# Survey of Network Traffic Models

**Balakrishnan Chandrasekaran,** bchandrasekaran@wustl.edu

**Abstract**
The need for communication networks capable of providing an ever increasing spectrum of services calls for efficient techniques for the analysis, monitoring, evaluation and design of the networks. Analysis are perpetually faced with incomplete & ever increasing user demands and uncertainty about the evolution of the network systems. To meet the requirements of users and to provide guarantees on reliability & affordability, system models must be developed that can capture the characteristics of the actual network load and yield acceptable precise predictions of performance of the system, in a reasonable amount of time. Traffic Analysis is a vital component to understand the requirements and capabilities of a network. The past years have seen innumerous traffic models proposed for understanding and analyzing the traffic characteristics of networks. Nevertheless, there is no single traffic model that can efficiently capture the traffic characteristics of all types of networks, under every possible circumstance. Consequently, the study of traffic models to understand the features of the models and identify eventually the best traffic model, for a concerned environment has become a crucial and lucrative task. Good traffic modeling is also a basic requirement for accurate capacity planning. This report attempts to provide an overview of some of the widely used network traffic models, highlighting the core features of the model and traffic characteristics they capture best.

## Table of Contents

## 1. Introduction

An accurate estimation of network performance is vital for the success of a network of any kind. Networks, whether voice or data, are designed around many different variables. Two of the most important factors that you need to consider in network design are service and cost. Service is essential for maintaining customer satisfaction. Cost is always a factor in maintaining profitability. One way that you can factor in some of the

service and cost elements in network design is to optimize circuit utilization.

Also to a large extent, the success of a network depends on the development of effective congestion control techniques that allow for optimal utilization of a network's capacity. Performance modeling is necessary for deciding the type of congestion control policies to be implemented. Performance models in turn, require very accurate traffic models that have the ability to capture the statistical characteristics of the actual traffic on the network.

If the underlying traffic models do not efficiently capture the characteristics of the actual traffic, the result may be the under-estimation or over-estimation of the performance of the network. This would totally impair the design of the network. Traffic Models are hence, a core component of any the performance evaluation of networks and they need to be very accurate. Depending upon the type of network and the characteristics of the traffic on the network, a traffic model can be chosen for modeling the traffic.

Traffic models are analyzed based on the number parameters required to describe the model, tractability, parameter estimation and how best the model captures the actual traffic, referred to as 'goodness-of-fit'. In order to evaluate based on goodness-of-fit, metrics should be defined that allow for quantifying how close the model is to the actual data. The metrics defined should also be directly related to the performance measures that are to be predicted from the traffic model. This report, attempts to survey some of the widely used network traffic models.

---

## 2. Need for Traffic Models

The design of robust and reliable networks and network services is becoming increasingly difficult in today's world. The only path to achieve this goal is to develop a detailed understanding of the traffic characteristics of the network.

### 2.1 Network Performance Management

Managing performance of networks involves optimizing the way networks function in an effort to maximize capacity, minimize latency and offer high reliability regardless of bandwidth available and occurrence of failures. Network performance management consists of tasks like measuring, modeling, planning and optimizing networks to ensure that they carry traffic with the speed, capacity and reliability that is expected by the applications using the network or required in a particular scenario.

Networks are of different types and can be categorized based on several factors. However, the factors that affect the performance of the different networks are more or less the same. These involve parameters like Latency, Packet Loss and Throughput. In order to design high performance networks or guarantee performance of any type of network detailed analysis of the above factors is a crucial step. Often the foremost step in such an analysis is the study of the traffic on the network. As a consequence the type of traffic model used to understand the flow of traffic in the network, and how closely the model depicts the real-time characteristics of the network, become vital parameters. Choosing a model that doesn't describe the real-time characteristics of the traffic in the network can be as disastrous as not analysis the traffic at all.

### 2.2 QOS Guarantees

The term Quality of Service, in the field of networking, refers to control procedures that can provide a guaranteed level of performance to data flows in accordance to requests from an application/user using the network. A network that provides supports QOS usually agrees on a traffic contract with an application and reserves a finite capacity in the network nodes, based on the contract, during the session establishment phase. While the session is in progress, the network strives to adhere to the contract by monitoring and ensuring that the QOS guarantees are met. The reserved capacities are released subsequently after the session.

There are several factors that might affect such QOS guarantees. Hence, to design a network to support QOS is not a easy task. The primary step is to once again have a clear understanding of the traffic in the network. Without a clear understanding of the traffic and the applications that might be using the network, QOS guarantees cannot be provided. Therefore, modeling of traffic becomes a crucial and necessary step.

---

## 3. Traffic Models

Analysis of the traffic provides information like the average load, the bandwidth requirements for different applications, and numerous other details. Traffic models enables network designers to make assumptions about the networks being designed based on past experience and also enable prediction of performance for future requirements. Traffic models are used in two fundamental ways: (1) as part of an analytical model or (2) to drive a *Discrete Event Simulation (DES)*.

Simple traffic comprises of single arrivals of discrete entities, viz., packets, cells, etc. This kind of traffic can be expressed mathematically as a *Point Process*. A point process consists of a sequence of arrival instants $T_1, T_2, T_3... T_n$ (by convention, $T_0 = 0$). Point processes can be described as a *Counting Process* or *Inter-Arrival Time (IAT) Process*. A counting process $N(t)$ is a continuous time, non-negative, integer-valued stochastic process, where $N(t) = max\{n:T_n \leq t\}$ denotes the number of (traffic) arrivals in the time interval $(0,t]$[Frost94]. An inter-arrival process is a non-negative random sequence $\{A_n\}$, where $A_n = T_n - T_{n-1}$ is the length of the time interval separating the nth arrival from the previous one[Frost94]. Discrete-time traffic processes are characterized by slotted time intervals. In other words, the random variables $A_n$ can assume only

integer values. Traffic

### 3.1 Poisson Distribution Model

One of the most widely used and oldest traffic model is the Poisson Model. The memoryless Poisson distribution is the predominant model used for analyzing traffic in traditional telephony networks[Frost94]. The Poisson process is characterized as a renewal process. In a Poisson process the inter-arrival times are exponentially distributed with a rate parameter $\lambda$: $P\{A_n \leq t\} = 1 - exp(-\lambda t)$. The Poisson distribution is appropriate if the arrivals are from a large number of independent sources, referred to as Poisson sources. The distribution has a mean and variance equal to the parameter $\lambda$.

The Poisson distribution can be visualized as a limiting form of the binomial distribution, and is also used widely in queueing models. There are a number of interesting mathematical properties exhibited by Poisson processes. Primarily, superposition of independent Poisson processes results in a new Poisson process whose rate is the sum of the rates of the independent Poisson processes. Further, the independent increment property renders a Poisson process memoryless. Poisson processes are common in traffic applications scenarios that comprise of a large number of independent traffic streams. The reason behind the usage stems from *Palm's Theorem* which states that under suitable conditions, such large number of independent multiplexed streams approach a Poisson process as the number of processes grows, but the individual rates decrease in order to keep the aggregate rate constant. Nevertheless, it is to be noted that traffic aggregation need not always result in a Poisson process.

The two primary assumptions that the Poisson model makes are:

1. the number of sources is infinite
2. the traffic arrival pattern is random.

The probability distribution function and density function of the model are given as:

$$F(t) = 1 - e^{-\lambda t}$$

$$f(t) = \lambda e^{-\lambda t}$$

There are also other variations of the Poisson distributed process that are widely used. There are for example, the Homogeneous Poisson process and Non-Homogeneous Poisson process that are used to represent traffic characteristics. An interesting observation in case of Poisson models is that as the mean increases, the properties of the Poisson distribution approach those of the normal distribution.
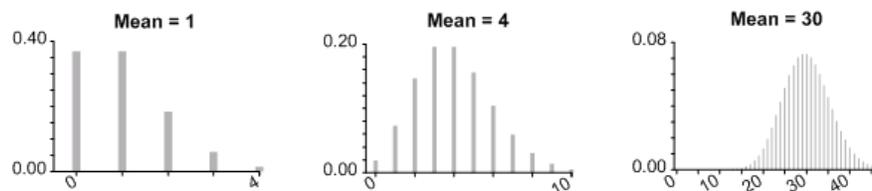


*Figure #1: Sample Poisson distributions (with increasing mean)*

The implication is that for Poisson distributions with means greater than 30 and subject to the accuracy required it is possible to use the normal distribution as an approximation.
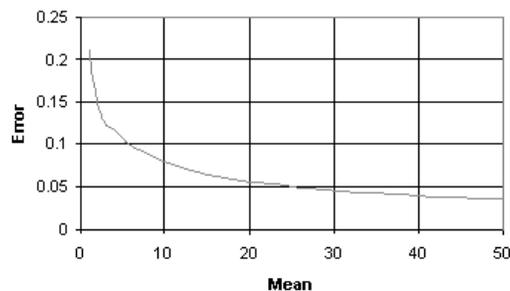


*Figure #2: Error of normal approximation to Poisson*

Bernoulli processes are the discrete time analog of Poisson processes. In a Bernoulli process the probability of an arrival in any time slot is $p$, independent of any other one. The time between arrivals corresponds to a Geometric distribution.

## 3.2 Pareto Distribution Process

The Pareto distribution process produces independent and identically distributed *(IID)* inter-arrival times[Adas97]. In general if $X$ is a random variable with a Pareto distribution, then the probability that $X$ is greater than some number $x$ is given by

$$P(X > x) = (x/x_m)^{-k} \text{ for all } x \geq x_m$$

*where k is a positive parameter and $x_m$ is the minimum possible value of $X_i$*

The probability distribution and the density functions are represented as:

$$F(t) = 1 - (\alpha/t)^{\beta} \text{ where } \alpha,\beta \geq 0 \ \& \ t \geq \alpha$$

$$f(t) = \beta\alpha^{\beta} t^{-\beta-1}$$

The parameters $\beta$ and $\alpha$ are the shape and location parameters, respectively. The Pareto distribution is applied to model self-similar arrival in packet traffic. It is also referred to as double exponential, power law distribution. Other important characteristics of the model are that the Pareto distribution has infinite variance, when $\beta \geq 2$ and achieves infinite mean, when $\beta \leq 1$
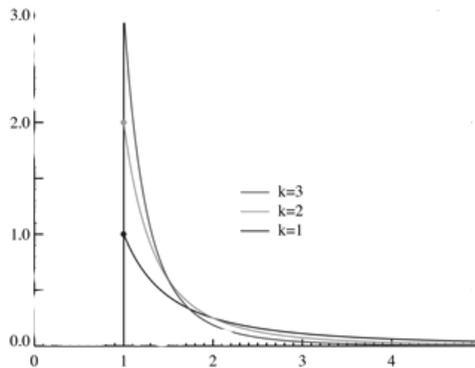


*Figure #3: Probability density function of a Pareto distribution*

## 3.3 Weibull Distribution Process

The Weibull distributed process is heavy-tailed and can model the fixed rate in ON period and ON/OFF period lengths, when producing self-similar traffic by multiplexing ON/OFF sources. The distribution function in this case is given by:

$$F(t) = 1 - e^{-(t/\beta)\alpha} \ t > 0$$

and the density function of the weibull distribution is given as:

$$f(t) = \alpha\beta^{-\alpha} t^{\alpha-1} e^{-(t/\beta)\alpha} \ t > 0$$

*where parameters $\beta \geq 0$ and $\alpha > 0$ are the scale and location parameters respectively.*

The Weibull distribution is close to a normal distribution. For $\beta \leq 1$ the density function of the distribution is L shaped and for values of $\beta > 1$, it is bell shaped[Adas97]. This distribution gives a failure rate increasing with time. For $\beta > 1$, the failure rate decreases with time. At, $\beta = 1$, the failure rate is constant and the lifetimes are exponentially distributed.

## 3.4 Markov and Embedded Markov Models

Markov models attempt to model the activities of a traffic source on a network, by a finite number of states. The accuracy of the model increases linearly with the number of states used in the model. However, the complexity of the model also increases proportionally with increasing number of states. An important aspect of the Markov model - the *Markov Property,* states that the next (future) state depends only on the current state. In other words the probability of the next state, denoted by some random variable $X_n+1$, depends only on the current state, indicated by $X_n$, and not on any other state $X_i$, where $i<n$.

The set of random variables referring to different states *{$X_n$}* is referred to as a *Discrete Markov Chain*. If the state transitions of the system under study, happens only at integral values *0,1,2,3...n,* then the *Markov chain (MC)* is discrete time and the random variable $X$ follows a geometric distribution; otherwise, it is continuous time, with the random variable taking an exponential distribution. In a simple Markov Traffic model, each of the state transition represents a new arrival process on the network. For modeling a continuous time system, the inter-arrival times times are

assumed to be exponentially distributed.

A *Semi-Markov* model is one that is obtained by allowing the time between state transitions to follow an arbitrary probability distribution. The time distribution between state transitions can also be ignored. In this model, the state transitions are then modeled as discontinuous entities with respect to time. The MC developed under such an assumption, is also referred to as an *Embedded* or Discrete Markov chain.

### 3.4.1 ON-OFF and IPP Models

The design & development of an ON-OFF model relies on a accurate description of traffic entities from link level to application level. The model is generally used, when it is necessary to capture the scaling behaviors of network traffic. For instance, analysis of the structure of IP traffic is performed predominantly using ON-OFF models. The ON-OFF model uses only two states, namely *ON & OFF*. The time spent between the ON & OFF states, commonly referred to as the *transition time,* is expected to follow an exponential distribution[Adas97]. The subsequent queueing analysis of multiplexed ON-OFF sources would detail the development of the model.

To understand the ON-OFF model, let us consider a queue/link in a network, shared by N different ON-OFF sources, as shown in Figure #. For the ON-OFF model to be used in this scenario, it is required that the sources are statistically identical and independent. The queue of size $M$ is served by a constant rate $C$, by the source. The ON-OFF source is characterized by $L$ - the mean number of packets/cells generated generated during the ON state, the peak rate $S$ when the source is ON, and the mean source rate $r$. These factors determine the mean durations of the ON & OFF periods of the source. The equilibrium probability in the ON phase of the source is calculated as, $\gamma = r/S$

With the ON-OFF periods exponentially distributed, the source can be modeled by a two-state Markov chain. The mean rate of packet/cell generation is assumed to be strictly greater then 1, $L >> $. The transition rates of the source, from the OFF state to the ON state and vice-versa, is calculated as,

$t_1$ *(from the Off to the On state): $\gamma S / (L(1-y))$*

$t_2$ *(from the On to the Off state): $S / L$*
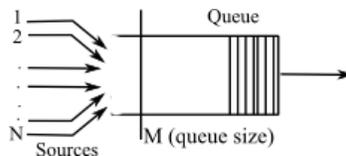


*Figure #4: Example queueing analysis for ON-OFF models*
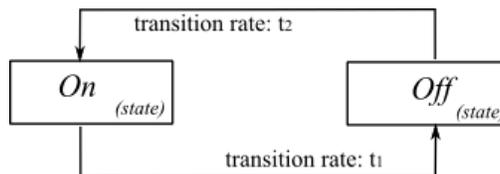


*Figure #5: Simple ON-OFF Model with transitional rates $t_1$ & $t_2$*

The *Interrupted Poisson Process (IPP)* is yet another two state process. The network channel is one of the two states, ON or OFF. In a discrete time IPP, a packet arrives in each of the time slots of the ON state, following a Bernoulli distribution. Though the IPP model is similar to the ON-OFF model, there is a slight variation that differentiates the two models. The difference is that in case of the IPP model, there is no traffic or in other words, no packets arrive during the OFF state.
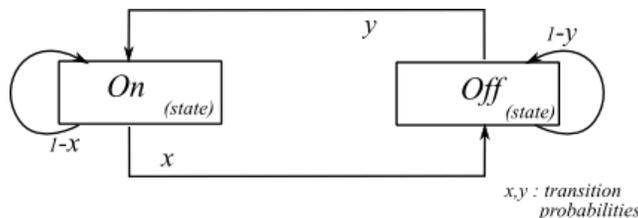


*Figure #6: IPP Model framework*

### 3.4.2 Alternating State Renewal Process

Conventional Markov models, though mathematically tractable, fail to fit actual traffic of high speed networks. In high-speed networks the packets are transmitted in a *packet train* fashion; once such a packet train is triggered, the probability that another packet will follow is very large. Further, the length of the packets exhibit a heavy-tail distribution. This observation led to the well known Alternating State Renewal Process *(ASRP)*.

ASRP is another two-state process used to model network traffic. Though there are only two states, $S_1$ and $S_2$, similar to the previous two-state models discussed, there is no *self-transition* in this model. The amplitude of the traffic in state $S_1$ is *0* and *1* in the state $S_2$. The mean time taken for transition between the two states, is denoted by $d_1$ & $d_2$ respectively. The ASRP model can be visualized as an *Embedded Markov Chain (EMC)* varying between the two states of the model. The probabilities for being in the individual states can be calculated using the simple formulae, $P_{S1} = d_1 / (d_1+d_2)$ and $P_{S2} = d_2 / (d_1+d_2)$.

### 3.4.3 Markov Modulated Poisson Process

The *Markov Modulated Poisson Process (MMPP)* is a widely used tool for analysis of teletraffic models. The model is preferred for its high versatility in qualitative behavior. It allows to capture network traffic sources that are bursty in nature. A *Markov Modulated Process (MMP)* employs an auxiliary Markov process, in which the current state of the Markov process controls the probability distribution of the traffic. MMPP is a variation of a Markov modulated process, where the auxiliary process used is a Poisson distributed process. In other words, MMPP is a doubly stochastic process where the intensity of a Poisson process is defined by the state of a MC[Smyth03][Michon03]. The MC can be visualized to modulate the Poisson process. The MMPP is also identified as a special case of the *Markovian Arrival Process (MAP)*.

MMPPs are classified by the number of states present in the modulating MC. An MC with two states and two different intensities is referred as MMPP-2. It is also called sometimes as a *Switched Poisson Process (SPP)*. An interesting feature to observe in this case, is that when the two intensities are equal, then the model transforms to a typical Poisson process. IPP is a special case of an MMPP, where either of the intensity is zero. Another important aspect of MMPP is that a superposition of MMPPs is also a MMPP. An MMPP with *M + 1* number of states, can be obtained by superposition of *M* identical and independent IPP sources[Smyth03].

The use of a Poisson distributed process in MMPP implies that the arrival rates of sources have Poisson distribution with a rate denoted by $\lambda_k$. MMPP model can be used for analyzing a mixture of voice and data traffic[Adas97]. In that case, the arrival rates are still assumed to exhibit a Poisson distribution. Therefore, the traffic is assumed to be comprised of voice and data packets, together adding up to the overall traffic on the network. However, though the arrival rates of both the types of traffic, voice and data are Poisson in nature, the rates of the individual packets can be different. Considering, hence, Poisson distribution for data packets with a rate $\lambda_d$ and assuming that voice packets also follow Poisson distribution, however, with a different rate $\lambda_V$, the resulting rate at any particular state $S_i$ is given by $S_i = \lambda_d + \lambda_v$.
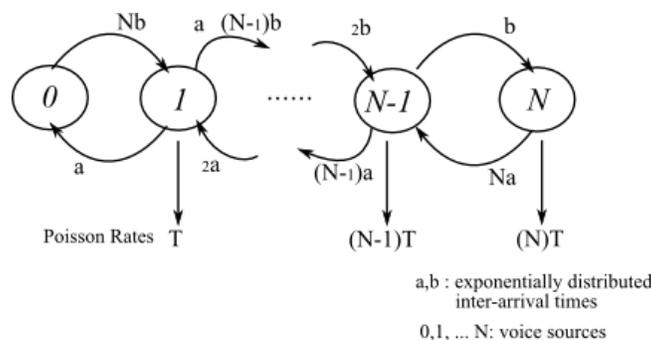


*Figure #7: Model graph of MMPP showing superposition of N voice sources*

### 3.4.4 Markov Modulated Fluid Models

Fluid flow models are conceptually simple. For instance, event simulation for an ATM multiplexer has several advantages, when fluid flow models are used for the simulation. Models other than the fluid flow models, that distinguish between the cells and consider the arrival of each cell as a separate event, typically consume huge amounts of memory and CPU time for the simulation. On the contrary, a fluid flow model that characterizes the incoming cells by a finite flow rate, require comparatively less resources[Adas97]. This is because in a fluid flow model, an event is generated only when the flow rate changes; and changes in flow rates are less frequent compared to the arrivals of cells. A fluid flow model as a consequence, utilizes lesser computing power and memory resources, compared to simulation using other models.

The basic feature of a fluid model is to characterize the traffic on a network as a continuous stream of input with a finite flow/stream rate. In other words, the incoming traffic rate is represented as a stream with a finite rate. By capturing the rate changes at the input, the models analyzes the different events that occur in the network. Because of the simple method of characterization of traffic, the fluid modes are analytically tractable and easier to simulate. Like any other Markov modulated process the *Markov Modulated Fluid Model (MMFM)*, uses an underlying MC that determines the rate of the sources. At any instant, the current state of the underlying MC determines the flow rate of the inputs.

### 3.5 Autoregressive Models

The Autoregressive model is one of a group of linear prediction formulas that attempt to predict an output $y_n$ of a system based on previous set of outputs $\{y_k\}$ *where* $k < n$ and inputs $x_n$ and $\{x_k\}$ *where* $k < n$. There exist minor changes in the way the predictions are computed based on which, several variations of the model are developed. Basically, when the model depends only on the previous outputs of the system, it is referred to as an auto-regressive model. It is referred to as a *Moving Average Model (MAM)*, if it depends on only the inputs to the system[Adas97]. Finally, Autoregressive-Moving Average models are those that depend both on the inputs and the outputs, for prediction of current output.

Autoregressive model of order $p$, denoted as *AR(p)*, has the following form:

$$X_t = R_1 X_{t-1} + R_2 X_{t-2} + ... + R_p X_{t-p} + W_t$$

where $W_t$ is the white noise, $R_i$ are real numbers and $X_t$ are prescribed correlated random numbers. The auto-correlation function of the *AR(p)* process consists of damped sine waves depending on whether the roots (solutions) of the model are real or imaginary. Discrete Autoregressive Model of order $p$, denoted as *DAR(p)*, generates a stationary sequence of discrete random variables with a probability distribution and with an auto-correlation structure similar to that of the Autoregressive model of order $p$.

---

## 4. Summary

The different traffic models each have its own pros and cons. The type of network under study and the traffic characteristics strictly influence the choice of the traffic model used for analysis. Traffic models that cannot capture or describe the statistical characteristics of the actual traffic on the network are to be avoided, since the choice of such models will result in under-estimation or over-estimation of network performance.

There is no one single model that can be used effectively for modeling traffic in all kinds of networks. For heavy-tailed traffic, it can be shown that Poisson model under-estimates the traffic[Paxson95]. In case of high speed networks with unexpected demand on packet transfers, Pareto based traffic models are excellent candidates since the model takes into the consideration the long-term correlation in packet arrival times[Adas97]. Similarly, with Markov models, though they are mathematically tractable, they fail to fit actual actual traffic of high-speed networks.

Other than the traffic models discussed in this report there are numerous other traffic models, that are used widely for traffic modeling. There are different categories of traffic models like stationary and non-stationary types. Stationary models can further be subdivided into models that are referred to as Short-range dependent and Long-range dependent types. Each model varies significantly from the other and is suitable for modeling different traffic characteristics.

A number of factors come into play while evaluating the efficiency of a traffic model. In general, the factor that differentiates one model from the other is the ability to model various correlation patterns and marginal distributions. Traffic models should have a manageable number of parameters, and parameter estimation should be simple; and, models that are not analytically tractable are preferred only for generating traffic traces.

---

## 5. References

1. [Frost94] Victor S. Frost and Benjamin Melamed, "Traffic Modeling for Telecommunications Networks", IEEE Communications, Mar. 1994.
http://ieeexplore.ieee.org/iel1/35/6685/00267444.pdf

2. [Adas97] Abdelnaser Adas, "Traffic Models in Broadband Networks", IEEE Communications Magazine, Jul. 1997.
http://ieeexplore.ieee.org/iel5/35/13111/00601746.pdf?isnumber=&arnumber=601746

3. [Smyth03] Steven L. Scott and Padhraic Smyth, "The Markov Modulated Poisson Process and Markov Poisson Cascade with Applications to Web Traffic Modeling"
http://www.datalab.uci.edu/papers/ScottSmythV7.pdf

4. [Yang01] X. Yang, A.P. Petropulu, "The Extended Alternating Fractal Renewal Process for Modeling Traffic in High-Speed Communication Networks," IEEE Trans. Sig. Proc., vol. 49, no. 7, July 2001.
http://citeseer.ist.psu.edu/cache/papers/cs/30369/http:zSzzSzwww.ece.drexel.eduzSzCSPLzSzpublicationszSzEAFRP-final.pdf/yang01extended.pdf

5. [Michon03] Gerard P. Michon, "Markov Modulated Poisson Processes", 2003.
http://www.itm.hk-r.se/~adrian/courses/modern_techniques_networking/assignments/tools/MMPP.pdf

6. [Paxson95] V. Paxson and S. Floyd, "Wide-area Traffic: The Failure of Poisson Modeling", IEEE/ACM Transactions on Networking, Jun. 1995.
http://www.cs.ucsb.edu/~ravenben/classes/276/papers/pf95.pdf

7. [Pruthi95] P. Pruthi and A. Erramilli. "Heavy-tailed ON/OFF source behavior and self-similar traffic", IEEE Transactions, 1995.
http://ieeexplore.ieee.org/iel3/3942/11415/00525209.pdf?arnumber=525209

8. [McMillan91] D. McMillan, "Traffic modeling and analysis for cellular mobile networks", 13th International Teletraffic Congress, Copenhagen,

Jun. 1991.

9. [Cock98] De Cock, K. and B. De Moor, "Stochastic system identification for ATM network traffic models: a time domain approach", 1998. http://ieeexplore.ieee.org/iel3/4040/11586/00527368.pdf?arnumber=527368

10. [Rose95] O. Rose, "Statistical Properties of MPEG Video Traffic and Their Impact on Traffic Modeling in ATM systems", Feb. 1995. http://ieeexplore.ieee.org/iel3/4040/11586/00527368.pdf?arnumber=527368

11. [Roughan] Matthew Roughan and Charles Kalmanek, "Pragmatic Modeling of Broadband Access Traffic"

12. [Rosenburg94] C. Rosenberg, G. H'ebuterne, "Dimensioning a Traffic Control Device in an ATM Network", Paris, 1994.

---

## 6. List of Acronyms

| | |
|---|---|
| **DES** | **Discrete Event Simulation** |
| QOS | Quality of Service |
| IAT | Inter-Arrival Time |
| ASRP | Alternating State Renewal Process |
| MC | Markov Chain |
| EMC | Embedded Markov Chain |
| IPP | Interrupted Poisson Process |
| MMPP | Markov Modulated Poisson Process |
| SPP | Switched Poisson Process |
| MAM | Moving Average Model |
| MMFM | Markov Modulated Fluid Model |
| FM | Fluid Model |

---

This report is available on-line at http://www.cse.wustl.edu/~jain/cse567-06/traffic_models3.htm
List of other reports in this series
Back to Raj Jain's home page