

## Research on the tasks arrive law of Satellite Information Application chain

HUANG Xuexiao<sup>1,a</sup>, LIU Gang<sup>2,b</sup>, XU Peide<sup>3</sup>

<sup>1</sup>Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China

<sup>2</sup>Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China

<sup>3</sup>Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China

<sup>a</sup>email: huangxuexiao91@163.com, <sup>b</sup>email:liugang1109@163.com

**Keywords:** Satellite information; Task arrives; PH distribution fitting

**Abstract.** This paper for the issue of Satellite Information Application chain input task describing and modeling, analyzes the basic process about tasks arising, And this paper will be divided into four satellite missions arrival process typically. Since the law has for mission to reach the general characteristics. So it is hard to described with general probability distribution. This paper proposes a method based on the EM algorithm fitting PH distributions, and uses PH distribution fitting the application of satellite information chain task input in different circumstances. At last, this paper gives experimental analysis to prove its feasibility.

### Introduction

In recent years, there is some incident aroused vigilance, such as earthquakes, floods and other natural disasters, as well as territorial disputes and maritime rights and interests. In response to these emergencies, people are increasingly using satellite image data to obtain the corresponding region, so as to better take appropriate measures to deal with these events [1]. As the satellite platform located lofty, not sovereignty restrictions, wide coverage, high reliability, has become an important tool for dealing with information security incidents. Information obtained through satellite platform under the pass, processing, integration and distribution, can be transformed into a suitable form, the user is actually available information. After acquiring the satellite image, it needs to be transmitted to the ground station, and after processing to users who needs it. This process is called satellite information application chain, but also the object of study. How to better respond to emergencies, which is the application of satellite information chain put forward higher requirements. Apart from increasing the number of satellite platforms, also need to study the design of satellite information application chain to ensure that the mass of the satellite information timely processing, distributed user.

In the study of the application of satellite information chain process, we met an obvious problem: How to be more realistic to describe the input law of Satellite Information Application chain? We only know the characteristics of the input and the input of the law, in order to assess the theoretical analysis of each design of Satellite Information Application chain, and ensure that the design of satellite information application chain can meet the actual demand. Satellite Information Application chain is similar with workflow. Currently, there are many studies about the workflow. Cook and Wolf[5] first proposed the idea about the reconstruction task flow in the software engineering literature, the proposed three kinds of flow reconstruction method. Agrawal[6] first introduced the idea of the reconstruction task flow in workflow management context. Research on PH distribution, the earliest dating back to the early work of Erlang (1918) . Professor Tian Naishuo [7] in his book devoted to the status of research on PH distribution. S.Asmussen [8] using the EM algorithm is generally in the form of data fitting PH distribution family, contributed to research data

about the PH distribution fitting areas.

This paper studies the task to reach the law of Satellite Information Application chain. Under the fixed number of satellites, the paper considers several typical Satellite Information Application chain task arrives law and the importing task intensity. We use PH distributions were fitted to Input task to support the follow-up analysis and evaluation work on the application of satellite information chain.

### Characteristics of satellite information needs

The satellite system as desk, if the user needs is a Poisson arrival and service obey the law of the negative exponential distribution, then enter the task of satellite information in the application chain is also Poisson arrival. But in reality, the satellite is not in line with the mandate to reach the Poisson arrival law. Using the Poisson distribution describes the task arrives, its apparent irrationality.

Existing satellite information needs analysis modeling method usually used exponential distribution or the like as a typical normal distribution assumption of random variables, and it will lead to poor model adaptability. In view of this situation, this paper will demand arrives, the probability distribution of the service time distribution is assumed to PH. In  $[0, +\infty)$  PH distributed to all probability distribution function with dense nature, can be used as general-purpose general representation of the distribution, so the use of PH distributed random variables modeling can relax the constraints and reduce the difficulty parsing model, and improve computable model. This part of the process of combining the characteristics of warfare and satellite information security, focusing on four different tasks demand model takes into account the strength of the different emergency situations. These four models were previously used other distributions other than the exponential distribution is difficult to resolve the problem of modeling, clearly showing PH distribution has the same exponential distribution and excellent analytical characteristics.

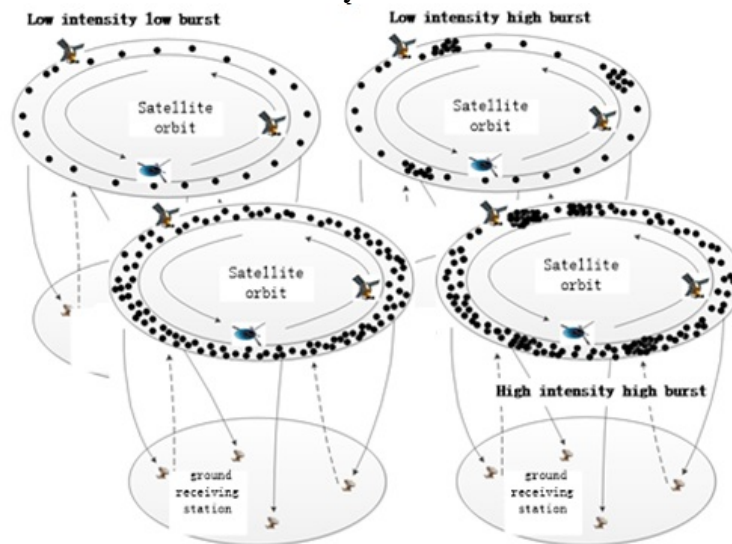


Fig.1. Four typical task arrives

The four typical tasks to achieve the law to reach the sudden intensity as the feature task flow and task flow is described, low-strength concrete can be divided into Low intensity low burst, Low intensity high burst, High intensity low burst, High intensity high burst.

Low intensity low burst for strategic reconnaissance satellites of everyday situations. Specific scenarios can be described as: satellite surveillance application information chain is mainly responsible for handling routine reconnaissance mission, the node hardware and software configuration is relatively small professional and technical staff, does not have the ability to handle high-intensity satellite application task flow.

Low intensity High burst, adapted to allow for the usual satellite reconnaissance certain period of time during insertion of the surveillance area at specific locations, and the task of data processing. Scenarios can be described as: satellite surveillance application information chain is mainly

responsible for the daily reconnaissance processing tasks, tasks for satellite applications in case of emergency have some buffering capacity. The number of satellite application tasks on time distribution has a significant burst characteristics over a period of time to reach the total of the small number of satellite application tasks, but the distribution of these tasks is not uniform in time, at some point in time on the satellite application tasks number will suddenly increase, namely the existence of shock phenomenon.

High intensity low burst for wartime ground control system for a period of time in advance of the satellite reconnaissance missions injected intensive instruction sequence and other tasks for the period strictly limited circumstances inserted. Scenarios can be described as: processing capacity of each node satellite reconnaissance information chain applications are very strong, the application can continue processing satellite high intensity task flow.

High intensity high burst for wartime or emergency situations launch small satellites for battlefield operations in the region to implement continuous reconnaissance of the situation. Scenarios can be described as: Desk configure the number of satellite reconnaissance application information chain more than adequate, can cope with satellite information security needs of various emergency situations and emergency situations. Satellite applications more intense wartime mission requirements, and all kinds of unexpected situations can cause satellite application tasks presented shock phenomenon in the time distribution.

### PH distribution data fitting based on EM Algorithm

PH distribution is a finite-state Markov process absorption and distribution of time, due to the different phase PH distributions fit the description of a complex multi-level random phenomenon and change parameters, it quickly became a powerful tool for the analysis of contemporary stochastic model. PH subclass distribution family, the hyper-Erlang distribution with good analytical nature, and in all probability distribution of non-negative random variables is dense. Therefore, this paper chooses of a hyper-Erlang distribution to fit the experimental data.

For fitting PH distributions, researchers often used two methods, matrix estimation and maximum likelihood estimation. Due to the higher of the solution matrix estimation requirements, we use EM algorithm parameters for maximum likelihood estimation, to get a better fitting effect. EM algorithm is an iterative algorithm posterior distribution of the latter in the case of incomplete data to calculate maximum likelihood estimation or, alternately perform two steps at each iteration of the loop process: E-step (Expectation step) and M step (Maximization step). E-step given in the case of complete data and parameters obtained in the previous iteration estimates, expectations of the condition of the log-likelihood function corresponding to the data to calculate exactly; M step maximization complete data log-likelihood function to determine parameters value, and the next step for the iteration. When algorithm parameters are less than a pre-given threshold end, return the parameter values.

For a hyper-Erlang distribution which is provided with I branch, wherein the number of the i-th branch stage, the parameters for the general Erlang distribution is  $\lambda_i$ . So the probability density function is:

$$f(x) = \sum_{i=1}^I a_i \times \frac{\lambda_i^{r_i} x^{r_i-1} e^{-\lambda_i x}}{(r_i - 1)!}, x \geq 0 \quad (1)$$

Wherein,  $a_i$  is the probability of the i-th branch, and  $\sum_{i=1}^I a_i = 1$

Each general Erlang distribution density function is:

$$f_i(x) = a_i \times \frac{\lambda_i^{r_i} x^{r_i-1} e^{-\lambda_i x}}{(r_i - 1)!}, x \geq 0, \lambda_i > 0, a_i \geq 0 \quad (2)$$

$$\Omega = \{f(x), f(x) = \sum_{i=1}^I a_i \times \frac{\lambda_i^{r_i} x^{r_i-1} e^{-\lambda_i x}}{(r_i-1)!}, x \geq 0, \lambda_i > 0, a_i \geq 0, r_i \geq 1, I \geq 1, \sum_{i=1}^I a_i = 1\}$$

Defined: (3)

Then  $\Omega$  is a collection of hyper-Erlang distribution density function. It can prove, density function of any non-negative random variables can be used to approximate the distribution of one set of hyper-Erlang distribution.

EM algorithm:

E-step: Known set K independent and identically distributed sample points is  $x = (x_1, x_2, \dots, x_N)$ . Parameters that need to fit hyper-Erlang distribution is  $\theta = (a_1, a_2, \dots, a_I; \lambda_1, \lambda_2, \dots, \lambda_I)$ . Unobservable data from the corresponding X is  $y = (y_1, y_2, \dots, y_N)$ . Get complete data likelihood function:

$$L(\theta|x, y) = \sum_{j=1}^N \ln(f(x_j, y_j | \theta)) \quad (4)$$

If we know k-1-step iteration value  $\theta^{k-1}$ , By Bayes formula we can get the probability density function of  $y_i, i=1, 2, \dots, K$ .

$$g(y | x_i, \theta^{k-1}) = \frac{f(x_i, y | \theta^{k-1})}{f(x_i | \theta^{k-1})} \quad (5)$$

Where  $f(x_i | \theta^{k-1})$  is the marginal density functions  $x_i$ , and we can get

$$E(\theta | \theta^{k-1}) = \sum_{i=1}^N E[\ln f(x_i, y_i | \theta)] = \sum_{i=1}^N \int [\ln f(x_i, y_i | \theta)] \bullet g(y | x_i, \theta^{k-1}) dy \quad (6)$$

M Step: Major solving  $\theta^k = \max_{\theta} E(\theta | \theta^{k-1})$ .

This article assumes that unobserved data  $y_i \in \{1, 2, \dots, I\}, i=1, 2, \dots, K$ . So if  $y_i = m$ , then the m-th point  $x_i$  can be regarded as a branch from the m-th Erlang mixed Erlang distribution. So:

$$g(y_i | x_j, \theta) = \frac{a_{yi} \lambda_i (\lambda_i x_j)^{r_i-1} e^{-\lambda_i x_j} / [(r_i-1)!]}{\sum_{m=1}^I a_{ym} \lambda_m (\lambda_m x_j)^{r_m-1} e^{-\lambda_m x_j} / [(r_m-1)!]} \quad (7)$$

$$E(\theta | \theta^{k-1}) = \sum_{i=1}^I \sum_{j=1}^N (\ln a_i) \bullet g(i | x_j, \theta^{k-1}) + \sum_{i=1}^I \sum_{j=1}^N \ln(f_i(x_j | \lambda_i)) \bullet g(i | x_j, \theta^{k-1}) \quad (8)$$

When the two values of the right side of the above formula 8, while taking the maximum, the equation takes the maximum value. This article reference document [10] processing method, calculated:

$$a_i^{k-1} = \frac{1}{N} \sum_{j=1}^N g(i | x_j, \theta^{k-1}), i=1, 2, \dots, I$$

$$\lambda_i^{k-1} = \frac{r_i \sum_{j=1}^N g(i | x_j, \theta^{k-1})}{\sum_{j=1}^N g(i | x_j, \theta^{k-1}) \bullet x_j}, i=1, 2, \dots, I \quad (9)$$

Program:

The first step: initialization parameters  $\theta^0 = \theta, x = (x_1, x_2, \dots, x_N), k=1, I, N, \varepsilon$

The second step:

E steps were calculated as follows:

$$f_i(x_n | \lambda_i^{k-1}) = \lambda_i (\lambda_i x_j)^{r_i-1} e^{-\lambda_i x} / [(r_i - 1)!]$$

$$f(x_n | \theta^{k-1}) = \sum_{i=1}^I a_i f_i(x_n | \lambda_i^{k-1})$$

$$g(i | x_j, \theta^{k-1}) = a_i f_i(x_n | \lambda_i^{k-1}) / f(x_n | \theta^{k-1})$$

M steps were calculated as follows:

$$a_i^{k-1} = \frac{1}{N} \sum_{j=1}^N g(i | x_j, \theta^{k-1}), i = 1, 2, \dots, I$$

$$\lambda_i^{k-1} = \frac{r_i \sum_{j=1}^N g(i | x_j, \theta^{k-1})}{\sum_{j=1}^N g(i | x_j, \theta^{k-1}) \bullet x_j}, i = 1, 2, \dots, I$$

The third step: *if*  $\|\theta^{k+1} - \theta^{k-1}\| \leq \varepsilon$ , *return*  $\theta^{k+1}$ , *else*,  $\theta^k = \theta^{k-1}$ , *return* to step 2. The program has been implemented on MATLAB.

### The mission arrives law fit of satellite information needs

In this paper, the fitting procedure has been achieved on the arrival of four typical tasks to fit. When hyper-Erlang distribution of the number of branches were chosen, for convenience, this article select the number of branches  $I = 5$ . After constantly adjust the number of stages  $r_i$ , to get the four tasks arrive fitting curve.

For low intensity low burst task arrives law, the paper get 100 experimental data values within a cycle. Because low intensity low burst task arrival time more evenly, through the EM algorithm to fit the data, we found that the basic laws of arrival presented as the mean of a normal distribution of 10, shown in Figure 2.

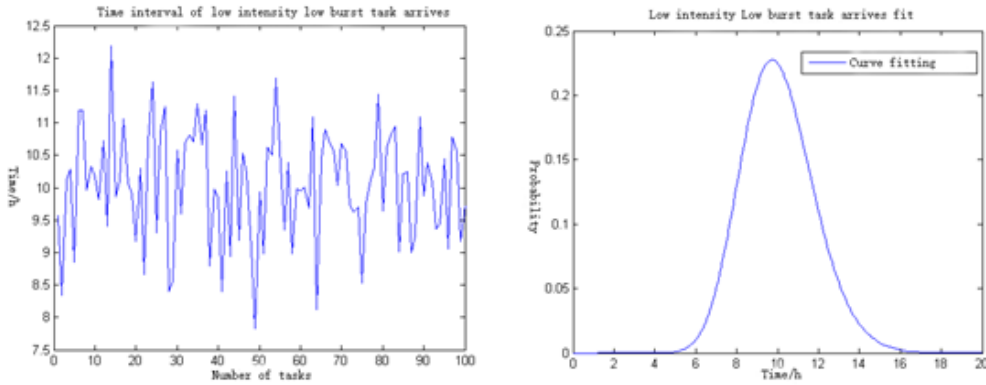


Fig.2. Low intensity Low burst task arrives fit

For Low intensity High burst task arrives law, It is within the period of normal operation is similar to low intensity low burst task arrives law, but in a special period of time will be a sudden increase in the number of tasks arrive. This paper get 100 experimental data values within a cycle, and the results obtained are similar to the actual description, there is a larger probability density near the mean of 4 and 10, as shown in Figure 3.

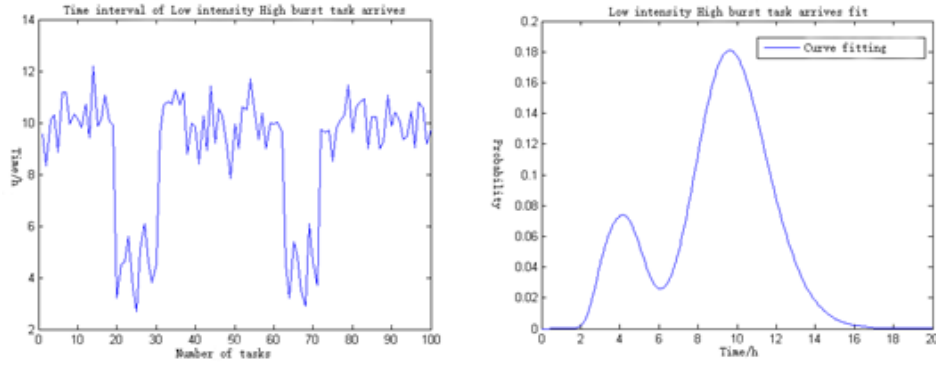


Fig.3. Low intensity High burst task arrives fit

High intensity low burst task arrives law and high intensity high burst task arrives law in a similar situation. The difference is that the time to reach the high intensity task interval will be shorter. Cure fitting shown in Figures 4 and 5.

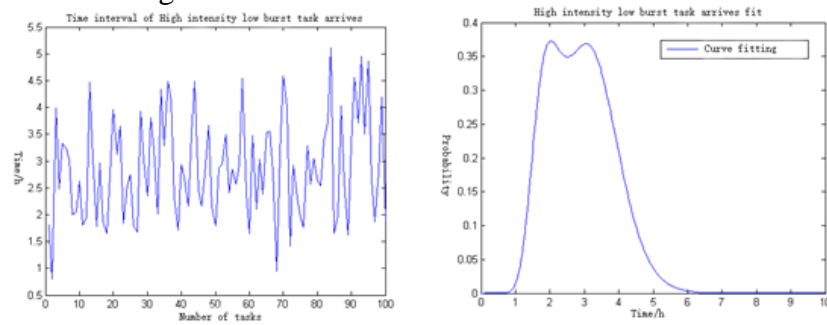


Fig.4. High intensity low burst task arrives fit

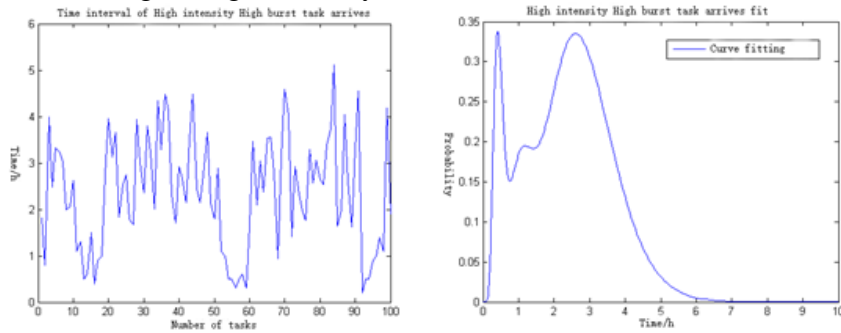


Fig.5. High intensity High burst task arrives fit

## Conclusion

When analyzing the Satellite Information Application chain, understand the task input of Satellite Information Application chain is a prerequisite. Only given the task law and features of satellite applications, in order to apply the theory to the satellite link design information to evaluate and ensure the program can meet the actual demand. According to the difference between wartime and everyday applications, satellite mission arrives law has its particularity, it is difficult simply using the general probability distribution to describe it. In this paper, PH distributed on the whole family has a dense probability distribution of the characteristics, the use of EM algorithm for data fitting satellite task arrives, get a better fitting effect. It is beneficial to further research. Meanwhile, the improved EM algorithm, such that it can be self-optimizing the number of branches and branch status in several areas, is future research directions.

## Acknowledgement

In this paper, the research was sponsored by the Nature Science Foundation of Henan Province (Project No. 201112400450401) and Youth Fund Project of Luoyang Institute of Science and

Technology (Project No. 2010QZ16).

## References

- [1] Guide, V.D.R. Production planning and control for remanufacturing: industry practice and research needs[J]. Journal of Operations Management, 2000, 18 (4): 467-483.
- [2] LI Zhi-meng, CHEN Jun-liang, MA Man-hao, et al. Design and Realization of System of Effectiveness Evaluation of Satellite Military Application based on Simulation[J]. Fire Control & Command Control, 2010, 35(10):157-159. (in Chinese)
- [3] KASLOW D. Planning and scheduling of earth observing satellites[C]//IEEE. IEEE Aerospace Conference. Malta. 2007:1-12.
- [4] WHITEFIELD D, RAJEEV G. Capacity enhancement with dynamic resource management for next generation satellite systems[C]//IEEE. IEEE MILCOM. New Jersey. 2005:761-767.
- [5] Gong Zhuo-jun. Research about Satellite Applications task decomposition [D]. Changsha: National University of Defense Technology, 2009.
- [6] COOK JE, WOLF AL. Software process validation: Quantitatively measuring the correspondence of a process to a model[J]. ACM Trans. on Software Engineering and Methodology, 1999, 8(2):147-176.
- [7] Tian Nai-shuo, QBD processes and matrix geometric solution [M]. Beijing: Science press, 2002.
- [8] S. Asmussen, O. Nerman, M. Olsson, Fitting Phase-Type Distributions via the EM Algorithm[J]. Scandinavian Journal of Statistics, 1996, 23 (4):419-441.
- [9] Asmussen S. Applied probability and queues[M]. Second Edition, Springer, 2003: 84-85.
- [10] Zhao Gou-xi, Wang Hong-wei. Mixed-Erlang Distribution Fitting for Heavy-tailed Data[J]. Operations Research and Management Science, 2013, 22(2):93-96.
- [11] AGRAWAL R, CRUZ RL, OKINO C, et al. Performance bounds for flow control protocols[J]. ACM Trans on Networking. 1999, 7(3):310-323.