermarket equation for web traffic on the Internet becomes, respectively:

$$E(x,t) = -\frac{1}{M}\frac{\partial f_o}{\partial t} \tag{18}$$

$$\frac{\partial f_o}{\partial x} = -\frac{\partial E}{\partial t} \tag{19}$$

$$\frac{\partial f_o}{\partial x} = \frac{1}{M}\frac{\partial^2 f_o}{\partial t^2} \tag{20}$$

Equation (18) defines a condition for the conservation of information on the Internet. The minus sign in Equation (18) is a point of debate, because it implies that information flows from sites of high densities to low densities (akin to classical diffusion) with the analogy to graph theory that the high point densities are the hubs in the network. This is a reasonable assumption in the context of inhomogeneous scale-free networks in graph theory with a power-law decline in connectivity.

5. Empirical Evidence for a RASTT Model for Internet Transactions

There are a number of hypotheses that are consequence of the application of the RASTT model to Internet transactions. These will be briefly reviewed in the context of recent geographic-specific Internet experiments and data sites (Padmanabhan and Subramanian, 2001; Internet Weather Report, 2001).

5.1 Is weak gravity interaction a feature of Internet flows?

The spatial solution to Equation (2) would suggest that gravity spatial interaction is still relevant to Internet traffic flows. Padmanabhan and Subramanian (2001) show an example of the gravity relationship between client and proxy sites for the America On-line network. The cumulative probability for clients follows the gravity distribution of a regional shopping mall except there is a difference in the order of magnitude for distance (Figure 5a and Figure 5b). This strongly suggests that the Internet exhibits very weak gravity interactions.

5.2 Is pause time (Dt) and surrogate measure for distance?

In Equation (2), one of the predictions of a time-based gaussian distribution is that the average transaction time (or delay or latency) taken by the transaction is defined by the mean square displacement $\overline{\Delta}t^2 = 2Mx$. In other words, one of the conditions of space-time convergence is that the transaction time (to and from a client from a site) is a function of the distance from the site. Padmanabhan and Subramanian (2001) undertook an experiment with a probe machine at Seattle, USA, measuring transaction delay in four categories (5-15ms; 25-35ms; 45-55ms; 65-75ms) relative to geographic distance (Figure 5c). The results support the prediction of a linear relationship between latency and distance (as suggested by Equation 6). Over 90% of small delay values (under 10ms) come within 300km from the source. The RASTT model therefore provides a clear mathematical relationship to test these experiments and suggest the existence of time-based gaussians.

5.3 Is the transport coefficient M a measure of congestion?

Congestion in the network is one of the problems in using time delay as a surrogate measure for distance. Padmanabhan and Subramanian (2001) overcome this problem by selecting minimums in delay samples (10 to 15 samples) between hosts to eliminate the effect of congestion. Equation (5) from the time-based gaussian suggests that the transfer constant M will be a function of the number of web sites and the transaction time. This transaction time can vary and so M can be a variable and the differential equation non-linear. What we could do is to seek M_{min} from a substantial sample size and set this as a constant in the differential equation and $M - M_{min}$ would then represent a measure of congestion in the system.

5.4 Are spatial demand waves of internet traffic observable?

Equation (2) predicts that there is a spatial demand wave circumnavigating the globe on a 24-hour cycle which, in our simulation (Figure 3), would most probably take the form of a time-based gaussian between k ~ 0.1 and k ~ 0.2. The Internet Weather Report (2001) provides an animation of this type of event, where variables of latency (the bigger the circle, the slower the return time trip) and the number of hosts at a given location (defined by a colour spectrum) varies over a 24-hour cycle.. The slowness of the latency (the time taken for a return trip between a host and a client) may be a function of demand and if this is the case, then the arrival of the Internet demand wave can be visualised in the USA. For example, on October 1, 2001 at 1.00am (CDT), there are few congestion nodes and smaller latencies compared to 4.00 pm (CDT) when there are peak demand times for both east and west coast interactions (Figure 6). The latency changes through a 24-hour cycle, where there are a few small guassian-type distributions in the early morning, but as the demand wave envelops, these grow in the peak afternoon period when the network comes under heavy load.

5.5 Is the nearest neighbourhood assumption justifiable?

In the derivation of a time-based random walk for Internet traffic (Equations 7 to 11), the exchange rate assumes nearest neighbour sites within the jumping between hosts. Is this justifiable? The WWW is a small-world network, which is a sparse network where nodes are connected to other nodes in their neighbourhood but otherwise the average distance between nodes is high (Watts and Strogatz,1998). Padmanabhan and Subramanian (2001) in their experiments found their geographic clustering algorithm (GeoCluster) produced the best results over varied data sets. This would be expected if time-dependent guassian distributions are generated at different scales (either for countries or individual hosts such as a university campus). This nearest neighbour assumption appears to have some credence, at least in mapping information generated from Web traffic.

6. Some Further Theoretical Thoughts

One of the basic assumptions in the time-dependent random walk is that it is equally likely that a transaction can move either forwards or backwards in relative time. This might be a satisfactory assumption for the internal USA situation or the Sydney-Perth-Auckland domain, but for a global model, there is going to be a higher probability of moving in either direction towards the high connectivity of USA sites. The jump frequencies can no longer be assumed to be isotropic (namely, the jump probabilities to the right or left have the same value $1/2\Gamma$). Therefore, the jumps forwards or backwards in time are therefore not equal for Internet sites because of the influence of USA hubs. In the derivation, the exchanges now do not vanish because of the asymmetry in jump frequencies between sites. A constant distribution is no longer a necessary condition for equilibrium, but the conservation law remains, as well as other assumptions in the derivation (such as, nearest neighbour and data ordering; Ghez,1988). This global Internet equation that can account for the asymmetry in transaction flows can be defined if the jump time distance is now $t_{i\pm 1/2} = (i \pm 1/2)a$ and there is introduced what is termed a drift velocity $v = Da$ into the model. Equations (18-20) can now be rewritten:

$$E(x,t) = -\frac{1}{M}\frac{\partial f_o}{\partial t} + v f_0$$

(21)

$$\frac{\partial f_o}{\partial x} = -\frac{\partial E}{\partial t} \tag{22}$$

$$\frac{\partial f_o}{\partial x} = \frac{1}{M}\frac{\partial^2 f_o}{\partial t^2} - v\frac{\partial f_0}{\partial t} \tag{23}$$

The drift velocity *v* through the network is proportional to the difference in the jump frequency between sites and is thus a measure in the bias that the USA sites introduce into the system, even if the households are evenly distributed among the sites. This drift velocity could simply be interpreted as the Internet's mean velocity field (which can fluctuate depending on the congestion or distance between sites within the network). Equation (23) represents a new differential equation which describes the Internet, where the first term may describe the hub exchange of transaction flow (the reverse diffusion), whilst the second is the overall movement throughout the network. It now appears to be non-linear and its solutions are beyond the current scope. Nevertheless, it allows a global view of the Internet and data flows.

There is an interesting connection with this Internet equation to the telegrapher's equation:

$$\frac{\alpha}{M}\frac{\partial^2 f}{\partial t^2}+\frac{\mu}{M}\frac{\partial f}{\partial t}=R\frac{\partial^2 f}{\partial x^2}+s\frac{\partial f}{\partial x} \qquad (24)$$

where α, μ , R and s are appropriate constants. In the early days of telegraphy, the signal diffusion reduced the data rate in long cables such as the early Atlantic cables (Montroll and West, 1979). The wave propagation over time was replaced by a diffusion packet. For the Internet equation, the time operators are the same as the telegrapher's equation (apart from the sign of the first time operator). However, in the Internet equation, there is no second order space operator ($\partial/\partial x^2$) and spatial interaction is exclusively defined by the gravity operator $(\partial/\partial x)$. The mathematical relationship between both differential equations is also an area for future inquiry.

7. Concluding Remarks

The application of a RASTT-type model (time-dependent differential equations with a second order time operator) to Internet transactions is explored and the results suggest that this type of model has some relevance to understanding the dynamics of the Internet and WWW as a geographical system. Such a model has a number of implications that warrant further consideration.

Firstly, this model involves the idea of relative time where flows of data can go either forwards or backwards in time relative to the rotation of the Earth. Such second order time operators in the RASTT model are a peculiarity and there is a corollary of the possibility of reversibility within time functions. The example of this process is in the USA elections where voters in California (because of different time zones) can change their vote according to events broadcast after the closure of polling in the eastern states. Relative time, therefore, has different characteristics to physical time where events are irreversible and assumed positively to infinity.

Secondly, in the graph network model of the Internet, the power-law evolution of the structure is not the same as a random-walk generation of linkages, where in the latter, the gaussian distribution (and negative exponential functions) is an integral part of the solutions of the evolution of connectivity. Perhaps this difference is significant, because the evolution of supply points might still follow random-walk generation and the gravity model might limit the physical accessibility of much of the power-law evolution of the WWW or Internet. It might be easy to get connected to the WWW, but to receive goods and services from e-tailers still relies on central places and points of distribution. This might be one reason for the difficulty of e-tailers in making profits from their WWW networks.

There are a number of encouraging results empirically that support predictions from the RASTT model.

1. There is an example of very weak gravity interaction in the relationship between the distance between client and proxy sites for the America On-line network.

2. The prediction from a guassian random walk that pause time is a function of distance to the site has been replicated in experiments by Padmanabhan and Subramanian (2001).
3. The postulated spatial demand wave can be observed across the USA and globally in the Internet Weather Report (2001).

Such results should provide motivation for further experimental work. The RASTT model provides a mathematical framework to look at the dynamics of the Internet and WWW. The bias from USA sites can be introduced into the model and this form has some similarity to the telegrapher's differential equation, which itself is an interesting area for further inquiry. The idea of a mean velocity field of Internet traffic flows has immediate applicability to geographical information systems and visual representations of the Internet.

The RASTT model has been successful at looking at distance trip behaviour to spatial centres of demand (shopping malls) between 10^0 and 10^1 orders of magnitude. There is also some indication that virtual trips with physical distances at 10^3 and 10^4 orders of magnitude through the Internet can be described by the same operators in the RASTT model. The implication is that these operators are scale invariant and supports the idea that operator-based modelling is a way to overcome problems of scale.

Finally, modelling the Internet is a new frontier for spatial interaction modelling. The equations from the RASTT model have no physical analogy. They are essentially geographic, where time zones and distance decay are fundamental to the dynamics. Operators are the key to this view of space-time processes. They allow 1960s concepts, such as, the space-time convergence at central places, to be equally applicable to the computer screen in the 2000s. Such opportunities occur because of the mathematical tractability of using partial differential equations to describe geographic processes.

Acknowledgements

I would like to thank Barry Boots for encouraging this research and to Mike Roach for his cartography.

References

Albert, R., Hawoong, J., and Barabasi A. (2000) "Error and attack tolerance of complex networks." *Nature*, **406**, 378-382.

Baker, R.G.V. (1994) "An assessment of the space-time differential model for aggregate trip behaviour to planned suburban shopping centres," *Geographical Analysis*, **26**, 341-362.

Baker, R.G.V. (1996) "Multi-purpose shopping behaviour at planned suburban shopping centres: a space-time analysis," *Environment and Planning A*, **28**, 611-630.

Baker, R.G.V. (2000) "Towards a dynamic aggregate shopping model and its application to retail trading hour and market area analysis," *Papers in Regional Science*, **79**, 413-434.

Barabasi, A. (2001) "The physics of the Web" Physics World, **14**, No.7.

Barabasi, A. and Albert R. (1999) "Emerging of scaling in ransom networks," *Science*, **286**, 509-512.

Blaut, J.M. (1961) "Space and process," *The Professional Geographer*, **13**, 1-7.

Boulding, K.E. (1985) "Regions of time," *Papers of the Regional Science Association*, **57**, 19-32.

Cheswick, B. (1999) Internet Mapping Project, http://www.cs.bell-labs.com

Cohen, R., Erez, K., Be-Avraham, D., and Havlin, S. (2000) " Resilience of the Internet to random breakdown," *Physical Review Letters*, **85**, 4626-4628.

Erdos, P. and Renyi, A. (1960) On the evolution of random graphs. *Publications of Mathematical Institute of Hungarian Academy of Sciences*, **5**, 17-60.

Faloutsos, M., Faloutsos, P., and Faloutsos C. (1999) “ On power law relationships of the Internettopology,”.Proceedings ACMSIGCOMM, http//www.can.org/sigcomm99/papers.

Forer, P. (1978) “Time-space and area in the city of the plains,” In *Timing Space and Spacing Time*,Vol.1, Eds. Carlstein, T., Parkes, D. and Thrift, N. (Edward Arnold, London).

Gatrell, A. (1983) *Distance and Space*, Clarendon Press, Oxford.

Ghez, R. (1988) *A Primer of Diffusion Problems*, John Wiley, New York.

Internet Weather Report (2001) hhttp://www.mids.org/weather

Interent Traffic Report (2001) http://internettrafficreport.com/

Janelle, D.G. (1968) “Central place development in a time-space framework,” *The Professional Geographer*, **20,** 5-10.

Janelle, D.G. (1969) “Spatial reorganisation: model and concept,” *Annals of the Association of American Geographers*, **59**, 348-364.

Jiang, B. and Ormeling, F. (2000) “Mapping Cyberspace: visualising, analysing and exploring virtual worlds.” *The Cartographic Journal*, **37**, 117-122.

Montroll, E.W and West, B.J. (1979) Stochastic processes. *Studies in Statistical Mechanics, VII, Fluctuation Phenomena*, Eds. Montroll E.W. and Lebowitz J.L (Elsevier: New York).

Padmanabhan, V.N. and Subramanian, L .(2001) “An investigation of geographic mapping techniques for Internet hosts” Proceedings of ACM SIGCOMM, San Diego, USA.

Roy, J.R. and Less, P.F. (1981) “On appropriate microstate descriptions of entropy modelling.” *Transportation Research B*, **15**, 85-96.

Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of ‘small world’ networks.” *Nature*, **393**, 440-442.

Wilson, A.G. (1967) “A statistical theory of spatial distribution models, ” *Transportation Research*, **1**, 253-269.

Figure 1 An example of the Internet as a geographic system is illustrated for UUNet (with 27% of the global market) and shows the dominance of the USA in site traffic (Source: UUNET Website, www.uu.net)

Figure 2 The generation of an Internet Web tree showing the aspatial connectivity from 100,000 Internet routers and the hierarchial structures that develop from a few highly connected nodes (Source: Cheswick, 1999).

Figure 3 (a) A range of possible space-time distributions that could apply to Internet demand are simulated for $\beta = 0.0001$, T=24 hours, $x_0 = 0$ to $x_0 = 10{,}000$ km and a scaled $\phi_{o\,max} = 10$ for a sequence of *k* values where *k* =0.1, 0.2, 0.3, 0.4, 0.5....1.0. (b) A three dimensional plot visualising a likely form of the demand wave for *k*= 0.1.

Figure 4 The equal likelihood of jumping forwards in time to sites in Auckland or backwards to Perth from the i^{th} Sydney site defines the underpinnings of the type of differential equations in Equations (18) to (20).

Figure 5 (a) The cumulative probability for a gravity-type distribution for the distance between client and proxy for America-Online (Source: Padmanabhan and Subramanian, 2001); (b) The cumulative probability for a gravity-type distribution for a regional shopping mall (Bankstown Square, 1998 afternoon distribution; Baker 2000); (c) The results of a probe machine at Seattle, USA, measuring transaction delay in four categories (5-15ms; 25-35ms; 45-55ms 65-75ms) relative to geographic distance. The results support the prediction of a linear relationship between latency and distance suggested by Equation 7. (Source: Padmanabhan and Subramanian , 2001)

Figure 6 Internet demand wave can be visualised in the USA on October 1, 2001, when at 1.00am (CDT) there are few congestion nodes and smaller latencies compared to 4.00 pm (CDT) when there are peak demand times for both east and west coast interactions. (Source: Internet Weather Report; 2001)