Proceedings of the 11th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'11), 19–22 October 2011, Riga, Latvia, p. 82-87. ISBN 978-9984-818-46-7
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

NETWORK TRAFFIC PRIORITIZATION USING MAP OF ARRIVALS

Jelena Revzina

Transport and Telecommunication Institute
Lomonosova str., 1, Riga, LV-1019, Latvia
Ph.: (+371) 67100584. E-mail: lena_revzina@tsi.lv

Recent advances in telecommunications and the proliferation of networking technologies dictate the need for their deeper and more accurate performance evaluation. Practically, both emerging and existing telecommunication services, including voice, video and data, demand the mechanisms for network traffic priorities allocation.

In this study we consider Markov-Bernoulli recording of MAPs of arrivals, for which the secondary process turns out to be a MAP of arrivals. Secondary recording of an arrival process is a method that from an original arrival process generates a secondary new arrival process.

In this paper we present the methods using Markov-Bernoulli recording of MAP of Arrivals which can be used for traffic prioritization modeling.

Key words: Markov additive process (MAP) of arrivals, Markov-Bernoulli recording, network traffic, prioritization

1. Applications of MAPs and their Derivatives

MAP has been widely used for modeling of many types of queueing systems. Thus, different types of MAP-based queueing models have been developed, including MAP/M/1 queue, MAP/G/1 queue and MAP/PH/1 queue. [1]

One major advantage of BMAP [2] over MAP is that batches associated with BMAP add to the modeling power and flexibility of MAP. This fact has been exploited by Klemm (2003) to model IP traffic. BMAP-based queueing systems have been extensively analyzed by many authors with respect to ATM traffic. Lucantoni (1993) provides a survey of the analysis of the BMAP/G/1 queue. As it is pointed out in Masuyama (2003), queues with batch Markovian arrival are so flexible that they can represent most of the queues studied in the past as special cases.

Many authors, including Heffes and Lucantoni (1986), Baiocchi (1991), Yamada and Sumita (1991), and Li and Hwang (1993), have used MMPP(2) to model the superposed ATM traffic. Their analysis deals with MMPP/M/1 or MMPP/G/1 queueing systems. Fischer AND Meier-Hellstern (1992) discuss other applications of MMPP.

Zhou and Gans (1999) consider an M/MMPP/1 queue, which can be used to model a system that processes jobs from different sources. The time to process jobs from each source (or job type) is exponentially distributed, but each job type has a different mean service time. Moreover, after a job completion, the choice of next job to be processed is governed by a Markov chain.

Muscariello (2005) uses a hierarchical MMPP traffic model that very closely approximates the long-range dependence (LRD) characteristics of the Internet traffic traces over relevant time scales. The LRD property of the Internet traffic means that values at any instant tend to be positively correlated with values at all future instants. This means that it has some sort of memory. However, long-term correlation properties, heavy tail distributions, and other characteristics are meaningful only over a limited time scale.

The overflow traffic has been modeled using an interrupted Poisson process (IPP) by Kuczura (1973) and Meier-Hellstern (1989). Min (2001) analyzed the adaptive worm-hole-routed torus networks with IPP traffic input. [3]

A survey of these traffic models is given in Bae and Suda (1991), Frost and Melamed (1994), Michiel and Laevens (1997) and Adas (1997).

2. MAP of Arrivals

We define the MAP (Markov additive process) of arrivals to be properly adapted to the multivariate setting. We use the definitions and notations of Pacheco, Tang and Prabhu. [4]

Firstly, we denote $\mathbf{R} = (-\infty, \infty)$, r a positive integer, and E a countable set.

A process $(\boldsymbol{X}, J) = \{(\boldsymbol{X}(t), J(t)), t \geq 0\}$ on the state space $\mathbf{R}^r \times E$ is an MAP if it satisfies the following two conditions:

(a) *(**X**,J)* is a Markov process;

(b) For $s, t \geq 0$, the conditional distribution of *(**X**(s+t) –**X**(s) J(s+t))* given *(**X**(s),J(s))* depends only on *J(s)*.

In this case, a MAP of arrivals is simply a MAP with the additive component taking values on the nonnegative (may be multidimensional) integers, i.e., $\mathbf{N}_+^r$ with $\mathbf{N}_+ = \{0,1,2,...\}$. For clarity we consider time-homogeneous MAPs, for which the conditional distribution of *(**X**(s+t) –**X**(s),J(s+t))* given *J(s)* depends only on *t*.

A MAP of arrivals *(**X**,J)* is simply a MAP taking values on $\mathbf{N}_+^r \times E$ and, as a consequence, the increments of ***X*** may then be associated with arrival events, with the standard example being that of different classes of arrivals into a queueing system. Accordingly, without loss of generality, we use the term arrivals to denote the events studied and make the interpretation

$X_i(t)$ = total number of class *i* arrivals in (0,*t*] (1)

for *i*=1,2,…,*r*.

We than call ***X***, the additive component in the terminology for MAPs, the arrival component of *(**X**,J)*. It follows from the definition of MAP that *J* is a Markov chain and, following the terminology for MAPs, we call it the Markov component of *(**X**,J)*.

Since a MAP of arrivals *(**X**,J)* is a Markov subordinator on $\mathbf{N}_+^r \times E$, it is a Markov chain. Moreover, since *(**X**,J)* is time-homogeneous it is characterized by its transition rates, which are invariant translation in the arrival component ***X***. Thus, it is sufficient to give, for $j, k \in E$ and $\mathbf{m}, \mathbf{n} \in \mathbf{N}_+^r$, the transition rate from *(***m***,j)* to *(***m***+***n***, k)*, which we denote simply (since the rate does not depend on **m**) by $\lambda_{jk}(\mathbf{n})$.

Whenever the Markov component *J* is in the state *j*, the following three types of transitions in *(**X**,J)* may occur with the respective rates:

(a) Arrivals without changes of state in *J* occur at rate $\lambda_{jj}(\mathbf{n}),\ \mathbf{n} > \boldsymbol{0}$;

(b) Change of state in *J* without arrivals occurs at rate $\lambda_{jk}(\mathbf{0}), k \in E, k \neq j$;

(c) Arrivals with change of state in *J* occur at rate $\lambda_{jk}(\mathbf{n}), k \in E, k \neq j, \mathbf{n} > \boldsymbol{0}$.

We denote $\boldsymbol{\Lambda}_{\mathbf{n}} = (\lambda_{jk}(\mathbf{n})), \mathbf{n} \in \mathbf{N}_+^r$, and say that *(**X**,J)* is a simple MAP of arrivals if $\Lambda_{(n_1,\dots n_r)} = 0$ if $n_l > 1$ for some *l*. In case **X** (=X) has a state space $\boldsymbol{N}_+$ we say that *(X,J)* is a univariate MAP of arrivals, an example of which is the Markov-Poisson process, which is known as MMPP (Markov-modulated Poisson process).

The identification of meaning of each of the above transitions is very important for applications. In particular, arrivals are clearly identified by transitions of types (a) and (c), so that we may talk about arrival epochs, interarrival times, and define the complex operations like thinning of MAPs of arrivals. Moreover, the parameterization of these processes is simple and, by being Markov chains, they have "nice" structural properties. These are especially important computationally since it is a simple task to simulate Markov chains.

3. Markov-Bernoulli Recording of MAPs of Arrivals

Secondary recording of an arrival process is a mechanism that generates a secondary arrival process from the original arrival process. A classical example of secondary recording is the Bernoulli thinning which records each arrival in the original process, independently of all others, with a given probability *p* or removes it with probability *1 – p*. For MAPs of arrivals *(**X**,J)* we consider secondary recordings that leave *J* unaffected.

A simple example for which the probability of an arrival being recorded varies with time is the case when there is a recording station (or control process) which is *on* and *off* from time to time, so that arrivals in the original process are recorded in the secondary process during periods in which the station is operational (*on*). The control process may be internal or external to the original process and may be of variable complexity (e.g. rules for access of customer arrivals into a queueing network may be simple or complicated and may depend on the state of the network at arrival epoch or not).

Another example is the one in which the original process counts arrivals of batches of customers into a queueing system, while the secondary process counts the number of individual customers, which is more important than the original arrival process in case service is offered to customers one at a time (this is a way in which the compound Poisson process may be obtained from the simple Poisson process). Here the secondary process usually counts more arrivals then the original process, whereas in the previous examples the secondary process always records fewer arrivals than the original process does (which corresponds to thinning).

Secondary recording is related to what is called marking the original process. Suppose that each arrival in the original process is given a mark from a space *M*, independently of the marks given to other arrivals (e.g. for Bernoulli thinning with probability *p* each point is marked "recorded' with probability *p* or "removed" with probability *1 – p*). If we consider the process that accounts only the arrivals which have marks on a subset *C* of *M*, it may be viewed as a secondary recording of the original process. A natural mark associated with cells moving in a network is the pair of origin and destination nodes of the cell. If the original process accounts for the traffic generated in the network, then the secondary process may represent the traffic generated between a given pair of nodes, or between two sets of nodes. The marks may represent *priorities*. These are commonly used in modeling access regulators to communication network system. A simple example of an access regulator is the one of which cells (arrivals) are judged in violation of the "contracts" between the network and the user are marked and may be dropped from service under the specific situations of congestion in the network; other cells are not marked and carried through the network. The secondary processes of arrivals of marked and unmarked cells are of obvious interest.

4. Considered Case

Consider a switch which receives the inputs from a number of sources and suppose that the original process accounts for inputs from those sources. If we are interested in studying one of the sources in particular, we should observe a secondary process which accounts only the input from this source; this corresponds to a margin of the original process. In case all sources generate the same time of input, what is relevant to the study of performance of the system it is a total input arriving to the switch; this is a secondary process which corresponds to the sum of the coordinates of the original arrival process. This shows that some of the transformations of arrival process may also be viewed as special cases of secondary recording of an arrival process.

Suppose that *(**X**,J)* is an MAP of arrivals on $\mathbf{N}_+^r \times E$ with $(Q,\{\Lambda_n\}_{n>0})$ – source. We consider a recording mechanism that independently records with probability $r_{jk}(\mathbf{n},\mathbf{m})$ an arrival in $\boldsymbol{X}$ of a batch of size **n** associated with a transition from *j* to *k* in *J* as an arrival of size **m** in the secondary process (with $\mathbf{m} \in \mathbf{N}_+^s$ for some $s \geq 1$). The operation is identified by the set *R* of recording probabilities

$$R = \{R_{(\mathbf{n},\mathbf{m})} = (r_{jk}(\mathbf{n},\mathbf{m})) : \mathbf{n} \in \mathbf{N}_+^r, \mathbf{m} \in \mathbf{N}_+^S\}. \tag{2}$$

Where $r_{jk}(\mathbf{n},\cdot)$ is a probability function on $\mathbf{N}_+^{\mathbf{s}}$, and $r_{jk}(\mathbf{0},\mathbf{m}) = \delta_{\mathbf{m0}}$. We call this recording Markov-Bernoulli recording with probabilities *R* and denote the resulting secondary process as $(\mathbf{X}^R, J)$. Thus $(\mathbf{X}^R, J)$ is a process on $\mathbf{N}_+^{\mathbf{s}} \times E$ which is non-decreasing $\mathbf{X}^R$, and which is increased only when $\boldsymbol{X}$ increases, i.e.

$$\mathbf{X}^R(t) = \mathbf{X}^R(T_p^\circ) \qquad (T_p^\circ \leq t \leq T_{p+1}^\circ). \tag{3}$$

Moreover, for **n**>**0**

$$r_{jk}(\mathbf{n},\mathbf{m}) = P(\mathbf{X}^R(T_{p+1}^\circ) - \mathbf{X}^R(T_p^\circ) = \mathbf{m} \mid A_{jk}(\mathbf{n})\} \tag{4}$$

with $A_{jk}(\mathbf{n}) = \{\mathbf{X}(T_{p+1}^\circ) - \mathbf{X}(T_p^\circ) = \mathbf{n}, J(T_{p+1}^\circ -) = j, J(T_{p+1}^\circ) = k\}$.

Theorem. Suppose that *(X,J)* is a MAP of arrivals on $\mathbf{N}_+^r \times E$ with $(Q, \{\Lambda_n\}_{n>0})$ – source and $R = \{R_{(\mathbf{n},\mathbf{m})}, \mathbf{n} \in \mathbf{N}_+^r, \mathbf{m} \in \mathbf{N}_+^S\}$ is a set of recording probabilities. Then:

(a) The process $(\mathbf{X}, \mathbf{X}^R, J)$ is a MAP of arrivals on $\mathbf{N}_+^r \times E$ with source

$$(Q, \{\mathbf{\Lambda_n} \bullet \mathbf{R_{(n,m)}}\}_{(\mathbf{n},\mathbf{m})>0}) \tag{5}$$

(b) The process $(\mathbf{X}^R, J)$ is a Map of arrivals on $\mathbf{N}_+^{r+s} \times E$ with source

$$(Q, \{\Lambda_{\mathbf{m}}^R\}_{\mathbf{m}>0}) = (Q, \{\sum_{\mathbf{n}>0} \mathbf{\Lambda_n} \bullet \mathbf{R_{(n,m)}}\}_{\mathbf{m}>0}) . \tag{6}$$

The proofs of the theorem adduced in Pacheco, Tang and Prabhu.

The special case of Markov-Bernoulli marking with the discrete space of marks *M* is considered too. Specifically, each arrival of a batch of size **n** associated with a transition from state *j* to state *k* in *J* is given a mark $m \in M$ with (marking) probability $c_{jk}(\mathbf{n}, m)$, independently of the marks given to other arrivals. If we assume without loss of generality that $M = \{0,1,..., K\} \subseteq \mathbf{N}_+$, Markov-Bernoulli marking becomes a special case of Markov-Bernoulli recording for which the marked process $(\mathbf{X}, \mathbf{X}^R, J)$ is such that $(\mathbf{X}^R, J)$ is a Markov-Bernoulli recording of *(X,J)* with recording probabilities

$$R_{(\mathbf{n},m)} = (r_{jk}(\mathbf{n}, m)) = (\mathbf{1}_{\{m \in M\}} c_{jk}(\mathbf{n}, m))$$

for **n**>**0** and $m \in N_+$.

Suppose that *X* is a Poisson process with rate λ, *J* is a stable finite Markov chain with generator matrix *Q*, and *X* and *J* are independent. We view *J* as a state of a recording station, and assume that each arrival in *X* is recorded, independently of all other arrivals, with probability p_j or non-recorded with probability $1 - p_j$, whenever the station is in state *j*. We let $X^R(t)$ be the number of recorded arrivals in (0,*t*].

Using theorems from [4] we may conclude that (X^R, J) is an MMPP with $(Q, (\lambda p_j \delta_{jk}))$ – source. In the special case where the probabilities p_j are either 1 or 0, the station is *on* and *off* from time to time, with the distributions of the on and off periods being Markov dependent.

5. Examples

We consider now two examples of Markov-Bernoulli recording. In the first example we see the overflow process from a state dependent M/M/1/K system as a special case of secondary recording of the arrival process of customers to the system. In the second example we see a more elaborated case of secondary recording.

5.1. Example 1

We consider a Markov-modulated *M/M/1/k* system with batch arrivals with (independent) size distribution $\{p_n\}_{n>0}$. When there is an arrival of a batch with *n* customers and only *m<n* positions are available, only *m* customers from the batch enter the system. Assume that the service rate is μ_j and the arrival rate of batches is α_j, whenever the number of customers in the system is *j*. We let *J(t)* be the number of customers arrivals in (0,*t*].

The process (*X,J*) is a MAP of arrivals on $\mathbf{N}_+ \times \{0,1,..., K\}$ with rates

$$\lambda_{jk}(n) = \begin{cases} \alpha_j p_n & k = \min(j+n, K) \\ \mu_j & k = j-1, n = 0 \\ 0 & otherwise \end{cases}$$

Its source $(Q,\{\Lambda_n\}_{n>0})$, where Q is obtained from the matrices $\{\Lambda_n\}_{n\ge 0}$ through () and (). If we define $X^R(t)$ as the overflow from the system in *(0,t]*, then it is readily seen that (X^R, J) is a Markov-Bernoulli recording of *(X,J)* with recording probabilities $r_{jk}(n,m)=\mathbf{1}_{\{m=n-(k-j)\}}$ for *n,m*>0. Thus (X^R, J) is a MAP of arrivals on $\mathbf{N}_+\times\{0,1,...,K\}$ with $(Q,\{\Lambda_m^R\}_{m>0})$ – source, where $\lambda_{jk}^R(m)=\alpha_j\delta_{kK}p_{m+(K-j)}$ for *m*>0.

5.2. Example 2

Suppose $\mathbf{X}=(X_1,...,X_r)$, $\mathbf{Y}=(Y_1,...,Y_s)$, and $(\mathbf{X},\mathbf{Y},J_2)$ is a MAP of arrivals on $\mathbf{N}_+^{r+s}\times E_2$ with source. We are interested in keeping only the arrivals in **X** which are preceded by an arrival in **Y** without a simultaneous arrival in **X**. The arrival counting process which we obtain by this operation is denoted by $\mathbf{X}^R$. For simplicity, we assume that E_2 is finite and $\Lambda_{(\mathbf{n},\mathbf{m})}=0$ if **n**>**0** and **m**>**0**.

We let $\{T_p^{\mathbf{X}}\}_{p>0}$ and $\{T_p^{\mathbf{Y}}\}_{p>0}$ be the successive arrival epochs in **X** and **Y**, respectively, denote $J=(J_1,J_2)$ with

$$J_1(t)=\begin{cases}1 & \max\{T_p^{\mathbf{Y}}:T_p^{\mathbf{Y}}\le t\}>\max\{T_p^{\mathbf{X}}:T_p^{\mathbf{X}}\le t\}\\ 0 & otherwise\end{cases}.$$

It can be checked in a routine fashion that ((**X**,**Y**),J) is a MAP of arrivals on $\mathbf{N}_+^{r+s}\times\{0,1\}\times E_2$ with $(Q^*,\{\Lambda^*_{(\mathbf{n},\mathbf{m})}\}_{(\mathbf{n},\mathbf{m})>0})$ – source (with the states of (J_1,J_2) ordered in lexicographic order), where $\Lambda^*_{(\mathbf{n},\mathbf{m})}$ =0 if **n**>**0** and **m**>**0**, and

$$Q^*=\begin{bmatrix}Q-\sum_{m>0}\Lambda_{(0,m)} & \sum_{m>0}\Lambda_{(0,m)}\\ \sum_{n>0}\Lambda_{(n,0)} & Q-\sum_{n>0}\Lambda_{(n,o)}\end{bmatrix},\ \Lambda^*_{(n,m)}=\begin{bmatrix}\Lambda_{(n,0)} & \Lambda_{(0,m)}\\ \Lambda_{(n,0)} & \Lambda_{(0,m)}\end{bmatrix}$$

if either **n**=**0** or **m**=**0**. It is easy to see that (X^R, J) is a Markov-Bernoulli recording of ((**X**,**Y**),*J*) with recording probabilities $r_{(p,j)(q,k)}(\mathbf{n},\mathbf{m}),\mathbf{l})=\mathbf{1}_{\{p=1,\mathbf{n}=\mathbf{1}\}}$, for (**n**,**m**)>**0**. Thus (X^R, J) is an MAP of arrivals on $\mathbf{N}_+^{r+s}\times\{0,1\}\times E_2$ with $(Q,\{\Lambda_n^R\}_{n>0})$ -source, where

$$\Lambda_{\mathbf{n}}^R=\begin{bmatrix}0 & 0\\ \Delta_{(\mathbf{n},\mathbf{0})} & 0\end{bmatrix}.$$

A particular case of the previous example with *r=1* and *Y* being a Poisson process was considered briefly by Neuts. In addition, He and Neuts considered an instance of the same example in which $(\mathbf{X},\mathbf{Y},J_2)=(\mathbf{X},Y,(K_1,K_2))$ results from patching together the independent MAPs $(\mathbf{X},K_1)$, a general MAP of arrivals, and (Y,K_2), a simple univariate MAP of arrivals. This leads to recording of the arrivals in **X** that are immediately preceded by an arrival on the simple univariate MAP of arrivals *X*.

6. Conclusions

In this paper we define a type of secondary recordings which include important special cases of marking and thinning. We consider in particular Markov-Bernoulli recording of MAP of arrivals, for which the secondary process turns out to be a MAP of arrivals.

We investigate two examples of behavior of the queueing systems. In the first example we consider the overflow process from the state dependent M/M/1/K system as a special case of secondary recording of the arrival process of customers to the system. In the second example we see a more complicated case of secondary recording.

This analysis of behavior of such systems helps the telecommunication networks to be maintained stable regardless of congestion using prioritization mechanisms. [5,6] We omit the numerical computation without information for the solution. The main purpose of this study is to show the usefulness and the flexibility of the proposed methods.

References

1. Oliver C. Ibe. Markov. *Processes for Stochastic Modeling.* Academic Press, 2009, 490 p.
2. Dudin A., Klimenok V. *Mass service systems with correlated flows*. Minsk, BGU, 2000 (in Russian).
3. G.Bolch, S.Greiner, H.de Meer, K.S.Trivedi. *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications.* A.JohnWiley&Sons, 2006, 878 p.
4. Pacheco A., Tang L.C., Prabhu N.U. *Markov-Modulated Processes & Semiregenerative Phenomena.* World Scientific, 2009, 224 p.
5. Revzina J. Stochastic Models of Data Flows in the Telecommunication Networks*, Journal Computer Modelling and New Technologies.* No 2, Vol. 14. Riga, 2010.
6. Scott S.L., Smyth P. *The Markov Modulated Poisson Process and Markov Poisson Cascade with Applications to Web Traffic Modeling.* Bayesian Statistics 7, Oxford University Press, 2003.