Estimation of the patient monitor alarm rate for a quantitative analysis of new alarm settings

Stijn de Waele, *IEEE Member*, Larry Nielsen, and Joseph Frassica.

Abstract — **In many critical care units, default patient monitor alarm settings are not fine-tuned to the vital signs of the patient population. As a consequence there are many alarms. A large fraction of the alarms are not clinically actionable, thus contributing to alarm fatigue. Recent attention to this phenomenon has resulted in attempts in many institutions to decrease the overall alarm load of clinicians by altering the trigger thresholds for monitored parameters. Typically, new alarm settings are defined based on clinical knowledge and patient population norms and tried empirically on new patients without quantitative knowledge about the potential impact of these new settings. We introduce alarm regeneration as a method to estimate the alarm rate of new alarm settings using recorded patient monitor data. This method enables evaluation of several alarm setting scenarios prior to using these settings in the clinical setting. An expression for the alarm rate variance is derived for the calculation of statistical confidence intervals on the results.**

I. INTRODUCTION

In environments where there are a large number of non-actionable alarms, hospital staff is at risk for developing alarm fatigue. Alarm fatigue can be thought of as a type of cognitive de-sensitization. As a result of this desensitization clinicians may not respond appropriately to critical device alarms. In one setting more than 100 alarms per patient per day were noted emanating from patient monitors alone. Among this large volume of alarms, as many as 80 – 99% have been identified as not clinically actionable [1]. This combination of high alarm rates and few actionable alarms has been found across a wide range of care settings, including the Intensive Care Unit (ICU) [2], Progressive Care Unit (PCU) [3] and Medical/Surgical floors [4].

Alarm fatigue may affect patient safety, because important clinical events may be missed. The Joint Commission has linked 98 reported sentinel events to alarms [5]. Moreover, the noise level may reduce the quality of sleep for patients, thus slowing the healing process. To stimulate improvements, the Joint Commission has published a national patient safety goal on Alarm Management [6].

There are several methods to address alarm fatigue. Some include improvements of work flow. For example, proper skin preparation of ECG electrodes improves the quality of the ECG signal, and thus reduces alarms caused by artifacts [7]. A second method which bears great promise for the longer term is the introduction of new advanced alarm-generating algorithms [8][9]. In this paper, we focus on another method to reduce the rate of clinically inactionable alarms by adjustment of the default alarm settings in patient monitors. Several studies have shown the potential to reduce the alarm rate by the adjustment of alarm settings [3][10].

We introduce a technology to support the introduction of new default alarm settings by estimating the alarm rate for new settings using recorded vital signs, alarms and patient information. This allows for estimation of the alarm rate for a specific patient population in a selected unit.

An accurate comparison of different settings is enabled by applying them to the vital signs of a single patient population, because the variability introduced by applying different settings to a new population is eliminated.

Alarm regeneration with new settings is discussed in section II; section III describes the data collection system. The estimator for the alarm rate and its confidence interval are derived in section IV. Illustrative results of alarm regeneration applied to the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) II database [11] are given in section V.

II. ALARM REGENERATION

With alarm regeneration, alarms corresponding to given alarm parameter settings are generated by applying the alarm algorithm with these settings to continuously recorded vital signs. The sampling frequency must be sufficient to allow for accurate application of the algorithm. The resulting alarms are referred to as <u>generated alarms</u>. Conversely, <u>recorded alarms</u> are alarms as they actually occurred during the data collection.

Alarm regeneration delivers an indication of expected alarm rate with new settings before applying them to new patients. Also, it allows for a quick review of several proposed alarm settings.

In the context of alarm regeneration, it is useful to categorize alarms by the way they can be configured, as follows:

- <u>Threshold alarms</u>: A vital signs exceeds a threshold value L, e.g. `**SpO2 94<96`. Threshold alarms typically are the majority of occurring alarms.

Stijn de Waele is with Philips Research North America, Briarcliff Manor, NY 10510, USA. Phone: 914 945 6039; email: stijn.de.waele@philips.com

Larry Nielsen and Joseph Frassica are with Philips Healthcare, Andover, MA 01810, USA.

- Tunable alarms: A parameter p exceeds a threshold P, where the parameter is not a vital sign, e.g. `***Apnea 0:21`. In this example, the tunable parameter is the apnea duration threshold. The apnea alarm is the major tunable alarm. It requires the respiratory waveform signal for accurate regeneration.
- On/off alarms: The algorithm has no user-adjustable parameters, e.g.: `*PAIR PVC`. Under the condition that an on/off alarm was 'on' during the data collection, alarm rate estimation for new settings is trivial for this type of alarms; the new setting 'off' simply means that the contribution of this alarm to the alarm rate is zero.

Alarms may include the following filtering mechanisms to reduce the alarm rate:

- Delay: Suppress the alarm if it lasts less than D seconds.
- Inhibition: Suppress the alarm if there was an alarm of the same type in the last I seconds.

Given the high prevalence of threshold alarms, and the trivial nature of alarm rate estimation for on/off alarms, the remainder of this paper will focus on threshold alarms.

To evaluate the performance of a new default alarm setting, the alarm algorithm is applied to all patients using this new setting. More complex alarm setting scenarios can be defined. For example, using recorded alarm setting data, a scenario can be defined where the default values that were active during the recording are replaced by new defaults, but any modifications made during the recording are maintained.

If only recorded alarms are available, approximate alarm regeneration can be performed by applying new thresholds to the extremum present in the alarm text, e.g. 94 in the alarm text `**SpO2 94<96`. A characteristic of this technique is that the results show the combination effect of vital signs data, and thresholds set during the recording. This approximation works well for pure threshold alarms, i.e. threshold alarms which do not involve alarm filtering (delay or inhibition), but can lead to poor results if filtering is present. Application of new thresholds to the extremum present in the alarm text can also be used for alarm regeneration of the apnea alarm, in the absence of waveform data.

III. Data Collection

To enable alarm regeneration as described above, the following data is -collected during a data collection period for all beds in the units of interest:

- **Patient monitor alarms**
  Alarm start time and alarm message. An example alarm text is `***TACHY 147>140`, which can be used to derive severity (`***`→Critical), originating modality (`TACHY`→Heart rate), the alarm threshold (`140`) and the extreme value (`147`). The extreme value must be taken over the entire duration of the alarm.
- **Vital signs**
  vital signs sampled at a 1 second sampling rate
- **Basic encounter information**
  used to associate alarms and vital signs to a patient/bed/unit

The size of the data set can be expressed by the number of patient days. A typical size is 1500 patient days, obtained during a 15 day data collection period for 100 beds.

A dedicated recording PC system has been developed to simultaneously collect this information from a large number of Philips patient monitors. The current system can handle up to 128 beds, and can be expanded to handle up to 1000 beds. This system is based on the Philips IntelliVue Information Center (PIIC) iX Mobility server [12], which is installed temporarily in the hospital site during the data collection period. The data is stored in a MariaDB database [13].

IV. Alarm Rate Estimation

A. Definition

The mean alarm rate is defined as the expectation of the alarm count c generated by a single patient divided by a time interval t:

$$r = \mathrm{E}\left[\frac{c}{t}\right], \tag{1}$$

With r is given in counts/patient/day. To arrive at the alarm rate as experienced by a clinician, the rate is multiplied by the number of patients under his care. When partitioning the time interval in N patient encounters, each with their individual alarm rate r_i and encounter duration (length of stay) d_i, the mean alarm rate can be estimated as:

$$\hat{r} = \frac{\sum_i c_i}{\sum_i d_i}, \tag{2}$$

Please note that this expression is in terms of the alert *count* c_i i.e. not normalized to an alarm rate. To express the alert rate in the encounter alert rate, the rate must be weighted proportionally by the encounter duration:

$$\hat{r} = \frac{\sum_i d_i r_i}{\sum_i d_i} = \frac{1}{N}\sum_i \frac{d_i}{D} r_i, \tag{3}$$

where D is the average encounter duration. The practical interpretation of the weight factor d_i/D is that reducing false alarm rate for a patient with a combination of a high alarm rate *and* a long stay results in the greatest reduction of alarm rate. Computationally, it means that the rate estimate is not equal to the normal (unweighted) average of the rate per encounter. The current definition of the alarm rate is motivated by the fact that it directly relates to the average rate as experienced by a clinician.

B. Confidence interval

For the calculation of the confidence interval, we assume that the estimate for the mean alarm rate $\hat{r}$ has a Gaussian distribution. This is motivated by the fact that $\hat{r}$ is an average of a large number of stochastic variables with a similar distribution. The distribution of this type of variable tends to a Gaussian distribution according to the central limit theorem [14, p. 194]. Assuming this distribution, the 95% confidence interval for $\hat{r}$ is given by $[\hat{r} - 1.96\hat{\sigma}, \hat{r} - 1.96\hat{\sigma}]$, where $\hat{\sigma}$ is the estimated standard deviation, calculated as the square root of the estimated variance $\hat{V}$.

A method for estimating the variance of an estimated mean is to divide the data into partitions [15]. Considering each partition as a sample, the variance is now derived from the sample mean. All encounters are concatenated on the time axis. The time axis is split into equal partitions of length t_σ. The alarm rate estimate within partition k is denoted $r_{t_\sigma,k}$, the number of partitions is M. The rate estimate $\hat{r}$ is the average rate over all partitions:

$$\hat{r} = \frac{1}{M}\sum r_{t_\sigma,k}, \quad (4)$$

Assuming independence between partitions, the variance of $\hat{r}$ can be estimated as:

$$\hat{V}^* = \frac{1}{M}\hat{V}_S, \quad (5)$$

where $\hat{V}_S$ is the partition sample variance:

$$\hat{V}_S = \frac{1}{M}\sum\left(r_{t_\sigma,k} - \hat{r}\right)^2. \quad (6)$$

Combining (5) and (6), we find:

$$\hat{V}^* = \frac{1}{M^2}\sum\left(r_{t_\sigma,k} - \hat{r}\right)^2. \quad (7)$$

The partition length t_σ must be chosen such that the dependence between partitions is negligible, as required for the validity of (5). However, a too large value will leave too few segments, resulting in an increased variance of the estimate. It was found that a partition length $t_\sigma = 5D$ is a reasonable choice. Because partition boundaries typically are not aligned with encounter boundaries, this estimator requires the knowledge of distribution of alarms over time within encounters. This is resolved by aligning partition boundaries with encounter boundaries, leading to the estimator $\hat{V}$:

$$\hat{V} = \frac{1}{N^2}\sum_k \frac{d_i}{D}(r_i - \hat{r})^2. \quad (8)$$

In practice, this estimator provides similar results as the segment based estimate (7) and is easier to use because it only depends on the alarm rates r_i and duration d_i per encounter. For small datasets, a model-based approach may be considered as described in [16] for time series, resulting in an accurate variance estimate also for small datasets.

V. Illustrative Results

The alarm rate estimators are applied to a subset of the publically available Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) II database [11]. The alarm load estimation has been applied to heart rate (HR), oxygen saturation (SpO_2) and respiratory rate (RR) data in the database, which has been sampled at 1 Hz frequency. These signals, along with recorded alarms, have been converted to the same MariaDB database used in the recording PC (see section IV). The analyzed data covers 500 encounters with an average length of stay of D =3.2 days.

Our analytics application produces a range of reports based on the recorded alarms, such as the top alarm types per severity, per unit, etc., as reported in literature [3][4][9][10]. A further useful report is the alarm rate as a function of the encounter, where the encounters are ordered from high to low alarm rate. Expressed in basic statistical functions, this is the inverse survival function s^{inv}:

$$s^{inv}(p) = X: P(x > X) = p,$$

where p is the encounter fraction, $P(x > X)$ is the survival function for the alarm rate X: the probability that $x > X$. The survival function is the reverse of the cumulative histogram:

$$P(x > X) = 1 - P(x \le X).$$

Figure 1 shows the inverse survival function for the alarm count per encounter c_i. Because of its direct contribution to the mean alarm rate (2), the encounter alarm *count* c_i is reported here, as opposed to the encounter alarm *rate* d_i. The distribution can be characterized by the fraction of encounters that cause a given fraction of the total alarms. For example, for this MIMIC-II subset, it is found that 50% of the alarms are caused by 12% of encounters with the highest alarm count. This statistic shows the potential alarm rate reduction that can be achieved by addressing a small percentage of patients.

Figure 2 shows the alarm rate estimates as a function of the SpO_2 alarm threshold using alarm regeneration, with the 95% confidence interval as calculated the variance estimate (5). These results allow for an assessment of the alarm rate reduction achieved by changes in default thresholds. The figure shows the example of decreasing the SpO_2 low alarm threshold from 90% to 86%, resulting in a predicted alarm rate reduction of 61%. Also shown is the histogram for recorded alarm thresholds. The total recorded SpO_2 Low rate is 11.6 alarms/patient/day.

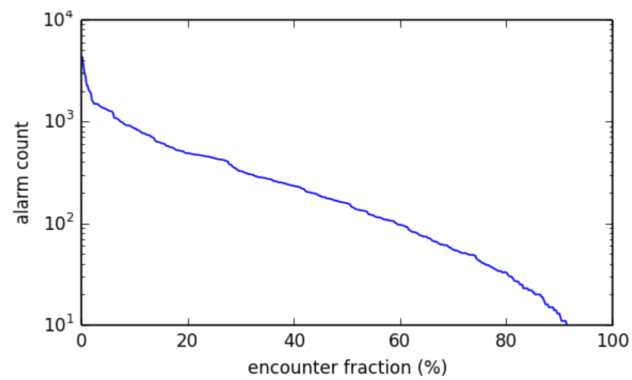

Figure 1 The inverse survival function for the alarm rate per encounter for the MIMIC-II subset. The inverse survival function shows the encounters ordered according to alarm rate, from high to low. 50% of the total alarm rate is caused by the 12% of encounters with the highest alert count.

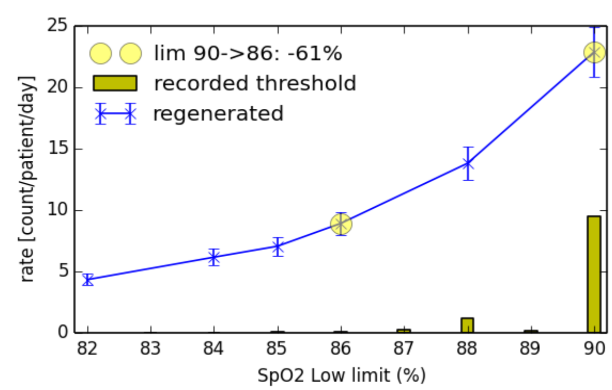

Figure 2 Alarm rate estimates as a function of the alarm threshold for the SpO_2 Low alarm. The scenario of a threshold reduction from 90% to 86% is also shown. The bars represent the frequency of thresholds in the recorded alarms.

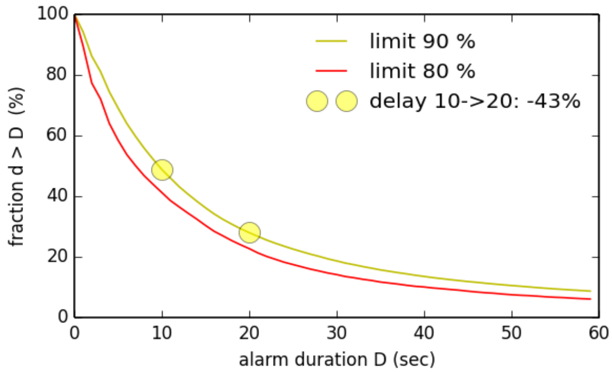

Figure 3 The survival function for the SpO_2 low alarm shows the fraction of alarms with a duration longer than D seconds. The alarm load reduction achieved by increasing the alarm delay can be directly derived from this graph.

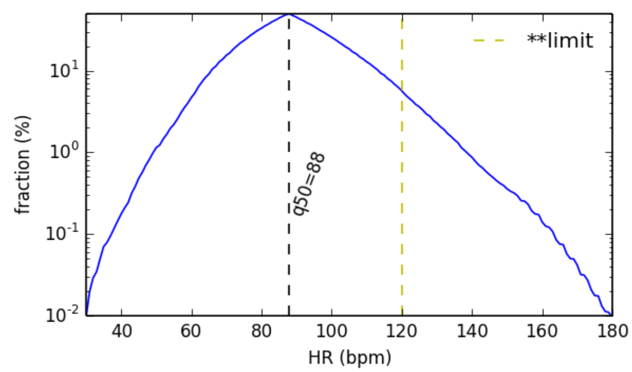

Figure 4 The folded cumulative histogram for heart rate (HR). To facilitate visual evaluation of the distribution for both low and high extreme values, the histogram F is replaced by the survival function 1-F for values greater than the median.

A further measure of interest is the alarm duration. The survival function for the alarm duration for the SpO_2 low alarm based on setting type A are given in figure 3. This graph shows the alarm rate reduction that can be achieved by increasing the alarm delay. The example scenario where we increase the alarm delay to 20 seconds results in an alarm rate reduction of 43% relative to an alarm delay of 10 seconds.

Besides duration per alarm, also the fraction of time an alarm condition is met across the entire population can be calculated. This can be derived efficiently from the folded cumulative histogram of the vital sign values – see figure 4 for the heart rate histogram.

For the MIMIC-II subset, the high threshold for heart rate of 120 bmp is violated for 5% of the time, or 3 minutes per hour. For High Priority alarms, the alarm continues to sound as long as the alarm condition is met, unless it is silenced earlier. Hence, the produced sound level for High Priority alarms is expected to be correlated to the alarm *duration*. Conversely, if an alarm condition is configured as Low Priority, an alarm sound is produced of a fixed duration. Consequently, for Low Priority alarms, it is expected that the sound level is most strongly correlated with the alarm *rate*.

VI. Conclusion

Alarm regeneration was introduced as a method to quantify the alarm rate with new alarm thresholds. This allows for a fast evaluation of several alarm settings, prior to applying them in the clinical setting. This will support clinicians in defining new settings aimed at reducing alarm fatigue, based on data from their own, unique patient population.

With this method clinicians can identify new thresholds that could significantly reduce alarm fatigue, while at the same time using their own judgment to set bounds on how wide thresholds can be safely set. Given sufficient resources, this judgment can be supported by recording the occurrence of clinical events along with vital signs and alarms. With this data, performance statistics (true/false alarm rate) can be calculated for both recorded and generated alarms.

References

[1] M. Cvach, "Monitor alarm fatigue – an integrative review", *Biomedical Instr. & Tech.*, July/August 2012, pp. 268-277.
[2] M. Görges *et al.*, "Improving alarm performance in the medical intensive care unit using delays and clinical context"
[3] K. C. Graham and M. Cvach, "Monitoring alarm fatigue: standardizing the use of physiological monitors and decreasing nuisance alarms", *American Journal of Critical Care*, January 2010, Vol 19, No. 1, pp. 28-34.
[4] B. Gross *et al.*, "Physiologic monitoring alarm load on medical/surgical floors of a community hospital", Horizon, Spring 2011, pp. 29-36.
[5] The Joint Commission, "Sentinel event alert", Issue 50, April 2013.
[6] Joint Commission, "2014 National Patient Safety Goal", *Joint Commission Perspectives*, 33(7), July 2013.
[7] American Association of critical-care nurses, "Practice alert alarm management", May 2013.
[8] M. Imhoff *et al.*, "Smart alarms from medical devices in the OR and ICU", *Best Practice & Research Clinical Anaesthesiology* 23, 2009, pp. 39-50.
[9] A. Aboukhalil *et al.*, "Reducing false alarm rates for critical arrhythmias: a large-scale MIMIC II study", *Journal of Biomedical Informatics*, June 2008; 41(3):442-451.
[10] D. A. Whalen *et al.*, "Novel Approach to Cardiac Alarm Management on Telemetry Units", *Journal of Cardiovascular Nursing*, December 2013.
[11] M. Saeed *et al.*, "Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database", *Crit. Care Medicine*, 39(5), 2011, pp. 952-60.
[12] IntelliVue Information Center iX – Service and Installation guide version A.02, 2014.
[13] MariaDB database: https://mariadb.org/
[14] A. M. Mood *et al.*, *Introduction to the theory of statistics*, McGraw-Hill, 1974.
[15] H. Flyvberg and H.G. Petersen, "Error estimates on averages of correlated data", *J. Chem. Phys.*, vol.91, pp. 461-466, 1989.
[16] P.M.T. Broersen, "Estimation of the Accuracy of Mean and Variance of Correlated Data", *IEEE Transactions on Instrumentation and Measurement*, vol. 47(5), 1998.