Optimized EWMA Control Charts in Function of Intrusion Detection

Petar Čisar

Telekom Srbija, Subotica, petarc@telekom.yu

Sanja Maravić Čisar

Subotica Tech, Subotica, sanjam@vts.su.ac.yu

Abstract: Intrusion detection is used to monitor and capture intrusions into computer and network systems which attempt to compromise the security of computer and network systems. Many intrusions manifest in dramatic changes in the intensity of events occuring in computer networks. Because of the ability of exponentially weighted moving average control charts to monitor the rate of occurrences of events based on their intensity, this technique is appropriate for implementation in intrusion detection systems.

Keywords: EWMA, control charts, optimization, network traffic

1 Introduction

The exponentially weighted moving average (EWMA) is a statistic for monitoring the process that averages the data in a way that gives less and less weight to data as they are further removed in time. For the EWMA control technique, the decision regarding the state of control of the process depends on the EWMA statistic, which is an exponentially weighted average of all prior data, including the most recent measurements.

By the choice of weighting factor λ, the EWMA control procedure can be made sensitive to a small or gradual drift in the process.

The statistic that is calculated is:

$EWMA_t = \lambda\, Y_t + (1-\lambda)\, EWMA_{t-1}$ for $t = 1, 2, ..., n$

where

- $EWMA_0$ is the mean of historical data (target)

- Y_t is the observation at time t
- n is the number of observations to be monitored including $EWMA_0$
- $0 < \lambda \leq 1$ is a constant that determines the depth of memory of the EWMA.

This equation is due to Roberts (1959).

The parameter λ determines the rate at which «older» data enter into the calculation of the EWMA statistic. A value of $\lambda = 1$ implies that only the most recent mesurement influences the EWMA. Thus, a large value of $\lambda = 1$ gives more weight to recent data and less weight to older data - a small value of λ gives more weight to older data. The value of λ is usually set between 0.2 and 0.3 (Hunter) although this choice is somewhat arbitrary.

In real situations, the exact value of the shift size is often unknown and can only be reasonably assumed to vary within a certain range. Such a range of shifts deteriorates the performance of existing control charts. The most usual quality control procedures used in intrusion detection systems, such as the cumulative sum (CUSUM) and EWMA charts are based on a mean shift with a given size. This shift can be caused by intrusion or attack, for example. Lucas and Saccucci (1990) have shown that although the smoothing factor λ used in an EWMA chart is usually recommended to be in the interval 0.05 to 0.25, in practice the optimally designed smoothing factor depends not only on the given size of the mean shift δ, but also on a given in-control Average Run Length (ARL). With shift = δ/σ_Q and ARL = 370, optimal choices are given by the following figure (an average line has been drawn for a lot of ARLs instead of one line for each ARL) [10].

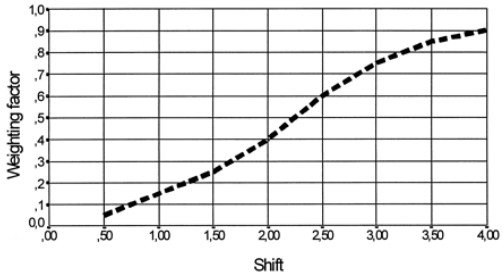

Figure 1
Weighting factor

The estimated variance of the EWMA statistic is approximately:

$\sigma^2_{EWMA} = (\lambda / (2 - \lambda)) \sigma^2$

when t is not small, where σ is the standard deviation calculated from the historical data.

The center line for the control chart is the target value or $EWMA_0$. The upper and lower control limits are:

$UCL = EWMA_0 + m\sigma_{EWMA}$

$LCL = EWMA_0 - m\sigma_{EWMA}$

where the factor m is either set equal 3 (the 3-sigma control limits) or chosen using the Lucas and Saccucci (1990) tables (ARL = 370):

Table 1
Choice of m

λ	0.05	0.1	0.2	0.3	0.4	0.5	0.75	1
m	2.49	2.70	2.86	2.93	2.96	2.98	3.00	3.00

To illustrate the construction of an EWMA control chart, consider a process with the following parameters calculated from historical data: $EWMA_0 = 50$ and $\sigma = 2.0539$, with λ chosen to be 0.3 so that $\lambda / (2- \lambda) = .3 / 1.7 = 0.1765$ and the square root = 0.4201. The control limits are given by:

UCL = 50 + 3 (0.4201) (2.0539) = 52.5884
LCL = 50 - 3 (0.4201) (2.0539) = 47.4115

Consider the following data consisting of 20 points:

52.0 47.0 53.0 49.3 50.1 47.0

51.0 50.1 51.2 50.5 49.6 47.6

49.9 51.3 47.8 51.2 52.6 52.4

53.6 52.1

These data represent control measurements from the process which is to be monitored using the EWMA control chart technique. The corresponding EWMA statistics that are computed from this data set are:

$EWMA_0 = 50$

$EWMA_1 = (0.3)52 + (1 - 0.3)50 = 50.60$

$EWMA_2 = (0.3)47 + (1 - 0.3)(50.6) = 49.52$

50.00 50.60 49.52 50.56 50.18

50.16 49.12 49.75 49.85 50.26

50.33 50.11 49.36 49.52 50.05

49.34 49.92 50.73 51.23 51.94

The control chart is given below.

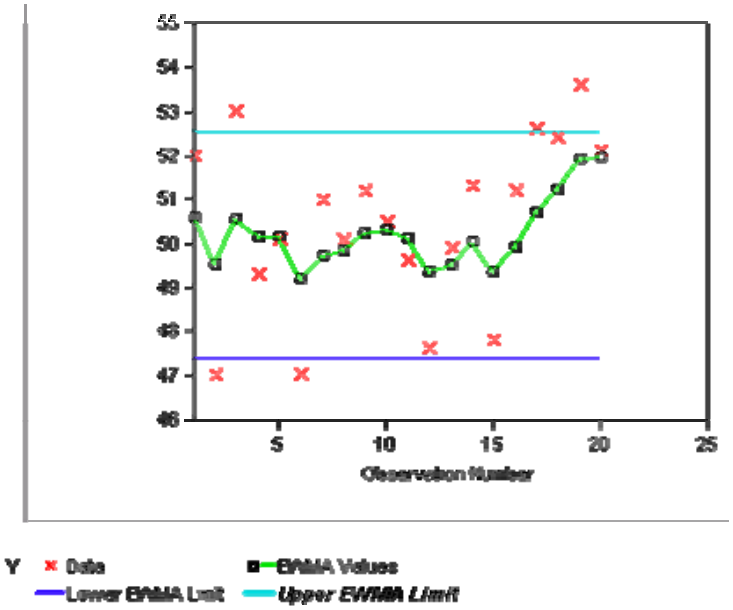

Figure 2
EWMA control chart

2 Exponential Smoothing

For any time period t, the smoothed value S_t is determined by computing:

$S_t = \lambda y_{t-1} + (1 - \lambda) S_{t-1}$ where $0 < \lambda \leq 1$ and $t \geq 3$

This is the basic equation of exponential smoothing. The formulation here is given by Hunter (1986).

This smoothing scheme begins by setting S_2 to y_1, where S_i stands for smoothed observation or EWMA, and y_i stands for the original observation. The subscripts refer to the time periods 1, 2, ..., n. For example, the third period is $S_3 = \lambda y_2 + (1 - \lambda) S_2$ and so on. There is no S_1. The optimal value for λ is the value which results in the smallest mean of the squared errors (MSE). Here is an illustration of this principle through an example. Consider the following data set consisting of n observations of data flow over time–for starting $\lambda = 0.1$:

Table 2
Smoothing scheme

Time	Flow (y_t)	S_t	Error ($y_t - S_t$)	Error squared
1	y_1			
2	y_2	y_1	E_2	E_{22}
3	y_3	S_3	E_3	E_{32}
...	...	...	...	...
n	y_n	S_n	E_n	E_{n2}
				SSE_n

The sum of the squared errors (SSE) is $SSE_{0.1}$. The mean of the squared errors is $MSE_{0.1} = SSE_{0.1} / (n - 1)$. After that, the MSE is calculated for $\lambda = 0.2$. If $MSE_{0.2} < MSE_{0.1}$ then $MSE_{0.2}$ is better value for λ. This iterative procedure is related to the range of λ between 0.1 and 0.9. In this way, the best initial choice for λ is determined and then, for getting more precise value, search optionally continues between $\lambda-\Delta\lambda$ and $\lambda+\Delta\lambda$.

The initial EWMA plays an important role in computing all the subsequent EWMA's. There are several approaches in defining this value:

1) Setting S_2 to y_1

2) Setting S_2 to the target of the process

3) Setting S_2 to average of the first four or five observations

It can also be shown that the smaller the value of λ, the more important is the selection of the initial EWMA. The user would be wise to try a few methods, before finalizing the settings.

2.1 Example 1

1) $S_2 = y_1$, data used from chapter 1

Table 3
Smoothing-the first iteration

Time	y_t	S_t ($\lambda = 0.1$)	Error ($y_t - S_t$)	Error squared
1	52.0			
2	47.0	52.0	-5	25
3	53.0	51.5	1.5	2.2
4	49.3	51.7	-2.4	5.5
5	50.1	51.4	-1.3	1.7
6	47.0	51.3	-4.3	18.3
7	51.0	50.9	0.1	0

8	50.1	50.9	-0.8	0.6
9	51.2	50.8	0.4	0.2
10	50.5	50.8	-0.3	0.1
11	49.6	50.8	-1.2	1.4
12	47.6	50.7	-3.1	9.5
13	49.9	50.4	-0.5	0.2
14	51.3	50.3	1	1
15	47.8	50.4	-2.6	6.9
16	51.2	50.2	1	1.1
17	52.6	50.3	2.3	5.5
18	52.4	50.5	1.9	3.6
19	53.6	50.7	2.9	8.5
20	52.1	51	1.1	1.3

SSE	92.6
MSE	4.874

λ	0.1	0.2	0.3	0.4	0.5	0.6
MSE	4.874	4.594	4.528	4.623	4.843	5.169

It is obvious that $\lambda = 0.3$ is the minimal value. Better precision is obtained by shortening the observation interval around 0.3.

Table 4
Smoothing-the second iteration

λ	0.27	0.28	0.29	0.3	0.31
MSE	4.529	4.5268	4.5265	4.528	4.531

So, the optimal calculated value for λ is 0.29. Now, $\lambda = 0.29 \rightarrow m = 2.93$.

The optimized control limits are given by:

UCL = 50 + (2.93)(0.4118)(2.0539) = 52.4782
LCL = 50 - (2.93)(0.4118)(2.0539) = 47.5218

2) $S_2 = 50.4$ - the mean of all samples

λ	0.1	0.2	0.3	0.4
MSE	3.895	3.897	3.967	4.128

λ	0.02	0.04	0.06	0.08	0.1
MSE	3.81	3.85	3.878	3.891	3.895

The value for S_2 is unacceptable because it leads to $\lambda \approx 0$.

3) $S_2 = 50.3$ - the mean of the first 4 samples

λ	0.1	0.2	0.3	0.4	0.5
MSE	3.88	3.87	3.94	4.11	4.35

λ	0.1	0.12	0.14	0.16	0.18	0.2
MSE	3.880	3.878	3.876	3.874	3.875	3.878

So, the optimal calculated value for λ is 0.16. Then, if λ = 0.16 --> m = 2.8.

The optimized upper control limit is given by:

UCL = 50 + (2.8)(0.295)(2.0539) = 51.697

The value for S_2 is unacceptable because it will generate a false alarm ($EWMA_{max}$ = 51,94).

2.2 Example 2 – Network Traffic

The smoothing method will be applied to authentical data obtained by the traffic analizer at the Polytechnical Engineering College in Subotica, whose grafical review of traffic intensity for a longer period of observation is given by the following diagram:

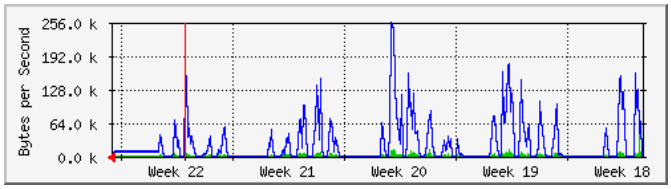

Figure 3
Monthly graph

Data from the following table represent local maximums which are determined by direct readout from the previous diagram.

Time	Flow (y_t)
1	60
2	205
3	100
4	70
5	254
6	125
7	170
8	180
9	115
10	155
11	85
12	78

According to these data: $EWMA_0 = 133.083$ and $\sigma = 60.3$.

These values will be used as historical values for further calculations in smoothing process on much shorter time interval.

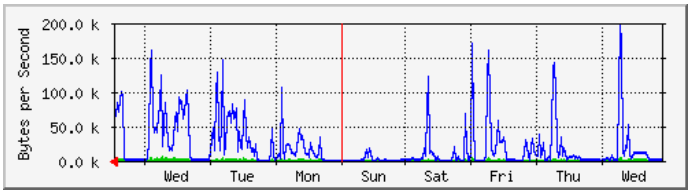

Figure 4
Weekly graph

1) $S_2 = y_1$

Time	Flow (y_t)	S_t ($\lambda = 0.1$)	Error ($y_t - S_t$)	Error squared
1	65			
2	165	65	100	10000
3	100	75	25	625
4	70	77.5	-7.5	56.3
5	200	76.8	123.3	15190.6
6	130	89.1	40.9	1674.9
7	170	93.2	76.8	5903.2
8	180	100.9	79.1	6264.6
9	115	108.8	6.2	38.9
10	155	109.4	45.6	2080.4
11	90	114	-24	576
12	35	111.6	-76.6	5860.7

The sum of the squared errors is $SSE_{0.1}$=48268.0. The mean of the squared errors is $MSE_{0.1} = SSE_{0.1}/11$=4388.0. In a similar way the following table is created:

λ	0.1	0.2	0.3	0.4	0.5	0.6
MSE	4388.0	3855.82	3653.44	3569.73	3570.91	3650.67

It is obvious that $\lambda = 0.4$ is the minimal value. Better precision is obtained by shortening the observation interval around 0.4.

λ	0.43	0.44	0.45	0.46	0.47
MSE	3561.59	3560.51	3560.25	3560.79	3562.13

So, the optimal calculated value for λ is 0.45. Then, $\lambda = 0.45$ --> $m = 2.97$.

The optimized control limits are given by:

UCL = 133.083 + (2.97)(0.5388)(60.3) = 229.58
LCL = 133.083 - (2.97)(0.5388)(60.3) = 36.59

Now, data and EWMA values are:

65 165 100 70

200 130 170 180

115 155 90 35

133.08 102.44 130.59 116.83

95.75 142.67 136.97 151.83

164.51 142.23 147.98 121.89

2) $S_2 = 122.92$ - the mean of all samples

λ	0.1	0.2	0.3	0.4
MSE	2599.96	2732.13	2815.12	2880.53

λ	0.02	0.04	0.06	0.08	0.1
MSE	2448.18	2489.26	2528.58	2565.58	2599.96

This value for S_2 leads to $\lambda \approx 0$.

3) $S_2 = 100$ - the mean of the first 4 samples

λ	0.1	0.2	0.3	0.4	0.5	0.6
MSE	2961.49	2975.71	3003.92	3039.31	3101.96	3206.91

λ	0.04	0.06	0.08	0.1
MSE	3024.50	2989.70	2970.44	2961.49

So, the optimal calculated value for λ is 0.1. Now, λ = 0.1 --> m = 2.7.

Then, the optimized control limits are:

UCL = 133.083 + (2.7)(0.229)(60.3) = 170.37

LCL = 133.083 - (2.7)(0.229)(60.3) = 95.80

Conclusions

This paper gives an overview of the EWMA control charting technique suitable for detecting intrusions in network systems. By the method of exponential smoothing it is shown how to compute the optimal value of parameter λ. This technique can be made more effective if the time component is also included. In that sense, an intrusion is detected if traffic intensity exceeds UCL with lasting of anomaly longer than a defined time period.

References

[1] D. Seibold: Enterprise Campus Security–Addressing the Imploding Perimeter, http://www.itsa.ufl.edu/2003/presentations/IntSec.ppt

[2] A Vasilios, S. and F. Papagalou: Application of Anomaly Detection Algorithms for Detecting SYN Flooding Attacks, http://www.ist-scampi.org/publications/papers/siris-globecom2004.pdf

[3] S. Sorensen: Competitive Overview of Statistical Anomaly Detection, White Paper, Juniper Networks, 2004

[4] Engineering Statistics Handbook–EWMA Control Charts, http://www.itl.nist.gov/div898/handbook/pmc/section3/pmc324.htm

[5] Y. Zhao, F. Tsung and Z. Wang: Dual CUSUM Control Schemes for Detecting a Range of Mean Shifts, IEEE Transactions, 2005., http://qlab.ieem.ust.hk/qlab/download/papers/paper%2035.pdf

[6] G. Fengmin: Deciphering Detection Techniques: Part II Anomaly–Based Intrusion Detection, White Paper, McAfee Security, 2003

[7] V. A. Mahadik, X. Wu and D. S. Reeves: Detection of Denial-of-QoS Attacks Based on $\chi 2$ Statistic And EWMA Control Charts, http://arqos.csc.ncsu.edu/papers/2002-02-usenixsec-diffservattack.pdf

[8] S.W. Roberts: Control Chart Tests Based on Geometric Moving Averages, Technometrics, 1959

[9] J. Viinikka and H. Debar: Monitoring IDS Background Noise Using EWMA Control Charts and Alert Information, http://viinikka.info/ViiDeb2004.pdf

[10] A. S. Neubauer: The EWMA Control Chart: Properties and Comparison with other Quality-Control Procedures by Computer Simulation, Clinical Chemistry, http://www.clinchem.org/cgi/content/full/43/4/594

[11] Engineering Statistics Handbook–Single Exponential Smoothing, http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc431.htm