Vol.2009-MPS-74 No.5
2009/7/13

Bayesian Forecasting of WWW Traffic on the Time Varying Poisson Model

Daiki Koizumi,[†1] Toshiyasu Matsushima[†2] and Shigeichi Hirasawa[†1]

Traffic forecasting from past observed traffic data with small calculation complexity has been one of important problems for planning of servers and networks. Focusing on World Wide Web (WWW) traffic as fundamental investigation, this paper would deal with Bayesian forecasting of network traffic on the time varying Poisson model from a viewpoint from statistical decision theory. Under this model, we would show that the forecasting estimate is obtained by simple arithmetic calculation with a known constant of time varying degree parameter and expresses real WWW traffic well from both theoretical and empirical points of view.

1. Introduction

Under network environment such as Internet, planning of servers and networks has been one of important problems for stable operation. For example, World Wide Web (WWW) server administrators have often made their operation plans by combination of their experience and intuition with log analysis tools[1)8)]. In this case, traffic forecasting rule is not clearly formulated and those summarized logs remain in the field of *descriptive statistics* from the statistical point of view.

On the other hand, researchers in the field of traffic engineering have been suggesting a lot of analysis models. Some of desirable conditions of traffic models are to express non-stationarity[5)3)], long-range dependence (LRD)[4)5)], and self-similarity[4)] of network traffic. For these characteristics, the point estimation of parameter is often performed at first[6)], then the point estimator is plugged into the parameter of model if the traffic forecasting is needed. This approach has been wide-spread in the field of *inferential statistics* from the statistical point of view.

However, substituting the estimated parameter as a constant for the model's parameter is not always suitable for forecasting problems. This is because there is often no guarantee that the assumptions under the parameter estimation of the model always hold for future unknown data set. Bayesian approach in terms of *statistical decision theory*[2)] , which assumes the prior distribution of parameter, is one of alternatives for this point.

This paper would deal with Bayesian forecasting of WWW traffic on the non-stationary i.e. time varying Poisson model. In this model, a random-walking type of transformation function of parameter is clearly defined to obtain the Bayes optimal prediction for WWW traffic. Another feature in the model is that the traffic forecasting value is obtained by simple arithmetic calculations with known constant $k\,(0 < k \leq 1)$ where k denotes time varying degree of parameter. Then the effectiveness of our approach would be evaluated with real WWW traffic data.

The rest of this paper is organized as the followings. Section 2 gives some definitions and explanations of the forecasting model with time varying Poisson distribution. Section 3 shows some analysis examples of real WWW traffic data to validate this paper's approach and Section 4 gives their discussions. Finally, Section 5 concludes this paper.

2. The Time Varying Poisson Model

2.1 Definitions

Let $t = 1, 2, \cdots$ be discrete time and $x_t = 0, 1, \cdots$ be number of WWW request arrivals at time t, respectively. This paper focuses on x_t for traffic analysis by assuming probability distribution $p(x_t \mid \theta_t)$ where $\theta_t > 0$ is a time varying density parameter at time t. On the time varying Poisson model, a sequence $x_1^t = x_1 x_2 \cdots x_t$ is taken as input and $\hat{x}_{t+1}$ is calculated as an output estimator where the prior distribution of parameter $p(\theta_t)$ and time variation rule of θ_t are known. The overview of the inferential process is depicted in **Fig.1**.

In Fig.1, x_t is assumed to be the Poisson distribution with a time varying density parameter θ_t as follows:

†1 Research Institute for Science and Engineering, Waseda University
†2 Department of Applied Mathematics, School of Fundamental Science and Engineering, Waseda University

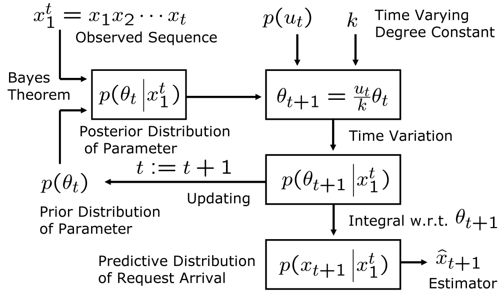

Fig. 1 The overview of inferential process.

For $x_t = 0, 1, 2, \cdots$,

$$p(x_t \mid \theta_t) = \frac{\exp(-\theta_t)}{x_t!} (\theta_t)^{x_t}, \tag{1}$$

where $\theta_t > 0$ is a time varying density parameter.

For density parameter θ_t, the following time varying model is assumed:

For $\theta_{t+1}, \theta_t > 0$,

$$\theta_{t+1} = \frac{u_t}{k}\theta_t, \tag{2}$$

where k is a constant such that $0 < k \leq 1$, and $0 < u_t < 1$ is a continuous random variable which is independent from θ_t.

(2) represents a transformation of θ_{t+1} from both θ_t and u_t under a known constant k. This transformation can be regarded as a kind of random-walk.

Furthermore, θ_t and u_t are assumed to be the following Gamma and Beta distributions, respectively:

$$\begin{cases} \theta_t \sim Ga(\alpha_t, \beta_t)\,; \\ u_t \sim Be(k\alpha_t, (1-k)\alpha_t)\,, \end{cases} \tag{3}$$

where α_t is the shape parameter, β_t is the scale parameter of Gamma distribution, and $k\alpha_t, (1-k)\alpha_t$ are also the shape and scale parameters of Beta distribution.

Finally, the initial condition of parameters in (3) at $t = 1$ is defined as follows:

$$\begin{cases} \alpha_1 = x_1\,; \\ \beta_1 = 1\,. \end{cases} \tag{4}$$

Remarks 2.1 In (2), a constant $0 < k \leq 1$ expresses time varying degree of θ_t. If $k = 1$, (2) simply becomes $\theta_{t+1} = u_t\theta_t$. In this case, θ_{t+1} does not vary since the variance of u_t , which equals to $k(1-k)/(\alpha_t+1)$ according to the nature of Beta distribution, becomes zero. This means that the Poisson distribution of x_t in (1) is stationary.

If $k < 1$, on the other hand, θ_{t+1} varies depending on the previous θ_t which expresses the time varying Poisson model. If $k = 0.5$, the time varying degree of θ_t becomes maximum since the variance of u_t takes the maximum value. Thus the proposed model defined in (1)–(3) includes a classical stationary Poisson distribution as a special case if $k = 1$.

Remarks 2.2 The initial condition defined in (4) corresponds to the following prior distribution in Bayesian context:

$$p(\theta_1) = \frac{1}{\theta_1}\,. \tag{5}$$

This is called *non-informative prior*[2)] which assumes no anomalies for objective traffic.

2.2 Updating Rule of Parameter θ_t

2.2.1 The Posterior Distribution of Parameter

As defined in (3), the prior distribution of parameter $p(\theta_t)$ is the Gamma distribution with parameter α_t, β_t. If new x_t is observed, the Bayes theorem gives the following posterior distribution:

$$p\left(\theta_t \mid x_1^t\right) = \frac{(\beta_t+1)^{\alpha_t+x_t}}{\Gamma(\alpha_t+x_t)} \exp[-\theta_t(\beta_t+1)](\theta_t)^{(\alpha_t+x_t)-1}\,, \tag{6}$$

where $t \geq 2$. In the denominator on the right side of (6), $\Gamma(x)$ is the Gamma function defined below:

$$\Gamma(x) = \int_0^\infty y^{x-1} \exp(-y)\, dy\,, \tag{7}$$

where $x > 0$.

(6) means that the posterior distribution $p\left(\theta_t \mid x_1^t\right)$ is also Gamma, which is

Vol.2009-MPS-74 No.5
2009/7/13

same distribution as (3), with parameter $(\alpha_t + x_t)$ and $(\beta_t + 1)$. This is because Gamma distribution is *the conjugate prior*[2] of the sample distribution of Poisson.

2.2.2 Time Variation of Parameter

To obtain $p\left(\theta_{t+1} \mid x_1^t\right)$, a time variation of density parameter defined in (2) is used. This is actually a transformation of random variables among u_t, θ_t, and constant k and its updating rule is obtained as follows:

$$p\left(\theta_{t+1} \,\middle|\, x_1^t\right) = \frac{\left[k\left(\beta_t+1\right)\right]^{k(\alpha_t+x_t)}}{\Gamma\left[k\left(\alpha_t+x_t\right)\right]} \exp[-\theta_{t+1}k\left(\beta_t+1\right)](\theta_{t+1})^{k(\alpha_t+x_t)-1}. \quad (8)$$

(8) means that the transformed distribution of θ_{t+1} becomes the Gamma distribution with the following parameters:

$$\begin{cases} \alpha_{t+1} = k\left(\alpha_t + x_t\right) ; \\ \beta_{t+1} = k\left(\beta_t + 1\right) . \end{cases} \quad (9)$$

If (9) is recursively applied with respect to t, the following equations are obtained:

$$\begin{cases} \alpha_{t+1} = k^t\alpha_1 + \sum_{i=1}^{t} k^{t+1-i}x_i ; \\ \beta_{t+1} = k^t\beta_1 + \sum_{i=1}^{t} k^{i-1} . \end{cases} \quad (10)$$

The above equations contribute drastic reduction of calculation complexity.

Remarks 2.3 Even if a transformation is newly defined after the posterior distribution in (6), the distribution family of $p(\theta_{t+1}|x_1^t)$ remains same as that of the conjugate prior distribution under certain class of transformations. Such class has been discussed under Simple Power Steady Model (S.P.S.M.)[7]. Therefore the time varying Poisson model in this paper with transformation function defined in (2) is included in S.P.S.M.

2.3 Output Estimator $\hat{x}_{t+1}$

In Fig.1, $\hat{x}_{t+1}$ is a prediction of number of request arrivals at time $(t+1)$. From the point of *statistical decision theory*[2], *the Bayes optimal prediction*[2] under the squared-error loss $L(\hat{x}_t, x_t) = (\hat{x}_t - x_t)^2$ is obtained as follows:

$$\hat{x}_{t+1} = \frac{k^t\alpha_1 + \sum_{i=1}^{t} k^{t+1-i}x_i}{k^t\beta_1 + \sum_{i=1}^{t} k^{i-1}} \quad ,0 < k \leq 1. \quad (11)$$

Remarks 2.4 In (11), $\hat{x}_{t+1}$ is obtained by simple arithmetic calculation. This point can be effective not only theoretical point of view but also the real implementation such as server log analysis software tools. The second term of numerator in (11) has a form of *Exponentially Weighted Moving Average*[7] with a time varying constant k. As k becomes larger in (11), the weighting of past observed sequence x_1^t increases. This means that k can be considered as a parameter of long-range dependence (LRD)[4][5].

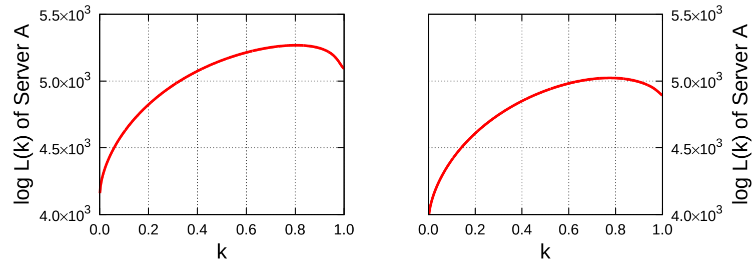

Fig. 2 Examples of log-likelihood functions. Each maximum likelihood estimator on the left is $\hat{k} = 0.804$, on the right is $\hat{k} = 0.775$, respectively.

3. Analysis Examples of WWW Traffic Data

3.1 Maximum Likelihood Estimation (MLE) for k

If real data is dealt with, k should be estimated. Taking the maximum likelihood estimation (MLE) of k, the objective likelihood function $L(k)$ becomes the following:

$$L(k) = p\left(x_1 \mid \theta_1\right)\prod_{i=2}^{t}\left[\frac{(\beta_i)^{\alpha_i}\Gamma(\alpha_i + x_i)}{(\beta_i+1)^{\alpha_i+x_i}\Gamma(\alpha_i)x_i!}\,\middle|\, \begin{array}{l} \alpha_i = k^{i-1}\alpha_1 + \sum_{j=1}^{i-1} k^{i-j}x_j \\ \beta_i = k^{i-1}\beta_1 + \sum_{j=1}^{i-1} k^{j-1} \end{array}\right]. \quad (12)$$

Note that α_i, β_i in (12), the previously obtained results in (10) should be applied.

Some plots of function $\log L(k)$ are shown in **Fig.2**. In (12), the solution can be obtained by numerical calculation. The interval $0 \leq k \leq 1$ is divided into 1,000 sub-intervals and value of $\log L(k)$ is calculated for each k to obtain the MLE solution numerically.

3.2 Point Estimation for WWW Traffic Forecasting

The real WWW data (from Server A on campus) was processed to evaluate the proposed model. For the performance comparison, the point estimates on the classical stationary Poisson model were also calculated. On the proposed model, each MLE of $\hat{k}$ was calculated from the previous day's log. To evaluate performance on both models, the mean squared error between each point estimate and observed value of request arrivals was calculated.

Table 1 shows mean squared error of proposed and stationary models. In the third row of Table 1, the MLEs of $\hat{k}$ from the previous day's logs are shown. **Fig.3** shows point and interval estimates v.s. observed values plot of server A on Mar. 25, 2005 where $\hat{k} = 0.804$. In Fig.3, the vertical axis is the number of request arrivals and the horizontal axis is time interval index t. The solid line, dotted line, and histogram represent the point estimates, the interval estimates (95% confidence limit) on the proposed and classical stationary Poisson model, and observed values of request arrivals, respectively.

Table 1 Mean squared error on server A.

Server A	Proposed Model	$\hat{k}$	Stationary Model
Mar. 19	2.829×10^2	0.716	4.988×10^2
Mar. 20	2.657×10^2	0.762	3.297×10^2
Mar. 21	3.816×10^2	0.753	4.802×10^2
Mar. 22	7.111×10^2	0.805	9.534×10^2
Mar. 23	8.202×10^2	0.759	1.335×10^3
Mar. 24	1.356×10^3	0.804	3.458×10^3
Mar. 25	9.523×10^2	0.804	2.062×10^3
Mar. 26	4.811×10^2	0.783	6.479×10^2
Mar. 27	8.596×10^2	0.771	1.239×10^3
Mar. 28	1.980×10^3	0.754	4.041×10^3
Mar. 29	1.657×10^3	0.777	4.019×10^3
Mar. 30	4.940×10^2	0.788	7.568×10^2
Mar. 31	8.088×10^2	0.787	1.218×10^3
Apr. 01	8.967×10^2	0.775	1.887×10^3
Apr. 02	1.258×10^1	0.753	1.220×10^1
Apr. 03	4.184×10^0	0.826	4.375×10^0
Apr. 04	4.206×10^1	0.914	4.317×10^1
Apr. 05	4.095×10^1	0.666	3.808×10^1
Apr. 06	3.612×10^1	0.710	4.723×10^1
Apr. 07	2.813×10^2	0.661	5.183×10^2
Apr. 08	2.295×10^1	0.786	2.325×10^1
Apr. 09	7.589×10^0	0.803	8.127×10^0

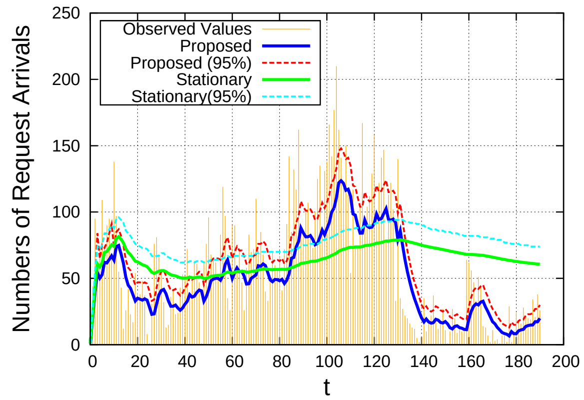

Fig. 3 Point and interval estimates v.s. observed values plot of server A on Mar. 25, 2005 ($\hat{k} = 0.804$).

3.3 Interval Estimation for WWW Traffic Forecasting

Table 2 shows interval estimation example of server A on Mar. 25, 2005. In Table 2, $t = 104$ is taken since it gives max observed value $x_{104} = 210$ as the numbers of request arrivals. The second, third, and forth rows show the expected value, 95% confidence limit, and 99% confidence limit, respectively on the proposed and stationary models.

Table 2 Interval estimation of server A on Mar. 25, 2005.

$t = 104$ on Server A	Proposed	Stationary
$\hat{x}_{104}$ (Expected Value)	111	69
$\hat{x}_{104}$ (95% Confidence Limit)	129	83
$\hat{x}_{104}$ (99% Confidence Limit)	136	89
x_{104} (Max Observed Value)	210	

4. Discussion

4.1 Maximum Likelihood Estimation for k

According to Fig.2, it is showed that there exist some cases where their likelihood functions of $\log L(k)$ are convex. Actually, all likelihood functions were convex to the best of numerical calculations in this paper. Fig.2 also shows that the absolute value of gradient in $\log L(k)$ around MLE of $\hat{k}$ becomes quickly larger as k increases beyond $\hat{k}$. This fact shows that the under estimation for k causes less error than its over estimation.

Table 3 shows Akaike Information Criterion (AIC) on server A. According to Table 3, most of AIC on the proposed model are smaller than those of stationary model to select the proposed model. Exception is that absolute values of AIC on the two models suddenly become smaller around Apr. 01, 2005 on server A. This is actually because the average traffic on server A drastically decreased to less than 1,000 request arrivals per a day. In this period, the traffic in each time interval stayed at lower level and could be regarded as the stationary Poisson model. As a whole, it can be concluded that the proposed model has stronger validity for real WWW traffic data than the stationary model from the viewpoint of model selection.

4.2 Point Estimation for WWW Traffic Forecasting

For point estimation of future request arrivals, Table 1 shows that the proposed model has the better performance than that of the stationary model in terms of mean squared error. Fig.3 also depicts that the point estimates on the proposed model are following more closely to the observed values than those on stationary model. As mentioned in **Remarks 2.1**, the proposed model contains the stationary model as a special case when $k = 1.000$. Therefore, regardless of stationarity or non-stationarity of WWW traffic, the proposed model can be applied to traffic forecasting and would help WWW server setting and network planning etc. among administrators.

However, it should be noted that this result strongly depends on the MLE performance of $\hat{k}$. In Table 1, each MLE of k during days in April often differs from $\hat{k} = 1.000$ in spite of stationarity on the real traffic. In such situation, the mean squared error on the proposed model becomes larger than that of stationary model. Another example is that if $k = 0.300$ on the proposed model with server A, the mean squared error of the proposed model becomes 5.41×10^3 where that of stationary model becomes 3.46×10^3. **Fig.4** depicts this poor performance of the proposed model. **Fig.5** is a plot of mean squared errors v.s. k. In interval of $k < 0.800$, the extremely smaller estimate of k could causes larger mean squared error. The under estimation near MLE of k, however, causes relatively smaller error than the over estimation as previously described in subsection 4.1. In Fig.5, mean squared errors takes minimum values around $k = 0.800$. In fact, Table 1 shows that corresponding maximum likelihood estimate is $k = 0.783$. This result suggests that data length in this simulation was sufficient for the maximum likelihood estimation for k.

4.3 Interval Estimation for WWW Traffic Forecasting

For interval estimation, time interval indices that give the maximum number of request arrivals are taken on Table 2. This is because one of administrators' concerns can be the maximum number of the request arrivals in terms of stable server operations. As a result, confidence limits of x_t derive larger values than those of point estimates of x_t (=expected values) and reduce the mean squared errors than point estimates. This effect on the proposed model would be stronger than that on the stationary model, since the performance of point estimates of x_t on the proposed model is superior to that of stationary model. Thus the advantage of Bayesian approach was observed.

Table 3 Akaike Information Criterion (AIC) on Server A.

AIC on Server A	Proposed	Stationary
Mar. 19, 2005	-3.562×10^3	-3.367×10^3
Mar. 20, 2005	-3.102×10^3	-3.011×10^3
Mar. 21, 2005	-5.085×10^3	-4.960×10^3
Mar. 22, 2005	-8.316×10^3	-8.108×10^3
Mar. 23, 2005	-1.052×10^4	-1.018×10^4
Mar. 24, 2005	-1.797×10^4	-1.715×10^4
Mar. 25, 2005	-1.088×10^4	-1.031×10^4
Mar. 26, 2005	-3.917×10^3	-4.649×10^3
Mar. 27, 2005	-9.023×10^3	-8.723×10^3
Mar. 28, 2005	-1.875×10^4	-1.795×10^4
Mar. 29, 2005	-1.869×10^4	-1.774×10^4
Mar. 30, 2005	-5.825×10^3	-5.610×10^3
Mar. 31, 2005	-1.003×10^4	-9.784×10^3
Apr. 01, 2005	-6.456×10^3	-5.783×10^3
Apr. 02, 2005	-1.538×10^1	-2.313×10^1
Apr. 03, 2005	$+1.763 \times 10^1$	$+8.860 \times 10^0$
Apr. 04, 2005	-1.655×10^2	-1.615×10^2
Apr. 05, 2005	-1.821×10^2	-1.803×10^2
Apr. 06, 2005	-1.771×10^2	-1.516×10^2
Apr. 07, 2005	-2.826×10^3	-2.675×10^3
Apr. 08, 2005	-1.170×10^2	-1.184×10^2
Apr. 09, 2005	$+6.273 \times 10^0$	$+4.767 \times 10^{-1}$

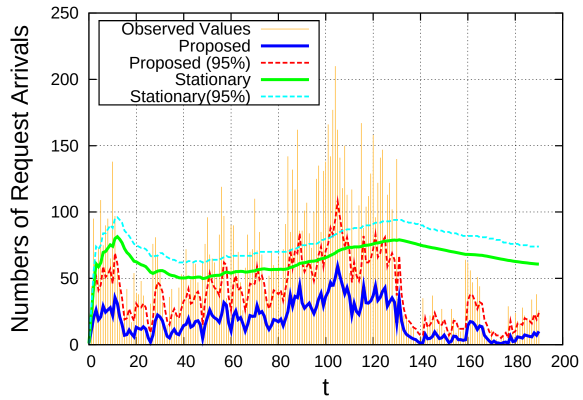

Fig. 4 Point and interval estimates v.s. observed values plot of server A on Mar. 25, 2005 ($k = 0.300$).

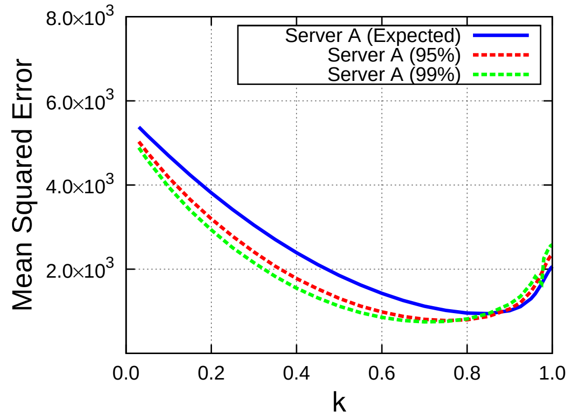

Fig. 5 Mean squared error v.s. k plot of server A on Mar. 25, 2005.

5. Conclusion

This paper showed Bayesian forecasting of WWW traffic on the time varying Poisson model. This model is obtained by defining a random-walk type of time varying parameter function on Simple Power Steady Model (S.P.S.M.). The forecasting estimator of this model guarantees the Bayes optimality in terms of statistical decision theory and is calculated by simple arithmetic calculation. The latter point especially can be effective for the real implementation such as server log analysis software tools.

Furthermore, the non-stationarity of traffic is expressed by a time varying degree constant k in the model. This paper pointed out that the constant k can be considered as a parameter of long-range dependent (LRD) for real traffic data and the model includes stationary Poisson model as a special case if $k = 1$.

For evaluation of Bayesian approach, the real WWW traffic data is applied to the model in this paper. The maximum likelihood estimation (MLE) method of k from real traffic data with sufficient length is also discussed. According to its result, the proposed model has stronger validity than classical stationary Poisson model in terms of model selection. Furthermore, under the estimated value of k, the point and interval estimates on the proposed model showed smaller mean squared error comparing to those on the stationary model for the traffic forecasting. Thus the advantage of the proposed model is shown from both theoretical and empirical points of view.

References

1) Bradford L. Barrett, The Webalizer. [Online]. http://www.webalizer.org/
2) Jose M. Bernardo and Adrian F. M. Smith, Bayesian Theory, John Wiley & Sons, Chichester, 2003.
3) T. Karagiannis, M. Molle, M. Faloutsos, A. Broido, “A Nonstational Poisson View of Internet Traffic, ” Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2004), Mar. 2004.
4) W.E.Leland, M.S.Taqqu, “On the Self-Similar Nature of Ethernet Traffic,” SIGCOMM’93, Sep. 1993.
5) Vern Paxon and Sally Floyd, “Wide Area Traffic: The Failure of Poisson Modeling, ” IEEE/ACM Trans. on Networking, vol.3, no.3, Jun. 1995.
6) Antoine Scherrer, Nicolas Larrieu, Philippe Owezarski, Pierre Borgnat, Patrice Abry, “Non-Gaussian and Long Memory Statistical Characterizations for Internet Traffic with Anomalies,” IEEE Trans. on Dependable and Secure Computing, vol.4, no.1, pp.56–70, 2007.
7) J. Q. Smith, “A Generalization of the Bayesian Steady Forecasting Model, ” Journal of the Royal Statistical Society, Series B, vol.41, pp.375–387, 1979.
8) Stephen Turner, Analog. [Online]. http://www.analog.cx/