An Statistical Analysis of Queries Receipt Flows in E-Governance Systems

Kyrylo Sazonov
Applied Mathematics Department
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
kir1218@hotmail.com

Mohamed Alkilani
Applied Mathematics Department
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
moh_alkilani@yahoo.com

Volodymyr Kobziev
Applied Mathematics Department
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
volodymyr.kobziev@nure.ua

Статистичний Аналіз Потоків Надходження Запитів у Системах Електронного Урядування

Кирило Сазонов
кафедра прикладної математики
Харківський національний університет радіоелектроніки
Харків, Україна
кir1218@hotmail.com

Мохамед АлКілані
кафедра прикладної математики
Харківський національний університет радіоелектроніки
Харків, Україна
moh_alkilani@yahoo.com

Володимир Кобзєв
кафедра прикладної математики
Харківський національний університет радіоелектроніки
Харків, Україна
volodymyr.kobziev@nure.ua

***Abstract*—The statistical approach to analyzing the flow of requests in e-governance systems is described. The function and density of the generalized distribution of the time interval between the receipt of requests is given. Two of the three possible boundary distributions do not allow the scale parameter to be evaluated as a dispersion.**

***Анотація*—Описаний статистичний підхід до аналізу потоків запитів у системах електронного урядування. Наведено функцію та щільність узагальненого розподілу інтервалу часу між надходженням запитів. Два з трьох можливих граничних розподілів не дозволяють оцінювати параметр масштабу у вигляді дисперсії.**

Keywords— e-government systems; flow of requests; statistical distribution of requests characteristics

Ключові слова—системи електронного урядування; потоки запитів; статистичний розподіл характеристик запитів

I. INTRODUCTION

The current state of information and communication technologies provides rich opportunities for the creation and development of e-government systems in many countries around the world [1]. Increasingly, information, administrative and other services at the state and local level are provided with the use of specialized information systems. The productivity of such systems depends on the qualitative organization of their information space and the coordinated work of qualified personnel and a complex of modern computer and communication equipment in the conditions of constantly growing flows of requests for services.

A large number of simultaneous incoming queries requires a sufficient personnel number, which do pretreatment of requests, documenting the stages of their passage in the system, fulfilling queries for their content, as well as the control of the accuracy and timing of queries processing. That is why it is important to carefully analyze the nature of the queries flows in e-governance systems.

II. A GENERAL DESCRIPTION OF THE QUERIES FLOWS

The work of E-gov systems with query flows can be considered on the model of the system of mass service (SMS), in which the unit of flow is called the application, and the service device is a channel. A characteristic feature of such systems is: the presence of a flow of applications coming from outside; transforming this flow in accordance with the current technological process in the system and the outflow of applications [2].

It is known that the formation of the SMS model consists of two subtasks: the formalization of structural transformation and the study of the structure and parameters of the input flow and flow of service. In accordance with the structural

structural transformation of information, such systems differ on several following features: the behavior of the application in the case when all channels are occupied (systems with failures and waiting systems); limited input flow (systems with limited or unlimited flow of applications); discipline of service applications (without prioritization, with relative or absolute priorities); consistent channel participation for a series of actions with applications (single or multiphase).

The analysis of the peculiarities of the flow of requests in e-government systems is based on the processing of pre-assembled arrays of relevant statistics. There are two fundamentally different analysis possibilities: to investigate the array of discrete random variables representing the number of requirements k distributed at intervals of time t or to investigate an array of continuous random variables representing the time intervals between the receipt of two neighboring applications in the SMS.

In the first (discrete) approach, the simplest flow model is most often considered (the probability that the arbitrary time interval t will be exactly k requirements) leads to the Poisson's law. In the second approach (in the case of the study of continuous random variables), the simplest flow is described by a continuous function (or density) of the distribution.

The most frequently considered stationary flow without aftereffects, for which the intervals between events are distributed according to the normal law (it is called the normal flow). However, the characteristic feature of real situations in E-gov systems is a combination of several stationary flows, that is, inflow of flows with different values of the parameters of continuous distribution.

III. Statistical approach to estimating flows characteritics

The large number of incoming requests in real E-gov systems allows us to proceed to consider the behavior of the statistical parameters of requests with a significant increase in their number. In the theory of extremal values, this situation leads to one of three variants of the limiting distributions of the largest values.

In [3] it is established that there are three types of boundary distributions, each of which has a specific behavior of the largest absolute values of a random variable. The first of the established classes covers the group of distributions of output quantities of the exponential type. The second class is defined for distributions of the type of Cauchy with heavy tails, and the output variables are unlimited for the largest values from below, for the smallest values above, or in both directions. The third class is valid for distributions of quantities that are bounded up or down.

It is also known [3] that a linear random variable function in each of the three extreme distributions is also an extreme value.

The generalized distribution of extreme values has the form:

$$F(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\},$$

where: μ - placement parameter, σ - scale parameter, $\xi > 0$ - form parameter.

The density function is given by the expression:

$$f(\xi,\mu,\sigma) = e^{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right)} \cdot \frac{1}{\sigma}\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-(1/\xi)-1}.$$

It is a generalization of the Humbel, Frechet, and Weibull distributions, which are used for the approximate modulation of the maxima of finite sequences of random variables.

By using the specified generalized distribution parameters, Euler constants and gamma functions, if necessary, one can determine the mathematical expectation, variance, median, mode, and the excess coefficient of the generalized distribution of extreme values.

Of the three mentioned types of limit distributions, only the first allows estimating the spread of values as a dispersion. The fact is that the integral, through which the variance is calculated, does not converge for other types of distributions.

Using these patterns, one can analyze the statistical distribution of request flows in E-gov systems. Moreover, the behavior of the average characteristics of the query flows can be estimated by standard methods, as a parameter of the scattering of values around the mean, you need to use the selective variation or the interval between the previously selected quantiles of the obtained distribution.

IV. Conclusions

It is shown that in the case of a large number of requests received for processing in E-gov systems, it is possible to use a statistical description of their flows. Depending on the characteristics of the unknown distribution of the time of receipt of requests, it is necessary to analyze the limit distribution of extreme values and take into account its type to estimate the average parameters and parameters of values around the average when processing real flows of processing requests in E-gov systems.

References

[1] AlKilani M. Almabrouk. Principles of accjnting requests in E-gov systems // Информационные системы и технологии: материалы 5-й Международ.науч.-техн. конф., Харьков, 12-17 сентября 2016 г.: тезисы докладов /[редкол.: А.Д. Тевяшев (отв. ред.)]. – Х.: НТМТ, 2016. – с.18.

[2] Хемди А. Таха. Глава 17. Системы массового обслуживания // Введение в исследование операций = Operations Research: An Introduction. — 7-е изд. — М.: «Вильямс», 2007. — С. 629—697.

[3] Гумбель Э. Статистика экстремальных значений. – М.: Мир, 1965. – 450 с.