

University of Würzburg
Institute of Computer Science
Research Report Series

**Improving the Timeliness of Rate
Measurements**

Ruediger Martin, Michael Menth

Report No. 331

April 2004

University of Wuerzburg, Germany
Institute of Computer Science
Department of Distributed Systems
Am Hubland, D-97074 Würzburg, Germany
phone: (+49) 931-8886652, fax: (+49) 931-8886632
{martin|menth}@informatik.uni-wuerzburg.de

Improving the Timeliness of Rate Measurements

Ruediger Martin, Michael Menth

University of Wuerzburg, Germany

Institute of Computer Science

Department of Distributed Systems

Am Hubland, D-97074 Würzburg, Germany

phone: (+49) 931-8886652, fax: (+49) 931-8886632

{martin|menth}@informatik.uni-wuerzburg.de

Abstract

Rate measurements are required for many purposes, e.g. for system analysis and modelling or for live systems that react to measurements. For off-line measurement all data is available in advance. Here, time delay between data collection and data analysis is not an issue. On-line measurement, however, measures rates on the fly. Thus, measurement algorithms that provide their output as timely as possible are required. We present three well known algorithms for rate measurement: The Disjoint Intervals method, the Moving Average, and the Exponentially Weighted Moving Average over Disjoint Intervals. We analyze and compare their properties and find problems like heavy time delay or over-reaction to random fluctuations. To address these problems, we derive a new algorithm called Time Exponentially Weighted Moving Average as a continuous version of the Exponentially Weighted Moving Average. Finally, we compare this algorithm to the other methods and show that it solves these problems.

Keywords: Measurement Methods

1 Introduction

System analysis and modelling often require the measurement of rates to obtain useful input parameters. Live systems that correlate measured rates to other system parameters and react to it accordingly are another example where rate measurements are necessary.

This motivates the distinction between on-line and off-line measurement. In the case of off-line measurement, first data are collected, then analyzed, and the results are used afterwards. In the case of on-line measurement, however, data collection, data analysis, and the usage of the results are linked and take place at virtually the same time. The time gap between the respective steps must remain as short as possible. On-line and off-line measurement have different requirements regarding measurement methods.

This work was funded by the Bundesministerium für Bildung und Forschung of the Federal Republic of Germany (Förderkennzeichen 01AK045) and Siemens AG, Munich. The authors alone are responsible for the content of the paper.

For off-line rate measurement, a common approach divides the time axis into fixed size intervals and counts the number of possibly weighted events per interval. Then, the measured rate is computed as the average number of events within each interval. This yields a timely value for the respective interval. If the same method is applied for on-line measurement, the rate value is obtained not before the interval is over. Hence, the measurements are delayed by one value. This delay between the events and the update of the measured rate seriously affects live systems and requires other measurement methods that exhibit better timeliness.

In this paper we briefly present methods designed for rate measurement based on Disjoint Intervals (DI), Moving Average (MA), and Exponentially Weighted Moving Average based on Disjoint Intervals (EWMA-DI) found for similar purposes in literature. We introduce a new algorithm called Time Exponentially Weighted Moving Average (TEWMA) as a continuous version of EWMA. It is designed to avoid problems that arise from the usage of the other methods. It is specifically intended to improve the timeliness of the measured rate and to allow for simple algorithmic implementation. We discuss the advantages and drawbacks of all algorithms, compare them, and make suggestions under which circumstances their application seems appropriate.

The work is structured as follows. Section 2 outlines fundamental concepts for rate measurement, Section 3 presents the algorithms under study in detail, Section 4 compares them, and Section 5 concludes this work.

2 Fundamental Rate Measurement Concepts

As mentioned above we distinguish two fundamentally different aspects of rate measurement, off-line and on-line measurement.

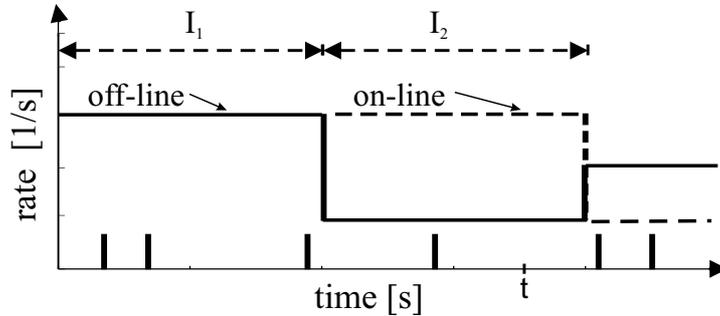


Figure 1: Off- and on-line measurement and their impact on timeliness.

In case of off-line measurement all data are available before they are analyzed. So as shown in Figure 1, the rate during interval I_2 is 1 event during this interval divided by the interval length of 2s. However, in case of on-line measurement, at time $t \in I_2$ the number of events in I_2 is still unknown. So the most current measured rate is the old value of interval I_1 .

Generally, when measuring off-line, the focus is laid upon long term characteristics of the data. Average rate and coefficient of variation on the long run are more important than short

term variations. Measuring input parameters for analytical or simulation models like the average rate of an arrival process is an example.

Opposed to that, when measuring on-line, short term variations, the evolution of the rate over time, are of specific interest. The measurements are required for live systems that correlate measured rates to other system parameters and react to it on the fly. Experience Based Admission Control (EBAC) [2] correlates admitted traffic volumes to measured traffic rates on the links of a communication network and derives suitable overbooking factors. It serves as an example for on-line measurement. The measurements should reflect the actual rate as timely as possible.

Other classifications of measurement types are possible as well and certainly no strict border can be drawn between both classes. However, we see one major challenge to measurements in stringent delay conditions and it is one criterion for which the TEWMA algorithm has been derived. Therefore, we analyze advantages and drawbacks of the rate measurement algorithms presented in this work in a general sense and at the same time we lay the focus of this work on the measurement delay.

In the following we observe a stochastic process $X(t), t \in \mathbb{R}_0$ with inter-arrival time A . Arrivals happen at times $t_i \in \mathbb{R}, i \in \mathbb{N}$. The variable $X(t_i) -$ short $X_i -$ denotes the size of the object, e.g. packet, arriving at time t_i . If there is no arrival at time t , then $X(t) = 0$. The object sizes are stochastic themselves. A_i denotes the inter-arrival time between the objects $i - 1$ and i , i.e. $A_i = t_i - t_{i-1}$. The random variables X_i and A_i may be generally distributed. Figure 2 shows an example of a packet arrival process to clarify the notation at one glance.

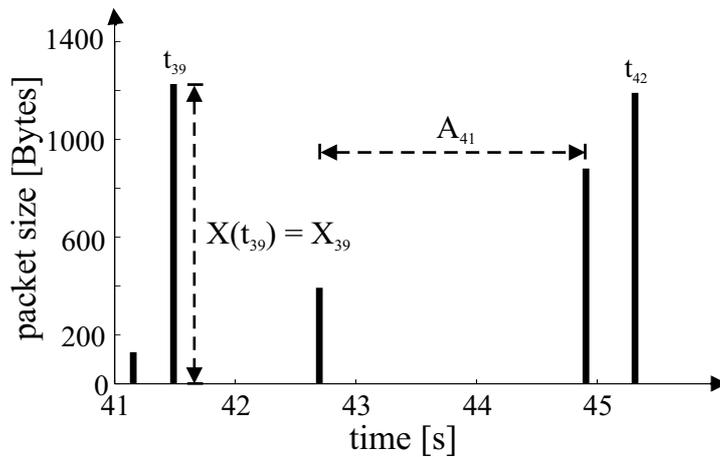


Figure 2: Packet arrival process.

A rate R is the (weighted) number of events Γ per time interval Δ and is constant or changing over time depending on the underlying process. In this work R stands for constant rates and $R(t)$ for rates that are a function of time. There is a difference between the generating rate R , the theoretical rate of the underlying process, and the empirical rate M , the rate observed by measurements. The generating rate of a process is usually unknown and the empirical rate is the only information about the process.

If we are just interested in the number of arrivals over time, we set the sizes of the arriving objects $X_i = 1$. Thus, we measure the pure arrival rate of a stochastic arrival process, i.e. the number of arrivals per time. In the case of packet arrivals, this is the number of packets arriving per time unit. Otherwise, if we take the size of the objects into account, the rate describes the number of object size units over time. In the case of a packet arrival process, this is the traffic volume, i.e. the number of bits per second. Assuming a constant rate for the example in Figure 2, 5 packets arrive within the presented interval of 5 seconds, which yields an empirical rate $M_P = \frac{1}{sec}$. With the packet sizes being 116, 1221, 397, 908, and 1198 *Bytes*, the traffic volume is $M_T = 768 \frac{Bytes}{sec}$. Of course, calculating the rate like that serves only as a simple example to point out the notation and concepts described in this section.

For the sake of simplicity, the figures in the following sections show only examples of arrival rate measurements, i.e., the object sizes are not take into account. However, they clarify the general case as well.

3 Rate Measurement Algorithms and their Characteristics

We present four distinct rate measurement algorithms that reveal different advantages and drawbacks.

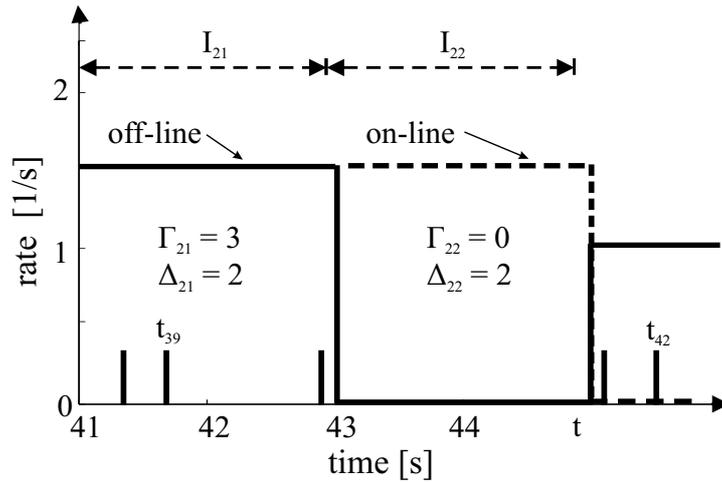


Figure 3: DI and its impact on off-line and on-line measurement.

Rate Measurement Based on Disjoint Intervals (DI) The algorithm based on disjoint intervals is a straightforward approach. The time axis is divided into disjoint (not necessarily equidistant) intervals $I_i = [t_i, t_{i+1})$ as shown in Figure 3. Let $t_k, t_{k+1}, \dots, t_{k+m}$ be the arrival times within I_i . Then $\Gamma_i = \sum_{j=k}^{k+m} X_j$ is the sum of the sizes of the corresponding objects. Given $\Delta_i = t_{i+1} - t_i$, DI meters the empirical rate

$$M_i = \frac{\Gamma_i}{\Delta_i}. \quad (1)$$

Thus, the rate M_i is determined by metering the object sizes in time interval I_i . The DI approach has several severe drawbacks.

- P1 If Δ_i are *too short*, the measured rate M_i alternates between zero and large values, i.e., it becomes very jerky.
- P2 If Δ_i are *relatively short*, random fluctuations are interpreted as rate changes although the properties of the traffic are unchanged.
- P3 The empirical rate M_i relates to the traffic behavior during I_i and is the most recent value during I_{i+1} for on-line measurements. Hence, the rate measurements are significantly delayed if Δ_i are *relatively long* (cf. Figure 3).
- P4 If Δ_i are *too long*, rate changes on a scale smaller than Δ_i can not be observed.

Now we consider equidistant intervals Δ . The interval size Δ serves as a memory to remember the state of the process. The algorithm knows the number of arrivals since the start of the current interval, but nothing about earlier arrivals. To compare the DI algorithm to the other methods presented in the following paragraphs, we define its equivalent memory

$$L = \Delta. \quad (2)$$

We define $M(t, j)$ as the contribution to the measured rate $M(t)$ by a single object j of size X_j with arrival time t_j in the case of on-line measurement. Assume that object j arrives in the interval $I = [t_I, t_I + \Delta]$, i.e. $t_j \in I$. The arrival j contributes to the rate measured by DI the value $\frac{X_j}{\Delta}$ in the interval $[t_I + \Delta, t_I + 2\Delta]$ – short $I + \Delta$ – due to the delay of Δ induced by on-line measurement here. Therefore, the measured rate at time t applied to the object j is

$$M(t, j) = \begin{cases} \frac{X_j}{\Delta} & : t \in I + \Delta \\ 0 & : t \notin I + \Delta \end{cases} \quad (3)$$

The rate integral over time applied to a single event yields

$$\int_0^{+\infty} M(\tau, j) d\tau = X_j. \quad (4)$$

Hence, the full object size is reflected by the rate.

Exponentially Weighted Moving Average Based on Disjoint Intervals (EWMA-DI) The empirical rates $\frac{\Gamma_i}{\Delta_i}$ are similarly determined based on disjoint intervals like in the preceding paragraph. To mitigate the effect of short-term rate fluctuations seen in measurements by the empirical rate, an exponentially weighted moving average (EWMA) takes the past to a certain degree into account.

In case of a uniform measurement interval length Δ , this can be written as

$$M_0 = 0 \quad (5)$$

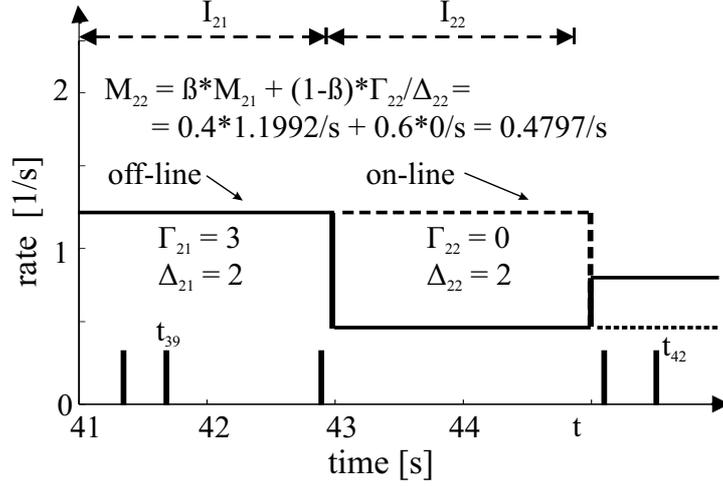


Figure 4: EWMA-DI and its impact on off-line and on-line measurement.

and

$$M_{i+1} = \beta \cdot M_i + (1 - \beta) \cdot \frac{\Gamma_{i+1}}{\Delta} \quad (6)$$

with $\beta \in (0; 1)$. An example is shown in Figure 4. The devaluation parameter β controls the influence of older measurement intervals on the current value. The impact of older measurements decays exponentially with β . Thus, the memory is controlled by the parameters β and Δ .

Solving Equation 6 recursively, we obtain

$$\begin{aligned}
 M_i &= \sum_{j=0}^{i-1} \beta^j \cdot (1 - \beta) \cdot \frac{\Gamma_{i-j}}{\Delta} + \beta^i \cdot M_0 = \\
 &= \sum_{j=0}^{i-1} \beta^j \cdot (1 - \beta) \cdot \frac{\Gamma_{i-j}}{\Delta}
 \end{aligned} \quad (7)$$

From Equation 7 we conclude that EWMA considers the values from the current interval I_i with weight $(1 - \beta)$. It further considers the values from the interval I_{i-j} with weight $\beta^j \cdot (1 - \beta)$ and therefore looks $(j + 1) \cdot \Delta$ back into the past. For $i \rightarrow \infty$ this leads to the following definition of the equivalent memory

$$\begin{aligned}
 L &= \sum_{j=0}^{\infty} (1 - \beta) \cdot \beta^j \cdot (j + 1) \cdot \Delta = \\
 &= \frac{\Delta}{1 - \beta}
 \end{aligned} \quad (8)$$

For the fair comparison with other measurement methods, β and Δ must be chosen to fulfill $\beta = 1 - \frac{\Delta}{L}$ that the same equivalent memory L is achieved.

Here, the contribution $M(t, j)$ to the measured rate $M(t)$ by a single object j in case of on-line measurement can be derived as follows. With EWMA-DI the object j of size X_j arrived at $t_j \in I = [t_I, t_I + \Delta]$ contributes to the measured rate the value

$$(1 - \beta) \cdot \beta^k \cdot \frac{X_j}{\Delta}$$

if $t \in [t_I + (k+1)\Delta, t_I + (k+2)\Delta]$. This is due to the weights of the form $\beta^k \cdot (1 - \beta)$ explained above (cf. Equations 7 and 8) and the time delay of Δ induced by on-line measurements.

$$M(t, j) = \begin{cases} (1 - \beta) \cdot \beta^k \cdot \frac{X_j}{\Delta} \\ \text{if } t \in I + (k+1)\Delta, \quad k \in \mathbb{N} \\ 0 \text{ otherwise} \end{cases} \quad (9)$$

The rate integral over time applied to a single event also yields

$$\int_0^\infty M(\tau, j) d\tau = X_j \quad (10)$$

The EWMA-DI was introduced by [3] and this mechanism has been studied quite intensively especially in the field of economics for chart analysis [4, 5, 6, 7, 8, 9]. The EWMA is also used in many technical documents of the IETF [10, 11], the most prominent one is probably the obsolete estimation of the round trip time for TCP in [12].

As this algorithm is designed not to react to random fluctuations by controlling the memory with the help of the additional parameter β , an appropriate value alleviates problem P2. The concerns P1, P3, and P4 are still valid.

Rate Measurement Based on Moving Average (MA) The measurement delay imposes severe problems on systems that depend on on-line measurements and must react to rate changes within a relatively small time gap. The following approach [13] reduces this delay.

Let $t_k, t_{k+1}, \dots, t_{k+m}$ be the arrival times within the interval $[t - \Delta, t]$. Then $\Gamma(t - \Delta, t) = \sum_{j=k}^{k+m} X_j$ is the sum of the sizes of the objects arriving within this interval. The measured rate $M(t)$ is determined by

$$M(t) = \frac{\Gamma(t - \Delta, t)}{\Delta}.$$

This algorithm moves a window over time and measures the average of the object sizes within this window. If $t < \Delta$, $M(t)$ is either undefined or $M(t) = \frac{\Gamma(0, t)}{t}$. Figure 5 shows the example from above using the MA method. This algorithm only considers the values within the current window of length Δ and therefore clearly has the equivalent memory $L = \Delta$.

As the MA method does not delay the impact of object j on the rate, its contribution $M(t, j)$ to the measured rate $M(t)$ is

$$M(t, j) = \begin{cases} \frac{X_j}{\Delta} & : t \in [t_j, t_j + \Delta] \\ 0 & : t \notin [t_j, t_j + \Delta]. \end{cases} \quad (11)$$

The rate integral over time applied to a single event similarly results in

$$\int_0^{+\infty} M(\tau, j) d\tau = X_j. \quad (12)$$

Even though this algorithm reduces problem P3, the concerns P1, P2, and P4 are still valid. In addition, another problem emerges.

P5 All events within the current position of the window must be recorded together with their arrival times. Hence, a history is required, which complicates the measurement apparatus.

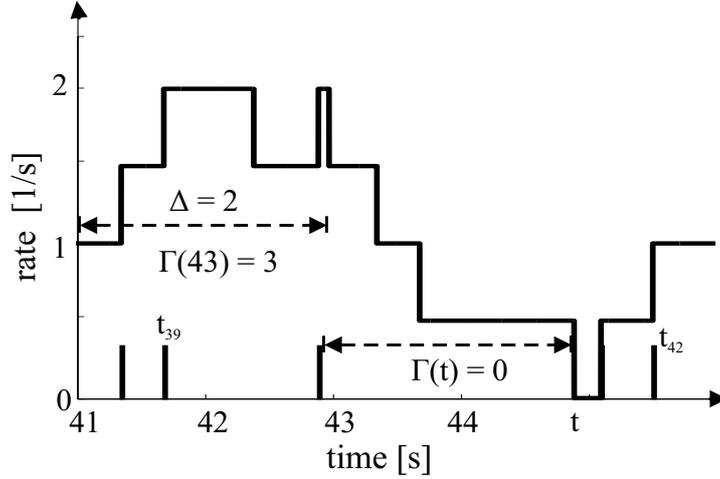


Figure 5: Rate measurement based on moving average with $\Delta = 2$.

Time Exponentially Weighted Moving Average (TEWMA) We finally derive a new algorithm called Time Exponentially Weighted Moving Average. We design this measurement approach to combine the insensitivity to random fluctuations of the EWMA-DI method with the timeliness of the MA method without the need to keep a history.

With the moving average all values in the history have the same importance. Values outside its window horizon are not considered at all. With the EWMA-DI the relevance of the values decays continuously with their age.

If we combine Equation 9 for the EWMA-DI algorithm

$$M(t, j) = (1 - \beta) \cdot \beta^k \cdot \frac{X_j}{\Delta}$$

for $t \in I + (k+1)\Delta$ with $t_j \in [t_I, t_I + \Delta]$, $k \in \mathbb{N}$ with the condition on the parameter β for EWMA - DI $\beta = 1 - \frac{\Delta}{L}$, we obtain the following result:

$$\begin{aligned} M(t, j) &= \left(1 - \left(1 - \frac{\Delta}{L}\right)\right) \cdot \left(1 - \frac{\Delta}{L}\right)^k \\ &\cdot \frac{X_j}{\Delta} = \frac{X_j}{L} \cdot \left(1 - \frac{\Delta}{L}\right)^k \end{aligned} \quad (13)$$

where k and Δ reflect the proceeding time.

Define $f(k \cdot \Delta) := \frac{X_j}{L} \cdot (1 - \frac{\Delta}{L})^k$ for fixed Δ . The function $f(k \cdot \Delta)$ reveals an exponential decay in k . The differential quotient

$$\begin{aligned} \frac{f(k \cdot \Delta) - f((k+1) \cdot \Delta)}{\Delta} &= \\ \frac{\frac{X_j}{L} \cdot (1 - \frac{\Delta}{L})^k - \frac{X_j}{L} \cdot (1 - \frac{\Delta}{L})^{k+1}}{\Delta} &= \frac{1}{L} f(k \cdot \Delta) \end{aligned}$$

motivates the substitution of $f(k \cdot \Delta)$ for $\Delta \rightarrow 0$ by the continuous function

$$f(t) = \frac{X_j}{L} \cdot e^{-\frac{1}{L} \cdot t},$$

which leads to a new measurement algorithm without measurement interval:

$$M(t, j) = \begin{cases} \frac{X_j}{L} \cdot e^{-\frac{1}{L} \cdot (t-t_j)} & : t \geq t_j \\ 0 & : \text{else.} \end{cases} \quad (14)$$

This continuous function finally yields our new algorithm TEWMA. Let

$$n(t) = \max\{j : j \in \mathbb{N}, t_j \leq t\}$$

be the number of arrivals before time t . Then the rate measured by TEWMA is

$$M(t) = \gamma \sum_{i=0}^{n(t)} X_i \cdot e^{-\gamma \cdot (t-t_j)} \quad (15)$$

with the devaluation factor $\gamma = \frac{1}{L} \in (0; 1)$. The TEWMA rate function $e^{-\gamma \cdot (t-t_j)} =: w(t, t_j)$ is an aging function that leads to an exponential decay of older values. A similar function has been applied for the estimation of the intensity of a doubly stochastic Poisson process in [14].

The property of the exponential function

$$e^{-\gamma \cdot (t-x)} = e^{-\gamma \cdot (t-y)} \cdot e^{-\gamma \cdot (y-x)}$$

for $y \in \mathbb{R}$ leads to a simple algorithmic formulation.

$$\begin{aligned} M(t_0) &= 0 \\ M(t_j) &= M(t_{j-1}) \cdot e^{-\gamma \cdot (t-t_j)} + \gamma \cdot X_j \\ M(t) &= M(t_n) \cdot e^{-\gamma \cdot (t-t_n)} \end{aligned} \quad (16)$$

with $n = n(t)$.

Obviously, no history is required and for each time instant a current value $M(t)$ can be determined. The derivation of the TEWMA measured rate function $M(t)$ as a continuous version of the EWMA rate function indicates the desired properties timeliness and insensitivity to random fluctuations of the observed process. The equivalent memory is $L = \frac{1}{\gamma}$ as can be seen from the above equations.

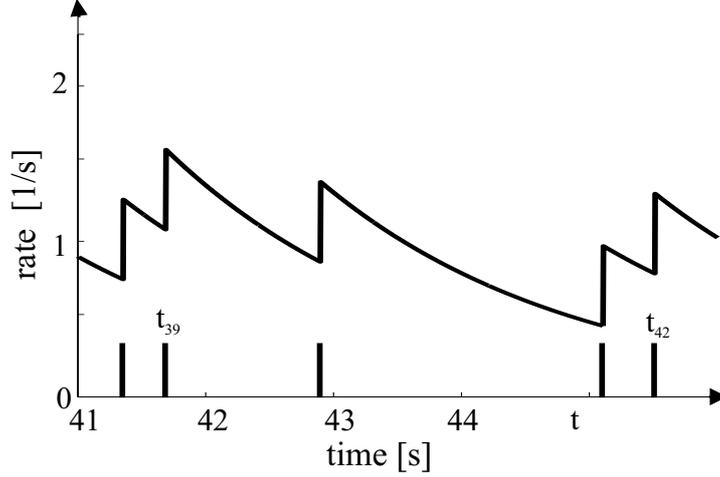


Figure 6: Rate measurement based on TEWMA with $\gamma = 0.5$.

A large devaluation parameter γ corresponds to a small memory. Therefore, the corresponding TEWMA is expected to produce jerky curves that ride closely on the current events of the process. Small devaluation factors γ yield a slow decay and, therefore, they are expected to produce smooth curves due to a long memory.

Figure 6 repeats the above example with TEWMA to give an impression of the algorithm beyond the formulas. Note that similar to MA the rate can be measured at all time instants. An object arrival increases the empirical rate. This is justified in the light of on-line measurement. We are interested in values as timely as possible without sensitivity to random fluctuations. Thus, arrivals give evidence of higher rates and the absence of arrivals gives evidence of lower rates. The increase due to arrivals and the exponential decay between arrivals is governed by γ .

After all, if γ is chosen inappropriately, this method reveals problems corresponding to P1 and P4. But those problems are inherent to measurements and reasonable parameters clearly avoid them without losing any of the positive aspects of this algorithm. Especially, the drawbacks P2 and P3 of the DI method are no issue as shown in the next section and no history is required (P5).

The rate integral over time applied to a single event

$$\begin{aligned}
 \int_0^{\infty} M(\tau, j) d\tau &= \int_{t_j}^{\infty} \gamma \cdot X_j \cdot e^{-\gamma \cdot (\tau - t_j)} d\tau \\
 &= X_j \cdot [-e^{-\gamma \cdot (t - t_j)}]_{t_j}^{\infty} \\
 &= X_j
 \end{aligned} \tag{17}$$

eventually proves that the TEWMA measurement method also respects the full object size X_j in the contribution $M(t, j)$ to the measured rate $M(t)$.

The relationship between DI, EWMA-DI, and TEWMA In the above paragraph we derived TEWMA as the continuous limit of EWMA - DI for $\Delta \rightarrow 0$. On the other hand, the DI

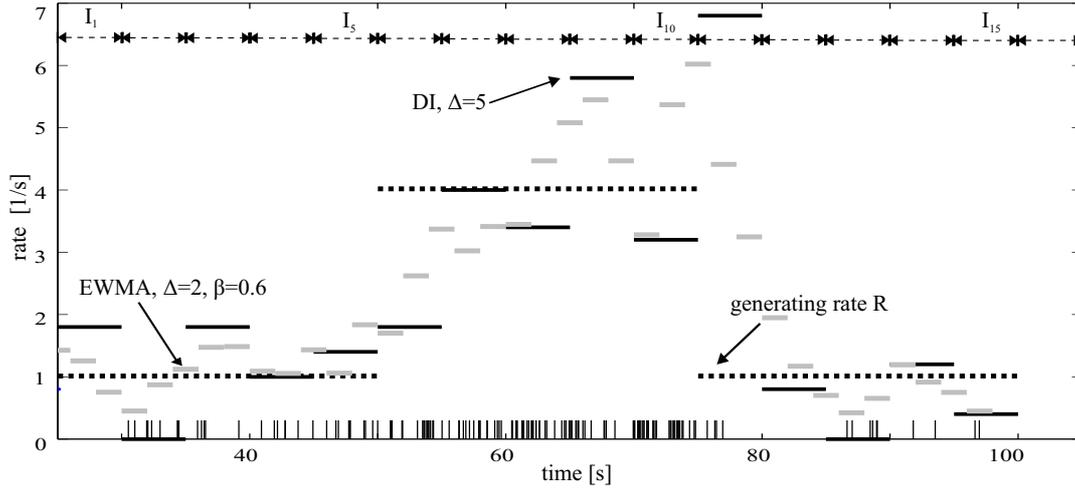


Figure 7: DI and EWMA-DI in comparison.

method is a special case of EWMA - DI with $\Delta = L$ and $\beta = 0$. This relationship between the algorithms shows that the respective methods have the same equivalent memory L , however, TEWMA yields the best time accuracy, i.e. timeliness.

4 Comparison of Rate Measurement Algorithms

The new TEWMA algorithm is designed to improve the timeliness of rate measurements without overreaction to random fluctuations. To demonstrate these properties in contrast to the other algorithms, we choose for the inter-arrival times an hyperexponential distribution with two phases and symmetry condition and relatively high coefficient of variation $c = 1.5$. Furthermore, we introduce sudden changes of the generating rate. This can be considered a worst case scenario.

We first set the parameters of the hyperexponential distribution to obtain an inter-arrival time with mean $m = 1s$ for the first 25 seconds, change the mean to $m = 0.25s$ for the next 25 seconds, and finally drop it to $m = 1s$ again for the last 25 seconds. This relates to a generating rate of $1\frac{1}{s}$ and $4\frac{1}{s}$, respectively. The generating rate is indicated by a dotted line in all figures as can be seen in Figure 7. The arrivals appear in the graphs as vertical lines at the bottom.

Figure 7 shows the scenario for DI and EWMA-DI with equivalent memory $L = 5s$, i.e. parameters $\Delta = 5s$ for DI and $\Delta = 2s$ and $\beta = 0.6$ for EWMA-DI. We plot the empirical rates M_i delayed as in the case of on-line measurement. So M_i is the freshest measured value during the next time interval I_{i+1} .

The DI method produces very jerky results. The 2nd and the 13th values go down to 0 because the measurement does not perceive any arrivals within the corresponding interval due to the variance of the underlying process. However, the generating rate is still $1\frac{1}{sec}$. The memory introduced by the parameter β of EWMA-DI relieves the problem. The result is

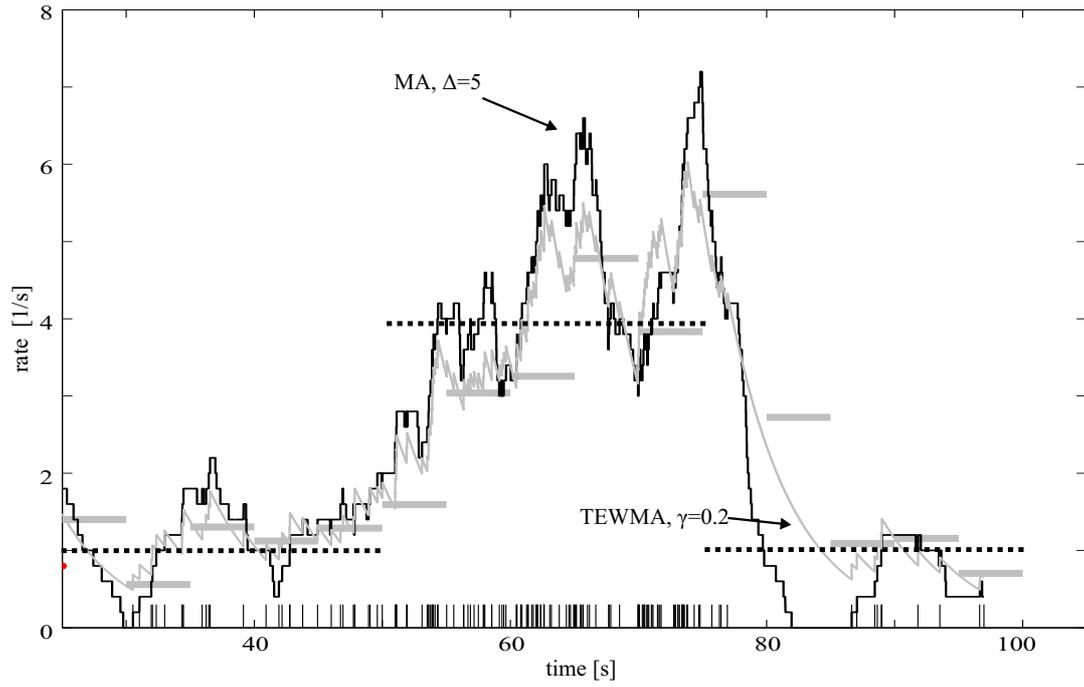


Figure 8: MA and TEWMA in comparison.

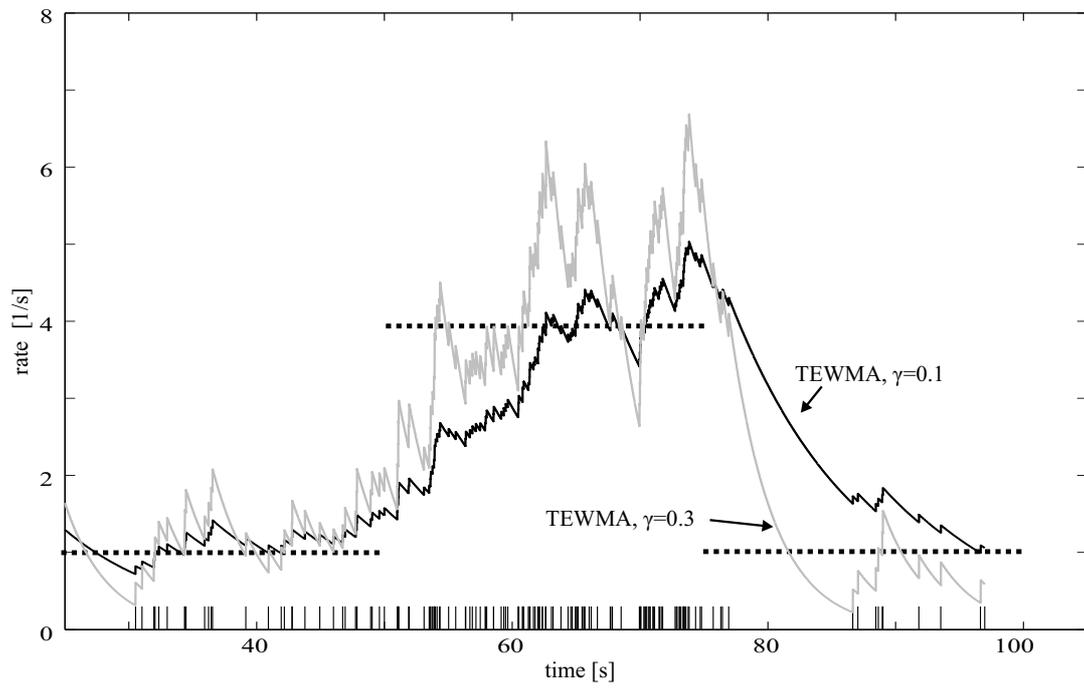


Figure 9: The effect of the devaluation factor γ on TEWMA.

less jerky and could be even improved with respect to this criterion by increasing β . But measurements should still distinguish between processes with a deterministic inter-arrival time of $1s$ and an inter-arrival time of $1s$ and a high variation. The first leads to exactly one arrival within a time window of 1 second, the latter leads to a random number of arrivals with mean 1 within the same interval. Thus, the result should not look too smooth by choosing a high value for β . Thus, on the one hand β improves rate measurement in contrast to DI by reducing the jerky course. On the other hand, β is a control parameter to adjust the sensitivity to variations.

Both methods, DI and EWMA-DI, are interval based methods and therefore their timeliness is inherently limited. Looking at sudden rate changes, the intervals I_5 to I_7 and I_{10} to I_{12} show that EWMA-DI reacts with delay in contrast to DI. This is due to the additional memory introduced by β , but it is tolerable considering the worst case character here.

Finally, EWMA-DI is already a good choice for off-line measurement, where the time gap is not so important. However, on-line measurement suffers from the delay introduced by the techniques based on intervals and the concerns mentioned in the previous section with regard to the proper selection of the time intervals are still valid.

Figure 8 shows the same scenario for MA with $\Delta = 5s$ and TEWMA with its corresponding parameter $\gamma = 0.2\frac{1}{s}$, i.e. $L = 5s$. The values for EWMA-DI are plotted here as well for the sake of comparison.

The results produced by both algorithms improve the timeliness significantly. The curves reflect the current state of the random process considering its high variance on a much finer granularity. However, MA produces jerky results and overreacts to random fluctuations. For example similar to the DI method, the empirical rate measured with MA decreases to 0 due to the lack of arrivals within an interval around 20 and 80 seconds. But as the generating rate is $R = 1$, this behavior is clearly too extreme. Opposed to that TEWMA reacts less severe to random fluctuations. Its output around the points of the sudden rate change is not severely delayed.

In comparison to the MA algorithm, the TEWMA algorithm produces smoother results. Particularly, having in mind that no history is needed, TEWMA clearly produces better results with a less complicated measurement apparatus.

In comparison to all the other algorithms, the TEWMA algorithm provides the most current values, does not overreact to random fluctuations, and allows for the easiest implementation. For each arrival at time t_i , the knowledge of the old value and the size of the arrival X_i is sufficient to update the value. Thus, TEWMA is certainly most suitable for on-line measurement.

To end with, Figure 9 shows the effect of different values for the devaluation parameter γ on TEWMA. As mentioned in the previous section, low values increase the memory and make the curve appear smoother. However, at the same time the sensitivity to the rate changes rises. So similar to the Δ and γ parameters of the other algorithms, a reasonable choice of this parameter is important.

Acknowledgment

The authors would like to thank Prof. Tran-Gia for the stimulating environment which was a prerequisite for that work.

5 Conclusion

The different time delay requirements of on-line and off-line measurements show the need for timely measurement methods. This work presented three rate measurement algorithms based on Disjoint Intervals (DI), on Moving Average (MA), and on Exponentially Weighted Moving Average over Disjoint Intervals (EWMA-DI) that can be found in literature. We introduced the notion of the equivalent memory to allow for a fair comparison of their properties with respect to timeliness and sensitivity to random fluctuations of the generating rate. While the timeliness of DI and EWMA-DI is inherently limited due to the usage of time intervals, MA overreacts to random fluctuations.

This led to the derivation of a new algorithm called Time Weighted Exponential Moving Average (TEWMA) as a continuous version of EWMA-DI. Keeping the equivalent memory of the EWMA-DI fixed, the DI method is obtained if the measurement interval size Δ approaches L , and for $\Delta \rightarrow 0$ the TEWMA method is obtained. The results illustrate that the TEWMA has the best timeliness without overreacting to random fluctuations and gives the best image of a continuous rate. Additionally, it allows for a simple algorithmic implementation as no history is required. Thus, it is the best choice for on-line rate measurement among the investigated methods.

References

- [1] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. Springer, New York, 1996.
- [2] M. Menth, J. Milbrandt, and S. Oechsner, "Experience Based Admission Control (EBAC)," *Proceedings of the The Ninth IEEE Symposium on Computers and Communications (ISCC2004)*, June 2004.
- [3] S. W. Roberts, "Control Chart Tests Based on Geometric Moving Averages," *Technometrics*, vol. 1, no. 3, pp. 239 – 250, 1959.
- [4] S. J. Hunter, "The Exponentially Weighted Moving Average," *Journal of Quality Technology*, vol. 18, no. 4, pp. 203 – 210, 1986.
- [5] S. V. Crowder, "A Simple Method for Studying Run-Length Distributions of Exponentially Weighted Moving Average Charts," *Technometrics*, pp. 401 – 407, 1987.
- [6] J. M. Lucas and M. S. Saccucci, "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," *Technometrics*, vol. 32, no. 1, pp. 1 – 29, 1990.
- [7] S. H. Steiner, "Exponentially Weighted Moving Average Control Charts with Time Varying Control Limits and Fast Initial Response," *Journal of Quality Technology*, vol. 31, pp. 75 – 86, 1999.

- [8] J. E. Wieringa, *Statistical Process Control for Serially Correlated Data*. PhD thesis, University of Groningen, Department of Econometrics, Feb. 1999.
- [9] A. Chen and E. A. Elsayed, "Design and Performance Analysis of the Exponentially Weighted Moving Average Mean Estimate for Processes Subject to Random Step Changes," *Technometrics*, vol. 44, no. 4, 2002.
- [10] D. O. Awduche, L. Berger, D.-H. Gan, T. Li, V. Srinivasan, and G. Swallow, "RFC3209: RSVP-TE: Extensions to RSVP for LSP Tunnels." <http://www.ietf.org/rfc/rfc3209.txt>, Dec. 2001.
- [11] Y. B. et al., "RFC3290: An Informal Management Model for Diffserv Routers." <http://www.ietf.org/rfc/rfc3290.txt>, May 2002.
- [12] I. S. Institute, "RFC793: Transmission Control Protocol." <http://www.ietf.org/rfc/rfc0793.txt>, September 1981.
- [13] J. Roberts, U. Mocci, and J. Virtamo, *Broadband Network Teletraffic - Final Report of Action COST 242*. Berlin, Heidelberg: Springer, 1996.
- [14] J. Virtamo, S. Aalto, and D. Down, "Window Based Estimation of the Intensity of a Doubly Stochastic Poisson Process," in *International IFIP/IEEE Conference on Broadband Communications*, (Canada), pp. 294 – 305, April 1996.