World Applied Sciences Journal 29 (12): 1605-1610, 2014
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2014.29.12.13987

Legitimate User Profiling Based on the Third Order Spline Approximation of the Initial Data Sequence

Jurij Aleksandrovich Gorbunov, Lev Nikolaevich Krotov and Elena L'vovna Krotova

Perm National Research Polytechnic University,
Komsomolsky Ave. 29, 614990 Perm, Russian Federation

Abstract: The article deals with an alternate solution of the system intrusion detection problem by developing the user's and the putative intruder's behavior profiles. The third order spline approximation of the initial data sequence is suggested to use as the legitimate user profiling method. Application of the large and small data sampling method is considered. The interval estimation method is proposed to use as the secondary method of defining the expected behaviour frontiers. The comparison of two alternative interval estimations-based on the maximum deviation from the mean value and the deviation at any specific time is adduced.

Key words: Intrusion detection systems · Statistical methods · User profile · Splines · Interval estimation

INTRODUCTION

Because of the remote access technologies and mobile device development the problem of the secure authentication and the detection of the illegal information system intrusions get to a new level. Users increasingly strive for mobility, giving new tasks to the information security officers. A mobile access to protected systems produces the user's convenience along with the new vulnerabilities of the information security system. In these cases the timely and precise detection of the system intrusion time becomes particularly important.

Now the system intrusion detection problem increasingly becomes the subject of many experts' and scientists' researches all over the world. Large companies set up the special units, which are not only responsible for their systems security, but conduct an analysis of this field. For example, IBM publishes an annual report which reflects the trends and risks in the field of the information security and which annually devotes attention to the intrusion and anomalies detection systems [1].

Some authors try to reflect the current state of the problem of the intruder detection in the information system and classify the existing solutions, rightly dividing the problem into the intrusion detection and the system anomaly detection [2]. They also point out at the absence of the computer models of the attack detection and its formal reasoning, which impedes the application of such systems.

There are many attempts to solve the intrusion detection problem in relation to its various aspects and application fields. For example, in the Mobile Ad Hoc Network (Mobileadhocnetworking-MANET) [3]. The intrusion detection system which applies the Markov Blanket and the Bayesian networks as the methods of selecting the system state data is suggested in the article.

Some authors point out at the complexity of using the traditional methods to solve the set task and suggest to apply the so-called soft computing methods based on the fuzzy logic, artificial neural networks, support vectors, probabilistic reasoning, genetic algorithms, etc. [4].

There are also the works in which the wavelet analysis well-proven in other fields is used to solve this task. When applying its methods to analyse the network traffic, the authors have managed to work out a prototype of the rather effective IDS, which effectiveness is shown in their article [5].

A separate group of works is devoted to the methods of assessing the effectiveness of the intrusion detection systems application, which are based on the anomaly [6] and signature [7] analysis.

Corresponding Author: Jurij Aleksandrovich Gorbunov, Perm National Research Polytechnic University, Komsomolsky Ave. 29, 614990 Perm, Russian Federation.

The model based on the analysis of the length of the vector of deviations from the common centre of mass-the system performance at the current time-could be used for a detection of the abnormal system activity [8].

In recent years we have also done a great work in the field of solving the task of the intruder detection in the information system.

Body: It is proposed to use the statistical methods of the intruder detection for solving this task [9, 10]. This article presents the legitimate user and putative intruder profiling method based on the third order spline approximation of the sequence of the data values which characterize their behavior in the system.

The profile describes the expected individual's behaviour and is based on the observations of the individual's actions in the past. Sampling of the initial data on the individual's behaviour (a legitimate user or an intruder) and time values, in which these data have been recorded, are required for profiling.

Let's assume that these values are spaced apart at equal intervals and are represented in a chronological order t_1 =(1 .. 32, [1, 2, 3, .., 31, 32]).

At the same time the true event time value is of no importance in this method, the compliance with the strict event chronology is more important. 12 randomly generated sets of 32 values length as the initial data (L_1..L_{12}). The values correspond to a conditional sequence of the event record time values. The arithmetic mean value has been calculated for each specific time value. The received U set of 32 arithmetic mean values characterizes the legitimate user's actions.

Only one set is taken as the initial data on the intruder's behaviour, as the intruder's detection is supposed soon after the beginning of its actions in the system.

A set of 5 values is additionally extracted from 32 values of the intruder's data for evaluating effectiveness of the method using small samplings of the data on the putative intruder (for a short time).

When third order spline approximating the U set of average values received at t_1 times we will get the time function f_L, which reflects the actions of the legitimate user at any time $1 < t < 32$.

It is necessary to estimate his expected behaviour for the legitimate user profiling. It could be done in several ways.

The maximum deviation of the initial data parameter value concerning the actions of the legitimate user from the arithmetic mean value of these data at any time is suggested to use for defining the expected behaviour frontiers in one of the methods. Let's qualify this method as the method of the maximum deviation.

As a result of calculations, we will obtain the maximum deviation value q. The upper and lower frontier of the expected behaviour is derived from the arithmetic mean value array by adding or subtracting the maximum deviation values respectively.

Since deviation q is a number and then the expected legitimate user's behaviour frontiers are derived from the y-shift of the function f_1 for + q and - q respectively. The received functions f_{Lu} (the upper frontier) and f_{Ld} (the lower frontier) characterize the expected behaviour of the legitimate user at any time $1 < t <32$ when using the method of the maximum deviation.

However, such definition of the permissible frontiers of the expected behaviour does not always allow an adequate evaluation of the expected behaviour. If there are significant "jumps" of the legitimate user behaviour parameter value, the frontiers of the expected behaviour will be too extended, which will serve as a flaw of the security system. The maximum deviation of the initial data values concerning the user L_1 .. L_{12} from the arithmetic mean value of U at each time is suggested for more adequate evaluation of the expected behaviour frontiers. This method will be qualified as the method of different deviations.

As a result of calculations, we will obtain a set of maximum deviations for each time U_D. The evaluation of the expected behaviour frontiers is carried out by elementwise adding (for the upper frontier) and subtracting (for the lower one), respectively, of elements of the arithmetic mean values array U and the array of the maximum deviations for each time U_D.

The received arrays L_{du} (the upper frontier) and L_{Dd} (the lower frontier) are third order spline approximated and we will obtain accordingly the functions of the expected legitimate user's behaviour frontiers f_{Du} and f_{Dd}, received by the method of different deviations.

The received splines f_1, f_{Lu}, f_{Ld}, f_{Du} and f_{Dd} are represented on the Figure 1.

Here and elsewhere the axis t-is the time axis, the axis y-is the axis values of the parameter of the initial data on the legitimate user. The expected behaviour of the legitimate user should be within the area limited by the upper and lower frontier for each method, respectively.

It is easy to notice that the use of deviations for each time significantly reduces the area of the expected user's behaviour.

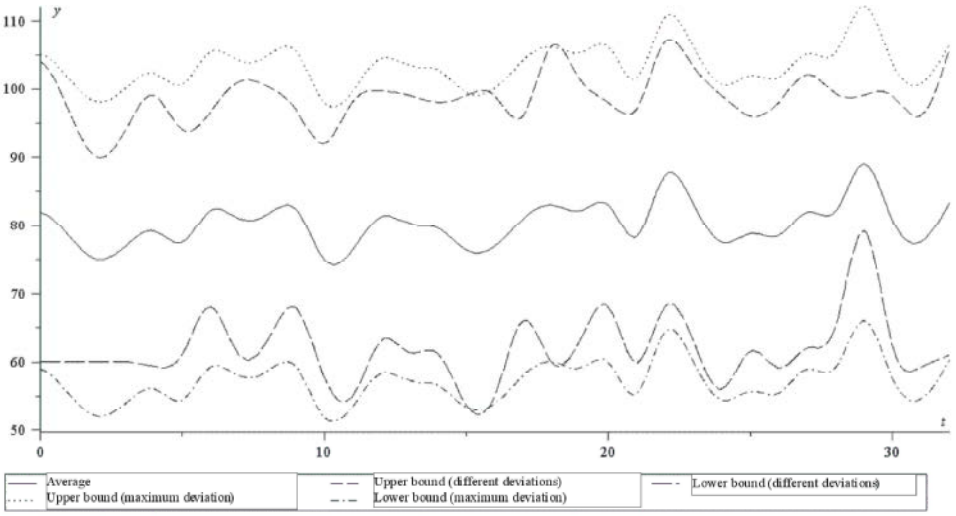

Fig. 1: Legitimate user's behaviour profile frontiers when using different methods of defining the expected behaviour

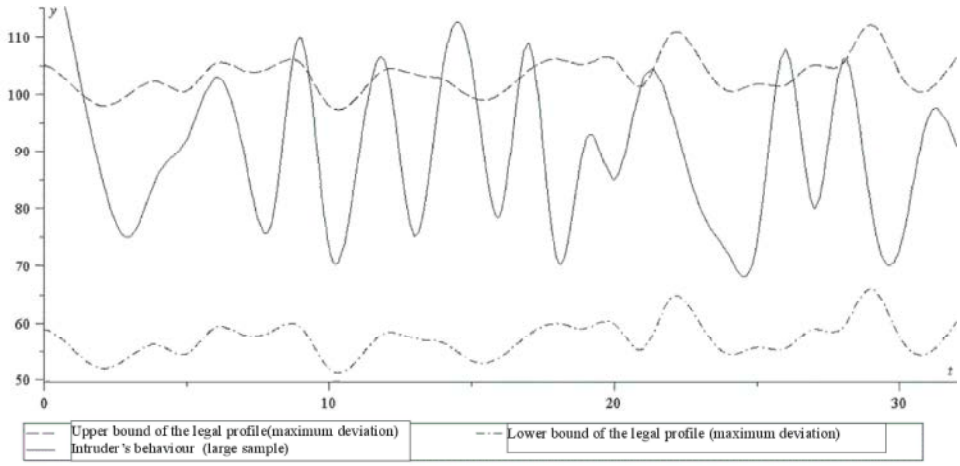

Fig. 2: Profile of the expected legitimate user's behaviour (maximum deviation) and the intruder's behaviour (large sampling)

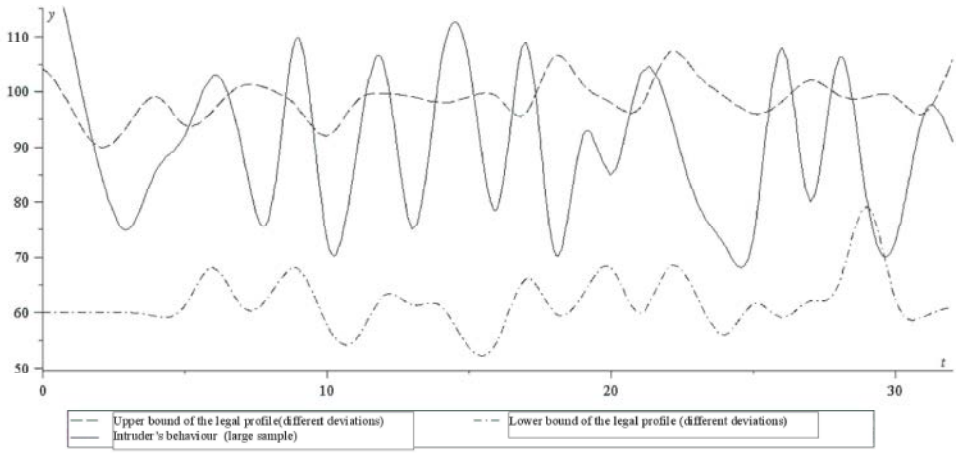

Fig. 3: Profile of the expected legitimate user's behaviour (different deviations) and the intruder's behaviour (large sampling)

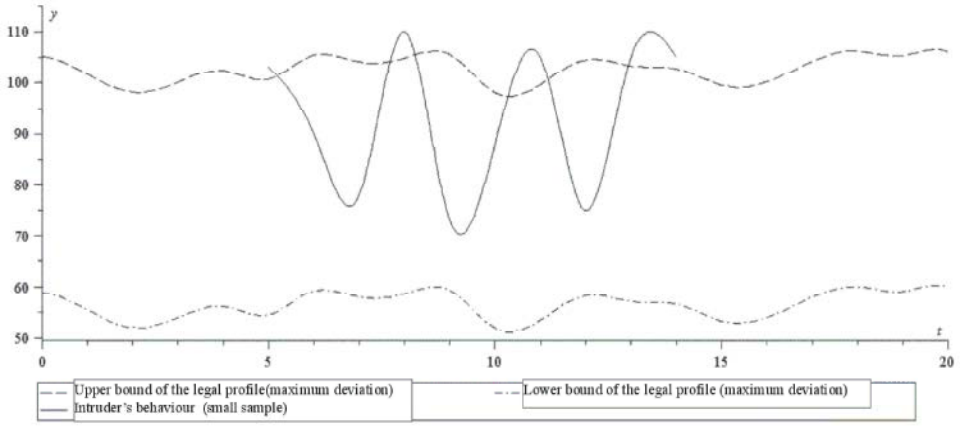

Fig. 4: Profile of the expected legitimate user's behaviour (maximum deviation) and the intruder's behaviour (small sampling)

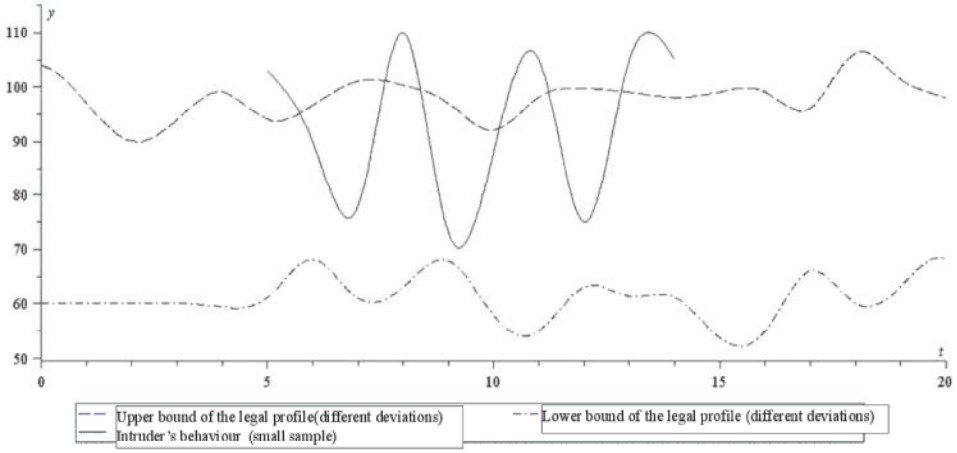

Fig. 5: Profile of the expected legitimate user's behaviour (different deviations) and the intruder's behaviour (small sampling)

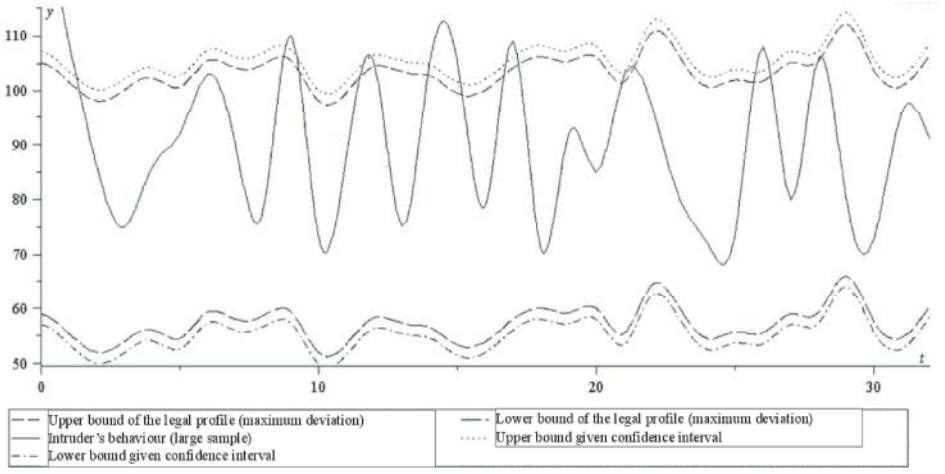

Fig. 6: Confidence intervals for the legitimate user's profile frontiers (maximum deviation, large sampling)

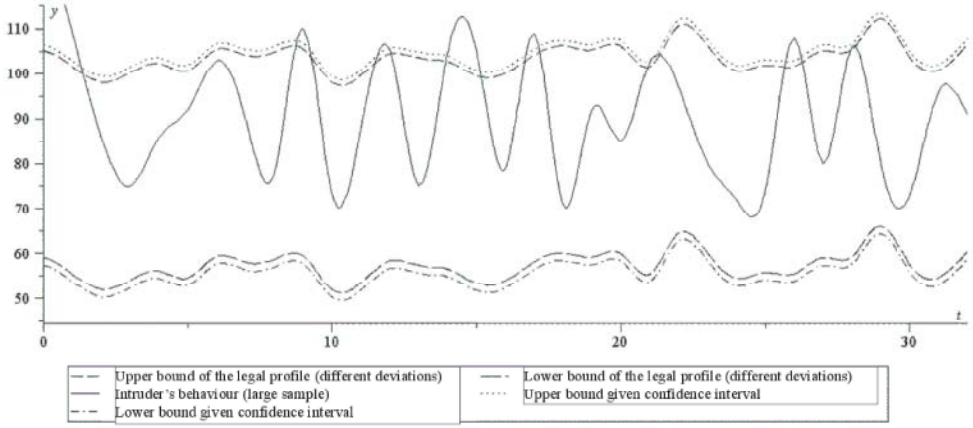

Fig. 7: Confidence intervals for the legitimate user's profile frontiers (different deviations, large sampling)

Let's take an array of 32 values K_1 and the array of 10 values K_2 as the initial data on the intruder's behaviour. When third order spline approximating these values we well obtain the functions f_{K1} and f_{K1} which characterize the behaviour of the putative intruder in time.

The Figure 2 shows the profile of the expected legitimate user's behaviour which frontiers were built using the maximum deviation method and the function of the intruder's behaviour is based on large sampling of the initial data on his actions (32 values).

The Figure 3 shows the profile of the expected legitimate user's behaviour which frontiers were built using the method of different deviations and the function of the intruder's behaviour is based on large sampling of the initial data on his actions.

The comparison clearly shows that when using the method of different variations, the intruder could be revealed earlier and with more certainty than when using the method of maximum deviations.

Let's similarly consider the case in which the actions of the putative intruder are known only at 10 times.

The profiles of the expected legitimate user's behaviour, developed by the maximum deviation method (Figure 4) and by the method of different deviations (Figure 5) and a spline of the intruder's behaviour based on small sampling of the data on his actions, are represented below.

The figures show that the intruder detection method based on the expected user's behaviour profile obtained by the third order spline approximation of the initial data operates the small sets of the data on the putative intruder and allows his detection in a short time after the system intrusion. In this case it is more preferable to use the method of different deviations for defining the expected behaviour frontiers, as it provides a higher protection level. The interval estimation could be applied to the expected user's behaviour frontiers for the implementation of the multi-level or taught (with constant updating of the profile of the expected legitimate user's behaviour) security system, based on a similar method. The confidence intervals being run outside the profile are of primary interest.

In case of entering the current value of the input data in these intervals the security system should record the event in its own log, but it should not sent out an alert signal.

When looking through the log the Security Manager decides if any given event complies with the actions of the legitimate user or the intruder, thereby editing the profile of the expected behavior.

The confidence intervals for each of the above represented methods for defining the legitimate user's profile frontiers are shown on the Figure 6 and the Figure 7, respectively.

CONCLUSIONS

Under the conducted researches we might as well say that the method reliably distinguishes the legitimate user's actions from the intruder's or any other anomaly actions in the system and provides the possibility to indicate the system state at certain intervals, what could be very useful when analysing the security incidents.

In the future we plan to consider the possibility of applying for each value of the initial data on the legitimate user's actions the relevant weight coefficient, which characterizes the probability of the given value introduction, which will allow the use of large data amounts for profiling.

REFERENCES

1. Gorbunov, Y.A., 2009. Using a Probabilistic Approach to Distinguish between Legitimate User's and Intruder's Profiles. Applied and Industrial Mathematics Review, 16(3).
2. Gorbunov, Y.A., 2009. Determining the Parameters of the Mathematical Model of the ARM User Profile. Applied and Industrial Mathematics Review, 17(2).
3. Nesterenko, V.A., 2006.Statistical Techniques for Detecting Network Security Violations, Information Processes, 6(3): 208-217.
4. Karaichev, G.V. and V.A. Nesterenko, 2010. Network Anomaly Detection through the Statistical Analysis of the IP Packet Headers. Proceedings of Higher Education Institutions. North Caucasus Region. Natural Sciences, 4: 13-17.
5. Almgren, M., 2003. Consolidation and Evaluation of IDS Taxonomies. Proceedings of the Eight Nordic Workshop on Secure IT Systems, NordSec.
6. Lad, M, 2006. PHAS: A Prefix Hijack Alert System. 15th USENIX Security Symposium.
7. Huston, G., M. Rossi and G. Armitage, 2011. Securing BGP-A Literature Survey. IEEE Communications Surveys and Tutorials, pp: 2.
8. Tumoian, E. and M. Anikeev, 2005. Network-based detection of passive covert channels in TCP/IP // LCN '05. IEEE Conf. on Local Computer Networks. Washington, DC.
9. Ariu, D., R. Tronci and G. Giacinto, 2011. HMMPayl: An intrusion detection system based on Hidden Markov Models. Computers and Security, 30(4): 221-241.
10. Internet Security Systems.IBM X-Force 2012 Trend and Risk Report. IBM Global Technology Services, 2013.