Chaos and Predictability of Internet Transmission Times

Joseph C. Park*
Zenith, Inc.,
3045 Church Hill Drive,
Boynton Beach, FL, USA

The Internet consists of millions of interconnected network nodes comprising a complex system utilized for information storage, processing, and transmission. It is demonstrated that the temporal evolution of transmission times for an information packet transmitted between two stable endpoints across the global network constitutes the chaotic time series of a dynamical system with positive Lyapunov exponents and fractal dimension. Examination of system invariants establishes the predictability of local dynamical variables and sets bounds on the ability to forecast the temporal evolution of the system variables. An artificial neural network (ANN) is invoked to learn and predict the Internet response times, wherein it is established that even though the system is dynamically divergent from a local system variable perspective, the ANN is fully capable of characterizing and predicting the macroscopic behavior of the global Internet network response times.

1. Introduction

The global phenomena collectively referred to as the Internet has become an instrumental feature in the evolution of modern society into an information-based structure. Almost certainly the evolution from an agrarian, to mechanistic, to technologically-centered paradigms would have taken place irrespective of the Internet, however, it is clear that the Internet has emerged as one of the central enabling technologies facilitating the shift of modern business and commerce from paper-based information technology into an electronic one. By its nature, the Internet is a complex system consisting of a massively interconnected, yet structurally diverse network of information processing systems and subnetworks of varying complexity. One could then expect interesting features to arise in the macroscopic view of Internet behavior. The central issue which this paper addresses is analysis and prediction of one particular feature: Internet packet transmission response times across the global network. The analysis demonstrates that the Internet response

*Electronic mail address: josephpark@bellsouth.net.

times constitute a chaotic time-series, and can therefore be viewed as arising from a nonlinear dynamical system. We then proceed to address the pragmatic issue of predictability of the network response times; an issue of considerable interest to the Internet community at-large, and the major service-providers and technical supporters in particular. The prediction is performed by invoking another complex system, an artificial neural network (ANN), which generically shares many of the topological features of the network that is being modeled. It is demonstrated that despite the fact that the Internet response times consist of a chaotic time-series with positive Lyapunov exponents, it is possible to employ the inherent nonlinear information processing power of an ANN to produce pragmatically useful Internet latency predictions over timescales of several days.

The Internet has experienced exponential growth in the past decade, largely directed by network resource requirements from local access providers. As a result, the network connectivity structure has evolved according to needs of information capacity, rather than by a deterministic hierarchy dictated by resource planning. It is estimated that as of April 1999, the Internet consisted roughly of 94,000 networks constituting the major interconnection points [1], and accounted for some 1.7 billion unique universal resource locators (URLs). It is estimated that by the time this paper reaches print, the Internet will support 3.2 billion URLs accessed by some 256 million users. Figure 1 presents a map of the Internet [2] depicting a snapshot of the network connectivity in January 1999. The mapping consists of frequent traceroute-style path probes, one to each registered Internet entity, from which a tree is built illustrating the connectivity of most of the networks on the Internet.

Even a cursory glimpse of this map strikes a remarkable resemblance to the topology of interglial connectivity exhibited in neural tissue slices from the cortical regions of mammalian brains. One must wonder whether such a coincidence speaks for an underlying structure in the evolution of information transmission and processing structures.

Another impression of Figure 1 is that the interconnectivity of the Internet constitutes a fractal. Not only does it appear to be fractal in the sense of a noninteger dimension, but it also adheres to the notion of scaling invariance in the sense of self-similarity when considering that each of the points represented on the map of Figure 1 is in itself a network consisting of potentially many nodes.

When an information packet is transmitted through the network of Figure 1, the path which is taken is influenced by random fluctuations in router loads and service levels. It is therefore consistent to expect that a time series of such information packets transmitted and received between two stable endpoints will exhibit a chaotic structure. To ascertain the validity of this expectation, several standard methods are implemented to estimate invariants of the topology and dynamics of a

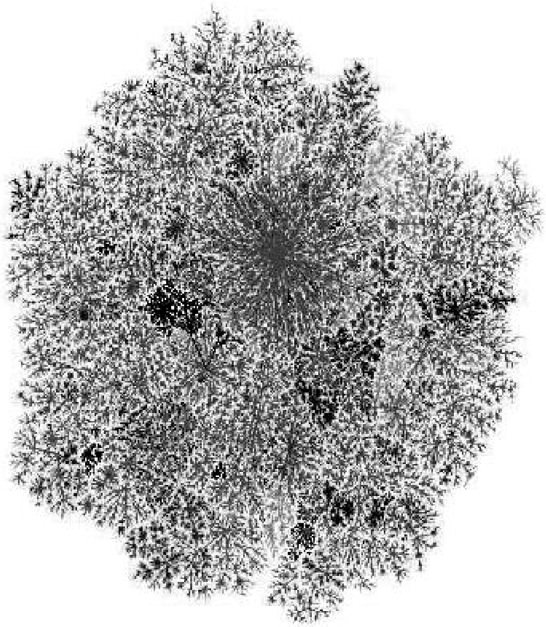

Figure 1. Map of the Internet.

phase-space attractor generated from the sampled time-series. In particular, we will examine the dimension of the attractor and the global Lyapunov exponents of its dynamics to demonstrate that the network response times are indeed chaotic. We will then turn to the question of how to pragmatically predict the performance of Internet traffic through application of a feedfoward perceptron neural network, and demonstrate that the inherent nonlinear processing capability of such networks is amenable to estimating the Internet response times.

2. Experimental data

Figure 2 details a time-series of ping times [3] from a server in the United States (Florida) to a domain in Japan (yomiura.net) for a duration of approximately eight weeks in February and March of 1999. The amplitudes are network response times in milliseconds, indicating the length of time required to propagate an information packet from the source, through the network to the destination, and back to the source, therefore providing a measure of the network packet transmission latency. The time-series displays a well-defined structure, with an obvious noise component. The upper plot shows a seven-day period starting with a Sunday. It is seen that the dominant signal energy follows a 24 hour period with the maximum amplitudes corresponding to business days in

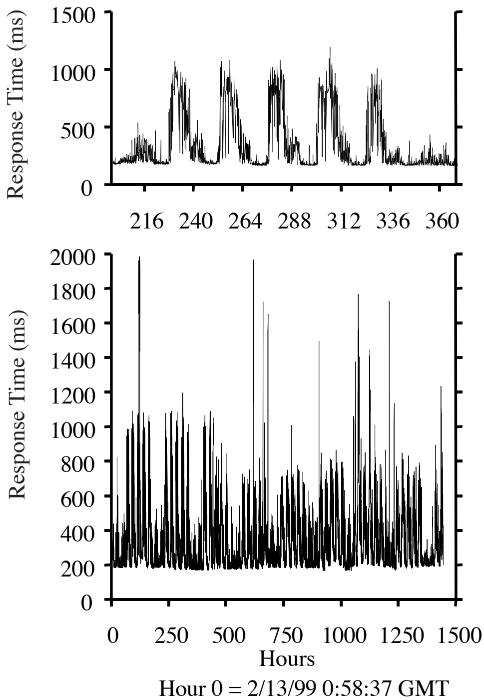

Figure 2. Internet packet transmission response times (latency).

North America and Western Europe. Further, it is clear that the signal constitutes a stable system, as there is a clear lack of divergence in the time-series amplitudes, indicating that the system dynamics occupy a finite state-space with closed orbital trajectories.

The ping times were recorded at an interval of five minutes for the duration of the data collection. Consistent with the expectation that the time-series is the result of observations of a chaotic system, the application of standard linear modeling techniques will not be applied. In this situation it is appropriate to examine the validity of the sampling interval in relation to the time-scales which will be utilized in the nonlinear data analysis. As indicated above, the dominant signal energy exhibits a quasiperiodic distribution with a primary period of 24 hours, and it is this diurnal variation which we seek to model. In section 5 we will subsample the time-series at a rate of one sample per 12 hours, in order to provide a demonstrable input set for the ANN predictions of Internet latency. In section 3 a phase-space reconstruction is performed from the original time-series with an embedding time-delay of approximately six hours. In comparison to the sampling period of 1/12 of an hour, the time-scales employed in the subsequent analysis are larger by a

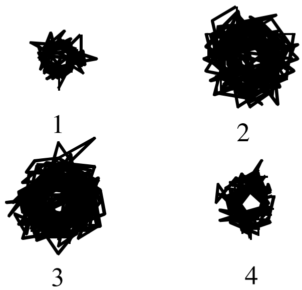

Figure 3. Two-dimensional state-space projection of Internet latency.

factor of 70. Since we do not attempt to analyze temporal components which approach the sampling interval, there is no cause for concern over observational fidelity.

The information provided by direct observation of the time-series is insufficient to determine whether the network propagation times are chaotic or deterministic, and in the event that the underlying dynamics are high-dimensional and fractal, to assess the system dimensionality. To address these questions, one can apply standard analysis techniques to extract the dimension of the time series and search for chaotic behavior. Prior to proceeding with the analysis, it is illustrative to examine an alternate geometric view of the system variance by projection of the semi-infinite time series into a closed orbital two-dimensional geometric space in order to gain a view of the temporal variance. This can be achieved through a simple mapping of the amplitude A and time t coordinates of the data in Figure 2 into a closed two-dimensional trajectory *via* the transformation: $x = A\cos(\omega t)$; $y = A\sin(\omega t)$ where ω represents the frequency of the signal period, assumed to be 24 hours. Figure 3 plots several contiguous time-slices consisting of 500 points each (roughly 41 hours).

These projections illustrate nicely the fact that the amplitude variance is a time-dependent phenomena, providing evidence that modeling of the system dynamics from a piecewise linear approach is sure to be an arduous undertaking. The compact nature of the system variations in this low-dimensional projection also provides evidence that the dynamics we seek to unravel are indeed confined to a closed system, and therefore are amenable to standard chaotic time-series analysis.

3. Dimension

An indication of higher-order complexity in the defining processes of the Internet latency would be that a large number of dimensions (system variables) are required to completely describe the temporal evolution of the time-series in a continuous differential or finite difference equation.

Further, if the dimension turns out to be noninteger then the process exhibits a fractal order and the process may exhibit chaotic behavior.

As there is not a universally accepted algorithm for establishing the dimension of an empirical data set, we will examine two methods for estimating the dimension: directly *via* false nearest neighbors (FNN), and through saturation of a system invariant. The FNN method provides a computationally efficient and robust algorithm for establishing an upper-bound on the set dimensionality, however, it does not provide noninteger estimates of dimension. Alternatively, examination of the slope of attractor interpoint correlation integrals can provide an estimate of the dimension in $\mathcal{R}$. Both of these methods rely upon access to a phase-space representation of the data at an arbitrary dimensionality; the particular level of dimensionality corresponding to a phase-space reconstruction is termed the *embedding dimension* d_E.

Here we use a time-delay embedding [4] which converts a discrete scalar time-series $s(n)$ into a d-dimensional vector time-series:

$$x(n) = [s(n), s(n+\tau), s(n+2\tau), \ldots, s(n+d\tau)]. \tag{1}$$

In order to select a reasonable time-delay τ for the embedding, one can simply compute the autocorrelation of the scalar dataset and use the value of the first zero-crossing of the autocorrelation. While this obviously cannot account for any nonlinearities in the statistical description of the interrelation between points in the time-series, it nonetheless provides a surprisingly robust initial estimate for an appropriate embedding delay as it prescribes on average the delay at which two points become statistically independent.

Figure 4 shows the autocorrelation of the time-series and indicates that a value of 70 points ($\tau = 350$ minutes) represents a delay at which the Internet latency becomes statistically independent in a linear sense. This is the value of τ that will be used in all subsequent embeddings and calculations concerning the embedded data.

3.1 False nearest neighbors estimate

The FNN method attempts to directly address the question: What embedding dimension is sufficient to eliminate false crossings of an orbit (phase space trajectory) with itself as a result of having projected the attractor into too low a dimensional space? The procedure is to define a function of nearness between adjacent points which depends solely on the geometrical arrangement imposed by the coordinate dimensions, and then iteratively increase the number of dimensions until one is satisfied that there are no "false" nearest neighbors [5]. That is, the closest neighboring point has a distance which is not an artifact of having projected the attractor into too low a dimensional phase-space.

If the distance function between the point in question $x(n)$ and it's nearest neighbor $x_{\mathrm{NN}}(n)$ is simply a euclidean distance, then the distance

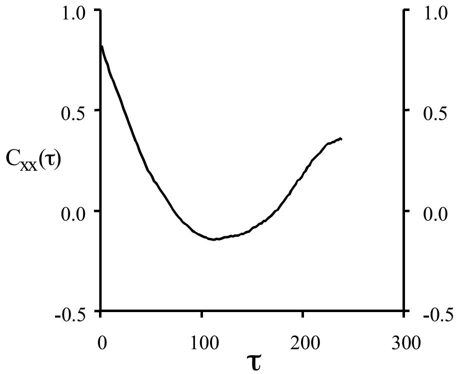

Figure 4. Autocorrelation of Internet latency.

in dimension d is:

$$D_d(n)^2 = [x(n) - x_{\rm NN}(n)]^2 + [x(n+\tau) - x_{\rm NN}(n+\tau)]^2 \\ + \cdots + [x(n+(d-1)\tau) - x_{\rm NN}(n+(d-1)\tau)]^2. \qquad (2)$$

As the data is embedded in the next higher dimension $(d+1)$, this nearest neighbor distance is changed due to the $(d+1)$ coordinates $x(n+d\tau)$ and $x_{\rm NN}(n+d\tau)$ to

$$D_{d+1}(n)^2 = D_d(n)^2 + [x(n+d\tau) - x_{\rm NN}(n+d\tau)]^2. \qquad (3)$$

If $D_{d+1}(n)$ is large, one can assume that the nearness of the two points is a result of the projection from some higher-dimensional attractor down to dimension d, since in going from dimension d to dimension $d+1$, we have unprojected these two points. One is then faced with establishing a criterion to decide when neighbors are false. A normalized distance metric can serve this purpose such that when

$$\frac{|x(n+d\tau) - x_{\rm NN}(n+d\tau)|}{D_d(n)} > D_{\rm FNN} \qquad (4)$$

the nearest neighbors at time index n are classified false. Here we use a threshold value of 25, which lies within the accepted span of $10 < D_{\rm FNN} < 50$ where the criteria is essentially constant.

Figure 5 presents the calculation of the FNN as a function of embedding dimension d_E. The results indicate that embedding the attractor of the scalar time-series into a five-dimensional phase-space will suffice to ensure that all trajectory orbits are valid representations of the attractor dynamics.

3.2 Correlation dimension estimate

Another procedure for establishing the dimension of a scalar time-series is to embed the time series into a multiple of higher dimensions [6]

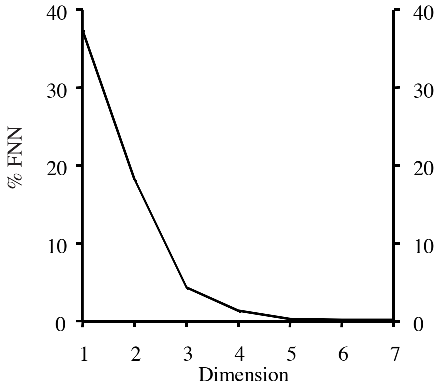

Figure 5. FNN of the Internet latency.

and then search for saturation of a system invariant as the embedding dimension increases. Such an invariant can in principle consist of any property associated with the attractor which depends on distances between points in the phase-space. A popular choice is an average over the attractor of moments of the number density. Define the number density, the number of points on the orbit within a radius of r of points y in the phase-space, as:

$$n(r,x) = \frac{1}{N}\sum_{n=1}^{N}\theta(r - |x(n) - y|) \tag{5}$$

with

$$\theta(u)\begin{cases} 0; & u < 0 \\ 1; & u > 0. \end{cases} \tag{6}$$

The average over all points of powers of $n(r,x)$ defines the well-known correlation integrals [8]:

$$C_q(r) = \frac{1}{M}\sum_{i=1}^{M}[n(r,x(i))]^{(q-1)} \tag{7}$$

which for $q = 2$ reduces to the familiar two-point correlation integral

$$C_2(r) = \frac{2}{N(N-1)}\sum_{i\neq j}^{N}\theta(r - |y(j) - y(i)|). \tag{8}$$

The correlation integral C_2 is assumed to be a geometric invariant of the system, and we compute values of C_2 as a function of embedding dimension in order to establish the dimension at which the integral saturates. Figure 6 presents computation of $\log(C_2)$ *versus* $\log(r)$ for d_E

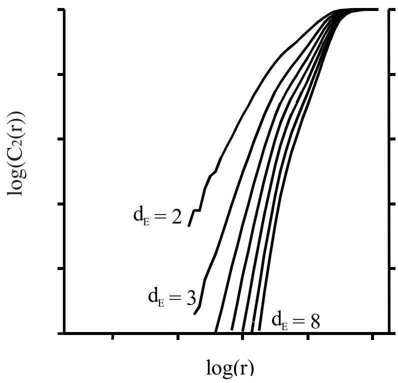

Figure 6. Correlation integrals for the Internet latency.

from 1 to 8. Examination of Figure 6 reveals that when d_E reaches 5, the slope of the curves considered over their linear extent is essentially constant, indicating that the attractor dimension is less than 5. For a more accurate assessment, the value of the slope of the $d_E = 5$ curve over the locally straight region produces a dimension estimate of 4.75. The fact that the resulting estimate is noninteger provides evidence that the attractor is indeed a fractal. It remains to be demonstrated that the attractor is a strange attractor, one with both fractal dimension and macroscopically closed orbits.

4. Lyapunov spectrum

Having established that the Internet latency times constitute a scalar representation of a process with a dimension of approximately 4.75, it is now appropriate to enquire as to whether or not the dynamics of the system can be properly described as chaotic. While the dimension is useful for characterizing the distribution of points in the phase-space, it sheds no light on the dynamics of evolving trajectories of such points. For this, one can turn to the Lyapunov exponents [7] λ_i which quantify the rate of growth of elemental subspaces in the phase-space. λ_1 relates the rate at which linear distances grow between two points on the attractor: two such points separated initially by an infinitesimal distance ϵ will, on average, have their separation grow as $\epsilon e^{\lambda_1 t}$. The sum $\lambda_1 + \lambda_2$ dictates the average growth rate of two-dimensional areas, and in general, the behavior of d-dimensional subspaces is described by the sum of the first d exponents: $\Lambda = \sum_{i=0}^{d} \lambda_i$. Since the λ_i govern the rate of attractor expansion and contraction, it is clear that for physically stable systems it not possible for Λ to be positive, this would indicate a globally unbounded behavior. The presence of positive exponents is however the hallmark of chaos. The exponential divergence of neighboring points

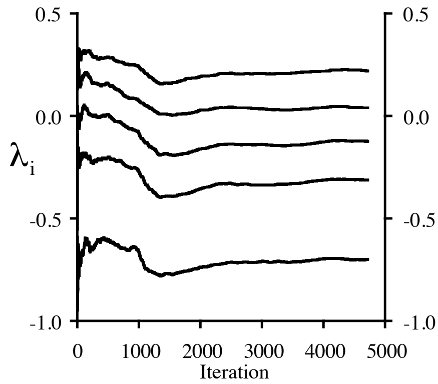

Figure 7. Lyapunov spectrum of the Internet latency.

along trajectories resulting from infinitesimal differences in initial conditions requires a positive exponent. Negative exponents arise when the trajectories of subspaces contract. If the system is a dissipative one, then Λ would be negative, ensuring that as time progresses the attractor will eventually collapse into a stable point. A hamiltonian system would exhibit $\Lambda = 0$, indicating an energy balance between the expansive and contractive dynamics. Further, one can in principle determine whether the embedded dynamics are describable as a differential equation or as finite time maps through the presence of a zero exponent.

The Lyapunov exponents are computed following the method in [9] for the data of Figure 2 embedded into a five-dimensional attractor. Figure 7 plots the Lyapunov spectrum of the embedded data and reveals both positive and negative components, as well as $\Lambda < 0$. As outlined before, the presence of positive exponents reveals the sensitive dependence on initial conditions and neighboring trajectory divergence which is associated with chaos. Further, since $\Lambda < 0$, it is known that the attractor is globally stable.

As the Lyaponov exponents provide a measure of the trajectory dynamics, a natural question to pose concerning the predictability of the nonlinear time-series is: Given a knowledge of the attractor characteristics, how far in the future can one reasonably expect to be able to make useful predictions of a particular trajectory? To assess this question we can draw on the realization that a dynamical system exhibiting chaotic behavior is an exact analog of Shannon's concept of an ergodic information source. Consider a discrete information source which produces a sequence of symbols drawn from an alphabet $s \in [S_1, S_2, \ldots, S_N]$, and denote the distribution of the realized sequence within s as $P(s)$. The amount of information that is transmitted by measurement of the sequence (or alternatively the amount of uncertainty removed at the receiver) can be quantified by an informational entropy as elucidated by Shannon in [10]. In the case where a joint set of measurements are performed, the entropy in bits is defined *via* the joint probability $P(s_1, s_2, \ldots, s_N)$ as:

$$H_N(s) = -\sum_{s_j} P(s_1, s_2, \ldots, s_N) \log[P(s_1, s_2, \ldots, s_n)]. \tag{9}$$

As N becomes large, this quantity normalized by N has a finite limit:

$$h(s) = \lim_{N \to \infty} \frac{H_N(s)}{N}. \tag{10}$$

Now consider the variable s to be the possible state-space coordinates of a dynamical system embedded within a finite-dimensional attractor. Let the measurements of this variable S_i define the distribution of attractor trajectories in the state-space. A joint measurement of the S_i then corresponds to following the trajectories of N orbits through the dynamics of the attractor. In this situation the source is essentially an abstract statement of the system producing the measurements, and the concept of entropy of a deterministic dynamical system, where no concept of probability is immediately evident, is much the same as in the information-theoretic setting. This connection was made by Kolomorogov in [11] who defined $h(s)$ as above, and this information measure is known as the Kolmogorov–Sinai (KS) entropy of the dynamical system. The KS entropy is an invariant of the system trajectories, and so is independent of the initial conditions or the specific trajectories observed. Further, it can be shown that the sum of positive Lyapunov exponents $\Lambda_+ = \sum_{i=0}^{d} \theta(\lambda_i) \cdot \lambda_i$ is equal to the KS entropy [12]. This connection allows one to make a statement regarding the limits of predictability for a nonlinear system in a local sense.

Consider an arbitrary phase-space for a stable dynamical system with sufficient dimension to completely unfold the attractor. Within this phase-space, any realizable statement regarding the system can only be specified to within a certain accuracy. Within that resolution cell we cannot distinguish between two distinct state-space points. However, any nonlinear system with a positive KS entropy has a degree of intrinsic instability, and the two points occupying the same resolution cell will, after a time T, move to disparate and individually resolvable cells in the state-space. A natural question concerning the predictability of a state-space point is then: How large can T grow before prediction of states is untenable? As a rough upper-limit, one can argue that when the number of states occupied by the system is equal to approximately the total number of states available for orbits of the system, that the ability to predict further state-evolution is lost. A measure of the total number of states that a d-dimensional system occupies after the passage of time T is $d^{h(s)T}$, therefore, a reasonable assessment then for the length of time a particular trajectory remains predictable would be $T \approx 1/h(s)$. Since $h(s) = \Lambda_+$, which in the case at hand is roughly in the range of 0.25 as indicated in Figure 7, we are drawn to the conclusion that a time interval of roughly four hours (240 minutes) defines the limit at which one could be expected to make reasonable estimates of a phase-space point evolution along a particular trajectory. It is interesting to note that this estimate for the predictability of an individual state point is less than the corresponding linear correlation estimate of statistical independence of 350 minutes, indicating that the nonlinearities of the system dynamics imposed through the Λ_+ increase the uncertainty of the evolution as compared to a linear-systems perspective.

5. Network response prediction

The question of how to best model the nonlinear dynamics of the Internet latency must be bounded by reasonable computational efforts and available resources. In the absence of any resource constraints one could be served by identifying the parameter coefficients of the d-dimensional differential equation, establishing the relevant initial conditions, and propagating the variational equation forward in time. This would be done under the provision that the presence of the positive Lyapunov exponents ensure that prediction accuracy is guaranteed to diverge exponentially as the system is propagated forward in time. Given that resource constraints are always present in prediction tasks, one must take a more resource reasonable approach, and we can turn to the inherent nonlinear estimation powers of ANNs in order to make a model-free estimate of the future latency times. As an additional motivation, one can note that the use of the variational equation with positive Lyapunov exponents as a predictor constitutes a state-evolution model tracking the microscopic morphology of individual state-points along trajectories. However, as often arises in complex systems there may exist global features of the collective dynamics which manifest in well-defined structures as a result of the negative exponents and natural tendencies to seek a basin of attraction. In this perspective, it may be appropriate to examine the global response of the dynamics independent of the individual trajectories, and seek out a structure in the overall system response. For this task, an ANN is well suited as it constitutes a model-free nonlinear estimation facility concerned with global system responses not constrained by microscopic state-point evolution.

The first question to address concerning the selection of an ANN as a candidate information processing and prediction architecture is: What type of ANN is to be used? Generally, there are two gross architectures: recurrent or feedforward networks, which refers to the information flow through the operational network as either incorporating feedback loops or sequential processing. Each of which can further be grouped into two classes of organizational algorithms: supervised or unsupervised, specifying whether or not an external error monitor is used to direct the training. The two most popular implementations are the perceptron, a supervised feedforward network, and the Kohonen, an unsupervised recurrent implementation. The former is typically employed in nonlinear estimation where a well-defined training set is available, while the latter is called to service in pattern classification tasks. In the task at hand, prediction of Internet latency times for which we have a sizable empirical data set, the perceptron is the natural choice and the one which we shall invoke to estimate future latency.

Having identified the ANN paradigm, the next issue to resolve is the exact configuration of the network in terms of the numbers of processing units, layers, connectivity, and algorithms. The number of inputs and outputs are determined by the available training data, and is derived in section 5.1.

5.1 Training data

As indicated earlier, the ANN is implemented to predict the future network latency for stable network endpoints as shown in Figure 2, and in accordance with this the ANN will contain a single output. The inputs can in principle consist of any information relevant to the state of the Internet at a given time. This could include state-information regarding the network loading, availability of routers and switching hubs, and even specific details concerning the hardware configuration and physical characteristics of the network endpoints and their interrelations. It would be expected that as the relevance of the input data increases in relation to the information it conveys regarding the Internet latency, that the training time and ultimate prediction fidelity of the network will improve. However, concerning the training-set available in Figure 2, the date-time are the only independent data available. An examination of Figure 2 indicates a strong coupling between the time-of-day, day-of-the-week and the resultant Internet latency, wherein there is a clear diurnal variation of increased latency centered on standard business days conducted in the continental United States time-zones. It is therefore consistent to expect that the three inputs consisting of: day-of-the-week, time-of-day, and business or holiday, constitute three parameters which have a direct bearing on the state of the Internet latency. Additionally the total time of the time-series provides a unique input which may be a relevant feature. Accordingly, for the purpose of training the network and subsequently making predictions, the ANN is configured to have four inputs corresponding to normalized $[0, 1]$ versions of these four parameters.

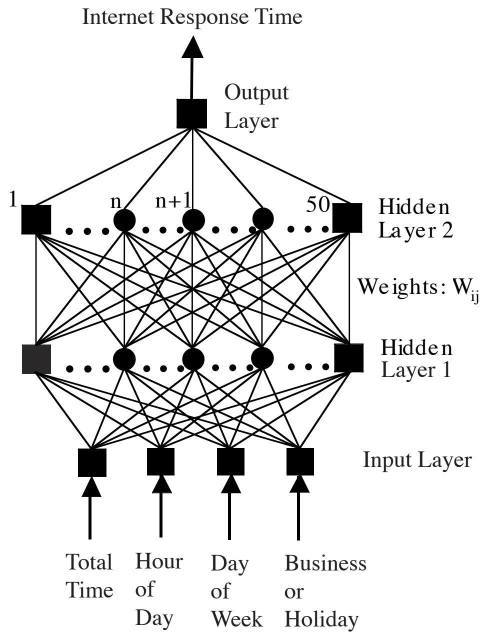

Figure 8. Perceptron ANN employed for Internet latency prediction.

5.2 Network architecture

Having identified the ANN input and output data in section 5.1, the basic input/output requirements of the ANN have been defined. The ANN therefore accommodates four input units and a single output. Further, the network is configured with two hidden layers to perform the nonlinear information transformation, with 50 neurons in each layer. Figure 8 depicts a schematic representation of the ANN, and emphasizes the full interconnection structure of the processing units (neurons) between subsequent layers.

As the perceptron ANN is a supervised paradigm, it requires a training phase prior to use as a predictive algorithm in order to organize the interconnection weight-states into a structure through which the relevant information extraction or recognition can be achieved. The process is essentially a search through the parametric information space encapsulated in the training data so as to produce a minimum error output dictated by the supervisor. In the present case, the supervisory function consists of the classic parametric euclidian error: $\epsilon = (T_i - O_i)$, where T indicates the supervisory output of the ith output neuron, and O the actual output of the ith unit during a training cycle. It is certainly possible to use other forms of supervisory feedback, and in fact there exists a substantial body of work illustrating the advantages of employing error functions based on the discrimination of information between the target and training output sets [13–15]. However, here we are concerned with demonstrating the predictive abilities of the perceptron, and as long as the training process presents no substantial difficulties the use of the euclidean error function is acceptable.

The activation function for the hidden layer neurons is a sigmoidal Bernoulli function with a zero-argument slope of 3.0, while the output layer units employ a simple linear scaling with a unit derivative. The ANN is trained with the backpropagation [16] gradient-descent algorithm which uses a fixed momentum parameter of $\lambda = 0.1$, and a learning rate of $\eta = 0.0035$. Details of the ANN internal computational algorithms can be found in [17].

5.3 Training results and prediction

The ANN was trained to predict the Internet latency times for two time-frames: the first consisting of a period of seven days, the second for a period of 14 days. The training sets consisted of sections of time-series extracted from the data of Figure 2, subsampled at a rate of once per 12 hours, therefore two state-points per-day are available for training. The input training data consisted of the hour, day, and business/holiday information corresponding to the center of a 12 hour period. The output training values consisted of an eight-hour moving-average of the sampled Internet latency values centered on the 12 hour period. The training sets were extracted starting at time index of hour 520 (Figure 2), and for the seven day time-frame employed seven days worth of training data, and likewise 14 days of training data for the 14 day time-frame.

Figure 9 plots the temporal evolution of the normalized supervisory error function during the training process for the seven day training set. It is observed that the network converges to a stable organizational state after approximately 400 iterations. Once the network has converged to a stable error-minimum, it can be used as a predictor for an unknown input set. The trained network corresponding to the seven day input data set was presented input data from time index hour 20, and the corresponding normalized network predictions are shown with the target values in Figure 10. The ANN performed a reasonable prediction for all days of the week except Wednesday, for which there appeared a substantial spike in the Internet latency about halfway through the day as is clearly evidenced in Figure 2. Nonetheless, even though the ANN was incapable of predicting this anomalous microfeature, which is not surprising considering the amplitude averaged limited training set to which it was exposed, it did faithfully predict the relative amplitudes of the Internet latency from one 12 hour period to the next.

The next task was to train the ANN for an attempt at predicting the Internet latency for a two week period. Figure 11 depicts the training

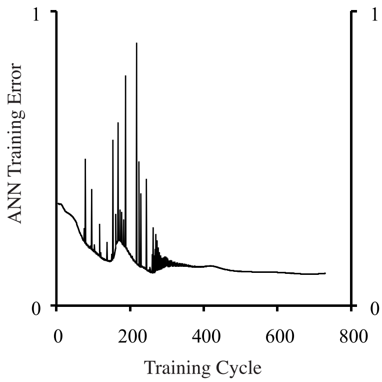

Figure 9. ANN training evolution for the seven day prediction.

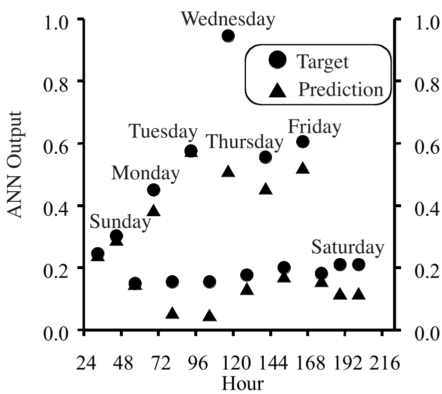

Figure 10. ANN Internet latency predictions for seven days.

curve for the 14 day data set, from which it is apparent that a considerable increase in organizational effort is required in relation to the seven day data set. The network reaches a stable state after approximately 6000 iterations, and is then used to predict latency for a 14 day period. The trained network is presented the input data starting at time index hour 20 from Figure 2, and the resulting network output is presented in Figure 12. It is evident that the protracted ANN organizational effort and increased training set diversity have resulted in a better network prediction than the seven day case, while it must noted that the prediction for the first Wednesday still exhibits the largest error.

Regarding the practical application of such predictions, the implemented ANN architecture provides a predictive filter of future Internet latency times based on a minimal input data set. In accordance with this minimally informative input set, the output predictions are not well suited to predicting anomalous events which might arise due to Internet

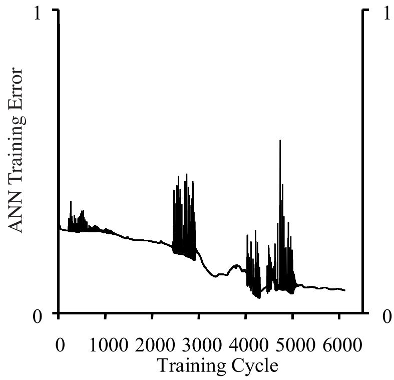

Figure 11. ANN training evolution for the 14 day prediction.

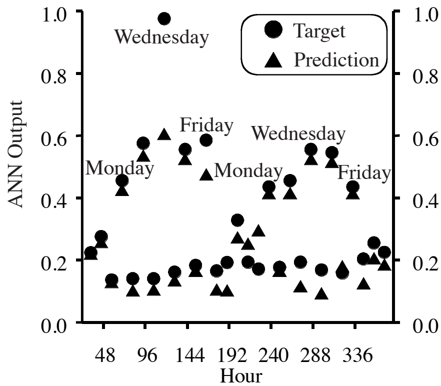

Figure 12. ANN Internet latency predictions for 14 days.

hardware performance variations and outages. To address this concern, it is possible to extend the network inputs to incorporate additional information relevant to the network delays such as router loads and availability, assuming that one would have available *a priori* knowledge of planned hardware outages. Another variable which could be tailored to short-term predictions would be utilization of a path probe ping response in the network inputs. The presence of such real-time network feedback could provide information pertinent to the short-term network latency and would serve to adjust the ANN predictions for rapid temporal responses. Given the myriad variations of input/output combinations and timescales for which the ANN can be configured, it is clear that alternative implementations can provide results applicable to the particular problem at hand, here we have demonstrated the plausibility and validity of a simple ANN implementation.

6. Conclusion

The explosive growth of the global Internet, driven by economic and local-use resource requirements, has resulted in a complex system currently consisting of roughly 10^5 interconnected networks, supporting an approximate 2×10^8 users able to access 3×10^9 unique information pages. Not surprisingly, the interconnection hierarchy across the global network for a packet of information transmitted between two endpoints is a time-varying fractal architecture. When one examines the statistical nature of a time-series of information packets transmitted through the network, it is revealed that the time-series constitutes a chaotic signal, indicating a fractal dimension to the underlying dynamical attractor. Further, the attractor state-space orbits exhibit positive Lyapunov exponents guaranteeing eventual divergence in the predictability of individual system variables. Notwithstanding the microscopic unpredictability of local system trajectories, there exists a well-defined and stable structure to the global system dynamics, as evidenced by the quasiperiodic appearance of the time-series depicted in Figure 2. One can then favorably pose the question as to whether or not an information processing structure can be configured to recognize and predict the macroscopic behavior of this dynamical evolution. Most conventional structures aimed at doing so are based on a statistical interpretation of the input/output data, proceeding to arrive at a best estimate for a particular input set stimulus based on a maximum-likelihood parametric estimation. More often than not, this estimation is based on a linear model, or superposition of linear models, targeted at extracting information from the statistical data. Further, such models are often constrained by limited state-information retention, constricting the available number of states amenable to prediction. To circumvent such limitations, the powerful class of computational paradigms comprising the ANNs may be applied, as they constitute a set of model-free information-processing architectures inherently suited to nonlinear information processing.

The ANN is itself a complex system, an agglomeration of nonlinear threshold-response detectors interconnected through an information-based hierarchy of weight-states. The weight-state of a trained network represents an optimum information processing structure derived from experiential refinement of desired network performance. This structure is analogous to the global interconnection hierarchy of the Internet, wherein the connection structure has grown in response to satisfaction of network information capacity, which can be viewed as an optimization of information bandwidth requirements under the constraints of economic dictums. In this light, the Internet can be viewed as generically analogous to the structure of an ANN, and both are recognized as information transmission structures evolving to fulfill an optimum functionality under the auspices of external constraints. If one accepts this association, then it is not surprising that the ANN can serve as a useful tool for predicting the complex macroscopic behavior of an intricate system such as the Internet.

In this paper, we have applied a perceptron ANN to the task of predicting Internet latency times across the global network, and have demonstrated the utility of such an ANN to predict these times averaged over 12 hour periods for durations of 14 days. This was achieved even though the Internet latency times are a chaotic time-series with positive Lyapunov exponents, which imposes severe restrictions on the local predictability of dynamical system variables. A perceptron ANN successfully predicted the macroscopic behavior of the latency dynamics, providing an alternative information processing structure to the usual maximum-likelihood prediction filters. Establishing the nature of the time-series is in itself an interesting finding, and provides a basis for further research concerning the predictability of Internet resources.

References

[1] B. Cheswick and H. Burch, "Mapping the Internet," *Computer*, **32**(4) (1999) 97.

[2] Bill Cheswick (Bell Labs) and Hal Burch (Carnegie Melon University), "Internet Mapping Project," http://cm.bell-labs.com/who/ches/map/index.html.

[3] The ping program is a standard network analysis and diagnostic tool that sends Internet Control Message Protocol (ICMP) ECHO_REQUEST datagram packets to network hosts in order to elicit an ICMP ECHO_RESPONSE from the host. ECHO_REQUEST datagrams (pings) have an 8 byte IP (Internet Protocol) and ICMP header, followed by a time data structure and an arbitrary number of pad data bytes used to fill out the packet. The default pad size is 56 bytes, which translates into 64 ICMP data bytes when combined with the 8 bytes of ICMP header data. If the data space is at least 8 bytes, ping uses the first 8 bytes of this space to include a timestamp, which it uses in the computation of round-trip times. The resulting round-trip time constitutes the amplitude of the time-series analyzed herein.

[4] R. Mane and F. Takens, "Dynamical Systems and Turbulence," in *Lecture Notes in Mathematics*, number 898, edited by D. Rand and L. S. Young (Springer-Verlag, Berlin, 1980).

[5] M. B. Kennel, R. Brown, and D. I. Abarbanel, "Determining Embedding Dimension for Phase-space Reconstruction using a Geometrical Construction," *Phys. Rev. A*, **45** (1992) 3403.

[6] H. Abarbanel, *et al.*, "The Analysis of Observed Chaotic Data in Physical Systems," *Rev. Mod. Phys.*, **65**(4) (1993) 1331.

[7] V. I. Oselecdec, "A Multiplicative Ergodic Theorem, Lyapunov Characteristic Numbers for Dynamical Systems," *Trudy. Mosk. Mat. Obsc. [Moscow Math. Soc.]*, **19** (1968) 197.

[8] P. Grassberger and I. Procaccia, "Characterization of Strange Attractors," *Phys. Rev. Lett.*, **50** (1983) 346.

[9] J. P. Eckmann and D. Ruelle, "Ergodic Theory of Chaos and Strange Attractors," *Rev. Mod. Phys.*, **57** (1985) 617.

[10] C. Shannon, "A Mathematical Theory of Communication," *Bell Systems Technical Journal*, **27** (1948) 623.

[11] A. N. Kolomorgorov, *Dokl. Akad. Nauk. SSSR*, **119** (1958) p861; *Math. Rev.*, **21** (1960) 386.

[12] Y. B. Pesin, "Lyapunov Characteristic Exponents and Smooth Ergodic Theory," *Usp. Mat. Nauk. [Russian Math. Survey]*, **32** (1977) 55.

[13] J. C. Park, P. S. Neelakanta, *et al.*, "Information-theoretic Based Error-metrics for Gradient Descent Learning in Neural Networks," *Complex Systems*, **9** (1995) 287.

[14] J. C. Park and S. Abusalah, "Symmetrization of Information-theoretic Error-measures Applied to Artificial Neural Network Training," *Complex Systems*, **11** (1997) 125.

[15] *Information Theoretic Aspects of Neural Networks*, edited by P. S. Neelakanta (CRC Press, Boca Raton, FL, 1999).

[16] P. D. Wasserman, *Neural Computing* (Van Nostrand Reinhold, New York, 1989).

[17] J. C. Park and S. Abusalah, "Maximum Entropy: A Special Case of Minimum Cross-entropy Applied to Nonlinear Estimation by an Artificial Neural Network," *Complex Systems*, **11** (1997) 289.