A New Approach in Tracking Efficiency of Anti-Spam Filter

Slobodan Mitrović[1], Valentina Radojičić[2]

***Abstract* – With the growing number of spam attacks, monitoring of Mail Transfer Agent (MTA) and anti-spam filter is required. Time series analysis and forecasting models for *received messages/identified spam* ratio could be used for performance assessment of some modules of the system. In this paper a simple approach in tracking efficiency of anti-spam filter is considered. Presented experimental results are based on the log data of the MTA and anti-spam filter of The Faculty of Transport and Traffic Engineering Computer Center in Belgrade.**

***Keywords* – monitoring, anti-spam filter, time series, efficiency.**

I. Introduction

During the last several years anti-spam protection has became equally important as protection against viruses or hacking, because of emerging growth of unsolicited email. Various techniques for filtering spam have been developed and implemented in software solutions. However, spam emails are frequently changing their forms in order to decrease efficiency of filtering, which can be assessed by monitoring process.

Based on this observation, a simple approach in tracking efficiency of anti-spam filter is proposed. MTA log files can be used for calculation of *received messages/identified spam* ratio which can be presented in time series form. This form is suitable for analysis and forecasting techniques can be applied as well. By assessment of real and predicted data, efficiency of given filter can be determined.

In Section II the concept of used MTA and anti-spam filter monitoring is described. Section III briefly presents prediction techniques applied to a given time series. Section IV presents our experimental results followed by corresponding conclusions.

II. Monitoring Email Statistics

Email servers based on LINUX/BSD platforms are suitable for integration with different kinds of anti-spam and anti-virus filters, as well as various types of logging and monitoring tools. One of the most popular MTAs is *Postfix* which can be successfully integrated with syslog engines such as *Syslog-NG*. In this way all activities related to email processing can be recorded in log files. The Faculty of Transport and Traffic Engineering Computer Center in Belgrade uses combination of *RRDTool* engine and *Mailgraph* script in order to collect statistics and presents them in graphical manner, Fig. 1.

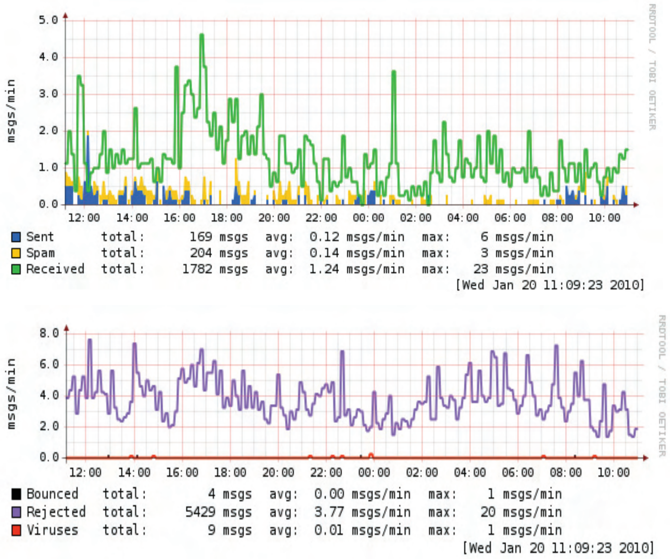

Fig. 1. Collected statistics presented by Mailgraph

Mailgraph is a very simple mail statistics RRDtool frontend for Postfix that produces daily, weekly, monthly and yearly graphs of received/sent and bounced/rejected mail. RRDtool has the ability to store and display data that changes over time in a *Round Robin Databases (*RRDs) that stays fixed in size over time.

In our case, *Postfix – Syslog-NG – RRDtool – Mailgraph* (PSRM) integration collects statistics which is based on the following rules:

1. Source server initiates session of sending email to our PSRM. In case of matching some SMTP restrictions, email will be rejected and session will be closed;
2. If there is no SMTP restrictions matching email will be accepted and sent to anti-spam and anti-virus filters;
3. In spam case or virus detection email will be marked, otherwise, it will be sent to recipient mailbox.

Hence, our RRDs contain statistics related to number of rejected emails, received emails, as well as spam and infected messages identified in bulk of received emails. RRD data can be extracted to XML file. Furthermore, it could be used in any other tool for time series analysis. In this case, time series are related to levels of received emails and level of identified spam in bulk of received emails. In this paper, other statistics are not significant for further analysis.

[1] Slobodan D. Mitrović is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: s.mitrovic@sf.bg.ac.rs

[2] Valentina Radojičić is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: valentin@sf.bg.ac.rs

III. Time series and prediction methods

A time series is a collection of time ordered observations $\{y_1, y_2, ..., y_n\}$, each one being recorded at a specific time t, appearing in a wide set of domains such as Finance, Production and Control [1]. In this case, time is independent variable used in analysis of trends (T), seasonal (S), cyclical (C) and irregular/randomized (I) variations that influence the demand data. General time series model can be described as [2]:

$$Y = T + S + C + I \tag{1}$$

Trends can be described by appropriate development, such as linear, exponential, logistic, polynomial, etc. While seasonal factors represent variations that repeating themselves in regular shorter intervals, up to one year, cyclical factors represent variations which interval of repeating is significantly longer.

Smoothing models is another group of time series models that can be considered, relays the assumption that behavior of relatively near future will be same or similar to relatively close history. Moving averages models as well as exponential smoothing models belong to this group [6].

For now, only models which are used in this paper are briefly presented.

A. Linear trend

Linear trend can be described as follows:

$$y = a + bt \tag{2}$$

$$b = \frac{\sum ty - n\bar{t}\,\bar{y}}{\sum t^2 - n\bar{t}^2} \qquad a = \bar{y} - b\bar{t} \qquad \bar{t} = \frac{\sum t}{n} \qquad \bar{y} = \frac{\sum y}{n}$$

Where are: y - item to be forecast (dependent variable); a, b - parameters to be calculated from historical data; t - point of time (independent variable) and n - number of pairs of values (t,y).

B. Moving average

The new value is calculated as the mean of a number of observed values (m):

$$F_{n+1} = \frac{1}{m} \sum_{i=n-m+1}^{n} X_i \tag{3}$$

Value m can be determined experimentally [3].

The purpose of this approach is to reduce irregularities caused by, for instance, seasonal variations [2].

C. Exponential smoothing

Whereas in the simple moving average the past observations are weighted equally, exponential smoothing assigns exponentially decreasing weights over time [2].

$$F_{n+1} = F_n + \alpha(X_n - F_n) = \alpha X_n + \alpha(1-\alpha)X_{n-1} + \alpha(1-\alpha)^2 X_{n-2} + ... \tag{4}$$

Value of α can be between 0 and 1 ($0 < \alpha < 1$). Recommended values are between 0.2 and 0.3 [3].

D. Examining systematic errors

Examining systematic errors is possible to check by the following procedure; denote the values on the fitted line corresponding to the recorded values of y by $\hat{y}$, and further denote the first readings by y_1, $\hat{y}$, the second by y_2, $\hat{y}$ and so on; calculate

$$w = \sum_{i=1}^{n-1} (y_i - \hat{y}_i)(y_{i+1} - \hat{y}_{i+1})$$

$$v = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{5}$$

and then check the value

$$2 - 2\frac{w}{v}$$

This statistic, known as the *Durbin-Watson statistic*, should lie between 1.5 and 2.5 and preferably between 1.7 and 2.3. If the value is outside this range, the data should be examined for indications of systematic deviation from the fitted curve [2].

IV. Tracking anti-spam filter efficiency

This approach is based on the idea that we could predict PSRM integration behavior in the near future. By comparing prediction results and fresh statistics we can track deviation of filter efficiency and do some corrections and tweaks.

Log data of presented PSRM integration has been collected for received email and detected spam from 12.07.2008. 10:00AM, to 15.10.2009. 02:00AM, in 56 hours snapshots. The collected data has been presented by Fig. 2.

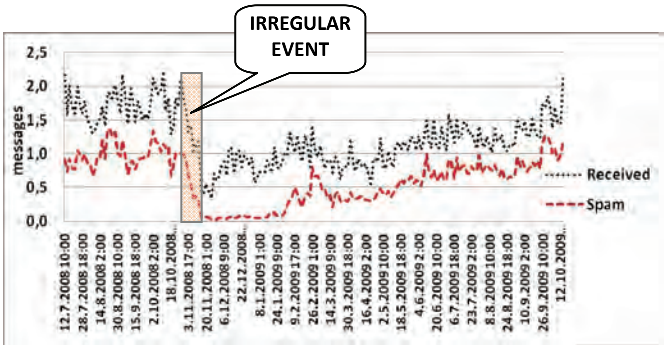

Fig. 2. The total received email and detected spam

From 1st to the 17th day of November, 2008, there have been some system changes. That event must be treated as *the irregular event*. In order to apply some forecasting methods, that event must be excluded, as it shown by Fig.3.

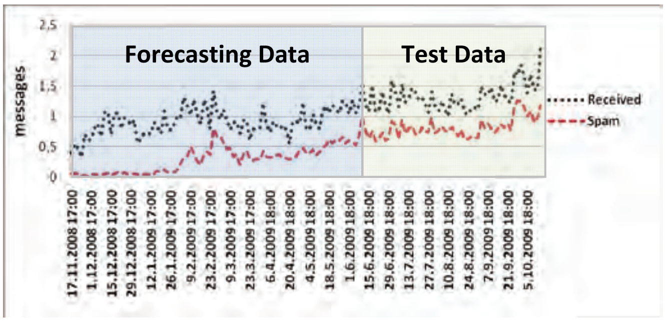

Fig. 3. The total received email and detected spam without the irregular event

It could be notice that each time series (*received* and *spam*) is divided in two parts – for creating forecasting model (from 17.11.2008. 17:00 to 6.7.2009. 18:00) and for purpose of testing (from 6.7.2009. 18:00:00 to 12.10.2009. 18:00).

For the purpose of forecasting procedure, three models have been chosen: moving average, exponential smoothing and linear regression, as shown in Fig. 4:

1. two moving average models have been created:
   - with step of 3 snapshots per 56 hours, which means 1 week (week*R(x3)* and week*S(x3)*)
   - with step of 12 snapshots per 56 hours, which means 1 month (month*R(x12)* and month*S(x12)*)
2. exponential smoothing with $\alpha = 0.2$ (in Fig 4. by legend *expo(R)* and expo*(S)*)
3. linear trend has been calculated by equation (2):
   - for received email - *linear(R)*:
     $$y_R = 0.70411 + 0.00487 \cdot t$$
   - for detected spam - *linear(S):*
     $$y_S = -0.03134 + 0.00743 \cdot t$$

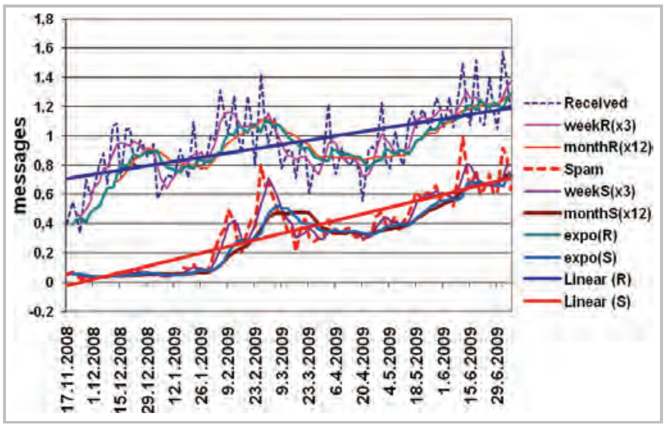

Fig. 4. Obtained forecasting results

Examination of systematic errors has been made by Durbin-Watson statistic. There were also calculated square errors sums and results are shown on Table I.

Durbin-Watson statistic shows no systematic errors in *received* series. The same statistic shows some systematic errors in *spam* series that could be explained by behavior of filter. After system change, new untrained anti-spam filter has been applied and level of detected spam has been decreased in the first 60 days.

TABLE I
DURBIN-WATSON STATISTIC AND SQUARE ERROR SUMS

Prediction method	Durbin-Watson statistic	Square error sums
weekR(x3)	2.547421012	1.81007311
weekR(x12)	1.654315532	2.77121651
expo(R)	1.779870971	3.36099602
linear(R)	1.344073907	3.74682626
weekS(x3)	1.724419852	0.44067631
weekS(x12)	0.709907847	1.21867089
expo(S)	0.87492502	1.19914342
linear(S)	0.642861566	1.47628343

The training process has been also followed by increased frequency of spam attacks which resulted intensive training and that can be visually noticed – after a number of days *spam* curve started to “follow” *received* curve. This fact is confirmed in Fig. 5. by values of deviations (*devR* and *devS*), as well as correlation coefficient – 0.73465.

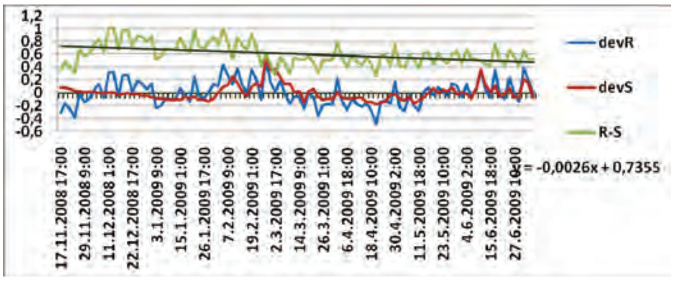

Fig. 5. Deviations of *received* and *spam* series

One of the most important characteristics is related to calculation of linear trend for each series. The slope of linear development of detected spam emails is greater than slope of linear development of received messages, which means that anti-spam filter is becoming more successful as time goes by. This characteristic is confirmed with difference between amounts of received messages and detected spam, as we can see in Fig 5. Also, it could be confirmed by levels of square errors in Fig. 6.

It is possible to conclude that liner trend implementation is the most convenient way for this kind of monitoring purposes and it will be used in the following consideration, only.

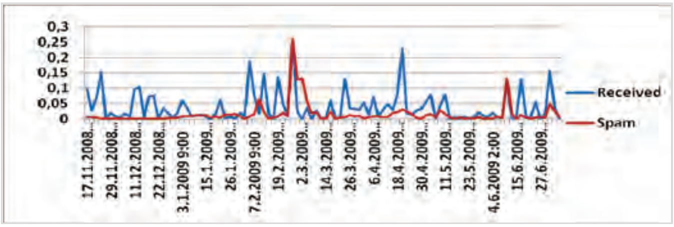

Fig. 6. Square error sums

Calculated linear trends have been tested with *Test data* series. Results are presented in Fig. 7.

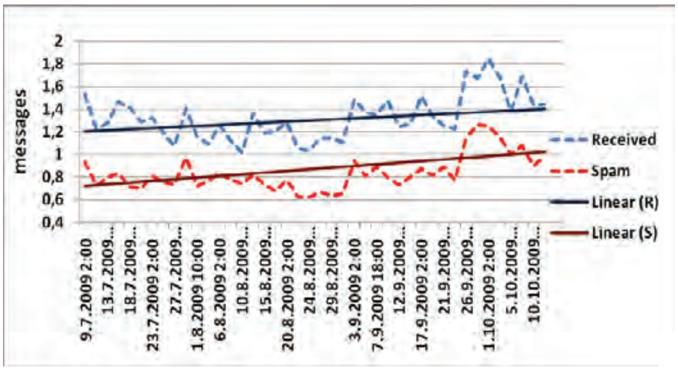

Fig. 7. Forecasted linear trends and *Test data* series

Testing has shown that forecasted linear trends successfully "followed" by curves of received messages and detected spam which is confirmed by correlation coefficient of 0.880009.

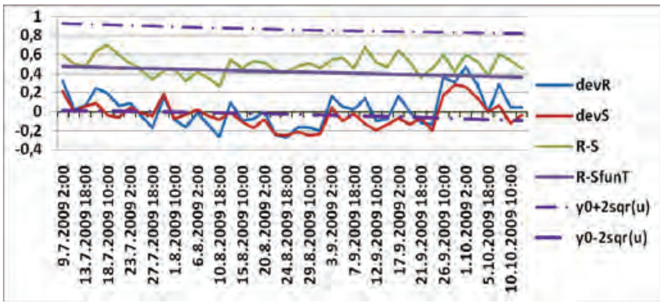

Fig. 8. *Test data* series - deviation of *received* and *spam* series, difference between received messages and detected spam and confidence interval of 95% *(y0+2sqr(u); y0-2sqr(u))*

Distance between received messages and detected spam continue to decrease which is presented by curve *R-S* in Fig. 8. This curve remains deeply inside of the corresponding confidence interval (95%).

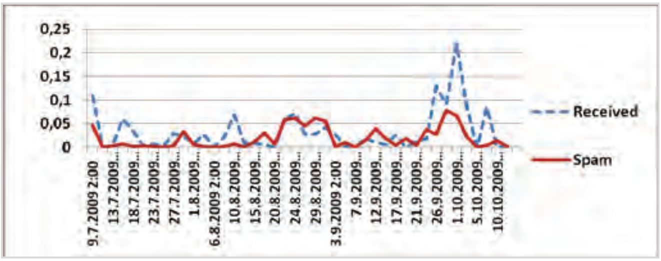

Fig 9. Square error sums

Also, it could be noticed that these square error sums related to difference of time series and corresponding linear trends remained on the same level, Fig 9.

V. Conclusion

Forecasting methods could be used for improvement of anti-spam filter efficiency tracking process, because of their simplicity and variety. Selection of suitable forecasting model depends on e-mail system time series characteristics.

In this paper linear trend has been selected because of advantages in process of visual monitoring of generated graphs, especially with presence of confidence interval.

Although there was no a huge spam attack or detection of anti-spam filter issues during test phase it is reasonably to suppose two different cases. First, higher level of spam attack could show higher level of difference between received messages and detected spam that could stay inside confidence interval. In this case regular tweaks of filter could be done in order to improve detection performance. On the other hand, important issues, such as *Bayesian poisoning*, could lead that level of difference between received messages and detected spam could arise out of confidence interval. In this case some important interventions are required.

Authors believe that for long-term forecasting, in the case of longer time series, linear trend have to be replaced with some more appropriate model such as logistic (Gompertz) trend. There is also possibility for presence of cyclical factors, caused by various factors, such as software replacement, etc.

Acknowledgement

This paper presents some of results in research project TR-11013, funded by Ministry of Science and Technological Development of Republic of Serbia.

References

[1] S. Makridakis, S. Weelwright, R. Hyndman, *Forecasting: Methods and Applications*. John Wiley & Sons, New York, USA, 1998.

[2] Leijon H., *Forecasting Theories*, PLANITU Doc. 61-E, ITU

[3] V. Radojičić, *Prognoziranje u telekomunikacijama*, *GND-Produkt*, Belgrade, Serbia, 2003.

[4] G. Sakkis, I. Androutsopoulos, G. Paliouras, V. Karkaletsis, Constantine D. Spyropoulos and P. Stamatopoulos, Stacking classifiers for anti-spam filtering of e-mail, Proceedings of the 6th Conference on Empirical Methods in Natural Language Processing (EMNLP 2001), Pittsburgh, USA, pp. 44–50, 2001.

[5] J. R. Evans, *Statistics, Data Analysis, and Decision Modeling*, James Robert Evans, Pearson/Prentice Hall, 2007.

[6] S. Mitrović, Neke tehnike detekcije spam poruka, Zbornik radova: SYM-OP-IS 2009, Ivanjica, 2009., str. 55-59

[7] S. Mladenović, S. Mitrović, S.Janković, Unapređenje implementacije modela anti-spam zaštite, Zbornik radova: PosTel 2008, Beograd, 2008., str. 279-288

[8] http://wiki.apache.org/spamassassin, (28.3.2010.)

[9] http://www.balabit.com/network-security/syslog-ng, (28.3.2010.)

[10] http://mailgraph.schweikert.ch, (28.3.2010.)

[11] http://oss.oetiker.ch/rrdtool/doc/index.en.html, (28.3.2010.)

[12] http://www.postfix.org, (28.3.2010.)