Journal of Theoretical and Applied Information Technology
20th May 2016. Vol.87. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

GENERALIZED PROBABILISTIC DESCRIPTION OF HOMOGENEOUS FLOWS OF EVENTS FOR SOLVING INFORMATIONAL SECURITY PROBLEMS

[1]YURI YURIEVICH GROMOV, [2]IVAN GEORGIEVICH KARPOV,
[3]YURI VIKTOROVICH MININ, [4]OLGA GENNADEVNA IVANOVA

[1,2,3,4.5] Tambov State Technical University
Sovetskaya Str., 106, Tambov, 392000, Russia
E-mail: [1]gromovtambov@yandex.ru, [3]yuri.minin@gmail.com, [4]olgaotd@yandex.ru

ABSTRACT

The method of producing a generalized probabilistic description of flows of events that are characterized by ordinariness, stationarity and lack of aftereffects for solving informational security problems is obtained. The examples of these flows are test queries to network addresses, attempts of available ports scanning, etc. The independence of increments property could be unrealized for these events, but the Markov lack of aftereffects property is necessarily realized. It consists that there is no influence to the time of appearing the future event of past events, excluding the last event. The authors obtained the distributions of flows of events for linear flow intensity function, Poisson, binominal, and negative binominal lows. We also obtained expressions for the distribution laws of k - th event occurrence and their main numerical characteristics.

Keywords: *Informational Security, Flow Of Events, Probability, Probability Distribution Density, Mathematical Expectation, Variance.*

1. INTRODUCTION

The majority of informational security systems are based on data analysis. The data is received from different components of IT-infrastructure, so the process of the event flows monitoring is quite significant in such systems.

Different flows of events are observed in real-time processing, transmission and control of data. A flow of events is normally understood as a sequence of homogeneous events, appearing one by one at random moments of time $t_0, t_1, t_2, \dots, t_k, \dots$, where $k = 0, 1, 2, \dots,$. The examples of such flows of events include a flow of test queries to network adresses, a flow of attempts of available ports scanning, etc. The simplest flow of events possesses the three properties - stationarity, ordinariness and independence of increments [1-9].

The property of stationarity means that the probability of occurrence of a given number of events in the time interval $(t, t+\Delta t)$ depends on Δt, but it doesn't depend on time t.

The property of ordinariness means that events appear one by one, rather than in groups. In this case, the probability of a single event in the interval duration Δt is equal to $\lambda\Delta t + o(\Delta t)$, where $\lambda > 0$ is the intensity of the simplest flow, and probability of occurrence of two or more events is equal to $o(\Delta t)$.

The independence of increments means the independence of occurrence of a certain number of events on disjoint intervals. In some studies, the independence of increments is equal to lack of aftereffect [2,9].

Currently, in queuing systems the input flow is the Poisson simplest flow of events or its generalizations caused by modification of the properties of ordinariness, stationarity and independence of increments, or waiver of these properties [2,4,5,8,9].

2. STATEMENT OF THE PROBLEM

The aim of this research is to produce a generalized probabilistic description of the flow of events, possessing the properties of stationarity and ordinariness. The property of independence of increments may not hold, but the Markov property of lack of aftereffect must be fulfilled. It means that past events apart from the last one have no effect on the time of occurrence of the future event.

3. SOLUTION OF THE PROBLEM

First we obtain equations for these flows of events. We denote by $P_k(t)$ the probability of occurrence of k events during the time interval $(0, t)$. To calculate the probability $P_k(t)$ we consider the three time intervals with duration of $t, \Delta t, t+\Delta t$. If during the time $t+\Delta t$ k events occurred, it might have happened under the following conditions:

1) k events occurred over the time t, but not a single event occurred during the time interval $(t, t+\Delta t)$;

2) k - 1 events occurred over the time t, and a single event occurred during the time interval $(t, t+\Delta t)$;

For sufficiently small values of Δt, the first of these two cases matches the probability $P_k(t)[1-\Lambda(k)\Delta t]+o(\Delta t)$, while the second one matches the probability $P_{k-1}(t)\Lambda(k-1)\Delta t+o(\Delta t)$.

Here, we denote $\Lambda(k)$ as a function of flow intensity, which depends in the general case on k events. Since the given cases are incompatible, by the sum law for $k \geq 1$ we obtain

$$P_k(t+\Delta t)=P_k(t)[1-\Lambda(k)\Delta t]+$$
$$+P_{k-1}(t)\Lambda(k-1)\Delta t+o(\Delta t).$$

We subtract the probability $P_k(t)$ from the right and left sides of the obtained formula, then divide it by Δt and proceed to limit $\Delta t \to 0$, to produce the differential equation

$$dP_k(t)/dt=-\Lambda(k)P_k(t)+\Lambda(k-1)P_{k-1}(t); \quad k=1, 2, 3, \ldots . \qquad (1)$$

For $k = 0$ only the first case is possible, therefore we have

$$dP_0(t)/dt=-\Lambda(k)P_k(t); \; k=0 . \qquad (2)$$

The differential equations (1) and (2) are solved under physically evident initial conditions

$$P_0(0)=1, P_k(0)=0, \; k=1, 2, 3, \ldots . \qquad (3)$$

If the total number of events is finite, then for $k=N$, instead of equation (1) it is expedient to use the equation

$$dP_N(t)/dt=\Lambda(N-1)P_{N-1}(t); \; k=N. \qquad (4)$$

It should be noted that for any t (including $t=0$) under a finite number of events one must observe the normalization condition

$$\sum_{k=0}^{N} P_k(t)=1. \qquad (5)$$

The solution of equation (2) with the initial condition (3) is

$$P_0(t)=\exp(-\Lambda(0)t); \; k=0. \qquad (6)$$

The solution of equation (1) can be found by variation of arbitrary constants in the form of the recurrence formula [10]

$$P_k(t)=\exp(-\Lambda(k)t)\int_0^t \Lambda(k-1)P_{k-1}(\tau)\exp(\Lambda(k)\tau)d\tau; \quad k=1, 2, 3, \ldots . \qquad (7)$$

Given the normalization condition (5) the solution of equation (4) under $N<\infty$ is

$$P_N(t)=1-\sum_{k=0}^{N-1} P_k(t); \; k=N. \qquad (8)$$

If the function $\Lambda(k)$ is a linear function of intensity, the solution of equation (7) given (6) for the flow of events can also be presented explicitly. At such, the following three cases are generic:

1. If $\Lambda(k)=\lambda$, we obtain the Poisson law [1-5]

$$P_k(t)=\frac{(\lambda t)^k}{k!}\exp(-\lambda t), \quad k=0, 1, 2, \ldots . \qquad (9)$$

We consider mathematical expectation $m(t)$, variance $D(t)$ and coefficient $K(t)=M_3(t)/D(t)$, where $M_3(t)$ is the third central moment, as the main numerical characteristics of distribution laws. For the Poisson distribution (9) they are determined by the expression

$$m(t)=D(t)=\lambda t; \quad K(t)=1. \qquad (10)$$

In [2] it is shown that the time of k - th event occurrence for the Poisson flow is subject to the Erlang distribution

$$p_k(t)=\frac{\lambda}{(k-1)!}(\lambda t)^{k-1}\exp(-\lambda t); \quad 0 \leq t<\infty, \; k \geq 1, \qquad (11)$$

with numerical characteristics

$$m=\frac{k}{\lambda}; \; D=\frac{k}{\lambda^2}; \; M_3=\frac{2k}{\lambda^3}. \qquad (12)$$

Probability density function (PDF) for time intervals between consecutive events is an exponential with a parameter of scale [1-5, 9]

$$p(\tau)=\lambda\exp(-\lambda\tau); \quad 0 \leq \tau<\infty. \qquad (13)$$

Numerical characteristics of the PDF (13) are equal to

$$m=\frac{1}{\lambda};\ D=\frac{1}{\lambda^2};\ M_3=\frac{2}{\lambda^3}. \qquad (14)$$

From a comparison of the expressions (12) and (14) it can be concluded that the numerical characteristics of the PDF (11) are k times greater than those of the PDF (13). This result is logical since

$$t_k=\sum_{i=1}^{k}\tau_i, \qquad (15)$$

where τ_i are independent and identically distributed random variables. In [3] it is shown that if for the flow of events the relations (13) and (15) are true, this flow has the property of lack of aftereffect.

The flow of events is stationary and ordinary, if the condition [1] holds

$$\Lambda=\lim_{t\to 0}\frac{1}{t}\sum_{k=0}^{\infty}P_k(t)=\lim_{t\to 0}\frac{1}{t}\sum_{k=1}^{\infty}k\,P_k(t), \qquad (16)$$

where $\Lambda=const$ is intensity of the flow. Substituting the expression (9) in (16), we find that for the Poisson flow $\Lambda=\lambda$. It also matches the property of independence of increments determined by the relation

$$P_k(t+\Delta t)=\sum_{i=0}^{k}P_i(t)P_{k-i}(\Delta t). \qquad (17)$$

2. If $\Lambda(k)=\lambda(N-k)$, then binominal distribution follows from (7) given (6) and (8)

$$P_k(t)=\frac{N!}{(N-k)!k!}(1-\exp(-\lambda\, t))^k\exp(-\lambda\, t)^{N-k}, \qquad (18)$$

$$0\le k\le N.$$

For it the numerical characteristics are equal to

$$\begin{aligned}&m(t)=N(1-\exp(-\lambda\, t));\\&D(t)=m(t)\exp(-\lambda\, t);\\&K(t)=2\exp(-\lambda\, t)-1;\ 0<K(t)<1.\end{aligned} \qquad (19)$$

At the same time, the law of the k - th event occurrence can be determined by [2] similarly to the Erlang distribution (11)

$$p_k(t)=\frac{N!\lambda}{(N-k)!(k-1)!}\times$$

$$\times(1-\exp(-\lambda\, t))^{k-1}\exp(-\lambda\, t)^{N-k+1}, \qquad (20)$$

$$0\le t<\infty,\ 1\le k\le N.$$

Its main numerical characteristics, given [7] are equal to

$$m=\sum_{i=0}^{k-1}\frac{\lambda^{-1}}{N-i};$$

$$D=\sum_{i=0}^{k-1}\frac{\lambda^{-2}}{(N-i)^2}; \qquad (21)$$

$$M_3=\sum_{i=0}^{k-1}\frac{2\lambda^{-3}}{(N-i)^3}.$$

The PDF of time intervals between consecutive events is an exponential with a parameter of scale $\Lambda(k)=\lambda(N-k)$

$$m=\frac{1}{\lambda(N-k)};$$

$$D=\frac{1}{(\lambda(N-k))^2}; \qquad (23)$$

$$M_3=\frac{2}{\lambda(N-k)^3}.$$

Since for the binomial flow of events the relations (22) and (15) given (21) and (23) are true, this flow has the property of lack of aftereffect. Substituting the expression (18) in (16), we find that the intensity of the binomial flow is $\Lambda=N\lambda$. Thus, the properties of statinarity and ordinariness are true for this flow, but the property of independence of increments determined by the relation (17) does not hold.

3. If $\Lambda(k)=\lambda(\alpha+k)$, then negative binominal distribution follows from (7) given (6)

$$P_k(t)=\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)k!}(1-\exp(-\lambda\, t))^k\exp(-\alpha\lambda\, t), \qquad (24)$$

$$0\le x\le N-1.$$

where $\Gamma(z)$ is gamma function.

For the probability distribution (24) the numerical characteristics are equal to

$$\begin{aligned}&m(t)=\alpha(\exp(\lambda\, t)-1);\\&D(t)=m(t)\exp(\lambda\, t)\\&K(t)=2\exp(\lambda\, t)-1;\\&K(t)>1.\end{aligned} \qquad (25)$$

The law of occurrence of the k - th event can be determined similarly to the Erlang distribution (11)

$$p_k(t)=\frac{\Gamma(\alpha+k)\lambda}{\Gamma(\alpha)\,(k-1)!}\times$$

$$\times(1-\exp(-\lambda\, t))^{l-1}\exp(-\alpha\lambda\, t), \qquad (26)$$

$$0\le t<\infty,\ k\ge 1.$$

Its main numerical characteristics given [11] are equal to

$$m=\sum_{i=0}^{k-1}\frac{\lambda^{-1}}{i+\alpha};\quad D=\sum_{i=0}^{k-1}\frac{\lambda^{-2}}{[i+\alpha]^2};\quad M_3=\sum_{i=0}^{k-1}\frac{2\lambda^{-3}}{[i+\alpha]^3}. \tag{27}$$

The PDF of time intervals between consecutive events is an exponential with a parameter of scale $\Lambda(k)=\lambda(\alpha+k)$

$$p(\tau)=\lambda(\alpha+k)\exp(-\lambda(\alpha+k)\tau);\quad 0\le\tau<\infty,\ k\ge 0. \tag{28}$$

Numerical characteristics of the PDF (28) are equal to

$$m=\frac{1}{\lambda(\alpha+k)};\quad D=\frac{1}{(\lambda(\alpha+k))^2};\quad M_3=\frac{2}{\lambda(\alpha+k)^3}. \tag{29}$$

Since for the flow of events the relations (28) and (15) given (27) and (29) are true, this flow has the property of lack of aftereffect. Substituting the expression (24) in (16), we find that the intensity of the binomial flow is $\Lambda=\alpha\lambda$. Thus, the properties of stationarity and ordinariness are true for this flow, but the property of independence of increments determined by the relation (17) does not hold.

If $\Lambda(k)$ is a quadratic function of the flow, the solution (7) for the flow of events cannot be produced in the explicit form. For the function $\Lambda(k)$ if $0\le k\le N-1$ one can use the following expressions:

1) negative hypergeometric law $\Lambda(k)=\lambda(N-k)(b+k)$, where $b>0$;

2) discrete uniform distribution $\Lambda(k)=\lambda(N-k)(1+k)$;

3) hypergeometric law $\Lambda(k)=\lambda(N-k)(b-k)$, where $b\ge N$.

In this case, the basic numerical characteristics of distribution laws of time of k - th event occurrence are determined by the expression

$$m=\sum_{i=0}^{k-1}\frac{1}{\Lambda(i)};\quad D=\sum_{i=0}^{k-1}\frac{1}{[\Lambda(i)]^2};\quad M_3=\sum_{i=0}^{k-1}\frac{2}{[\Lambda(i)]^3},\quad k\ge 1. \tag{30}$$

The PDF of time intervals between consecutive events for the considered flow is an exponential with a parameter of scale

$$p(\tau)=\Lambda(k)\exp(-\Lambda(k)\tau);\quad 0\le\tau<\infty. \tag{31}$$

The numerical characteristics of distribution (31) are equal to

$$m=\frac{1}{\Lambda(k)};\quad D=\frac{1}{(\Lambda(k))^2};\quad M_3=\frac{2}{(\Lambda(k))^3};\quad 0\le k\le N-1. \tag{32}$$

From a comparison of the expressions (30) and (32) it follows that they are connected by the relation, similar to expression (15). Hence, the given flows of events also have the property of lack of aftereffect. The property of independence of the increments does not hold for them.

4. CONCLUSIONS

Thus, we produced a generalized probabilistic description for the flow of events possessing the properties of ordinariness, stationarity and lack of aftereffect. We showed that the obtained distributions for the flow of events in a linear function of the flow intensity can be distinguished either by the value $K(t)$, or the values of expectation and variance. For the Poisson flow of events the expectation value coincides with variance and $K(t)=1$, for the binomial flow $m(t)>D(t)$ and $K(t)<1$, and for the negative binomial flow $m(t)<D(t)$ and $K(t)>1$. We also obtained the expressions for the distribution laws of time of k - th event occurrence and their main numerical characteristics

REFRENCES:

[1] Sedjakin N.M. Elements of theory of random pulse flows. - M.: Sov. radio, 1965. (Rus)

[2] Tihonov V.I., Mironov M.A. Markov processes. - M.: Sov. radio, 1977. (Rus)

[3] Ventcel' E.S., Ovcharov L.A. The theory of stochastic processes and its engineering applications. - M.: Nauka, 1991. (Rus)

[4] Shahtarin B.I. Random processes in radio engineering. - M.: Radio i svjaz', 2000. (Rus)

[5] Gnedenko B.V., Kovalenko I.N. Introduction to queuing theory. - M.: KomKniga, 2005. (Rus)

[6] Yu.Yu. Gromov, I.G. Karpov, Laws of distribution of continuous random variable with maximum entropy. Generalized Method of Moments // Nauchno-tehnicheskie vedomosti Sankt-Peterburgskogo gosudarstvennogo politehnicheskogo universiteta. Informatika. Telekommunikacii. Upravlenie. 2009. Vol. 1. No 72. Pp. 37-42. (Rus)

[7] Yu.Yu. Gromov, I.G. Karpov, Further development of existing ideas about the basic forms of distribution laws and numerical characteristics of random variables for solving problems of information security / Informacija i bezopasnost'. 2010. Vol. 13. No 3. Pp. 459-462. (Rus)

[8] Yu.Yu. Gromov, I.G. Karpov, G.N. Nurutdinov, D.K. Proskurin, Generalized probabilistic model of conditional Poisson flow // Radiotehnika. 2010. No 12. Pp. 21-25. (Rus)

[9] Yu.Yu. Gromov, I.G. Karpov Generalized probabilistic description of flows of homogeneous events // Informacija i bezopasnost'. 2012. Vol. 15. No 1. Pp. 43-48. (Rus)

[10] Zajcev V.F., Poljanin A.D. Handbook of Differential Equations. - M.: Fizmatlit, 2001. (Rus)

[11] Prudnikov A. P., Brychkov Ju. A., Marichev O.I. Integrals and series. Elementary functions. - M.: Nauka, 1984. (Rus)