

待ち行列シンポジウム
「確率モデルとその応用」
報文集

PROCEEDINGS
OF
THE QUEUEING SYMPOSIUM
STOCHASTIC MODELS AND THEIR APPLICATIONS

2018 年 1 月 17 日（水）～ 19 日（金）
大阪大学会館 アセンブリー・ホール（大阪府豊中市）

実行委員長：塩田 茂雄（千葉大学）

主催 待ち行列シンポジウム実行委員会

2017年度待ち行列シンポジウム「確率モデルとその応用」

2018年1月17日(水)～19日(金)
大阪大学会館 アセンブリーホール(大阪府豊中市)

セッションの種類は緑=一般, 青=学生, 赤=ショート
質疑応答5分を含め一般1人あたり30分, 学生1人あたり25分, ショート1人あたり20分
*=発表者(複数著者の場合)

1月17日(水)

14:00 開会にあたって 塩田 茂雄(千葉大学)

Session W1 (14:10-15:40) 一般セッション 座長: 増山 博之(京都大学)

- 14:10 A Note on Quantitative Security Evaluation of VM-Based Intrusion Tolerant Systems with MAP
Zheng Junjun, *岡村 寛之, 土肥 正(広島大学)
- 14:40 Congestion Control for a System with Parallel Stations and Homogeneous Customers using Priority Passes
*増田 靖(慶應義塾大学), 辻 暁(キヤノン)
- 15:10 物理対経済体のサンドイッチ場考: 待ち行列法によるベアマップ仮説と波動方程式化例
松井 正之(神奈川大学)

Session W2 (16:05-17:35) 一般セッション 座長: 笠原 正治(奈良先端科学技術大学院大学)

- 16:05 Cross Moments Approaches for Mean-Bounded Finite Random Classes
高田 寛之(長崎大学)
- 16:35 Distribution of the waiting and service time in an M/M/m preemptive-resume priority queue with impatient customers
高木 英明(筑波大学)
- 17:05 A Toy Model of Preferentially-Attached Networks with Break-Ups
*豊泉 洋, 下邨 貴裕(早稲田大学)

懇親会(18:00-20:00) 司会: 増山 博之(京都大学)

1月18日(木)

Session T1 (9:30-11:10) 学生セッション 座長: 岸 康人(松蔭大学)

- 9:30 レベル依存する G/G/1 型マルコフ連鎖における条件付き定常分布の誤差評価付き数値計算法
*木村 雅俊, 滝根 哲哉(大阪大学)
- 9:55 フロー到着とクロスファイックを考慮したSDN の待ち行列モデルとその解析
*牛垣 龍一(東京工業大学), フンドックトウアン(筑波大学), 河西 憲一(群馬大学),
Yuan-Cheng Lai(National Taiwan University of Science and Technology), Ying-Dar Lin(National Chiao Tung University)
- 10:20 着陸待ち行列
*石井 宏明, 佐藤 真史, 佐藤 賢, 豊泉 洋(早稲田大学)
- 10:45 Cooperation area を考慮した高密度セルラ移動体ネットワークの性能解析
原口 昌太(東京工業大学)

Session T2 (11:35-12:35) ショートペーパーセッション 座長: 井家 敦(神奈川工科大学)

- 11:35 エスカレーターは歩くべきか、止まるべきか
*鶴田 裕太郎, 豊泉 洋(早稲田大学)
- 11:55 雀荘における待ち行列に関する考察
*井手 広康, 奥田 隆史(愛知県立大学)
- 12:15 マルコフモデルを用いたテニスプレーヤーの分類
*大久保 幸美, 豊泉 洋(早稲田大学)

12:35-14:05 昼休み

Session T3 (14:05-15:20) 学生セッション 座長: フンドックトゥアン(筑波大学)

- 14:05 セルラネットワークにおける移動体のデータ量解析とハンドオフの機会制限モデルの考察
徳山 喜一(東京工業大学)
- 14:30 リードタイムが一般の分布にしたがう(r,Q) 方式在庫管理システムの解析
高木 秀(東京工業大学)
- 14:55 GI/M^b/1 モデルによるビットコイン・トランザクション承認時間解析
*河瀬 良亮, 笠原 正治(奈良先端科学技術大学院大学)

Session T4 (15:45-17:25) 学生セッション 座長: 木村 達明(NTT)

- 15:45 Probabilistic model checking for biochemical reaction systems
*Ty Ratana, 藤田 憲悦, 河西 憲一(群馬大学)
- 16:10 複合ポアソン過程を用いた企業価値のモデル化とリスク評価
井田 倫宏(東京工業大学)
- 16:35 サーバブロックの起動時間を持つデータセンターの待ち行列モデルによる性能解析
*日山 慎人, フンドックトゥアン(筑波大学)
- 17:00 マルコフ型集団到着過程と裾の軽いサービス分布を持つ無限サーバ待ち行列の安定性の十分条件
*矢島 萌子(東京工業大学), フンドックトゥアン(筑波大学), 増山 博之(京都大学)

1月19日(金)

Session F1 (9:30-11:10) 学生セッション 座長: 河西 憲一(群馬大学)

- 9:30 BGPネットワークの障害状況のマッピング
*伊藤 翔馬, 小林 香(富山県立大学)
- 9:55 行列式点過程が作る無線干渉場の巨視的極限
*油山 拓生, 三好 直人(東京工業大学)
- 10:20 BGPデータを使ったネットワーク障害のパターン推定
*奥寺 諒, 小林 香(富山県立大学)
- 10:45 Ginibre cluster process and its application in wireless networks
*小林 拓矢, 三好 直人(東京工業大学)

Session F2 (11:35-12:35) ショートペーパーセッション 座長: 加藤 憲一(神奈川大学)

- 11:35 2次元反射型ランダムウォークにおける定常分布の積率母関数の収束領域: 解析的証明
*小林 正弘(東海大学), 宮沢 政清(東京理科大学), 小沢 利久(駒澤大学)
- 11:55 連続時間非斉時マルコフ連鎖の特別なクラスに対する一様化法
井上 文彰(大阪大学)
- 12:15 強相関近似による複雑ネットワーク上の情報拡散過程の解析
*南川 雅人, 中島 圭佑, 塩田 茂雄(千葉大学)
- 12:35 待ち行列研究部会 研究奨励賞 受賞者発表 塩田 茂雄(千葉大学)
閉会にあたって 笠原 正治(奈良先端科学技術大学院大学)

目 次

A Note on Quantitative Security Evaluation of VM-Based Intrusion Tolerant Systems with MAP ... 1	
Zheng Junjun, 岡村 寛之, 土肥 正(広島大学)	
Congestion Control for a System with Parallel Stations and Homogeneous Customers using Priority Passes	11
増田 靖(慶應義塾大学), 辻 暁(キヤノン)	
物理対経済体のサンドイッチ場考:待ち行列法によるペアマップ仮説と波動方程式化例	16
松井 正之(神奈川大学)	
Cross Moments Approaches for Mean-Bounded Finite Random Classes	23
高田 寛之(長崎大学)	
Distribution of the waiting and service time in an M/M/m preemptive-resume priority queue with impatient customers	33
高木 英明(筑波大学)	
A Toy Model of Preferentially-Attached Networks with Break-Ups	43
豊泉 洋, 下邨 貴裕(早稲田大学)	
レベル依存する G/G/1 型マルコフ連鎖における条件付き定常分布の誤差評価付き数値計算法	51
木村 雅俊, 滝根 哲哉(大阪大学)	
フロー到着とクロストラフィックを考慮した SDN の待ち行列モデルとその解析	61
牛垣 龍一(東京工業大学), フン ドック トゥアン(筑波大学), 河西 憲一(群馬大学), Yuan-Cheng Lai(National Taiwan University of Science and Technology), Ying-Dar Lin(National Chiao Tung University)	
着陸待ち行列	71
石井 宏明, 佐藤 真史, 佐藤 賢, 豊泉 洋(早稲田大学)	
Cooperation area を考慮した高密度セルラ移動体ネットワークの性能解析	78
原口 昌太(東京工業大学)	
エスカレーターは歩くべきか、止まるべきか	88
鶴田 裕太郎, 豊泉 洋(早稲田大学)	
雀荘における待ち行列に関する考察	90
井手 広康, 奥田 隆史(愛知県立大学)	
マルコフモデルを用いたテニスプレイヤーの分類	92
大久保 幸美, 豊泉 洋(早稲田大学)	

セルラネットワークにおける移動体のデータ量解析とハンドオフの機会制限モデルの考察.....	94
徳山 喜一(東京工業大学)	
リードタイムが一般の分布にしたがう(r;Q) 方式在庫管理システムの解析.....	102
高木 秀(東京工業大学)	
GI/M ^b /1 モデルによるビットコイン・トランザクション承認時間解析	109
河瀬 良亮, 笠原 正治(奈良先端科学技術大学院大学)	
Probabilistic model checking for biochemical reaction systems.....	119
Ty Ratana, 藤田 憲悦, 河西 憲一(群馬大学)	
複合ポアソン過程を用いた企業価値のモデル化とリスク評価	129
井田 倫宏(東京工業大学)	
サーバブロックの起動時間を持つデータセンターの待ち行列モデルによる性能解析.....	135
日出山 慎人, Phung-Duc Tuan(筑波大学)	
マルコフ型集団到着過程と裾の軽いサービス分布を持つ無限サーバ待ち行列の安定性の十分条件	145
矢島 萌子(東京工業大学), フン ドック トゥアン(筑波大学), 増山 博之(京都大学)	
BGP ネットワークの障害状況のマッピング	153
伊藤 翔馬, 小林 香(富山県立大学)	
行列式点過程が作る無線干渉場の巨視的極限.....	161
油山 拓生, 三好 直人(東京工業大学)	
BGP データを使ったネットワーク障害のパターン推定.....	171
奥寺 諒, 小林 香(富山県立大学)	
Ginibre cluster process and its application in wireless networks.....	180
小林 拓矢, 三好 直人(東京工業大学)	
2次元反射型ランダムウォークにおける定常分布の積率母関数の収束領域:解析的証明.....	190
小林 正弘(東海大学), 宮沢 政清(東京理科大学), 小沢 利久(駒澤大学)	
連続時間非斉時マルコフ連鎖の特別なクラスに対する一様化法.....	192
井上 文彰(大阪大学)	
強相関近似による複雑ネットワーク上の情報拡散過程の解析	194
南川 雅人, 中島 圭佑, 塩田 茂雄(千葉大学)	

A Note on Quantitative Security Evaluation of VM-Based Intrusion Tolerant Systems with MAP

Junjun Zheng Hiroyuki Okamura Tadashi Dohi
Department of Information Engineering
Graduate School of Engineering
Hiroshima University

Abstract Intrusion tolerance is an ability to keep the correct service by masking the intrusion based on fault-tolerant techniques. With the rapid development of virtualization, the virtual machine (VM) based intrusion tolerance scheme has been developed according to the concept of state machine replication with Byzantine fault tolerant technique. This paper presents the quantitative security evaluation of the VM-based intrusion tolerant system with the time to security failure. We assume that the arrival stream follows a Markovian arrival process which is one of the most general stochastic processes, and analytically derive the Laplace-Stieltjes transform of time to security failure based on the analysis of the $MAP/G/1/\infty$ queue.

1. Introduction

Computer security is the protection of computing systems and data that are stored or accessed, and has become an increasingly important hot topic in computer and communication industry, since it is important to support critical business processes and to protect personal and sensitive information. In general, computer security is categorized to three aspects; confidentiality, integrity and availability (CIA). Confidentiality is to make sure that the data is available only for intended and authorized persons. Integrity aims at maintaining and assuring the accuracy and consistency of data. The function of availability is to make sure that the data, network resources or network services are continuously available for the legitimate users. Highly-secure system should be designed to guarantee high levels of these criteria.

With the rapid development of computer networks, computer systems face various types of security threats, such as deny-of-service (DoS), computer virus, and intrusion. These threats cause security failures that degrade the CIA performance of system. This paper focuses on the intrusion which is a major cause of degrading both confidentiality and integrity of the system. Once the intrusion occurs, unauthorized users access to the system by exploiting security holes. The intrusion detection system (IDS) is well known as a countermeasure against the intrusion. The IDS is to detect misuse or abnormality of system automatically by monitoring the system and to send a warning to system operators. The disadvantage of IDS is that the restoration should be operated by humans. Thus the IDS is not always appropriate for highly-distributed systems consisting of a large number of computing agents used in recent computational environments. The intrusion tolerance system (ITS) is another countermeasure against the intrusion. The ITS keeps the correct service by masking the intrusion based on fault-tolerant techniques. Generally, the intrusion is regarded as a Byzantine failure which is defined as arbitrary deviations of an intruded process from its normal behavior. Thus the ITS utilizes the Byzantine fault tolerance technique. The ITS is more appropriate for the distributed system than IDS. Also the ITS is promising for the cyber physical system.

Since the early of 2000s, the ITS has been discussed regarding its implementation and performance evaluation in the enterprise-like system. For example, Wang et al. [1] developed a scalable intrusion tolerant architecture for the distributed service with COTS (commercial-off-the-shelf) servers. Zhao [2] proposed a web server with Byzantine fault tolerance framework. Madan et al. [3] discussed the quantitative security measures for the ITS based on COTS servers by using discrete-time semi-Markov models. In recent years, the ITS has been drastically enhanced by using the virtualization technique. Junior et al. [4] presented an intrusion tolerant system with virtual

machines (VMs), where they used the technique, which is called shared memory based intrusion tolerance. Although $3f + 1$ replicas are needed to tolerate f Byzantine failures in [2], the Junior et al.'s scheme requires only $2f + 1$ VMs to tolerate f Byzantine failures because the VM monitor is never intruded. Also, Lau et al. [5] developed a consensus protocol with the multi-stage algorithm in the shared memory based intrusion tolerance. To tolerate f Byzantine failures, $f + 1$ VMs are required at the first stage, and $2f + 1$ VMs are required at the second stage.

In this paper, we focus on the quantitative security evaluation of VM-based ITS. Zheng et al. [6] discussed the performance measure for a VM-based ITS as the success (failure) probability for processing one task. However, the success probability is insufficient to evaluate the system performance, because the success probability ignores external environment factors such as arrival stream of access. On the other hand, In [3], Madan et al. proposed the mean time to security failure (MTTSF) for a COTS-based ITS as a standard performance index of system security, which depends on external environments as well as internal ones. In the paper, we also consider the MTTSF for the VM-based ITS. Recently, Zheng et al. [7] formulated the MTTSF for the VM-based ITS under the Poisson arrival stream. This paper is an extension of [7]. In particular, we consider a Markovian arrival process (MAP) [8] as the arrival stream of tasks. Generally, it is well known that the arrival stream strongly affects the system performance, thus it is significant to consider the arrival stream of tasks even in the design of intrusion tolerance scheme. The MAP is one of the fairly general point processes which is defined by a continuous-time Markov chain (CTMC) and has a capability of representing a wide class of arrival processes. The MAP can approximate any point process accurately, and includes the Poisson process as a special case. Also, the MAP includes as a sub-class the Markov modulated Poisson processes (MMPPs) and the superposition of phase-type (PH) renewal processes. Thus the formulation of this paper completely includes the result of [7] mathematically.

2. VM-Based Intrusion Tolerant System

Consider an intrusion tolerant system in the virtualized environment as shown in Fig. 1(a). In the system, we suppose that there are n (≥ 2) VMs and one component agreement service (AS) running on a hypervisor. The VMs are identical, provide the same services to clients, and communicate with each other through a shared memory (SM) in the physical host where any VM can read and write values. The AS is assumed to be reliable, and is kept isolated. When the system receives a task from a client, all VMs process the task independently, and store the responses in SM. Then AS makes a consensus under majority voting. That means, the consensus will succeed if the number of intruded VMs is less than $n/2$. Let f be the tolerance level, namely, the maximum number of intruded VMs that are tolerated to the extent that the system behaves normally. From the argument of Byzantine fault tolerance, the inequalities $n \geq 2f + 1$ and $f \leq (n - 1)/2$, hold. If consensus succeeds, AS puts a signature into the agreed response and then the system sends the reply to the client. If consensus fails, AS activates r suspended VMs, and all VMs including already-intruded VMs process the task at the subsequent round. In general, the intruded VMs can be regarded as VMs with Byzantine failures. Also, the maximum number of rounds is given as m . Thus, for one service process, AS tries to make a consensus within m rounds using n initially activated VMs and r additionally activated VMs at each round. In this case, the maximum tolerance level of system is given by $n + r(m - 1) \geq 2f + 1$ and $f = \lfloor (n + r(m - 1) - 1)/2 \rfloor$. In addition, since the AS can make the consensus at the first round if the number of intruded VMs is less than $n/2$, thus it is clear that the tolerance level at the first round is $f = \lfloor (n - 1)/2 \rfloor$ which is more severe than $f = \lfloor (n + r(m - 1) - 1)/2 \rfloor$. That is, the intrusion tolerance improves as the number of rounds increases. Besides, a security failure occurs when the consensus fails for all the rounds.

Figure 1(b) depicts the possible behavior of one service process in the system design: $n = 3, r = 2$ and $m = 3$. In the figure, the initially activated VMs are VM1, VM2, and VM3. Each VM processes the task independently after a task arrives at the system. VM1 and VM3 are intruded during the task processing. Thus AS fails to make a consensus at the end of the first round, since

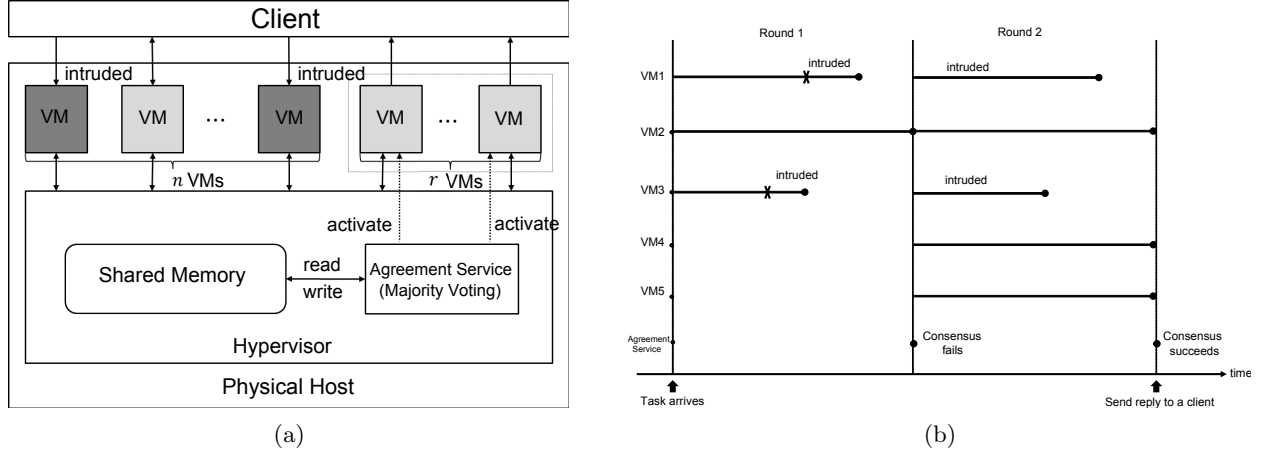


Figure 1: VM-based ITS: (a) System architecture, (b) Possible behavior of one service process.

the majority of VMs are intruded. AS activates VM4 and VM5, and then all VMs process the task at the second round. The consensus eventually succeeds at the second round. In such case, the service process for one task is succeeded within two rounds. If the consensus still fails at the second round, r suspended VMs will be activated, then the service process goes to the third round. The security failure occurs when the consensus cannot be made at the third round.

3. Markovian Arrival Process

Suppose that the arrival stream to the VM-based ITS follows a Markovian arrival process (MAP). The MAP is an extension of Poisson process where arrival rate varies according to a continuous-time Markov chain (CTMC), and is commonly used to model arrival streams in communication networks because it can represent the bursty traffic.

Let $N(t)$ be the cumulative number of arrivals following a MAP before time t . When the underlying CTMC is given by $J(t)$ on the finite state space, the MAP is represented by the two-dimensional point process $(N(t), J(t))$ which is a CTMC with the following infinitesimal generator:

$$Q_{\text{MAP}} = \begin{pmatrix} D_0 & D_1 & & \\ & D_0 & D_1 & \\ & & \ddots & \ddots \end{pmatrix}, \quad (1)$$

where D_0 is an infinitesimal generator without the occurrence of arrival and D_1 is a transition rate matrix when an arrival occurs. The state of $J(t)$ is called the phase of MAP. Note that the infinitesimal generator of phase process $J(t)$ becomes $D_0 + D_1$. That is,

$$(D_0 + D_1)\mathbf{1} = 0, \quad (2)$$

where $\mathbf{1}$ is a column vector whose elements are all equal to 1. In the paper, we assume the underlying CTMC is irreducible. Thus there exists a stationary probability (row) vector π_s such that

$$\pi_s(D_0 + D_1) = 0, \quad \pi_s\mathbf{1} = 1. \quad (3)$$

Using the stationary probability vector, the average number of arrivals per unit time, called the *fundamental arrival rate* of the MAP, is given by

$$\lambda = \pi_s D_1 \mathbf{1}. \quad (4)$$

Here we suppose that the service times of tasks are independent and identically distributed (i.i.d.) random variables with cumulative distribution function (c.d.f.) $F(t)$. Let ρ be the traffic intensity.

Then we have

$$\rho = \lambda \int_0^\infty t dF(t). \quad (5)$$

The MAP leads to an embedded Markov chain on time instants when arrivals occur, and the transition probability matrix of the embedded Markov chain is given by $\mathbf{T} = (-\mathbf{D}_0)^{-1}\mathbf{D}_1$. The stationary probability vector for the embedded Markov chain $\boldsymbol{\pi}_d$ becomes

$$\boldsymbol{\pi}_d = \boldsymbol{\pi}_d(-\mathbf{D}_0)^{-1}\mathbf{D}_1, \quad \boldsymbol{\pi}_d \mathbf{1} = 1. \quad (6)$$

Let $\hat{X}$ be an inter-arrival time of MAP in the steady state. Then $\hat{X}$ follows the phase-type (PH) distribution with the following probability density function:

$$f_{PH}(t) = \boldsymbol{\pi}_d \exp(\mathbf{D}_0 t) \mathbf{D}_1 \mathbf{1}. \quad (7)$$

Therefore, the k -th moment of $\hat{X}$ can be obtained as

$$\mathbb{E}[\hat{X}^k] = k! \boldsymbol{\pi}_d (-\mathbf{D}_0)^{-k} \mathbf{1}. \quad (8)$$

Note that $\mathbb{E}[\hat{X}] = 1/\lambda$ where λ is the fundamental arrival rate. Also the squared coefficient of variation of $\hat{X}$, c_v^2 , and the lag- k autocorrelation, r_k , of inter-arrivals are given by [9]

$$c_v^2 = \frac{\mathbb{E}[\hat{X}^2] - \mathbb{E}[\hat{X}]^2}{\mathbb{E}[\hat{X}]^2} = \frac{2\boldsymbol{\pi}_d(-\mathbf{D}_0)^{-2}\mathbf{1}}{(\boldsymbol{\pi}_d(-\mathbf{D}_0)^{-1}\mathbf{1})^2} - 1, \quad r_k = \frac{\mathbb{E}[\hat{X}]^{-2} \boldsymbol{\pi}_d(-\mathbf{D}_0)^{-1} \mathbf{T}^k (-\mathbf{D}_0)^{-1} \mathbf{1} - 1}{c_v^2}, \quad (9)$$

which is the correlation coefficient between the current inter-arrival time and the inter-arrival time after the k -th lag. If the arrival stream has a long-term dependency, the lag- k autocorrelation slowly decrease with k .

Finally, we present formulas of MAP which are used in our analysis. Define the matrix $\mathbf{P}(n, t)$ whose (i, j) -element is given by

$$[\mathbf{P}(n, t)]_{i,j} = P(N(t) = n, J(t) = j | N(0) = 0, J(0) = i), \quad (10)$$

Then we obtain the following differential-difference equations:

$$\frac{d}{dt} \mathbf{P}(0, t) = \mathbf{P}(0, t) \mathbf{D}_0, \quad \frac{d}{dt} \mathbf{P}(n, t) = \mathbf{P}(n, t) \mathbf{D}_0 + \mathbf{P}(n-1, t) \mathbf{D}_1, \quad n = 1, 2, 3, \dots \quad (11)$$

Also, we define the z-transform of $\mathbf{P}(n, t)$ as follows.

$$\mathbf{P}^*(z, t) = \sum_{n=0}^{\infty} z^n \mathbf{P}(n, t). \quad (12)$$

Thereby we have

$$\mathbf{P}^*(z, t) = \exp((\mathbf{D}_0 + z\mathbf{D}_1)t). \quad (13)$$

Letting $z = 1$ yields

$$\sum_{n=0}^{\infty} \mathbf{P}(n, t) = \exp((\mathbf{D}_0 + \mathbf{D}_1)t). \quad (14)$$

4. Analysis

In this section, we formulate the LS transforms of one service process and time to security failure. The LS transform of a nonnegative random variable Y with c.d.f. $F_Y(t)$, is defined as

$$\mathbb{E}[e^{-sY}] = \int_0^\infty e^{-st} dF_Y(t), \quad (15)$$

where s is a complex number. Since the LS transform involves the information of distribution completely, we can derive the moments from the LS transform. For example, the mean value (the first moment) of Y is obtained from the first derivative of LS transform, i.e.,

$$\mathbb{E}[Y] = -\frac{d}{ds} \mathbb{E}[e^{-sY}] \Big|_{s \rightarrow 0}. \quad (16)$$

Then, we derive the LS transforms of one service process and time to security failure through the following sections.

4.1. Probability of Security Failure

Let n_i be the number of VMs at the beginning of the i -th round. Then it follows that

$$n_i = n_1 + r(i-1), \quad i = 1, \dots, m. \quad (17)$$

Thus, the tolerance level (i.e., the maximum number of VMs that tolerate the intrusion) at the i -th round, f_i , is obtained from the argument of Byzantine fault tolerance,

$$f_i = \lfloor \frac{n_i - 1}{2} \rfloor, \quad i = 1, \dots, m. \quad (18)$$

Suppose that the intrusion for one VM occurs according to a Poisson process with rate η while the VM processes a task. The processing time for a task on a VM is a random variable S with the c.d.f. $G(t) = P(S \leq t)$. Note that the times for making consensus and activating VMs are ignored. The probability that the process is completed before the VM is intruded becomes

$$p_S = \int_0^\infty e^{-\eta t} dG(t). \quad (19)$$

Then the probability that the VM is intruded before the process is completed becomes

$$p_F = 1 - \int_0^\infty e^{-\eta t} dG(t). \quad (20)$$

Let N_i be the number of normal (non-intruded) VMs at the end of the i -th round. The probability mass function for N_i is

$$P(N_i = y | N_{i-1} = x) = \binom{x+r}{y} p_S^y p_F^{x+r-y}. \quad (21)$$

The behavior of one service process is described by a discrete-time Markov chain (DTMC) $\{N_i; i \geq 1\}$. In the DTMC, the state space consists of m subspaces, each of which is defined by $\{(0, i), \dots, (n_i - f_i - 1, i)\}$, ($1 \leq i \leq m$) at one round, since the consensus succeeds when the number of normal VMs becomes greater than or equal to $n_i - f_i$. That is, the state space depends on the number of rounds. Thus, each state of the DTMC is labeled as (v, i) , indicating that there are v normal VMs at the end of the i -th round. Also, the state space changes as the number of round increases.

If the consensus at the i -th round fails, the (x, y) -element of state transition probability matrix $\bar{\mathbf{P}}_i$ is defined by

$$[\bar{\mathbf{P}}_i]_{x,y} = P(N_i = y | N_{i-1} = x), \quad x = 0, \dots, n_{i-1} - f_{i-1} - 1, \quad y = 0, \dots, n_i - f_i - 1. \quad (22)$$

On the other hand, if the consensus at the i -th round is successful, then we define

$$[\mathbf{P}_i]_{x,y} = P(N_i = y | N_{i-1} = x), \quad x = 0, \dots, n_{i-1} - f_{i-1} - 1, \quad y = n_i - f_i, \dots, x + r. \quad (23)$$

Note that the state transition matrix of DTMC is not a square matrix, and the (x, y) -element represents the transition probability from state $(i-1, x)$ to state (i, y) .

Next we consider the probability (row) vector for N_1 . The probability vector is also defined as twofold. If the consensus fails at the first round, the element of probability vector is

$$[\bar{\mathbf{p}}_1]_y = P(N_1 = y), \quad y = 0, \dots, n_1 - f_1 - 1, \quad P(N_1 = y) = \binom{n_1}{y} p_S^y p_F^{n_1-y}, \quad (24)$$

and $[\cdot]_y$ represents a probability vector whose elements are computed while $N_1 = y$. Otherwise, if the consensus succeeds at the first round, we have

$$[\mathbf{p}_i]_y = P(N_1 = y), \quad y = n_1 - f_1, \dots, n_1. \quad (25)$$

Therefore, the probability that the security failure occurs during one service process is given by

$$P_{\text{failure}} = \bar{\mathbf{p}}_1 \bar{\mathbf{P}}_2 \cdots \bar{\mathbf{P}}_m \mathbf{1}, \quad (26)$$

and the probability that the security failure does not occur during one service process is

$$P_{\text{success}} = \mathbf{p}_1 \mathbf{1} + \bar{\mathbf{p}}_1 \mathbf{P}_2 \mathbf{1} + \cdots + \bar{\mathbf{p}}_1 \bar{\mathbf{P}}_2 \cdots \bar{\mathbf{P}}_{m-1} \mathbf{P}_m \mathbf{1} = 1 - \bar{\mathbf{p}}_1 \bar{\mathbf{P}}_2 \cdots \bar{\mathbf{P}}_m \mathbf{1}. \quad (27)$$

4.2. LS Transform of One Service Process

Consider the LS transform of the time length of one round. During one task processing, two events are considered; the VM is not intruded and the VM is intruded. Let τ_i denote the random variable to represent the time length of the i -th round. Also, define the c.d.f.'s of processing time by a VM under the intrusion as follows.

$$G_S(t) = P(S \leq t, \{\text{the VM is not intruded}\}) = \int_0^t e^{-\eta t} dG(t), \quad (28)$$

$$G_F(t) = P(S \leq t, \{\text{the VM is intruded}\}) = \int_0^t (1 - e^{-\eta t}) dG(t). \quad (29)$$

Then we have the c.d.f. for the time length of the i -th round given x and y :

$$F_{xy}^{(i)}(t) = \binom{x+r}{y} G(t)^{n_{i-1}-x} G_S(t)^y G_F(t)^{x+r-y}. \quad (30)$$

Therefore, the LS transform of the time for the i -th round is obtained;

$$\mathbb{E}[e^{-s\tau_i} I(N_i = y) | N_{i-1} = x] = \int_0^\infty e^{-st} dF_{xy}^{(i)}(t), \quad (31)$$

where $I(\cdot)$ denotes the indicator function.

Using Eq. (31), we define the following LS transform matrices:

- The consensus at the i -th round fails

$$\begin{aligned} [\bar{\mathbf{L}}_i(s)]_{x,y} &= \mathbb{E}[e^{-s\tau_i} I(N_i = y) | N_{i-1} = x], \\ x &= 0, \dots, n_{i-1} - f_{i-1} - 1, \quad y = 0, \dots, n_i - f_i - 1. \end{aligned} \quad (32)$$

- The consensus at the i -th round succeeds

$$\begin{aligned} [\mathbf{L}_i(s)]_{x,y} &= \mathbb{E}[e^{-s\tau_i} I(N_i = y) | N_{i-1} = x], \\ x &= 0, \dots, n_{i-1} - f_{i-1} - 1, \quad y = n_i - f_i, \dots, x + r. \end{aligned} \quad (33)$$

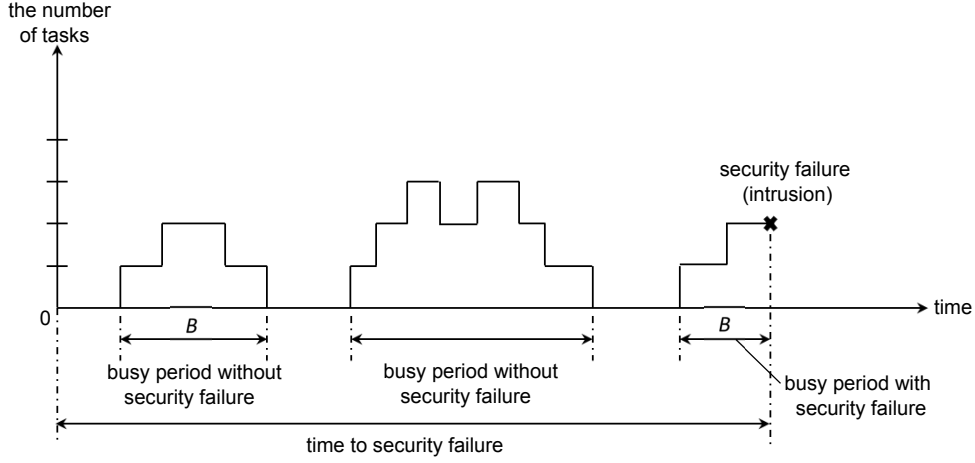


Figure 2: Queueing process.

Also, we denote the LS transform (row) vectors for τ_1 :

- The consensus at the first round fails

$$[\bar{l}_1(s)]_y = E[e^{-s\tau_1} I(N_1 = y)], \quad y = 0, \dots, n_1 - f_1 - 1. \quad (34)$$

- The consensus at the first round succeeds

$$[l_1(s)]_y = E[e^{-s\tau_1} I(N_1 = y)], \quad y = n_1 - f_1, \dots, n_1, \quad (35)$$

where

$$E[e^{-s\tau_1} I(N_1 = y)] = \int_0^\infty e^{-st} dF_y^{(1)}(t), \quad F_y^{(1)}(t) = \binom{n_1}{y} G_S(t)^y G_F(t)^{n_1-y}. \quad (36)$$

Let τ_S be the time length of one service process. Then the LS transform of τ_S when the security failure does not occur is given by

$$\begin{aligned} & E[e^{-s\tau_S} I(\{\text{security failure does not occur}\})] \\ &= l_1(s)\mathbf{1} + \bar{l}_1(s)L_2(s)\mathbf{1} + \dots + \bar{l}_1(s)\bar{L}_2(s) \dots \bar{L}_{m-1}(s)L_m(s)\mathbf{1}. \end{aligned} \quad (37)$$

Also, the LS transform of τ_S when the security failure occurs is

$$E[e^{-s\tau_S} I(\{\text{security failure occurs}\})] = \bar{l}_1(s)\bar{L}_2(s) \dots \bar{L}_m(s)\mathbf{1}. \quad (38)$$

The LS transform of one service process is obtained as

$$\begin{aligned} E[e^{-s\tau_S}] &= E[e^{-s\tau_S} I(\{\text{security failure does not occur}\})] \\ &\quad + E[e^{-s\tau_S} I(\{\text{security failure occurs}\})]. \end{aligned} \quad (39)$$

4.3. LS Transform of Time to Security Failure

Suppose that tasks arrive at the system according to a MAP with $(\mathbf{D}_0, \mathbf{D}_1)$. Tasks are served sequentially under FCFS (first come first serve) discipline. The security failure may occur while the server is processing tasks. Once the security failure occurs, the system stops. Based on these assumptions, the number of tasks in the buffer behaves a $MAP/G/1/\infty$ queueing process with a catastrophic failure (security failure) shown in Fig. 2.

First we consider the LS transform of one busy period. Let B be the time length of a busy period which is defined as the time period between idle time (see Fig. 2). Also, when the security failure occurs, B is defined as the time until the failure occurs.

Here we define the LS transforms regarding the busy period

$$B_{i,j}(s) = E[e^{-sB} I(J(B) = j, \{\text{security failure does not occur}\}) | J(0) = i], \quad (40)$$

$$Z_{i,j}(s) = E[e^{-sB} I(J(B) = j, \{\text{security failure does occurs}\}) | J(0) = i], \quad (41)$$

where $J(t)$ is defined as the phase of MAP in Section 3, and $J(B) = j$ means that the phase of MAP is j at time B . Also, $\mathbf{B}(s)$ and $\mathbf{Z}(s)$ denote the matrices whose (i, j) -element is given by $B_{i,j}(s)$ and $Z_{i,j}(s)$, respectively.

Taking account of the first task in the busy period, we have

$$\mathbf{B}(s) = P_{\text{success}} \int_0^\infty \sum_{n=0}^\infty \mathbf{P}(n, t) e^{-st} \mathbf{B}(s)^n dH_1(t), \quad (42)$$

where $\mathbf{P}(n, t)$ is defined in Eq. (10) and $H_1(t)$ is the c.d.f. of τ_S provided that the security failure does not occur. According to the argument of the analysis of $MAP/G/1/\infty$ queue [10], we define

$$\mathbf{A}_n(s) = \int_0^\infty e^{-st} \mathbf{P}(n, t) dH_1(t). \quad (43)$$

Then the Eq. (42) is rewritten in the form

$$\mathbf{B}(s) = P_{\text{success}} \sum_{n=0}^\infty \mathbf{A}_n(s) \mathbf{B}(s)^n. \quad (44)$$

Similarly, in the case where the security failure occurs during the process of first task, we have

$$\begin{aligned} \mathbf{Z}(s) &= P_{\text{failure}} \int_0^\infty \sum_{n=0}^\infty \mathbf{P}(n, t) e^{-st} dH_2(t) \\ &+ P_{\text{success}} \int_0^\infty \sum_{n=1}^\infty \mathbf{P}(n, t) e^{-st} \sum_{l=1}^n \mathbf{B}(s)^{l-1} \mathbf{Z}(s) dH_1(t), \end{aligned} \quad (45)$$

where $H_2(t)$ is the c.d.f. of τ_S provided that the security failure occurs. By substituting Eqs. (14), (43) and (44) into Eq. (45), we obtain

$$\mathbf{Z}(s) = (\mathbf{I} - \mathbf{B}(s))(\mathbf{I} - \mathbf{G}_1^*(s))^{-1} \mathbf{G}_2^*(s), \quad (46)$$

where $\mathbf{I}$ is the identity matrix and

$$\mathbf{G}_1^*(s) = P_{\text{success}} \int_0^\infty e^{-st} \exp((\mathbf{D}_0 + \mathbf{D}_1)t) dH_1(t), \quad (47)$$

$$\mathbf{G}_2^*(s) = P_{\text{failure}} \int_0^\infty e^{-st} \exp((\mathbf{D}_0 + \mathbf{D}_1)t) dH_2(t). \quad (48)$$

Finally, we consider the LS transform of time to security failure. Define the matrix

$$\mathbf{Y}(s) = \int_0^\infty e^{-st} \exp(\mathbf{D}_0 t) dt \mathbf{D}_1. \quad (49)$$

The (i, j) -element of $\mathbf{Y}(s)$ corresponds to the LS transform of an idle period when the phase is j at the end of idle period provided that the phase is i at the beginning of idle period. Also, let X be the system life (i.e., the time until the security failure occurs), and $\mathbf{X}(s)$ denotes the matrix whose (i, j) -element is the following LS transform:

$$X_{i,j}(s) = E[e^{-sX} I(J(X) = j) | J(0) = i]. \quad (50)$$

Since the system stops due to the security failure, we obtain the LS transform of the time to security failure by

$$\begin{aligned} \mathbf{X}(s) &= \mathbf{Y}(s) \sum_{n=0}^\infty (\mathbf{B}(s) \mathbf{Y}(s))^n \mathbf{Z}(s) \\ &= \mathbf{Y}(s) (\mathbf{I} - \mathbf{B}(s) \mathbf{Y}(s))^{-1} (\mathbf{I} - \mathbf{B}(s)) (\mathbf{I} - \mathbf{G}_1^*(s))^{-1} \mathbf{G}_2^*(s). \end{aligned} \quad (51)$$

5. Performance Measures

5.1. Mean Time to Security Failure

Based on the property of LS transform, we can derive MTTSF by taking the first derivative of $\mathbf{X}(s)$ with respect to s and letting $s = 0$. Let $\boldsymbol{\pi}_0$ be the probability vector for phases at time $t = 0$. Then the MTTSF is obtained as

$$\text{MTTSF} = \mathbb{E}[X] = -\boldsymbol{\pi}_0 \mathbf{X}'(0) \mathbf{1}, \quad (52)$$

where $\mathbf{X}'(s)$ is the first derivative of $\mathbf{X}(s)$ with respect to s . From Eq. (51), the MTTSF can be expressed as

$$\text{MTTSF} = \frac{\mathbb{E}[\tau_S]}{P_{\text{failure}}} + \boldsymbol{\pi}_0 (\mathbf{I} - \mathbf{Y}(0) \mathbf{B}(0))^{-1} (-\mathbf{Y}'(0)) \mathbf{1}. \quad (53)$$

The matrices $\mathbf{Y}(0)$ and $\mathbf{Y}'(0)$ are the transition matrices of phases during the idle period and the negative of the mean time of idle period, respectively. Since the idle period of $MAP/G/1/\infty$ queue follows the PH distribution, these matrices becomes

$$\mathbf{Y}(0) = (-\mathbf{D}_0)^{-1} \mathbf{D}_1, \quad \mathbf{Y}'(0) = -(-\mathbf{D}_0)^{-2} \mathbf{D}_1. \quad (54)$$

On the other hand, $\mathbf{B}(0)$ is the matrix whose (i, j) -element gives the probability that the security failure does not occur in a busy period when arrival phase changes from i to j , namely,

$$[\mathbf{B}(0)]_{i,j} = P(J(B) = j, \{\text{security failure does not occur}\} | J(0) = i). \quad (55)$$

This can be computed by the recursive approach using Eq. (44).

In particular, when considering the MTTSF of system under the Poisson arrival stream [7], and letting the arrival rate be λ , it is easily found

$$\mathbf{Y}(0) = \mathbf{1}, \quad \mathbf{Y}'(0) = -\frac{1}{\lambda}. \quad (56)$$

Then the Eq. (53) is reduced to

$$\text{MTTSF} = \frac{\mathbb{E}[\tau_S]}{P_{\text{failure}}} + \frac{1}{\lambda(1 - B(0))} \quad (57)$$

with $B(0)$ that is defined as

$$B(0) = P(\{\text{security failure does not occur}\}), \quad (58)$$

and is computed by solving the following recursive formula through an iterative method,

$$B(0) = P_{\text{success}} \tilde{H}_1^*(\lambda(1 - B(0))), \quad (59)$$

where $\tilde{H}_1^*(t)$ is the LS transform of $H_1(t)$.

5.2. Traffic Intensity

An attribute of a queueing system which forms the simplest measure of its performance is the traffic intensity (denoted by ρ), which is a measure of how busy a system is, and is defined as

$$\rho = \frac{\text{mean service time}}{\text{mean interarrival time}}. \quad (60)$$

Since the intrusion tolerance scheme causes overhead to the system, this paper considers two traffic intensities; fundamental and effective traffic intensities.

5.2.1. Fundamental Traffic Intensity

The fundamental traffic intensity is the traffic intensity without the intrusion tolerance scheme under the assumption where the intrusion never occurs. According to the queueing theory, the fundamental traffic intensity is given by

$$\rho_0 = \lambda E[S], \quad (61)$$

where λ is the fundamental arrival rate of MAP in Eq. (4).

5.2.2. Effective Traffic Intensity

The effective traffic intensity is the traffic intensity considering the overhead of intrusion tolerance. Then the effective traffic intensity is defined as

$$\rho_e = \lambda E[\tau_S], \quad (62)$$

where $E[\tau_S]$ is the mean time of one service process. From Eq. (39), we can compute

$$E[\tau_S] = -\frac{d}{ds} E[e^{-s\tau_S}] \Big|_{s \rightarrow 0}. \quad (63)$$

References

- [1] F. Wang, F. Gong, C. Sargor, K. Goseva, K. S. Trivedi, and F. Jou: Scalable intrusion tolerance architecture for distributed server, in *Proceedings of the 2nd IEEE SMC Information Assurance Workshop* (2001).
- [2] W. Zhao: BFT-WS: A Byzantine fault tolerance framework for Web services, in *Proceedings of the 11th International IEEE Conference Enterprise Distributed Object Computing Conference Workshop (EDOCW'07)* (2007), 89–96.
- [3] B. B. Madan, K. Goseva-Popstojanova, K. Vaidyanathan, and K. S. Trivedi: A method for modeling and quantifying the security attributes of intrusion tolerant systems, *Performance Evaluation*, **56**(1/4) (2004), 167–186.
- [4] V. S. Junior, L. C. Lung, M. Correia, JD. S. Fraga, and J. Lau: Intrusion tolerant services through virtualization: a shared memory approach, in *Proceedings of The 24th IEEE International Conference on Advanced Information Networking and Applications (AINA'10)* (2010), 768–774.
- [5] J. Lau, L. Barreto, and JD. S. Fraga: An infrastructure based in virtualization for intrusion tolerant services, in *Proceedings of The 19th IEEE International Conference on Web Services (ICWS 2012)* (2012), 170–177.
- [6] J. Zheng, H. Okamura, and T. Dohi: Survivability analysis of VM-based intrusion tolerant systems, *IEICE Transactions on Information & Systems (D)*, **E98-D**(12) (2015), 2082–2090.
- [7] J. Zheng, H. Okamura, and T. Dohi: Performance evaluation of VM-based intrusion tolerant systems with Poisson arrivals, in *Proceedings of The 4th International Symposium on Computing and Networking (CANDAR 2016)* (2016), 181–187.
- [8] S. Asmussen and G. Koole: Marked point processes as limits of Markovian arrival streams, *Journal of Applied Probability*, **30**(2) (1993), 365–372.
- [9] G. Casale, E. Z. Zhang, and E. Smimi: Trace data characterization and fitting for Markov modeling,” *Performance Evaluation*, **67**(2) (2010), 61–79.
- [10] M. F. Neuts: *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, Marcel Dekker, New York, (1989).

Congestion Control for a System with Parallel Stations and Homogeneous Customers using Priority Passes: Summary of the Results

Yasushi Masuda
Keio University

Akira Tsuji
Canon Inc.

Abstract

We consider a game theoretic congestion model with parallel nodes and homogeneous customers. The purpose of this paper is to examine how the priority passes improve social welfare for such a system. To this end, we prove the existence of an equilibrium. The system with no priority pass has a unique equilibrium. With the introduction of priority passes, the uniqueness of the equilibrium may be destroyed. We provide a sufficient condition under which the system with priority passes outperforms the system with no priority passes. The problem is explored numerically as well. In this short memo, we provide the introduction and the summary of the results of the paper. The pdf files of the presentation slides and the full papers are available at:

- <http://www.ae.keio.ac.jp/lab/soc/masuda/articles/PriorityPass-beamer.pdf>
- <http://www.ae.keio.ac.jp/lab/soc/masuda/articles/PriorityPass.pdf>

Key words: Congestion game; Priority; Wardrop equilibrium; Capacitated network; Theme park

1. Introduction

A priority system is a powerful mechanism for making efficient use of congestion-prone facilities when customers are heterogeneous. Specifically, when some customers have a higher opportunity cost of time and/or lesser service requirements than others, then, for the benefit of the customers as a whole it makes sense to give higher priority to these customers; see e.g., Heyman and Sobel (1982). On the other hand, the fact that a priority system is effective even when the population is homogeneous has been demonstrated by Hassin (1985). Hassin shows that for an observable queueing system with Markovian arrival, FIFO is the worst queueing discipline in terms of social welfare of all customers; LIFO is the best, and any priority system assigned randomly or based on irrelevant data is an improvement relative to FIFO. The argument of Hassin is based on an insight into the congestion externality arising from individually optimal decisions as to when to join/leave the queue.

This paper is motivated by priority passes distributed in a theme park, in particular, the FastPass issued by Disneyland. We consider a stylized congestion-prone system where a fixed number of priority passes are allocated to each customer, free of charge, upon arrival, and explore how the priority scheme can improve social welfare when the customers are homogeneous. Formally, we consider parallel stations providing substitutable services with a homogeneous population of nonatomic customers; i.e., a single customer does not influence the overall congestion but a mass of customers does. Each customer maximizes his/her benefit from visiting a station minus the time cost spent in stations, subject to a constraint on the total amount of

time spent in the theme park. Each customer who decides to use the pass at a service station gains priority over all customers that do not, but still has to wait for others who used the pass and arrived before him/her.

As a starting point, we offer a rough intuition as to why the priority pass scheme tends to ease congestion in a system with homogeneous customers who visit each attraction once at most. Consider a ride-type popular attraction with a small capacity, which generates large congestion externalities. Suppose that the priority scheme is introduced and that the customers start using a priority pass at this popular attraction with a positive probability. Then, visiting this attraction as a non-priority customer becomes less attractive because a non-priority customer has to wait longer in the presence of priority customers. Thus, the total probability of visiting this popular attraction becomes small with the introduction of the priority scheme, and the problem of over-congestion at this attraction is alleviated. Since the customers reduce the probability of visiting the popular attraction, they have extra time available. What happens to this extra time? If a time constraint based on the limited park hours is not binding, customers leave the park early (or take a long lunch break), and the extra time does no harm. When the time constraint is binding, however, complications arise because the extra time is *valuable* for the customers, and therefore they make a new plan to utilize it. The customers may increase their probability of visiting an under-occupied theatrical attraction with a large capacity, which generates small congestion externalities. Then, the priority pass system alleviates the overall congestion problem. However, the customers may increase their probability of visiting a very popular attraction. Thus, if the time constraint is binding, the overall effect of the priority pass on the welfare of the customers is ambiguous.

There are two streams of research on game theoretic congestion problems that are relevant to this paper. The first stream is the congestion game formulated in a deterministic/static setting, as initiated by Rosenthal (1973) and further studied by Monderer and Shapley (1996), among others. In the deterministic/static setting, customers use only the average system behavior and do not use any dynamically changing system information. See Altman and Wynter (2004), Altman et al. (2006) and Patriksson (2015) for a survey of congestion games, potential games, and Wardrop equilibrium. In terms of our modeling approach, our work is closely related to the literature on routing demand to parallel servers providing substitutable services, see Section 8.1 of Hassin (2016). In particular, El Azouzi and Altman (2003) show that in a network of parallel servers with atomic customers and no priority, if there are flow capacity constraints, then the uniqueness of the equilibrium may be destroyed; see also Orda et al. (1993) and Correa et al. (2004) on this point. Within this research stream, the price of anarchy (PoA), defined by the ratio between the socially optimal solution and the worst equilibrium, is discussed by Roughgarden and Tardos (2002). Various extensions of PoA exist; see e.g., Anselmi et al. (2011) and Dube and Jain (2014). In particular, Anselmi et al. (2011) consider a parallel server system with a continuum of users, where each node is a profit maximizer, and show that the node competition improves social welfare. Correa et al. (2008) formally define equilibrium in capacitated networks and clarify the relationship between the equilibrium and a capacitated traffic assignment problem (see also Patriksson (2015), Larsson and Patriksson (1999), Marcotte et al. (2004) and Stier-Moses (2004)). The sensitivity analysis of the traffic equilibrium is discussed often in the context of Braess's paradox, and has important implications in capacity management, see Do Chung et al. (2014), Kameda (2009), Hwang and Cho (2016), Masuda and Whang (2002), and their references.

The second stream of research on game theoretic congestion problems, initiated by Naor (1969), examines the stochastic/dynamic aspect of congestion. That is, customers incorporate the instantaneous system state information in their decision-making process. Along this line, Hassin (1985) shows the optimality of LIFO queues. The readers are referred to Hassin and Haviv (2003) and Hassin (2016) for extensive overviews of the literature.

We examine customers' equilibrium behavior in a deterministic/static setting. Thus, our modeling approach follows the first stream of research on congestion problems. Our motivation for analyzing the effect of priority on homogeneous customers, however, is closely related to Hassin (1985) in the second stream of research. In particular, we focus on how the priority scheme improves social welfare when the customers are homogeneous.

2. Summary of the Results

1. We consider the effect of priority passes on the performance of a congested system with a finite number of parallel stations and a homogenous population of customers. A state of the system is described by a vector of probabilities, one for each station. The probability of station j is the probability of each customer in the population visiting station j . Each customer strives to maximize the benefit from visiting stations minus the time cost of spending in stations, subject to a constraint on the total amount of time spent in the theme park. The service time of each station is assumed to be an increasing function of the probability of each customer visiting it. We prove that this model is a potential game with a unique equilibrium.
2. We generalize this model assuming that a fixed number of priority passes is handed out to each customer. A state of the system is described by two probability vectors, for visiting each station with and without a priority pass, respectively. We show that if the time constraint is not binding, then the equilibria in the model with priority passes are not worse than the equilibrium in the model with priority passes in terms of social welfare. We provide sufficient conditions under which the model with the priority pass strictly outperforms the model without the priority pass.
3. We explore the relationship between our models and the standard Wardrop-style model.
4. Numerical examples show that: i) the proposed priority system sometimes achieves the social optimum when there are an appropriate number of priority passes; ii) multiple equilibria often arise in the priority system, which reflects the "follow the crowd" behavior of customers (Hassin and Haviv (1997)); and iii) the model with priority passes sometimes has equilibria that are worse than the model without priority passes in terms of social welfare.

There are three limitations in our model;

1. The customers are homogeneous: all customers assign the same benefit to visiting each station, have the same opportunity costs for time, and have the same time constraints.
2. Both the time constraint and the priority pass constraint are based on expectations, not on the realized value.
3. The model is deterministic/static and ignores the stochastic/dynamic nature of customers' behavior.

Admittedly, these are all strong assumptions. To counter the first point, however, we emphasize that the priority pass scheme can be effective for a system even with highly homogeneous customers, so that it would be even more effective for a system with heterogeneous customers. An argument to counter the second point is that the models incorporating actual paths appear to be intractable, thus we employ the models with relaxed constraints. Although the third point is an important concern, it is a standard setting in queueing/congestion games (see e.g. Roughgarden (2005)). We leave the analysis of the consequences of relaxing these limitations to future research.

References

- Altman, E., Boulogne, T., El-Azouzi, R., Jiménez, T. and Wynter, L. (2006), ‘A survey on networking games in telecommunications’, *Computers & Operations Research* **33**(2), 286–311.
- Altman, E. and Wynter, L. (2004), ‘Equilibrium, games, and pricing in transportation and telecommunication networks’, *Networks and Spatial Economics* **4**(1), 7–21.
- Anselmi, J., Ayesta, U. and Wierman, A. (2011), ‘Competition yields efficiency in load balancing games’, *Performance Evaluation* **68**(11), 986–1001.
- Correa, J. R., Schulz, A. S. and Stier-Moses, N. E. (2004), ‘Selfish routing in capacitated networks’, *Mathematics of Operations Research* **29**(4), 961–976.
- Correa, J. R., Schulz, A. S. and Stier-Moses, N. E. (2008), ‘A geometric approach to the price of anarchy in nonatomic congestion games’, *Games and Economic Behavior* **64**(2), 457–469.
- Do Chung, B., Cho, H.-J., Friesz, T. L., Huang, H. and Yao, T. (2014), ‘Sensitivity analysis of user equilibrium flows revisited’, *Networks and Spatial Economics* **14**(2), 183–207.
- Dube, P. and Jain, R. (2014), ‘Bertrand equilibria and efficiency in markets for congestible network services’, *Automatica* **50**(3), 756–767.
- El Azouzi, R. and Altman, E. (2003), ‘Constrained traffic equilibrium in routing’, *IEEE Transactions on Automatic Control* **48**(9), 1656–1660.
- Hassin, R. (1985), ‘On the optimality of first come last served queues’, *Econometrica* **53**(1), 201–02.
- Hassin, R. (2016), *Rational queueing*, CRC Press, Boca Raton.
- Hassin, R. and Haviv, M. (1997), ‘Equilibrium threshold strategies: The case of queues with priorities’, *Operations Research* **45**(6), 966–973.
- Hassin, R. and Haviv, M. (2003), *To queue or not to queue: Equilibrium behavior in queueing systems*, Kluwer, Boston.
- Heyman, D. P. and Sobel, M. J. (1982), *Stochastic Models in operations research. Vol. 1 Stochastic Processes and Operating Characteristics.*, McGraw-Hill, New York.
- Hwang, M.-C. and Cho, H.-J. (2016), ‘The classical braess paradox problem revisited: a generalized inverse method on non-unique path flow cases’, *Networks and Spatial Economics* **16**(2), 605–622.
- Kameda, H. (2009), ‘Coincident cost improvement vs. degradation by adding connections to noncooperative networks and distributed systems’, *Networks and Spatial Economics* **9**(2), 269–287.
- Larsson, T. and Patriksson, M. (1999), ‘Side constrained traffic equilibrium models—analysis, computation and applications’, *Transportation Research Part B: Methodological* **33**(4), 233–264.
- Marcotte, P., Nguyen, S. and Schoeb, A. (2004), ‘A strategic flow model of traffic assignment in static capacitated networks’, *Operations Research* **52**(2), 191–212.

- Masuda, Y. and Whang, S. (2002), ‘Capacity management in decentralized networks’, *Management Science* **48**(12), 1628–1634.
- Monderer, D. and Shapley, L. S. (1996), ‘Potential games’, *Games and economic behavior* **14**(1), 124–143.
- Naor, P. (1969), ‘The regulation of queue size by levying tolls’, *Econometrica* **37**(1), 15–24.
- Orda, A., Rom, R. and Shimkin, N. (1993), ‘Competitive routing in multiuser communication networks’, *IEEE/ACM Transactions on Networking* **1**(5), 510–521.
- Patriksson, M. (2015), *The traffic assignment problem: models and methods*, Courier Dover Publications.
- Rosenthal, R. W. (1973), ‘A class of games possessing pure-strategy nash equilibria’, *International Journal of Game Theory* **2**(1), 65–67.
- Roughgarden, T. (2005), *Selfish routing and the price of anarchy*, MIT press, Cambridge.
- Roughgarden, T. and Tardos, É. (2002), ‘How bad is selfish routing?’, *Journal of the ACM* **49**(2), 236–259.
- Stier-Moses, N. E. (2004), Selfish versus coordinated routing in network games, PhD thesis, Massachusetts Institute of Technology.

物理対経済体のサンドイッチ場考：待ち行列法によるペアマップ仮説と波動方程式化例

松井正之

神奈川大学、電気通信大学

概要：地球上の人類が築いてきている社会の種々の人工体は、一般に 3M&I(ヒト、モノ、カネ、情報)系またはその複合体からなっている。この待ち行列法による人工体の研究は、最近筆者ら(2016)によって、サンドイッチ原理とバランシング原理が示されている。本論ではこれをベースに、人工体のマクロ対ミクロの止揚解であるサンドイッチ原理(場)を対象にして、その中核になるペアマップ(オイラー公式)が人工体の基底であるという仮説を提示して、検証を試みる。検証は、止揚解(波動方程式)の存在例と物理対経済体の符号合わせ(結合対比例定数等)による整合性で検証される。

1. はじめに

近代科学は、ニュートンの動力学に始まるが、人工体動学に関するニュートン力学は科学的管理法の父である F. W. Taylor の時間研究、電話技師 A. K. Erlang の待ち行列研究であると考えられる。現在までに、このテイラーらの科学以降の人工体動学の発展は多種多様で、あまり巨視的に見える化されていない[1]。

例えば、ノーベル賞の H. A. Simon は、「人偽性の科学」を著して人工体研究への注目を向けさせた[2]。その後、これをもとに「人工体と人偽性科学」が出版された[3]。我々は、コンベヤ系 CSPS 体の待ち行列研究から、1983 年に製造・販売コラボレーションの利益最大化研究を始めたが、その際にゲーム理論にある双行列を発展させたペア行列(入力と出力のペア)を用いた[4]。その後、種々の製版タイプのペア行列表研究から、ペアマップの原形を抽出して[5][6]、需給体等におけるその仕組み、深底に見える化してきた[7]。

本論は、この一連の研究成果により、自然対人工体が深遠で繋がっているというサンドイッチ($S=W$)場におけるペアマップ仮説の更なる検証を試みている。検証は、マクロ(上位レベル)対ミクロ(下位レベル)の止揚解(波動方程式)の存在と、物理対経済体の符号合わせ(結合対比例定数等)による整合性で示されている。

2. 人工体のサンドイッチ論

2.1 3M & I 系と人工体科学

本論では、3M&I 系からなる人工体を対象として議論をする。3M とは、ヒト(huMan), モノ(Material/machine), カネ(Money)であり、I は情報(Information)である。この分野は、3M&I 系の術として、IE、OR、QC などのエンジニアリングで知られており、工場、事業体、サプライチェーンなどが主たる対象として取り上げられている[5][6]。

この人工体科学のベースとして、種々の待ち行列法が用いられている。これまでの主な成果として、松井の式($W = ZL$) [8][9]とムダ対効率基準(カメレオン基準) [10][11]が見られる。また、近著[1]では、人工体の基礎として、次の2つの原理が挙げられている。

A. サンドイッチ原理 マクロ対ミクロの止揚解

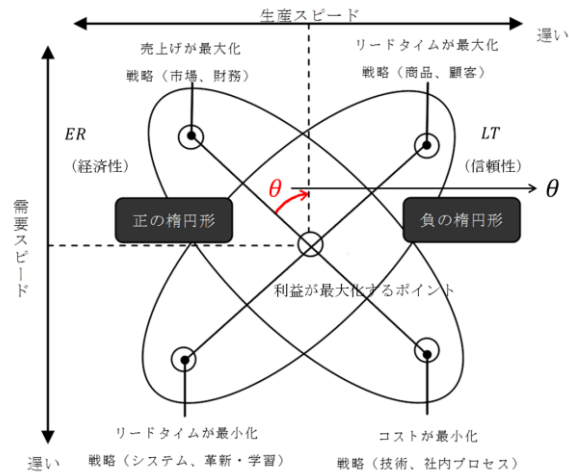
B. バランシング原理 $Z_1 L_1 = Z_2 L_2 (= W)$, for $n = 2$

サンドイッチ原理とは、マクロ(上位レベル)対ミクロ(下位レベル)の止揚である、マクロとミクロとの間に、人工体の腰($S=W$ 場)に相当するペアマップ(オイラー公式)の存在(止揚)を仮説とする原理である。ペアマップは、入力(需要)と出力(供給)からなるペア行列表に現れる楕円交叉

形からなるマップであり、製販協力ゲームの研究から生まれている。

2.2 S=W 場と基本ペアマップ

ペアマップの基本形は、文献[12][13]等に見られる。現在の形は、[12]にライフサイクルを加味したものがあり、図1として与えられている[10]。図1は需要スピード(d)と生産スピード(m)の2次元表に、経済軸(収益対コスト)と信頼軸(リードタイム)の楕円形が交叉しており、交叉点がバランス点(利益最大化)である。



? 3M&I系: $(d, m) \rightarrow (Z, L)$, 介在角度 θ

ただし、 d :需要スピード, m :生産スピード

図1. 基本ペアマップの概要 ([11]参照)

すなわち、図1は、まず2次元表上に以下のように、両側(左右)を1対1対応させて出来ている。

$$\begin{array}{ccc} \text{ペア行列}(d, m) & \xleftrightarrow[\text{(1対1)}]{\text{対応}} & \text{ペア行列: } (ER(d, m), LT(d, m)) \\ \text{<下位レベル>} & & \text{<上位レベル>} \end{array} \quad (2.1)$$

下位レベルは通常の待ち行列系(ニュートン動力学)である。上位レベルは、表1のように、楕円交叉形の4極(中心点)を、バランス・スコアカード(BSC)の4視点・アプローチと符号合わせてしている。ここでは、ペアマップが企業のような事業体のライフサイクル様態を3M&I系がとる4極系で表現できることを暗に仮定していることが注意される。

表1. ペアマップ対BSC表(併・事業コンセプト)

4極 対応	<Z> 経済軸(ER)		<L> 信頼軸(LT)	
	ペア1(ER)	ペア2(ER)	ペア1	ペア2
ペアマップ	収益最大化	コスト最小化	リードタイム最大	リードタイム最小
BSC	財務の視点	社内プロセス視点	顧客の視点	革新・学習の視点
事業コンセプト	市場	技術	商品	システム

3. ペアマップと波動方程式

3.1 ペアマップ図解：プロセスビュー

一般に、上下位レベルの挙動はそれぞれ異なっているので、統一的な扱いには、ペアマップ図1の3次元表示が有用で、図類は幾何学的表示方法[12][14]にみられる。最近、この幾何学的表示のプロセス視点から、波動方程式による定式化が可能であることが見えてきた。この3次元ビューは、図2で与えられる。

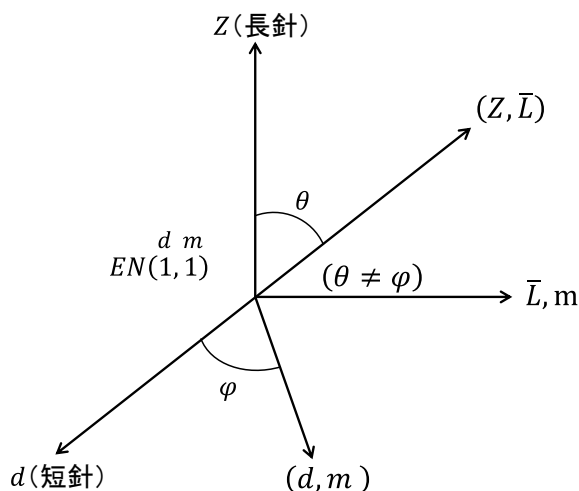


図2. ペアマップの3次元ビュー（骨子）

他方、この3次元体は、形式的には2次元でのオイラー公式によるペアマップ図解が可能である。図3は、ペアマップ(S=W解)の右回りの回転形を与えているが、オイラー公式により付随した波形表示より、物理対と同様に波動方程式として以下に与えることができる。

$$Z \exp(i\theta) = Z \exp(it \cdot R \cos \varphi). \quad (3.1)$$

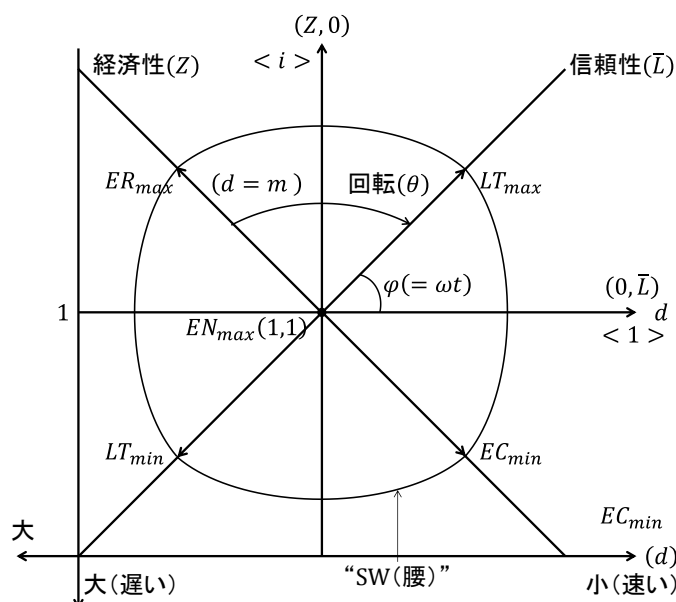


図3. オイラー公式によるペアマップ図解

$$\left[\begin{array}{l} \text{“回転(SW)” } Z = e^{i\theta} \leftrightarrow \text{“波形” } \bar{L} \cos \theta + iZ \sin \theta, \\ \text{ただし、} \theta = dt, Z(=\bar{L}) \rightarrow 1 \text{ とする。} \end{array} \right]$$

3.2 波動方程式の導出

図2に基づいて、長針(θ)と短針(φ)による統一的な波動方程式を考える。長針と短針との関係は、図4のそれぞれa)図、b)図で与えられる。いま、松井の式: $W = ZL$ において、 $Z = \bar{L}/\cos\theta$ から次式が得られる。

$$ZR = \bar{L}/\{\sin\theta \cdot \cos\varphi\}. \quad (3.2)$$

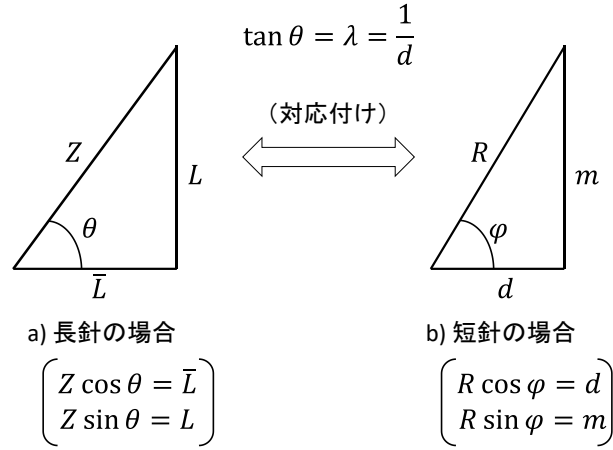


図4. 長針対短針の関係

式(3.2)より、三角関数の加法定理から、

$$\begin{aligned} \bar{L} &= (ZR) \sin\theta \cdot \cos\varphi \\ &= (ZR) \cdot \frac{1}{2} \{\sin(\theta + \varphi) + \sin(\theta - \varphi)\}. \end{aligned} \quad (3.3)$$

となり、次の関係が求まる。

$$\frac{\bar{L}}{Z} = \frac{R}{2} \{\sin(\theta + \varphi) + \sin(\theta - \varphi)\}. \quad (3.4)$$

ここで、特に右辺の u 、すなわち、式(3.4)：

$$u = \frac{1}{2} \{\sin(\theta + \varphi) + \sin(\theta - \varphi)\}. \quad (3.5)$$

に注目する。 $\theta = x, \varphi = \omega t (\omega > 0, t > 0)$ とおくと、このとき式(3.4)は次式のように書ける。

$$u = u(t, x) = \frac{1}{2} \{\sin(x + \omega t) + \sin(x - \omega t)\}. \quad (3.6)$$

一般に、式(3.5)の u に対して、 t, x の一次偏微分、二次偏微分は可能であり、この結果から次の波動方程式が導出される。すなわち、式(3.5)の u は式(3.6)をみたとす。

$$\frac{\partial^2 u}{\partial t^2} = \omega^2 \frac{\partial^2 u}{\partial x^2}, \quad (3.7)$$

境界条件は、 $u(t, 0) = u(t, \pi)$ である。

4. 人工体のペアマップ仮説

4.1 物理対経済体のペアマップ形

物理体と経済体のペアマップは、図1都の符号合わせによる対応付けで、それぞれ図5,6に示されている。物理・素粒子系は量子力学により、電磁気、重力、強い力、弱い力の4様態からなり、国内経済体はマクロ経済学により財政、消費、市場、金融の4つから説明できると仮定している。

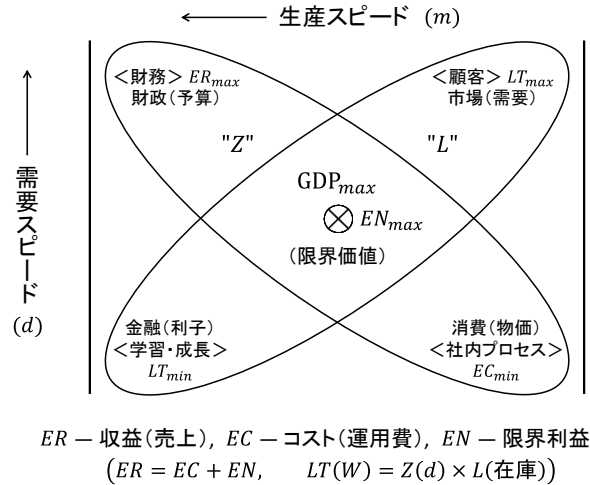


図5. 国内経済体のペアマップ概要

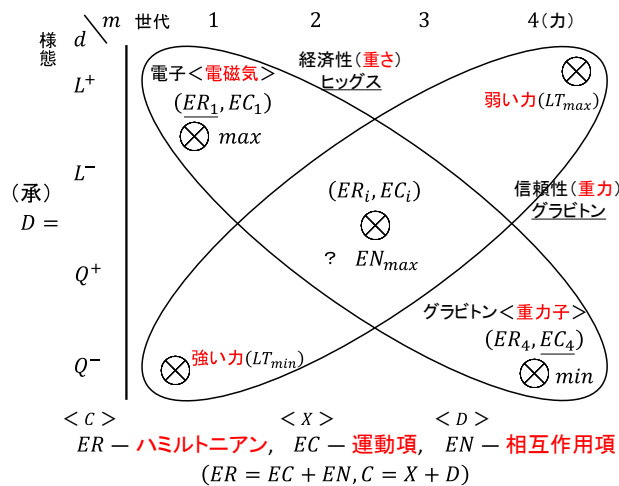


図6. 物理・素粒子体のペアマップ概要

4.2 ペアマップのパターン類型例

自然(物理)・人工(経済)体のペアマップの形として、表2はパターン類型例からペアマップマップの仮説を検証している。符号合わせ(結合対比例定数等)により、この検証は仮説の整合性を示している。また、表2では、ペアマップ仮説(回転系)と入出力管理式(波動系)の対応関係が与えられており、これはオイラー公式と呼応している。

表 2. ペアマップのパターン類型例

基本マップ(4様態)		企業BSC	国内経済性	物理素粒子体		入出力系対応 ⁺⁺	
経済性	ER_{max} (収益最大)	財務	財政(予算)	重さ	電磁気 (ハミルトン)	次期発注量(O)	入力
	EC_{min} (コスト最小)	社内 プロセス	消費(物価)		重力(運動項)	期末在庫量(I)	
信頼性	LT_{max} (リードタイム最大)	顧客	市場(需要)	力	弱い力 (フェルミ相互作用)	次期需要量(D)	出力
	LT_{max} (リードタイム最大)	学習・ 成長	金融(利子)		強い力 (量子色力学)	基準在庫量(MP)	
EN_{max} (バランス点) ⁺ < 限界利益最大 >			限界価値 (GDP_{max})	結合定数 (運動項対相互作用項)		$O^* = I^* = D^* = MP^*$	

+ バランス点: $ER^*/2 = EC^* = LT^*(= EN_{max})$

in 人工対自然体: $(ER = EN + EC, LT(W) = Z(d) \times L(I))$, d : 入力スピード

++ 流動数管理式: 次期発注量(O) = 次期需要量(D) + 基準在庫量(MP) - 期末在庫量(I)
($\longleftrightarrow$ 波動方程式)

5. まとめと今後

本論は、1983 年の学会発表で提示したペアマップの原形が、その後の研究成果により、自然対人工体の基底としての見える化から、その存在性(普遍化)を明らかにしてきた。これはまた、併立する松井の方程式の(折りたたみ式の)起承転結系展開プロセスの中核(S=W 場)としても重要である。

今後は、本論のペアマップ仮説のより一層の検証、理論化が望まれる。これらにより、種々の人工体行動の定式化、管理運用法、自働管理ロボットへの展開が可能となる。本研究は、今後の高度 ICT 管理化社会における人工体のあり方、その経済性、仕組みの構築に有用と考えられる。

参考文献

- [1] Matsui, M., *Fundamentals and Principles of Artifacts Science: 3M&I-Body System*, Springer Briefs in Business, 2016
- [2] Simon, H. A., *The Sciences of Artificial*, M. I. T. Press, 1969
- [3] Dahlbom, B., Beckman, S., Nilsson, G. B., *Artifacts and Artificial Science*, Almqvist & Wiksell International—Stockholm, 2001
- [4] 松井正之: 受注選択と能力切換えの結合方策について、日本経営工学会誌、Vol. 39, No. 2, pp. 83-87 (1988)
- [5] Matsui, M., *Manufacturing and Service Enterprise with Risks: A Stochastic Management Approach*, International OR&MS series, Springer, 2008
- [6] 松井正之: 生産企業のマネジメントー利益最大化と工場理学、共立出版、2005
- [7] 松井正之: 松井の方程式の SW 体図解: 物理素粒子体式例、企業体・物理体の双対系鎖と需給格子解構造、第 8 回横幹コンファレンス、立命館大学。2017. 12
- [8] Matsui, M., CSPA Model: Look-Ahead Controls and Physics, *International Journal of*

Production Research, Vol. 43, No. 10, pp. 2001-2025 (2005)

[9] 松井正之：松井の式体系とその周辺、神奈川大学工学研究所所報、第 34 号、pp. 2-5 (2011)

[10] 松井正之：効率対ムダ世界におけるカメレオン基準とメジアン・コントロール法、神奈川大学工学研究所報、第 36 号、pp. 40-45 (2013)

[11] Matsui, M., *Manufacturing and Service Enterprise with Risks II: The Physics and Economics of Management*, International OR&MS series, Springer, 2014

[12] Matsui, M., A Management Game Model: Economic Traffic, Leadtime and Pricing Settings, *Journal of Japan Industrial Management Association*, Vol. 53, No. 1, pp. 1-9 (2002)

[13] 松井正之、藤川裕晃：「ペア戦略チャートが明らかにする製版コラボレーションの利益最大化理論」、DIAMOND ハーバード・ビジネスレビュー、第 30 巻第 1 号、pp. 2-13 (2005)

[14] 松井、中島、高野倉、林、石井、山田：企業ロボットの開発—自働経営管理とアルキメデス(分散)対ソクラテス(集中)方式、日本経営工学会春季大会予稿集、早稲田大学、pp. 128-130 (2016)

Cross Moments Approaches for Mean-Bounded Finite Random Classes

Hiroyuki Takada, Nagasaki University, htakada@cis.nagasaki-u.ac.jp

Abstract We consider a distribution-free stochastic class consists of mean-bounded finite random processes. We develop an interval reduction approach and an expectation inequality on the maximum statistic for the class, in order to understand the evaluation accuracy limit of this class. It is shown that the upper bound can not be improved as long as the class is fixed. The inequality is equivalent to the Boole's inequality, when we apply it to the union probability, while the bound is equal to or smaller than that of the maximum-sum inequality. We also apply them to the single queue with a stream. It produces a stationary bound of backlog average smaller than that of the worst case under a stability condition.

1. Introduction

The technology guaranteeing quality of service (QoS), it is required for real time applications on the Internet. Best effort can not protect streams that should not be delayed and lost from traffic congestion, whereas this keeps stable communication environment for them, by provision of network resources. The procedure is similar to securing lanes to implement marathon events in the road transport network. If there is no such mechanism, marathon runners must run through the crowd. Therefore, this is the important function for the applications.

Network calculus[1, 2] is known as the key theory for the QoS technology. The research field has been started by the seminal work [3], which introduced the simple traffic model so called the (σ, ρ) -regulator. The (min,plus)-filter theory [4] enables systematic evaluation of large scale networks. The performance metrics like backlog and delay are written by the combination and the input functions. The differences of service principles (e.g., FIFO, priority, and processor sharing) and network topologies (e.g., tandem and tree) appears as that of the combination of the basic operators on (min,plus)-algebra; maximum, minimum, plus, and minus. Thus, we can systematically evaluate the worst values of the metrics. Due to the simplicity of the traffic definition and the systematic evaluation framework, The theory has been widely accepted by IntServ [5] and other applications found in [6].

We derive the outstanding features of the deterministic model. The traffic definition assumes the quantitative condition instead of the qualitative ones like statistical independence, Markov property (conditional independence), and martingale property. It is hard to guarantee the qualitative ones. The quantitative assumption can be applied to the evaluation of queueing model with dependent sources. In addition, the output traffic is also characterized as the regulator. The symmetry is similar to the quasi-reversibility found in the studies of product form networks. The deterministic approaches can easily evaluate non-feed-forward networks[1]. On the other hand, the deterministic approaches potentially can not capture the statistical multiplexing effects. This causes a large overestimation, which accelerates resource exhaustion [7]. In fact, IntServ was not popular due to low network utilization. Thus, the deterministic one has advantages in evaluation feasibility and implementation, but it has a drawback in evaluation tightness.

Stochastic network calculus (e.g., see surveys and books [1, 8–13] and their references) is the probabilistic extension in order to overcome the drawback. The input and output functions, backlog, and delay are considered here as random variables. We often evaluate the error bounds, which are defined as the upper bounds of the tail probabilities, instead of the worst values for the metrics. Exponential bounded burstiness (EBB)[14, 15] and its derivations [10, 16–18] directly define the input error bounds. Another branch of the traffic definition (e.g., $(\sigma(\theta), \rho(\theta))$ -regulator [1] and the regulated traffic class[19]) is based on the moment generating function (MGF) and Chernoff inequality. Boole' inequality is often used to evaluate the tail probability of the maximum statistics. Fidler [8] has proposed fundamental inequalities for the MGFs of the maximum, minimum, and summation statistics. These inequalities are simple, however they can lost the correlation information among components.

In recent years, three kinds of new approaches to consider correlation information have been proposed. Jiang [20] has assumed martingale property and used Doob's martingale inequalities. Ciucu *et. al.*[21, 22] have assumed Markov modulated on off model and performed the error bound by the martingale technique. Copula approaches [23–29] have been imported. Copula function has the full correlation information of the joint distribution. The author's previous work [30] has provided the joint MGF approach for the minimum and summation statistics on a mean-bounded finite random class. The approach is based on the multivariate generalization of the special case of Bennett' inequality (see Corollary 2.4.5, Section 2.4.1, pages 56–57, [31])

and the generalized Chernoff inequality (Section 7.2.4, pages 379-380, [32]). The upper bound improves that by the MGF one in exponential term for an many-flow asymptotic evaluation to a FIFO multiplexer.

The motivation of [30] is as follows. The regulated traffic model seems to be good in the viewpoint of realization; implementation and low cost metering. Its accuracy limit is unknown. If the correlation information works for the tightness, it may be sufficient to develop inequalities that does not lose it.

The aim of the paper is to build a non asymptotic approach for the maximum statistics on the class. We will develop an expectation inequality for $\max_{i=1}^n X_i$. In addition, we will introduce a deterministic method which decreases the number of components based on the absorption property of the maximum operator in order to avoid the semi-deterministic problem [33], which claims that the evaluated tail probability takes only 0 and 1 for a class of stationary and ergodic processes. The new upper bound will be compared with those of Boole's inequality and the max-sum one for nonnegative components. We will apply the approach to a single flow queueing model and characterize the output process in terms of the traffic class.

The remainder of the paper is organized as follows. We overview the previous works in Section 2. The main theorem is given in Section 3. A part of the proof is omitted in the paper. The examples and the application are derived in Section 4. In Section 5, we conclude the paper and describe the remaining the works.

2. Related Works

We overview the previous related works in the section. The section consists of three subsections for the traffic definitions, the typical metrics, and the evaluation tools.

2.1. Traffic Definitions

We overview the traffic definitions; the (σ, ρ) -regulator, the regulated traffic class, and the mean-bound finite random class in our notation. Through the paper, the probability space $(\Omega, \mathcal{F}, \Pr)$ is fixed and the time system is discrete, i.e., $t = 0, 1, 2, 3, \dots$

We introduce the general class denoted by $\mathcal{A}$ of random input functions.

Definition 1. The random input function $A(t)$ is monotone increasing with $A(0) = 0$, namely

$$\mathcal{A} = \{A; \Pr[A(s, s+t) \geq 0] = 1 \ (s, t \geq 0), \Pr[A(0) = 0] = 1\},$$

where $A(s, s+t) = A(s+t) - A(s)$.

The (σ, ρ) -regulator is the subclass of $\mathcal{A}$ and denoted by $\mathcal{A}_0(b, r)$, where parameters σ, ρ are replaced to b, r , respectively.

Definition 2 ([3]). Let b and r be positive real values.

$$\mathcal{A}_0(b, r) = \{A \in \mathcal{A}; \Pr[A(s, s+t) \leq rt + b] = 1, (s, t \geq 0)\}.$$

The regulated traffic class is the subclass of $\mathcal{A}_0(b, r)$ and denoted by $\mathcal{A}_1(b, r, \eta)$.

Definition 3 ([19]). Let b, r and η be positive real values.

$$\mathcal{A}_1(b, r, \eta) = \{A \in \mathcal{A}; \Pr[A(s, s+t) \leq rt + b] = 1, E[A(s, s+t)] = \eta, (s, t \geq 0)\}.$$

We relax the conditions. The reason is that it is difficult to evaluate the exact expectation in a short time. The non-negative condition will be removed.

Definition 4 ([30]). Let $\alpha(t)$ and $\beta(t)$ be non-decreasing function which takes value 0 at $t = 0$.

$$\mathcal{A}_1(\alpha, \beta) = \{A \in \mathcal{A}; \Pr[A(s, s+t) \leq \alpha(t) = 1, E[A(s, s+t)] \leq \beta(t) \ (s, t \geq 0)\}.$$

2.2. Typical Metrics

We overview the typical metrics for a single flow model to explain why we focus on the maximum statistic. It is supposed for simplicity that the systems has a work-conserving server with constant service rate $c > 0$.

The typical metrics are given or upper bounded by some maximum statistics. The typical metrics of the single flow model with input function $A \in \mathcal{A}$ is the backlog $Q(t)$, the delay $D(t) = Q(t)/c$, and the output function $B(s, s+t) = B(s+t) - B(s)$, where $B(t) = A(t) - Q(t)$.

$$Q(t) = \max_{u=0}^t A(t-u, t) - cu, \quad B(s, s+t) = B(s+t) - B(s) \leq \min \left(ct, \max_{u=0}^s A(s-u, s+t) - cu \right).$$

Therefore, these metrics are written by the maximum statistics in terms of net flow processes $X_u = A(t-u, t) - cu$ and $X_u = A(s-u, s+t) - cu$. In particular, their expectation bounds are required to characterize the output process by $\mathcal{A}_1$.

The maximum and minimum operators have the following properties.

- Lemma 1.** 1. (commutative property) Identities $\max(\max(x, y), z) = \max(x, \max(y, z)) = \max(y, \max(x, z))$ hold for real values x, y and z . The value is denoted by $\max(x, y, z)$. It is inductively generalized to the case of n variables. This also applies to the minimum operator.
2. (absorption property) If $x \leq y$, then $\max(x, y) = y$ and $\min(x, y) = x$ hold.
3. (maximum-minimums and minimum-maximums identities (Proposition 2.2, Section 7.2.2, pages 313-314, [34])) Suppose that $a_1, a_2, \dots, a_n$ are real values. Let N and 2^N denote $\{1, \dots, n\}$ and the power set of N .

$$\max_{i=1}^n a_i = \sum_{K \in 2^N, K \neq \emptyset} (-1)^{\#K+1} \min_{i \in K} a_i, \quad (2.1)$$

$$\min_{i=1}^n a_i = \sum_{K \in 2^N, K \neq \emptyset} (-1)^{\#K+1} \max_{i \in K} a_i, \quad (2.2)$$

In particular, we have $x + y = \min(x, y) + \max(x, y)$ for $n = 2$.

2.3. Probability Inequalities and Identities

We overview the evaluation tools for the classes.

Boole's inequality has been often used to evaluation the union probability.

Lemma 2. Let $A_1, \dots, A_n$ be events.

$$P[\cup_{i=1}^n A_i] \leq \sum_{i=1}^n P(A_i). \quad (2.3)$$

The inequality is identical with

$$E\left[\max_{i=1}^n 1_{A_i}\right] \leq \sum_{i=1}^n E[1_{A_i}], \quad (2.4)$$

where 1_A is the indicator function, which takes value 1 if A occurs, 0 otherwise.

The max-sum inequality for non-negative components is widely used to evaluate the expectation of a maximum statistic. Inequality (2.3) is a special case of (2.5).

Lemma 3. Suppose that $X_1, \dots, X_n$ be non-negative random variables.

$$E\left[\max_{i=1}^n X_i\right] \leq \sum_{i=1}^n E[X_i]. \quad (2.5)$$

Chernoff inequality connects the tail probability to the MGF.

Lemma 4.

$$\Pr(X > x) \leq \inf_{\theta > 0} E[e^{\theta X}] e^{-x}. \quad (2.6)$$

The generalized Chernoff inequality is the multivariate generalization of Chernoff one and the solution of the definite convex optimization problem [32](see Section 7.4.2, pages 379-380),

Lemma 5. Suppose that $\mathbf{X} = (X_1, \dots, X_n)$ be an n -dimensional random vector. Then, for $C \in \mathbb{R}^n$,

$$\Pr[(X_1, \dots, X_n) \in C] \leq \exp \left\{ \inf_{(\theta_1, \dots, \theta_n) \in \mathbb{R}^n} \left(E \left[\exp \left(\sum_{i=1}^n \theta_i X_i \right) \right] - \inf_{(z_1, \dots, z_n) \in C} \sum_{i=1}^n \theta_i z_i \right) \right\}. \quad (2.7)$$

The corollary of Bennett's inequality found in [31] (see Corollary 2.4.5, Section 2.4.1, pages 56-57) has been used to evaluate the effective bandwidth of the regulated traffic class [19].

Lemma 6. Fix $a < b$. Suppose that $a \leq X \leq b$ is a real valued random variable with $\bar{x} = E(X)$. Then, for any $\theta \in \mathbb{R}$,

$$E[\exp(\theta X)] \leq \frac{\bar{x} - a}{b - a} e^{\theta b} + \frac{b - \bar{x}}{b - a} e^{\theta a}. \quad (2.8)$$

The author [30] has been proposed the following result based on (2.1), (2.2), and Maclaurin's expansion for a multivariate function. The inequality is a multivariate generalization of (2.8) with zero lower boundaries.

Lemma 7. Let $N = \{1, \dots, n\}$. Fix $0 < \xi_{a,i} < \xi_{1,i}$ for $i \in N$. Suppose that $0 \leq X_i \leq \xi_{1,i}$ is a real valued random variable with $E(X_i) \leq \xi_{a,i}$ for $i \in N$. Let $\mathbf{p}_a$ be an n -dimensional vector, the i -th component denoted by $p_{a,i}$ is given by $\xi_{a,i}/\xi_{1,i}$. Put $x^+ = \max(0, x)$ for $x \in \mathbb{R}$. For $\mathbf{p} = (p_1, \dots, p_n) \in [0, 1]^n$ and $K \in 2^N$, we define a discrete type distribution function denoted by $\pi(K; \mathbf{p})$ by

$$\pi(K; \mathbf{p}) = \begin{cases} 1 - \max_{i \in N} p_i & K = \emptyset \\ (\min_{i \in K} \min_{j \in K^c} (p_i - p_j))^+ & K \neq \emptyset, N \\ \min_{i \in N} p_i & K = N \end{cases} \quad (2.9)$$

Let $\mathbf{e}(K)$ be an n -dimensional real vector, the i -th component of which takes value $\xi_{1,i}$ if $i \in K$, 0 otherwise, for $K \in 2^N$. Suppose that $\mathbf{Z} = (Z_1, \dots, Z_n)$ is a multivariate Bernoulli type random vector given by $\Pr[\mathbf{Z} = \mathbf{e}(K)] = \pi(K; \mathbf{p}_a)$. Then, for any $\theta_i > 0$ ($i = 1, \dots, n$),

$$E \left[\exp \left(\sum_{i=1}^n \theta_i X_i \right) \right] \leq E \left[\exp \left(\sum_{i=1}^n \theta_i Z_i \right) \right]. \quad (2.10)$$

3. General Result

We introduce a mean-bounded finite random class and present an inequality for the maximum statistic on the class.

We introduce the mean-bounded finite random class. Suppose that ξ_j be n -dimensional real vectors, the i -th component of which is denoted by $\xi_{j,i}$ for $j = 0, 1, a$. We call ξ_0, ξ_a and ξ_1 the lower, mean, and upper boundary vectors, respectively. The boundary vectors are the parameter of the class. Inequalities between vectors are always component-wise. We assume that

$$\xi_0 \leq \xi_a \leq \xi_1. \quad (3.1)$$

We put $\xi_d = (\xi_0, \xi_1)$ and $\xi = (\xi_0, \xi_a, \xi_1)$ for simplicity. Let $\mathbf{X} = (X_1, \dots, X_n)$ be a n -dimensional random vector. Its component-wise expectation vector is denoted by $\mathbf{E}[\mathbf{X}] = (E[X_1], \dots, E[X_n])$. We use the following conditions to define the class.

$$\Pr(\xi_0 \leq \mathbf{X} \leq \xi_1) = 1, \quad (3.2)$$

$$\mathbf{E}[\mathbf{X}] \leq \xi_a. \quad (3.3)$$

Definition 5. We say that $\mathbf{X}$ is a finite random process if it satisfies (3.2). The whole set of $\mathbf{X}$ is denoted by $\mathcal{T}_0(\xi_d)$ (or simply $\mathcal{T}_0$). The mean-bound finite random process class is denoted by $\mathcal{T}_1(\xi)$ (or simply $\mathcal{T}_1$) and defined by

$$\mathcal{T}_1(\xi) = \{\mathbf{X} \in \mathcal{T}_0(\xi_d); \mathbf{X} \text{ satisfies (3.3)}\}.$$

We assume an additional condition to the parameter ξ . We consider the maximum statistic denoted by $V(\mathbf{X}) = \max_{i \in N} X_i$ for $\mathbf{X} = (X_1, \dots, X_n) \in \mathcal{T}_1(\xi)$. Let $\xi_{1,0} = \max_{i \in N} \xi_{0,i}$. We can assume that

$$\xi_{1,0} \leq \xi_{1,1} \leq \dots \leq \xi_{1,n}, \quad i \in N. \quad (3.4)$$

without loss of generality. It is because that if there exists some index $k \in N$ such that $\xi_{1,k} < \xi_{1,0}$, then $X_k \leq \xi_{1,k} < \xi_{1,0} = \xi_{0,j} \leq X_j$ holds, where $j \in N$ is some index with $\xi_{0,j} = \max_{i \in N} \xi_{0,i}$, hence, by the absorption property of max, we can reduce the corresponding components to k , moreover, the monotone property is established by sorting and relabeling. The reduction method is referred to as *the interval reduction*. We put $N_0 = \{0, 1, 2, \dots, n\}$.

We introduce the normalized vector $\mathbf{p} = (p_1, \dots, p_n)$ of $\mathbf{x} = (x_1, \dots, x_n)$ which satisfies $\xi_0 \leq \mathbf{x} \leq \xi_1$.

$$p_i = \begin{cases} \frac{x_i - \xi_{0,i}}{\xi_{1,i} - \xi_{0,i}} & \xi_{0,i} < \xi_{1,i} \\ 1 & \xi_{0,i} = \xi_{1,i} \end{cases}. \quad (3.5)$$

Definition (3.5) directly yields equation

$$x_i = \xi_{0,i} * (1 - p_i) + \xi_{1,i} * p_i, \quad i \in N. \quad (3.6)$$

The normalized vectors of $\mathbf{X} \in \mathcal{T}_1$ and ξ_a are denoted by $\mathbf{P}$ and $\mathbf{p}_a$, respectively. Note that condition (3.3) is equivalent to

$$\mathbf{E}[\mathbf{P}] \leq \mathbf{p}_a \quad (3.7)$$

We present the main result for the maximum order statistic.

Theorem 1. Assume (3.1) and (3.4). Define $q_i = 1 - \sum_{j=i}^n p_{a,j}$ with $q_0 = 0$ and $q_{n+1} = 0$, and $\hat{\pi}(i) = q_{i+1}^+ - q_i^+$ for $i \in N_0$. Then, (i) $\hat{\pi}(i)$ is a probability function on N_0 . Moreover, suppose that Y is a discrete random variable defined by $\Pr[Y = i] = \hat{\pi}(i)$. Then, (ii) for any $\mathbf{X} \in \mathcal{T}_1$,

$$\mathbb{E}[V(\mathbf{X})] \leq \mathbb{E}[Y]. \quad (3.8)$$

In addition, (iii) there exists $\mathbf{W} = (W_1, \dots, W_n) \in \mathcal{T}_1$ such that

$$\mathbb{E}[V(\mathbf{W})] = \mathbb{E}[Y]. \quad (3.9)$$

We describe some notations, concepts, and lemmas to show Theorem 1. Let us define

$$\mathbf{e}(K)_i = \begin{cases} \xi_{1,i} & i \in K \\ \xi_{0,i} & i \in K^c \end{cases}, \quad K \in 2^N. \quad (3.10)$$

It is interpreted as the K -th extreme point of cuboid $\prod_{i \in N} [\xi_{0,i}, \xi_{1,i}]$.

We also provide another form of π defined in (2.9) for the intuitive comprehension.

Lemma 8. For consistent, put $p_0 = 0$ and $p_{n+1} = 1$. If

$$0 \leq p_1 \leq \dots \leq p_n \leq 1, \quad (3.11)$$

then, we have

$$\pi(K) = \begin{cases} p_k - p_{k-1} & \begin{cases} K = \{k, k+1, \dots, n\} \\ k = 1, \dots, n, n+1 \end{cases} \\ 0 & \text{otherwise} \end{cases}. \quad (3.12)$$

Proof. It is easy to confirm the equality between (2.9) and (3.12) for the case that K forms $\{k, k+1, \dots, n\}$. Otherwise, we can find a pair of indices $k \in K$ and $l \in K^c$ such that $p_k \leq p_l$. It implies that $\pi(K) = 0$, because $0 \leq (\min_{i \in K} \min_{j \in K^c} (p_i - p_j))^+ \leq (p_k - p_l)^+ = 0$. $\square$

We derive corollaries from (3.12). With (3.12), it is trivial that $\{\pi(K); K \in 2^N\}$ is a division of interval $[0, 1]$, hence π is a probability function on 2^N .

Lemma 9. We also have the convex combination

$$\mathbf{x} = \sum_{K \in 2^N} \pi(K; \mathbf{p}) \mathbf{e}(K). \quad (3.13)$$

Hence, it yields

$$F(\mathbf{x}) \leq \sum_{K \in 2^N} \pi(K; \mathbf{p}) F(\mathbf{e}(K)), \quad (3.14)$$

for convex function $F : \mathbb{R}^n \rightarrow \mathbb{R}$.

Proof. The former identity is justified by the following component-wise identity. For $k \in N$, the k -th component of the right hand side of (3.13) is given by $\sum_{K \ni k} \pi(K) \mathbf{e}(K)_k + \sum_{K \not\ni k} \pi(K) \mathbf{e}(K)_k$. It becomes $\sum_{K \ni k} \pi(K) \xi_{1,k} + \sum_{K \not\ni k} \pi(K) \xi_{0,k}$ by (3.10). From (3.12), the coefficients of $\xi_{1,k}$ and $\xi_{0,k}$ are p_k and $1 - p_k$, respectively. From (3.6), $\pi_k \xi_{1,k} + (1 - p_k) \xi_{0,k}$ is equal to x_k .

The latter inequality follows to the definition of convex functions. $\square$

We introduce a power difference operator for simple description.

Definition 6. Let F and G be functions from 2^N to $\mathbb{R}$. Define Δ by

$$\Delta F(K) = \sum_{J \in K} (-1)^{\#J} F(J + K^c), \quad (3.15)$$

respectively.

The linear operator is unitary.

Lemma 10.

$$\sum_K \Delta F(K) G(K) = \sum_K F(K) \Delta G(K). \quad (3.16)$$

Proof. The left hand side of (3.16) is given by $\sum_K \sum_{J \subset K} (-1)^{\#J} F(J + K^c) G(K)$ by (3.15). By the exchange of the summation order, we have $\sum_J \sum_{K \supset J} (-1)^{\#J} F(J + K^c) G(K)$. By the change of variable K to $L = J + K^c$, this becomes $\sum_J \sum_{L \supset J} (-1)^{\#J} F(L) G(J + L^c)$. It is equal to $\sum_L F(L) \sum_{J \subset L} (-1)^{\#J} G(J + L^c)$, which is the right hand side of (3.16) from (3.15). $\square$

We introduce the rectangle-antitone property. The super rectangle-antitone property is so called the Δ -antitone one in [35](Definition 3.3.14 (c)(95 page)).

Definition 7. We say that $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is super (resp. sub) rectangle-antitone, if

$$\Delta f_{\mathbf{e}}(K) \geq (\text{resp. } \leq) 0, \quad \text{for all } K \neq \emptyset. \quad (3.17)$$

where $f_{\mathbf{e}}(K) = f(\mathbf{e}(K))$.

We omit the proof of the following lemma.

Lemma 11. V is sub rectangle-antitone.

The probability function π given in (2.9) is represented in terms of $\{\max_{i \in L} p_i; L \in 2^N\}$ from (2.2).

Lemma 12. For given $\mathbf{p} \in [0, 1]^n$, and $K \in 2^N$, we have the following form.

$$\pi(K) = 1[K = \emptyset] - \Delta \max_{\mathbf{p}}(K), \quad (3.18)$$

where $\max_{\mathbf{p}}(K) = \max_{i \in K} p_i$.

We give a trivial upper bound of $E[\max_{i \in K} P_i]$ for $K \neq \emptyset$. Recall that $E[\mathbf{P}] \leq \mathbf{p}_a$ holds. Define $S_{\mathbf{p}}(K) = \min(1, \sum_{i \in K} p_i)$. Then, we have from the max-sums inequality (2.5)

$$E[\max_{\mathbf{p}}(K)] \leq S_{\mathbf{p}_a}(K). \quad (3.19)$$

Due to (3.19) and (3.16), we will require a new function $\tilde{\pi}(\cdot; \mathbf{p}) : 2^N \rightarrow \mathbb{R}$ defined by

$$\tilde{\pi}(K; \mathbf{p}) = 1[K = \emptyset] - \Delta S_{\mathbf{p}}(K). \quad (3.20)$$

for $K \in 2^N$. Symbol $\tilde{\pi}(K; \mathbf{p})$ is simplified by $\tilde{\pi}(K)$ so long as there is no risk of confusion.

Function $\tilde{\pi}$ is not always a probability one. For $n = 2$, it is easy to check that $\tilde{\pi}$ is a probability function.

$$\tilde{\pi}(\emptyset) = (1 - p_1 - p_2)^+, \quad \tilde{\pi}(\{1\}) = \min(p_1, 1 - p_2), \quad \tilde{\pi}(\{2\}) = \min(p_2, 1 - p_1), \quad \tilde{\pi}(\{1, 2\}) = (p_1 + p_2 - 1)^+.$$

We present a counter example for the statement that $\tilde{\pi}$ is a probability function.

Example 1. Set $n = 3$, $p_1 = 0.3$, $p_2 = 0.4$, and $p_3 = 0.5$. Then, we have

$$\begin{aligned} \tilde{\pi}(\emptyset) &= 0, & \tilde{\pi}(\{1\}) &= 0.1 & \tilde{\pi}(\{2\}) &= 0.2, & \tilde{\pi}(\{1, 2\}) &= 0.2 & \tilde{\pi}(\{3\}) &= 0.3, & \tilde{\pi}(\{1, 3\}) &= 0.2, \\ \tilde{\pi}(\{2, 3\}) &= 0.2, & \tilde{\pi}(N) &= -0.2 \end{aligned}$$

Thus, $\tilde{\pi}$ is not a probability function, but a negative one.

Let us associate $\tilde{\pi}$ with $\hat{\pi}$. Put $\mathcal{N}(k) = \{K; \max K = k\}$ for $k \in N_0$, where it is supposed that $\mathcal{N}(0) = \{\emptyset\}$ for consistent. Clearly, family $\{\mathcal{N}(k); k \in N_0\}$ is a division of 2^N , namely,

$$\sum_{k \in N_0} \mathcal{N}(k) = 2^N, \quad \mathcal{N}(i) \cap \mathcal{N}(j) = \emptyset \quad (i \neq j). \quad (3.21)$$

We have from (3.10) and (3.4) that

$$V(\mathbf{e}(K)) = \xi_{1,k}, \quad K \in \mathcal{N}(k), k \in N_0. \quad (3.22)$$

The following lemma indicates the coefficient of $\xi_{1,k}$ in $\sum_K \tilde{\pi}(K) V(\mathbf{e}(K))$. The proof is omitted.

Lemma 13. We have

$$\hat{\pi}(k) = \sum_{K \in \mathcal{N}(k)} \tilde{\pi}(K), \quad k \in N_0. \quad (3.23)$$

The goal of the subsection is to concretely produce the random vector $\mathbf{W}$, or equivalently, to find a non-negative solution $\{\tilde{\pi}(K); K \in 2^N\}$ of linear equations

$$\sum_{K; K \ni k} \tilde{\pi}(K) = p_{a,k}, \quad k \in N, \quad (3.24)$$

$$\sum_{K \in \mathcal{N}(k)} \tilde{\pi}(K) = \hat{\pi}(k), \quad k \in N_0, \quad (3.25)$$

With the solution, let $\mathbf{W}$ be a multivariate Bernoulli type random vector which takes value $\mathbf{e}(K)$ with probability $\check{\pi}(K)$ for $K \in 2^N$. It is clear that $\mathbf{W} \in \mathcal{T}_0$. We also have $\mathbf{W} \in \mathcal{T}_1$, because the following equations hold.

$$\begin{aligned} E[W_k] &= \Pr[W_k = \xi_{0,k}] \xi_{0,k} + \Pr[W_k = \xi_{1,k}] \xi_{1,k} = \sum_{K \not\ni k} \check{\pi}(K) \xi_{0,k} + \sum_{K \ni k} \check{\pi}(K) \xi_{1,k} \\ &= (1 - p_{a,k}) \xi_{0,k} + p_{a,k} \xi_{1,k} = \xi_{a,k}, \quad k \in N, \end{aligned}$$

from (3.24) and (3.6). In addition, we obtain that (3.9) from (3.25), (3.22) and the definition of Y . We therefore concentrate to provide the solution $\check{\pi}$.

We give a numerical example to show that the uniqueness of the solution does not always hold.

Example 2. Let $n = 3$ and $\mathbf{p} = (0.9, 0.8, 0.5)$. Then $\hat{\pi}$ is given by

$$\hat{\pi}(0) = 0, \quad \hat{\pi}(1) = 0, \quad \hat{\pi}(2) = 0.5, \quad \hat{\pi}(3) = 0.5.$$

The non-negative solution $\check{\pi}$ is given by

$$\begin{aligned} \check{\pi}(\emptyset) &= 0, \quad \check{\pi}(\{1\}) = 0, \quad \check{\pi}(\{2\}) = s, \quad \check{\pi}(\{1, 2\}) = 0.5 - s, \quad \check{\pi}(\{3\}) = t, \quad \check{\pi}(\{1, 3\}) = 0.2 - t \\ \check{\pi}(\{2, 3\}) &= 0.1 - s - t, \quad \check{\pi}(N) = 0.2 + s + t, \end{aligned}$$

for $s, t \geq 0$ with $s + t \leq 0.1$. In particular, if $s = 0$ and $t = 0.1$, then

$$\begin{aligned} \check{\pi}(\emptyset) &= 0, \quad \check{\pi}(\{1\}) = 0, \quad \check{\pi}(\{2\}) = 0, \quad \check{\pi}(\{1, 2\}) = 0.5, \quad \check{\pi}(\{3\}) = 0.1, \quad \check{\pi}(\{1, 3\}) = 0.1 \\ \check{\pi}(\{2, 3\}) &= 0, \quad \check{\pi}(N) = 0.3. \end{aligned}$$

We derive notation used in the subsection. Symbols $p_{a,i}$ and $\mathbf{p}_a$ are simply denoted by p_i and $\mathbf{p}$, respectively. Each subset K of N is corresponding to $x = b(K) = \sum_{i \in K} 2^{i-1}$. For example of $n = 3$,

$$\begin{aligned} b(\emptyset) &= 0 = 000_{(2)}, \quad b(\{1\}) = 1 = 001_{(2)}, \quad b(\{2\}) = 2 = 010_{(2)}, \quad b(\{1, 2\}) = 3 = 011_{(2)}, \\ b(\{3\}) &= 4 = 100_{(2)}, \quad b(\{1, 3\}) = 5 = 101_{(2)}, \quad b(\{2, 3\}) = 6 = 110_{(2)}, \quad b(\{1, 2, 3\}) = 7 = 111_{(2)}. \end{aligned}$$

Inversely, $b^{-1}(x)$ denotes x 's corresponding subset K for integer $x \in [0, 2^n - 1]$.

The solution $\hat{\pi}$ is inductively defined by Algorithm 1. See also Table 1 for the execution result to Example 2.

Algorithm 1 Allocation to $\check{\pi}$

Input: Integer $n \geq 1$ and $\mathbf{p} \in [0, 1]^n$.

- 1: $\check{\pi}(\emptyset) \leftarrow \hat{\pi}(0) = (1 - \sum_{i=1}^n p_i)^+$.
- 2: $\alpha(i) \leftarrow p_i$ for all $i = 1, \dots, n$.
- 3: **for** $k = 1$ to n **do**
- 4: $\alpha(0) \leftarrow \hat{\pi}(k) = \min(p_k, (1 - \sum_{i=k+1}^n p_i)^+)$.
- 5: **for** $x = 2^k - 1$ to 2^{k-1} **do**
- 6: $K \leftarrow b^{-1}(x)$.
- 7: $\check{\pi}(K) \leftarrow \min_{i \in K + \{0\}} \alpha(i)$.
- 8: $\alpha(i) \leftarrow \alpha(i) - \check{\pi}(K)$ for all $i \in K + \{0\}$.
- 9: **end for**
- 10: **end for**

Output: $\check{\pi}(K)$ for all $K \in 2^N$.

We show the non-negativity of $\check{\pi}$. Due to the complexity of the procedure order of K in Algorithm 1, we avoid the inductive computation of $\check{\pi}$. We interpret $\alpha(i)$ as the residual amount of resource i . We also interpret $\check{\pi}(K)$ as the amount of product K . The same amount of resources i for $i \in K + \{0\}$ are consumed to produce a unit of product K (steps 7 and 8). Resource 0 is filled up to $\min(p_k, q_{k+1}^+)$ at step 4 for every k . For $i \in N$, resource i is once filled up to p_i at step 2. In the view of resource allocation, it is trivial that $\alpha(i)$ and $\check{\pi}(K)$ are preserved to be non-negative.

We observe the exhaustion of the resources to show (3.24) and (3.25). For each k , inequality $\alpha(k) \geq \alpha(0)$ is always preserved, because inequality $p_k \geq \min(p_k, q_{k+1}^+)$ holds at step 4, moreover, resources 0 and k are consumed by the same amount for each loop on K . At $K = \{k\}$, there is a chance to dry up resource 0. That is, resource 0 is exhausted earlier than or simultaneously resource k . It implies that (3.25). On the other hand, total amount of consumed resource 0 is $1 - q_1^+ = \min(1, \sum_{i \in N} p_i)$, which is greater than or equal to p_i for every $i \in N$. Hence, all resources are exhausted earlier than or simultaneously resource 0. Total amount of products with resource k is equal to $\sum_{K \in 2^N; K \ni k} \check{\pi}(K)$, which must be equal to p_k . It implies (3.24).

Table 1: Variable transition of Algorithm 1: Input parameters are $n = 3$ and $\mathbf{p} = (0.9, 0.8, 0.5)$ same as Example 2. Except for the last line, the values for $k = 0$ and $k \geq 1$ are respectively captured just after step 2 and step 7, respectively. The last line shows the values just after step 10.

k	K	x	binary	$\tilde{\pi}(K)$	$\alpha(0)$	$\alpha(3)$	$\alpha(2)$	$\alpha(1)$
0	$\emptyset$	0	000 ₍₂₎	0	-	0.5	0.8	0.9
1	$\{1\}$	1	001 ₍₂₎	0	0	0.5	0.8	0.9
2	$\{1, 2\}$	3	011 ₍₂₎	0.5	0.5	0.5	0.8	0.9
	$\{2\}$	2	010 ₍₂₎	0	0	0.5	0.3	0.4
3	$\{1, 2, 3\}$	7	111 ₍₂₎	0.3	0.5	0.5	0.3	0.4
	$\{2, 3\}$	6	110 ₍₂₎	0	0.2	0.2	0	0.1
	$\{1, 3\}$	5	101 ₍₂₎	0.1	0.2	0.2	0	0.1
	$\{3\}$	4	100 ₍₂₎	0.1	0.1	0.1	0	0
Values just after step 11					0	0	0	0

4. Application

In the section, we consider three examples.

4.1. vs Boole's inequality

Put $X_i = 1_{A_i}$, $\xi_0 = \mathbf{0}$, $\xi_a = \mathbb{E}[1_{A_i}] = \Pr[A_i]$, and ξ_1 . Then, the maximum bound $\mathbb{E}[Y]$ is given by $\mathbb{E}[Y] = 0 * q_1^+ + \sum_{i \in N} 1 * (q_{i+1}^+ - q_i^+) = 1 - q_1^+ = \min(1, \sum_{i \in N} p_i)$, which is identical with the Boole's bound.

4.2. vs max-sum inequality

We present the example with $n = 2$, $\xi_0 = \mathbf{0}$, $\xi_a = (0.25, 1.75)$, and $\xi_1 = (1, 2)$. $\mathbb{E}[X_1] + \mathbb{E}[X_2] \leq 2$, while $\mathbb{E}[Y] = 1.875$. Hence, the new bound is better than that of (2.5).

4.3. A single flow queue

We consider a discrete time queue with a single regulated flow and a work conserving constant rate server with rate $c > 0$. Assume that the input function $A(t)$ belongs to $\mathcal{A}_1(rt + b, rt)$, a stability condition $r < c$. If $b + r \leq c$, then $Q(t) = 0$, hence $\mathbb{E}[Q(t)] = 0$. Assume $b + r > c$. Put $X_s = A(t) - A(s) - c(t - s)$. $Q(t) = \max_{s=0}^t X_s$. From the assumptions, we have $0 \leq A(t) - A(s) \leq r(t - s) + b$ and $\mathbb{E}[A(t) - A(s)] \leq r(t - s)$. Put

$$\xi_{0,s} = -c(t - s), \quad \xi_{1,s} = -(c - r)(t - s) + b, \quad \xi_{a,s} = -(c - r)(t - s), \quad p_{a,s} = \frac{r(t - s)}{r(t - s) + b}.$$

Since $\max_{s=0}^t \xi_{0,s} = 0$, $\xi_{1,s} \geq 0$ holds for $s \geq t - m$, where $m = \lfloor \alpha \rfloor$, $\alpha = b/(c - r)$. We have

$$Q(t) = \max_{s=t-m}^t X_s.$$

Put $q_s = 1 - \sum_{u=s}^t p_{a,u}$, $\alpha_0 = \alpha - m$. We apply Theorem 1 to $\mathbb{E}[Q(t)]$.

$$\begin{aligned} \mathbb{E}[Q(t)] &= \mathbb{E} \left[\max_{s=t-m}^t X_s \right] \leq \sum_{s=t-m}^t \xi_{1,s} (q_{s+1}^+ - q_s^+) = \dots \\ &= b - (c - r) \left(\alpha_0 \left(1 - \sum_{x=1}^m \frac{rx}{rx + b} \right)^+ + \sum_{y=0}^{m-1} \left(1 - \sum_{x=1}^y \frac{rx}{rx + b} \right)^+ \right). \end{aligned}$$

Let b' be the bound.

- If $b = 50M$ [bits], $r = 20M$ [bpms], $c = 25M$ [bpms] then $b' = 40.6M$ [bits]
- If $b = 50M$ [bits], $r = 5M$ [bpms], $c = 25M$ [bpms] then $b' = 4.4M$ [bits]
- If $b = 10M$ [bits], $r = 20M$ [bpms], $c = 25M$ [bpms] then $b' = 3.3M$ [bits]

5. Conclusion and Future Works

We have introduced the mean-bounded finite random class, and obtained the maximum expectation of the maximum statistic on the class. The bound is identical with the Boole's bound, while it is better than that of the max-sum inequality. For a single flow queueing model, the stationary bound of the backlog expectation is evaluated smaller than the burst tolerance. Thus, the approach can capture the statistical multiplexing effect in the mean of the expectation.

We describe the future works. The bound is not the ideal bound like $\max_{\mathbf{X} \in \mathcal{T}_1} \Pr(\mathbf{X} \in C)$. In the mean, we have a room to improve the inequality. It is needed to confirm how good the new bound is. If it is not suitable for the actual system, then we should add assumptions to capture the correlation information explicitly. We would like to apply the result to non-feed-forward network models after the output characterization of a FIFO multiplexer.

References

- [1] C.-S. Chang, *Performance guarantees in communication networks*. Springer Science & Business Media, 2000.
- [2] J.-Y. Le Boudec and P. Thiran, *Network calculus: a theory of deterministic queuing systems for the internet*. Springer Science & Business Media, 2001, vol. 2050.
- [3] R. L. Cruz, “A calculus for network delay. i. network elements in isolation,” *Information Theory, IEEE Transactions on*, vol. 37, no. 1, pp. 114–131, 1991.
- [4] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat, *Synchronization and linearity: an algebra for discrete event systems*. John Wiley & Sons Ltd, 1992.
- [5] R. Braden, D. Clark, and S. Shenker, “Rfc 1633: Integrated services in the internet architecture: an overview, june 1994,” *Status: Informational*, 1994.
- [6] F. Ciucu, M. Fidler, J. Liebeherr, and J. Schmitt, “Network Calculus (Dagstuhl Seminar 15112),” *Dagstuhl Reports*, vol. 5, no. 3, pp. 63–83, 2015. [Online]. Available: <http://drops.dagstuhl.de/opus/volltexte/2015/5269>
- [7] J. B. Schmitt, F. Zdarsky, M. Fidler *et al.*, “Delay bounds under arbitrary multiplexing: When network calculus leaves you in the lurch...” in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, IEEE, apr 2008, pp. 1669–1677.
- [8] M. Fidler, “An end-to-end probabilistic network calculus with moment generating functions,” in *Quality of Service, 2006. IWQoS 2006. 14th IEEE International Workshop on*, IEEE. IEEE, jun 2006, pp. 261–270.
- [9] S. Mao and S. Panwar, “A survey of envelope processes and their applications in quality of service provisioning,” *Communications Surveys & Tutorials, IEEE*, vol. 8, no. 3, pp. 2–20, 2006.
- [10] Y. Jiang and Y. Liu, *Stochastic network calculus*. Springer, 2008, vol. 1.
- [11] M. Fidler, “Survey of deterministic and stochastic service curve models in the network calculus,” *Communications Surveys & Tutorials, IEEE*, vol. 12, no. 1, pp. 59–86, 2010.
- [12] Y. Jiang, “Stochastic network calculus for performance analysis of internet networks - an overview and outlook,” in *Computing, Networking and Communications (ICNC), 2012 International Conference on*. IEEE, 2012, pp. 638–644.
- [13] M. Fidler and A. Rizk, “A guide to the stochastic network calculus,” *Communications Surveys & Tutorials, IEEE*, vol. 17, no. 1, pp. 92–105, 2015.
- [14] O. Yaron and M. Sidi, “Performance and stability of communication networks via robust exponential bounds,” *Networking, IEEE/ACM Transactions on*, vol. 1, no. 3, pp. 372–385, 1993.
- [15] —, “Generalized processor sharing networks with exponentially bounded burstiness arrivals,” in *INFOCOM’94. Networking for Global Communications., 13th Proceedings IEEE*. IEEE, 1994, pp. 628–634.
- [16] D. Starobinski and M. Sidi, “Stochastically bounded burstiness for communication networks,” in *INFOCOM’99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 1. IEEE, 1999, pp. 36–42.
- [17] Q. Yin, Y. Jiang, S. Jiang, and P. Y. Kong, “Analysis on generalized stochastically bounded bursty traffic for communication networks,” in *Local Computer Networks, 2002. Proceedings. LCN 2002. 27th Annual IEEE Conference on*. IEEE, 2002, pp. 141–149.
- [18] Y. Jiang, “A basic stochastic network calculus,” *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 4, pp. 123–134, 2006.
- [19] F. Kelly, “Notes on effective bandwidths. stochastic networks: Theory and applications, royal statistical society lecture notes series, vol. 4,” 1996.
- [20] Y. Jiang, “A note on applying stochastic network calculus,” 2010.
- [21] F. Ciucu, F. Poloczek, and J. Schmitt, “Sharp bounds in stochastic network calculus,” in *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 1. ACM, 2013, pp. 367–368.

- [22] —, “Sharp per-flow delay bounds for bursty arrivals: The case of fifo, sp, and edf scheduling,” in *INFOCOM, 2014 Proceedings IEEE*. IEEE, 2014, pp. 1896–1904.
- [23] F. Dong, K. Wu, and V. Srinivasan, “Copula analysis for statistical network calculus,” in *Computer Communications (INFOCOM), 2015 IEEE Conference on*. IEEE, 2015, pp. 1535–1543.
- [24] F. Sun and Y. Jiang, “On power and quality of service tradeoff in device-to-device communication,” in *Communication Workshop (ICCW), 2015 IEEE International Conference on*. IEEE, 2015, pp. 614–619.
- [25] —, “Further properties of wireless channel capacity,” *ArXiv e-prints*, Jan. 2015.
- [26] L. Santinelli, “Probabilistic component-based analysis for networks,” *ACM SIGBED Review*, vol. 13, no. 3, pp. 65–72, 2016.
- [27] F. Sun, Y. Jiang, and L. Li, “A study on further properties of wireless channel capacity,” in *Communications in China (ICCC), 2016 IEEE/CIC International Conference on*. IEEE, 2016, pp. 1–6.
- [28] P. Nikolaus and J. Schmitt, “On Per-Flow Delay Bounds in Tandem Queues under (In)Dependent Arrivals,” in *Proceedings of 16th IFIP Networking 2017 Conference (NETWORKING’17)*, Stockholm, Sweden, 2017, iSBN 978-3-901882-94-4.
- [29] F. Sun and Y. Jiang, “Towards a calculus for wireless networks,” *arXiv preprint arXiv:1705.03714*, 2017.
- [30] H. Takada, “Joint moment generating function approach for stochastic network calculus,” in *the Queue Symposium 2015*, jan 2016.
- [31] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Springer Science & Business Media, 2009, vol. 38.
- [32] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [33] F. Ciucu and J. Schmitt, “Perspectives on network calculus: no free lunch, but still good value,” in *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication*. ACM, 2012, pp. 311–322.
- [34] S. Ross, *A first course in probability*. Pearson, 2012.
- [35] A. Müller and D. Stoyan, “Comparison methods for stochastic models and risks. 2002,” *John Wiley&Sons Ltd., Chichester*, 2002.

Distribution of the waiting and service time in an M/M/ m preemptive-resume priority queue with impatient customers

Hideaki Takagi

Professor Emeritus, University of Tsukuba

Abstract We study an M/M/ m preemptive-resume priority queue with impatient customers where customers of the same priority class are served on FCFS and preemptive-resume (PR-) LCFS basis in a unified framework. The presentation is focused on the FCFS case with reference to the correspondence with the PR-LCFS case. We consider the time interval from arrival to either service completion or abandonment of an arbitrary customer of each class in the steady state. The problem is formulated as a combination of two one-dimensional birth-and-death processes representing the service and waiting states of a tagged customer. We provide explicit expressions in terms of Laplace-Stieltjes transform of the distribution function for the first passage time to service completion and abandonment which is decomposed into the waiting and the service time of an arbitrary customer of each priority class.

1. Introduction

We consider a queueing system with m servers and an infinite capacity of the waiting room with several priority classes of customers. Customers of class p arrive in a Poisson process with rate λ_p (> 0) independently of customers of all other classes. Every customer requests a service which has the exponential distribution with mean $1/\mu$ irrespective of his class. Classes are indexed $1, 2, \dots$ such that customers of class p have preemptive priority for service over customers of class q if $p < q$. As soon as a server becomes available, one of the customers of the highest priority class among those in the waiting room is called in for service. When the service is resumed, a new sample of the service time is set up from the exponential distribution with mean $1/\mu$, irrespective of the amount of service given to him previously. Thus customers of class q ($> p$) have no effect on customers of class p . This service discipline is phrased “preemptive-resume priority” in the title of the paper.

With respect to priority for customers of the same class, we can deal with the following cases in a unified framework as shown in Table 1 given at the end of the paper:

- (i) An arriving customer joins the tail of all customers of the same class. A customer is chosen for service on the first-come, first-served (FCFS), naturally nonpreemptive basis.
- (ii) An arriving customer preemptively comes ahead of all customers of the same class. A customer is chosen on the last-come, first-served (PR-LCFS) basis.

In the presentation below, we focus on case (i) referring the reader to Table 1 for symmetric treatment of case (ii).

In addition, we assume that each customer in the waiting room leaves the system (abandons waiting) with probability $\theta\Delta t$ during a short interval $(t, t + \Delta t)$. In other words, the patience time of each customer is exponentially distributed with mean $1/\theta$ irrespective of his class. Customers in service never leave the system. Let us define

$$\tau := \frac{\theta}{\mu} ; \quad \lambda_p^+ := \sum_{k=1}^p \lambda_k ; \quad \rho_p^+ := \frac{\lambda_p^+}{m\mu}$$

A brief review of relevant literature is given as follows. In the case without impatient customers, the mean waiting and response time in an M/M/ m preemptive-resume priority queue is studied by Brosh [1], Segal [4], and Buzen and Bondi [2]. Then, the distribution of the waiting and response time is considered by Tatashev [11], Zeltyn et al [12], and Takagi [8] by assuming the FCFS discipline for customers of the same class, and by Takagi [6] by assuming the LCFS discipline for customers of the same class. In the case with impatient customers, Takagi [5, 10] and Takagi [7, 9]

derive the moments and distribution, respectively, for the waiting and service time in an M/M/ m preemptive-resume LCFS queue with impatient customers without priority class.

The present paper, for the first time, deals with the distribution, in a form of Laplace-Stieltjes transform of the distribution function (LST of the DF), for the waiting and service time in an M/M/ m preemptive-resume priority queue with impatient customers. The explicit derivation for the corresponding moments of distribution can be found in [9]. Due to the page limit, we omit the numerical results, which will be presented elsewhere.

2. Number of customers present in a system in the steady state

According to the PASTA (Poisson arrivals see time averages) property, in a queueing system with Poisson arrivals, an arriving customer finds the number of customers present in the system whose distribution is the same as that at an arbitrary time. In an FCFS-based system, those customers of class $1, 2, \dots, p$ found in the system by an arriving tagged customer of class p are served before him. Therefore, we consider the number of customers of class $1, 2, \dots, p$ present in the system at an arbitrary time in the steady state, whose probability distribution is defined by

$$Q_{k,p}^+ := P\{\mathcal{N}_p^+ = k\} \quad k = 0, 1, 2, \dots$$

From the balance equations in a birth-and-death process $\mathcal{N}_p^+$, we get

$$\begin{aligned} \lambda_p^+ Q_{k-1,p}^+ &= k\mu Q_{k,p}^+ & 1 \leq k \leq m, \\ \lambda_p^+ Q_{k-1,p}^+ &= [m\mu + (k-m)\theta] Q_{k,p}^+ & k \geq m+1. \end{aligned}$$

The solution is given by

$$Q_{k,p}^+ = \begin{cases} Q_{0,p}^+ \frac{(m\rho_p^+)^k}{k!} & 1 \leq k \leq m, \\ Q_{m,p}^+ \frac{(\rho_p^+)^{k-m}}{\prod_{j=1}^{k-m} (1+j\tau/m)} & k \geq m+1 \end{cases} \quad k \geq m+1,$$

where

$$\frac{1}{Q_{0,p}^+} = \sum_{k=0}^m \frac{(m\rho_p^+)^k}{k!} + \frac{(m\rho_p^+)^m}{m!} \sum_{k=1}^{\infty} \frac{(\rho_p^+)^k}{\prod_{j=1}^k (1+j\tau/m)} < \infty.$$

The mean number of customers of class $1, 2, \dots, p$ present in the entire system is given by

$$E[\mathcal{N}_p^+] = Q_{0,p}^+ \left[\sum_{k=1}^m \frac{(m\rho_p^+)^k}{(k-1)!} + \frac{(m\rho_p^+)^m}{m!} \sum_{k=1}^{\infty} \frac{(m+k)(\rho_p^+)^k}{\prod_{j=1}^k (1+j\tau/m)} \right].$$

The mean number of customers of class $1, 2, \dots, p$ present in the waiting room is given by

$$E[\mathcal{L}_p^+] = Q_{0,p}^+ \frac{(m\rho_p^+)^m}{m!} \sum_{k=1}^{\infty} \frac{k(\rho_p^+)^k}{\prod_{j=1}^k (1+j\tau/m)}.$$

The mean number of customers of class $1, 2, \dots, p$ being served is given by

$$E[\mathcal{S}_p^+] = Q_{0,p}^+ \left[\sum_{k=1}^m \frac{(m\rho_p^+)^k}{(k-1)!} + \frac{(m\rho_p^+)^m}{(m-1)!} \sum_{k=1}^{\infty} \frac{(\rho_p^+)^k}{\prod_{j=1}^k (1+j\tau/m)} \right].$$

For a customer of class p , the respective mean numbers are given by

$$E[\mathcal{N}_p] = E[\mathcal{N}_p^+] - E[\mathcal{N}_{p-1}^+] \quad ; \quad E[\mathcal{L}_p] = E[\mathcal{L}_p^+] - E[\mathcal{L}_{p-1}^+] \quad ; \quad E[\mathcal{S}_p] = E[\mathcal{S}_p^+] - E[\mathcal{S}_{p-1}^+].$$

From Little's theorem we obtain the mean time until departure $E[\mathcal{T}_p]$, the mean waiting time $E[\mathcal{W}_p]$, and the mean service time $E[\mathcal{H}_p]$ as follows.

$$E[\mathcal{N}_p] = \lambda_p E[\mathcal{T}_p] \quad ; \quad E[\mathcal{L}_p] = \lambda_p E[\mathcal{W}_p] \quad ; \quad E[\mathcal{S}_p] = \lambda_p E[\mathcal{H}_p].$$

3. First passage time to service preemption or completion from state k , $0 \leq k \leq m-1$

We focus on a tagged customer of class p in state k , signifying that there are k other customers who arrived before him, where $k = 0, 1, 2, \dots$. We first consider a finite-state birth-and-death process of state transitions for a tagged customer in state k , $0 \leq k \leq m-1$, in which he is being served. The service to this customer, with probability one, is eventually either preempted by a customer who arrives after him or completed without preemption.

3.1. Customer behavior until service preemption or completion

The state transition diagram for the discrete-time one-dimensional birth-and-death process modeling the behavior of a tagged customer in service is shown in Fig. 1. This process has m transient states $\{0, 1, 2, \dots, m-1\}$ and two absorbing states denoted by “Pr” (state m) and “Sr”, representing the service preemption and completion, respectively. State transitions occur when a customer arrives and the service to one of customers is completed. Therefore, the state transition probabilities and the LST of the DF for the time spent by a tagged customer in state k are given by

$$\alpha_{k,p-1} = \frac{k\mu}{\lambda_{p-1}^+ + (k+1)\mu} \quad ; \quad \beta_{k,p-1} = \frac{\mu}{\lambda_{p-1}^+ + (k+1)\mu} \quad ; \quad B_{k,p-1}^*(s) = \frac{\lambda_{p-1}^+ + (k+1)\mu}{s + \lambda_{p-1}^+ + (k+1)\mu}.$$

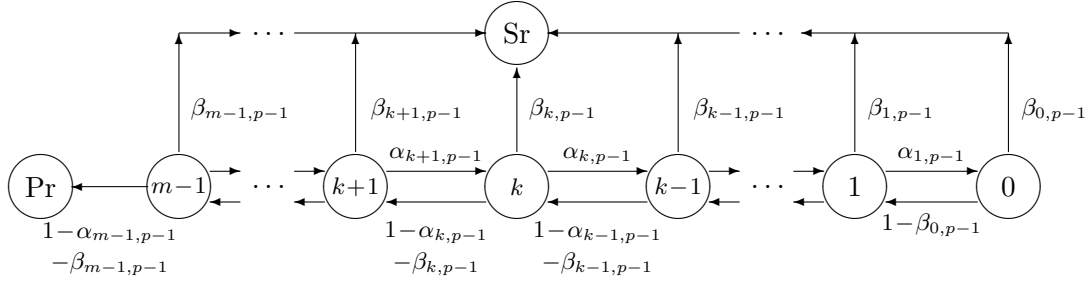


Figure 1: State transitions for a customer being served until service preemption or completion.

3.2. LST of the DF for the first passage time to service preemption or completion

The joint probability of reaching service preemption “Pr” and LST of the DF for the time until preemption from state k is denoted by $H_{k,p-1}^*(s, \text{Pr})$, $0 \leq k \leq m-1$. They satisfy

$$\begin{aligned} H_{0,p-1}^*(s, \text{Pr}) &= B_{0,p-1}^*(s)(1 - \beta_{0,p-1})H_{1,p-1}^*(s, \text{Pr}), \\ H_{k,p-1}^*(s, \text{Pr}) &= B_{k,p-1}^*(s)[\alpha_{k,p-1}H_{k-1,p-1}^*(s, \text{Pr}) + (1 - \alpha_{k,p-1} - \beta_{k,p-1})H_{k+1,p-1}^*(s, \text{Pr})] \\ &\quad 1 \leq k \leq m-2, \\ H_{m-1,p-1}^*(s, \text{Pr}) &= B_{m-1,p-1}^*(s)[\alpha_{m-1,p-1}H_{m-2,p-1}^*(s, \text{Pr}) + (1 - \alpha_{m-1,p-1} - \beta_{m-1,p-1})]. \end{aligned}$$

The joint probability of reaching service completion “Sr” and LST of the DF for the time until service completion from state k is denoted by $H_{k,p-1}^*(s, \text{Sr})$, $0 \leq k \leq m-1$. They satisfy

$$\begin{aligned} H_{0,p-1}^*(s, \text{Sr}) &= B_{0,p-1}^*(s)[\beta_{0,p-1} + (1 - \beta_{0,p-1})H_{1,p-1}^*(s, \text{Sr})], \\ H_{k,p-1}^*(s, \text{Sr}) &= B_{k,p-1}^*(s)[\alpha_{k,p-1}H_{k-1,p-1}^*(s, \text{Sr}) + \beta_{k,p-1} + (1 - \alpha_{k,p-1} - \beta_{k,p-1})H_{k+1,p-1}^*(s, \text{Sr})] \\ &\quad 1 \leq k \leq m-2, \\ H_{m-1,p-1}^*(s, \text{Sr}) &= B_{m-1,p-1}^*(s)[\alpha_{m-1,p-1}H_{m-2,p-1}^*(s, \text{Sr}) + \beta_{m-1,p-1}]. \end{aligned}$$

The solution to the above sets of equations is obtained in terms of $\{h_{k,p-1}^*(s); 0 \leq k \leq m\}$ as

$$H_{k,p-1}^*(s, \text{Pr}) = \frac{h_{k,p-1}^*(s)}{h_{m,p-1}^*(s)} \quad ; \quad H_{k,p-1}^*(s, \text{Sr}) = \frac{\mu}{s + \mu} \left[1 - \frac{h_{k,p-1}^*(s)}{h_{m,p-1}^*(s)} \right] \quad 0 \leq k \leq m,$$

with $H_{m,p-1}^*(s, \text{Pr}) \equiv 1$ and $H_{m,p-1}^*(s, \text{Sr}) \equiv 0$. Then, $h_{k,p-1}^*(s)$ is a k th degree polynomial in s , the coefficient of s^k being $(1/\lambda_{p-1}^+)^k$, $1 \leq k \leq m$, that satisfies the following recurrence relations:

$$\begin{aligned} h_{0,p-1}^*(s) &= 1 \quad ; \quad s + \lambda_{p-1}^+ + \mu = \lambda h_{1,p-1}^*(s), \\ [s + \lambda_{p-1}^+ + (k+1)\mu] h_{k,p-1}^*(s) &= k\mu h_{k-1,p-1}^*(s) + \lambda_{p-1}^+ h_{k+1,p-1}^*(s) \quad 1 \leq k \leq m-1. \end{aligned}$$

4. First passage time to service resumption or abandonment from state k , $k \geq m$

We next consider another infinite-state birth-and-death process of state transitions for a tagged customer in state k , $k \geq m$, in which he is staying in the waiting room. With probability one, this customer, eventually, either is called in to resume his service or abandons waiting.

4.1. Customer behavior until service resumption or abandonment

The state transition diagram for the discrete-time one-dimensional birth-and-death process modeling the behavior of a tagged customer in the waiting room is shown in Fig. 2. The process has an infinite number of transient states $\{m, m+1, \dots\}$ and two absorbing states denoted by “Rs” (state $m-1$) and “Ab”, representing service resumption and abandonment, respectively. State transitions occur when a customer arrives, the service to one of customers is completed, and a waiting customer before the tagged one abandons waiting. Therefore, the state transition probabilities and the LST of the DF for the time spent by a tagged customer in state k are given by

$$\begin{aligned} \alpha'_{k,p-1} &= \frac{m\mu + (k-m)\theta}{\lambda_{p-1}^+ + m\mu + (k+1-m)\theta} \quad ; \quad \beta'_{k,p-1} = \frac{\theta}{\lambda_{p-1}^+ + m\mu + (k+1-m)\theta}, \\ B'_{k,p-1}(s) &= \frac{\lambda_{p-1}^+ + m\mu + (k+1-m)\theta}{s + \lambda_{p-1}^+ + m\mu + (k+1-m)\theta}. \end{aligned}$$

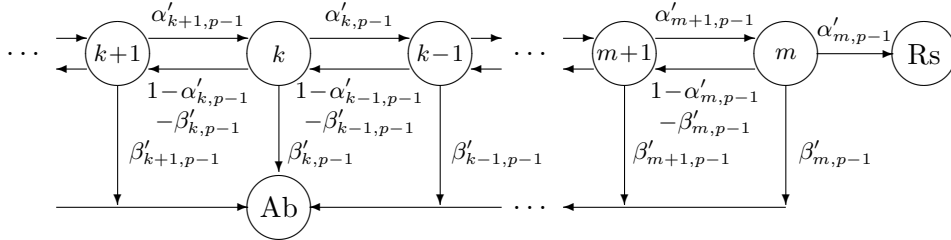


Figure 2: State transitions for a waiting customer until service resumption or abandonment.

4.2. LST of the DF for the first passage time to service resumption or abandonment

The joint probability of reaching service resumption “Rs” and LST of the DF for the time until resumption from state k is denoted by $W_{k,p-1}^*(s, \text{Rs})$, $k \geq m$. They satisfy

$$\begin{aligned} W_{m,p-1}^*(s, \text{Rs}) &= B'_{m,p-1}(s)[\alpha'_{m,p-1} + (1 - \alpha'_{m,p-1} - \beta'_{m,p-1})W_{m+1,p-1}^*(s, \text{Rs})], \\ W_{k,p-1}^*(s, \text{Rs}) &= B'_{k,p-1}(s)[\alpha'_{k,p-1}W_{k-1,p-1}^*(s, \text{Rs}) + (1 - \alpha'_{k,p-1} - \beta'_{k,p-1})W_{k+1,p-1}^*(s, \text{Rs})] \\ &\quad k \geq m+1. \end{aligned}$$

The joint probability of reaching abandonment “Ab” and LST of the DF for the time until abandonment from state k is denoted by $W_{k,p-1}^*(s, \text{Ab})$, $k \geq m$. They satisfy

$$\begin{aligned} W_{m,p-1}^*(s, \text{Ab}) &= B'_{m,p-1}(s)[\beta'_{m,p-1} + (1 - \alpha'_{m,p-1} - \beta'_{m,p-1})W_{m+1,p-1}^*(s, \text{Ab})], \\ W_{k,p-1}^*(s, \text{Ab}) &= B'_{k,p-1}(s)[\alpha'_{k,p-1}W_{k-1,p-1}^*(s, \text{Ab}) + \beta'_{k,p-1} + (1 - \alpha'_{k,p-1} - \beta'_{k,p-1})W_{k+1,p-1}^*(s, \text{Ab})] \\ &\quad k \geq m+1. \end{aligned}$$

The solution to the above sets of equations is obtained in terms of $\{G_{k,p-1}^*(s); k \geq m\}$ as

$$W_{k,p-1}^*(s, \text{Rs}) = G_{k,p-1}^*(s + \theta) \quad ; \quad W_{k,p-1}^*(s, \text{Ab}) = \frac{\theta}{s + \theta} [1 - G_{k,p-1}^*(s + \theta)] \quad k \geq m,$$

where $G_{k,p-1}^*(s)$ is the LST of the DF for the *busy period* started with k ($\geq m$) customers in an M/M/ m queue with arrival rate λ_{p-1}^+ and service rate $m\mu$. It satisfies

$$\begin{aligned} (s + \lambda_{p-1}^+ + m\mu)G_{m,p-1}^*(s) &= \lambda_{p-1}^+ G_{m+1,p-1}^*(s) + m\mu, \\ [s + \lambda_{p-1}^+ + m\mu + (k - m)\theta]G_{k,p-1}^*(s) &= [m\mu + (k - m)\theta]G_{k-1,p-1}^*(s) + \lambda G_{k+1,p-1}^*(s) \quad k \geq m + 1. \end{aligned}$$

According to Iravani and Balcioglu [3], we have

$$G_{k,p-1}^*(s) = \frac{\frac{m\mu}{s + m\mu} + \sum_{i=1}^{\infty} (-1)^i \psi_{i,k-m}(\lambda_{p-1}^+/\theta) \left[\prod_{j=0}^{i-1} \left(1 - \frac{m\mu}{s + m\mu + j\theta} \right) \right] \frac{m\mu}{s + m\mu + i\theta}}{1 + \sum_{i=1}^{\infty} \frac{(\lambda_{p-1}^+/\theta)^i}{i!} \prod_{j=0}^{i-1} \left(1 - \frac{m\mu}{s + m\mu + j\theta} \right)} \quad k \geq m,$$

where

$$\psi_{i,k}(x) := \sum_{j=\max\{0, i-k\}}^i \frac{(-x)^j}{j!} \binom{k}{i-j} \quad i \geq 1, \quad k \geq 0$$

with $\psi_{0,k}(x) \equiv 1$ for $k \geq 0$. Thus we get

$$G_{k,p-1}^*(\theta) = \sum_{i=0}^{\infty} \frac{i!(-\tau/m)^i \psi_{i,k-m}(\lambda_{p-1}^+/\theta)}{\prod_{j=1}^{i+1} (1 + j\tau/m)} \bigg/ \sum_{i=1}^{\infty} \frac{(\rho_{p-1}^+)^i}{\prod_{j=0}^i (1 + j\tau/m)}.$$

5. Distribution of the waiting and service time until departure

We are now in a position to consider the distribution of the time until departure (either abandonment or service completion) for the tagged customer in the combined birth-and-death processes whose state transitions are shown in Figs. 1 and 2. We note that state “Pr” in Fig. 1 is common to state m in Fig. 2, whereas state “Rs” in Fig. 2 is common to state $m - 1$ in Fig. 1. The detailed derivation can be found in [9].

5.1. Waiting and service time until abandonment

The joint probability of abandonment and LST of the DF for the waiting and service time until abandonment from state k is given by

$$\begin{aligned} \mathcal{T}_{k,p-1}^*(s, s', \text{Ab}) &= \frac{H_{k,p-1}^*(s', \text{Pr})W_{m,p-1}^*(s, \text{Ab})}{1 - W_{m,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s', \text{Pr})} \\ &= \frac{\theta}{s + \theta} \cdot \frac{h_{k,p-1}^*(s')[1 - G_{m,p-1}^*(s + \theta)]}{h_{m,p-1}^*(s') - h_{m-1,p-1}^*(s')G_{m,p-1}^*(s + \theta)} \quad 1 \leq k \leq m - 1, \\ \mathcal{T}_{k,p-1}^*(s, s', \text{Ab}) &= W_{k,p-1}^*(s, \text{Ab}) + \frac{W_{k,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s', \text{Pr})W_{m,p-1}^*(s, \text{Ab})}{1 - W_{m,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s', \text{Pr})} \\ &= \frac{\theta}{s + \theta} \left\{ 1 - \frac{[h_{m,p-1}^*(s') - h_{m-1,p-1}^*(s')]G_{k,p-1}^*(s + \theta)}{h_{m,p-1}^*(s') - h_{m-1,p-1}^*(s')G_{m,p-1}^*(s + \theta)} \right\} \quad k \geq m. \end{aligned}$$

The joint probability of abandonment and LST of the DF for the time until abandonment from state k is given by

$$\begin{aligned}
\mathcal{T}_{k,p-1}^*(s, \text{Ab}) &= \frac{H_{k,p-1}^*(s, \text{Pr})W_{m,p-1}^*(s, \text{Ab})}{1 - W_{m,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s, \text{Pr})} \\
&= \frac{\theta}{s + \theta} \cdot \frac{h_{k,p-1}^*(s)[1 - G_{m,p-1}^*(s + \theta)]}{h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)G_{m,p-1}^*(s + \theta)} \quad 1 \leq k \leq m-1, \\
\mathcal{T}_{k,p-1}^*(s, \text{Ab}) &= W_{k,p-1}^*(s, \text{Ab}) + \frac{W_{k,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s, \text{Pr})W_{m,p-1}^*(s, \text{Ab})}{1 - W_{m,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s, \text{Pr})} \\
&= \frac{\theta}{s + \theta} \left\{ 1 - \frac{[h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)]G_{k,p-1}^*(s + \theta)}{h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)G_{m,p-1}^*(s + \theta)} \right\} \quad k \geq m.
\end{aligned}$$

Similarly, we have the joint probability of abandonment and LST of the DF for the waiting time until abandonment from state k

$$\mathcal{W}_{k,p-1}^*(s, \text{Ab}) = \begin{cases} \frac{\theta}{s + \theta} \cdot \frac{h_{k,p-1}^*(0)[1 - G_{m,p-1}^*(s + \theta)]}{h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)G_{m,p-1}^*(s + \theta)} & 0 \leq k \leq m-1, \\ \frac{\theta}{s + \theta} \left\{ 1 - \frac{[h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)]G_{k,p-1}^*(s + \theta)}{h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)G_{m,p-1}^*(s + \theta)} \right\} & k \geq m \end{cases}$$

and the joint probability of abandonment and LST of the DF for the service time until abandonment from state k

$$\mathcal{H}_{k,p-1}^*(s, \text{Ab}) = \begin{cases} \frac{h_{k,p-1}^*(s)[1 - G_{m,p-1}^*(\theta)]}{h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)G_{m,p-1}^*(\theta)} & 0 \leq k \leq m-1, \\ 1 - \frac{[h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)]G_{k,p-1}^*(\theta)}{h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)G_{m,p-1}^*(\theta)} & k \geq m. \end{cases}$$

The probability of abandonment from state k is given by

$$\mathcal{P}_{k,p-1}\{\text{Ab}\} = \begin{cases} \frac{h_{k,p-1}^*(0)[1 - G_{m,p-1}^*(\theta)]}{h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)G_{m,p-1}^*(\theta)} & 0 \leq k \leq m-1, \\ 1 - \frac{[h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)]G_{k,p-1}^*(\theta)}{h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)G_{m,p-1}^*(\theta)} & k \geq m. \end{cases}$$

5.2. Waiting and service time until service completion

The joint probability of service completion and LST of the DF for the waiting and service time until service completion from state k is given by

$$\begin{aligned}
\mathcal{T}_{k,p-1}^*(s, s', \text{Sr}) &= H_{k,p-1}^*(s', \text{Sr}) + \frac{H_{k,p-1}^*(s', \text{Pr})W_{m,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s', \text{Sr})}{1 - W_{m,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s', \text{Pr})} \\
&= \frac{\mu}{s' + \mu} \left\{ 1 - \frac{h_{k,p-1}^*(s')[1 - G_{m,p-1}^*(s + \theta)]}{h_{m,p-1}^*(s') - h_{m-1,p-1}^*(s')G_{m,p-1}^*(s + \theta)} \right\} \quad 0 \leq k \leq m-1, \\
\mathcal{T}_{k,p-1}^*(s, s', \text{Sr}) &= \frac{W_{k,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s', \text{Sr})}{1 - W_{m,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s', \text{Pr})} \\
&= \frac{\mu}{s' + \mu} \cdot \frac{[h_{m,p-1}^*(s') - h_{m-1,p-1}^*(s')]G_{k,p-1}^*(s + \theta)}{h_{m,p-1}^*(s') - h_{m-1,p-1}^*(s')G_{m,p-1}^*(s + \theta)} \quad k \geq m.
\end{aligned}$$

The joint probability of service completion and LST of the DF for the time until service completion from state k is given by

$$\begin{aligned}
\mathcal{T}_{k,p-1}^*(s, \text{Sr}) &= H_{k,p-1}^*(s, \text{Sr}) + \frac{H_{k,p-1}^*(s, \text{Pr})W_{m,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s, \text{Sr})}{1 - W_{m,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s, \text{Pr})} \\
&= \frac{\mu}{s + \mu} \left\{ 1 - \frac{h_{k,p-1}^*(s)[1 - G_{m,p-1}^*(s + \theta)]}{h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)G_{m,p-1}^*(s + \theta)} \right\} \quad 0 \leq k \leq m-1, \\
\mathcal{T}_{k,p-1}^*(s, \text{Sr}) &= \frac{W_{k,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s, \text{Sr})}{1 - W_{m,p-1}^*(s, \text{Rs})H_{m-1,p-1}^*(s, \text{Pr})} \\
&= \frac{\mu}{s + \mu} \cdot \frac{[h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)]G_{k,p-1}^*(s + \theta)}{h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)G_{m,p-1}^*(s + \theta)} \quad k \geq m.
\end{aligned}$$

Similarly, we have the joint probability of service completion and LST of the DF for the waiting time until service completion from state k

$$\mathcal{W}_{k,p-1}^*(s, \text{Sr}) = \begin{cases} 1 - \frac{h_{k,p-1}^*(0)[1 - G_{m,p-1}^*(s + \theta)]}{h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)G_{m,p-1}^*(s + \theta)} & 0 \leq k \leq m-1, \\ \frac{[h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)]G_{k,p-1}^*(s + \theta)}{h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)G_{m,p-1}^*(s + \theta)} & k \geq m \end{cases}$$

and the joint probability of service completion and LST of the DF for the service time until service completion from state k

$$\mathcal{H}_{k,p-1}^*(s, \text{Sr}) = \begin{cases} \frac{\mu}{s + \mu} \left\{ 1 - \frac{h_{k,p-1}^*(s)[1 - G_{m,p-1}^*(\theta)]}{h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)G_{m,p-1}^*(\theta)} \right\} & 0 \leq k \leq m-1, \\ \frac{\mu}{s + \mu} \cdot \frac{[h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)]G_{k,p-1}^*(\theta)}{h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)G_{m,p-1}^*(\theta)} & k \geq m. \end{cases}$$

The probability of service completion from state k is given by

$$\mathcal{P}_{k,p-1}\{\text{Sr}\} = \begin{cases} 1 - \frac{h_{k,p-1}^*(0)[1 - G_{m,p-1}^*(\theta)]}{h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)G_{m,p-1}^*(\theta)} & 0 \leq k \leq m-1, \\ \frac{[h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)]G_{k,p-1}^*(\theta)}{h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)G_{m,p-1}^*(\theta)} & k \geq m. \end{cases}$$

5.3. Waiting and service time until departure

The unconditional joint LST of the DF for the waiting and service time to departure (either by abandonment or service completion) from state k is given by

$$\begin{aligned}
\mathcal{T}_{k,p-1}^*(s, s') &= \mathcal{T}_{k,p-1}^*(s, s', \text{Ab}) + \mathcal{T}_{k,p-1}^*(s, s', \text{Sr}) = 1 - \frac{s}{\theta} \mathcal{T}_{k,p-1}^*(s, s', \text{Ab}) - \frac{s'}{\mu} \mathcal{T}_{k,p-1}^*(s, s', \text{Sr}) \\
&= \begin{cases} \frac{\mu}{s' + \mu} + \left(\frac{\theta}{s + \theta} - \frac{\mu}{s' + \mu} \right) \frac{h_{k,p-1}^*(s')[1 - G_{m,p-1}^*(s + \theta)]}{h_{m,p-1}^*(s') - h_{m-1,p-1}^*(s')G_{m,p-1}^*(s + \theta)} & 0 \leq k \leq m-1, \\ \frac{\theta}{s + \theta} + \left(\frac{\mu}{s' + \mu} - \frac{\theta}{s + \theta} \right) \frac{[h_m^*(s') - h_{m-1}^*(s')]G_k^*(s + \theta)}{h_{m,p-1}^*(s') - h_{m-1,p-1}^*(s')G_{m,p-1}^*(s + \theta)} & k \geq m. \end{cases}
\end{aligned}$$

Then the unconditional LST of the DF for the time to departure from state k is given by

$$\begin{aligned} \mathcal{T}_{k,p-1}^*(s) &= \mathcal{T}_{k,p-1}^*(s, \text{Ab}) + \mathcal{T}_{k,p-1}^*(s, \text{Sr}) = 1 - \frac{s}{\theta} \mathcal{T}_{k,p-1}^*(s, \text{Ab}) - \frac{s}{\mu} \mathcal{T}_{k,p-1}^*(s, \text{Sr}) \\ &= \begin{cases} \frac{\mu}{s+\mu} + \left(\frac{\theta}{s+\theta} - \frac{\mu}{s+\mu} \right) \frac{h_{k,p-1}^*(s)[1 - G_{m,p-1}^*(s+\theta)]}{h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)G_{m,p-1}^*(s+\theta)} & 0 \leq k \leq m-1, \\ \frac{\theta}{s+\theta} + \left(\frac{\mu}{s+\mu} - \frac{\theta}{s+\theta} \right) \frac{[h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)]G_{k,p-1}^*(s+\theta)}{h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)G_{m,p-1}^*(s+\theta)} & k \geq m. \end{cases} \end{aligned}$$

Similarly, we have the LST of the DF for the waiting time until departure from state k

$$\begin{aligned} \mathcal{W}_{k,p-1}^*(s) &= \mathcal{W}_{k,p-1}^*(s, \text{Ab}) + \mathcal{W}_{k,p-1}^*(s, \text{Sr}) \\ &= \begin{cases} 1 - \frac{s}{s+\theta} \cdot \frac{h_{k,p-1}^*(0)[1 - G_{m,p-1}^*(s+\theta)]}{h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)G_{m,p-1}^*(s+\theta)} & 0 \leq k \leq m-1, \\ \frac{\theta}{s+\theta} + \frac{s}{s+\theta} \cdot \frac{[h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)]G_{k,p-1}^*(s+\theta)}{h_{m,p-1}^*(0) - h_{m-1,p-1}^*(0)G_{m,p-1}^*(s+\theta)} & k \geq m \end{cases} \end{aligned}$$

and the LST of the DF for the service time until departure from state k

$$\begin{aligned} \mathcal{H}_{k,p-1}^*(s) &= \mathcal{H}_{k,p-1}^*(s, \text{Ab}) + \mathcal{H}_{k,p-1}^*(s, \text{Sr}) \\ &= \begin{cases} \frac{\mu}{s+\mu} + \frac{s}{s+\mu} \cdot \frac{h_{k,p-1}^*(s)[1 - G_{m,p-1}^*(\theta)]}{h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)G_{m,p-1}^*(\theta)} & 0 \leq k \leq m-1, \\ 1 - \frac{s}{s+\mu} \cdot \frac{[h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)]G_{k,p-1}^*(\theta)}{h_{m,p-1}^*(s) - h_{m-1,p-1}^*(s)G_{m,p-1}^*(\theta)} & k \geq m. \end{cases} \end{aligned}$$

We then obtain the following relations between the moments for state $k \geq 0$

$$\begin{aligned} E[\mathcal{W}_{k,p-1}] &= \frac{1}{\theta} \mathcal{P}_{k,p-1}\{\text{Ab}\} \quad ; \quad E[\mathcal{H}_{k,p-1}] = \frac{1}{\mu} \mathcal{P}_{k,p-1}\{\text{Sr}\}, \\ E[\mathcal{W}_{k,p-1}^\ell] &= \frac{\ell}{\theta} E[\mathcal{W}_{k,p-1}^{\ell-1}, \text{Ab}] \quad ; \quad E[\mathcal{H}_{k,p-1}^\ell] = \frac{\ell}{\mu} E[\mathcal{H}_{k,p-1}^{\ell-1}, \text{Sr}] \quad \ell = 2, 3, \dots, \\ E[\mathcal{W}_{k,p-1}^\ell \mathcal{H}_{k,p-1}^{\ell'}] &= \frac{\ell}{\theta} E[\mathcal{W}_{k,p-1}^{\ell-1} \mathcal{H}_{k,p-1}^{\ell'}, \text{Ab}] + \frac{\ell'}{\mu} E[\mathcal{W}_{k,p-1}^\ell \mathcal{H}_{k,p-1}^{\ell'-1}, \text{Sr}] \quad \ell, \ell' = 1, 2, \dots, \\ E[\mathcal{T}_{k,p-1}] &= E[\mathcal{W}_{k,p-1}] + E[\mathcal{H}_{k,p-1}] = \frac{1}{\theta} \mathcal{P}_{k,p-1}\{\text{Ab}\} + \frac{1}{\mu} \mathcal{P}_{k,p-1}\{\text{Sr}\}, \\ E[\mathcal{T}_{k,p-1}^\ell] &= \frac{\ell}{\theta} E[\mathcal{T}_{k,p-1}^{\ell-1}, \text{Ab}] + \frac{\ell}{\mu} E[\mathcal{T}_{k,p-1}^{\ell-1}, \text{Sr}] \quad \ell = 2, 3, \dots \end{aligned}$$

6. Waiting and Service Time of an arriving customer

By unconditioning the number of competing customers found in the system upon arrival, we obtain the LST of the DF for the waiting and service time of an arbitrary customer of class p . The way of unconditioning depends on the service discipline for prioritizing the customers of the same class as mentioned in Section 1.

6.1. FCFS for customers of the same class

Those customers of class $1, 2, \dots, p$ found in the system by an arriving tagged customer of class p are served before him. Those customers of class $1, 2, \dots, p-1$ who arrive after the tagged customer of class p are served before him. Therefore we obtain

$$\begin{aligned} \mathcal{W}_p^*(s, \text{Sr}) &= \sum_{k=0}^{\infty} Q_{k,p}^+ \mathcal{W}_{k,p-1}^*(s, \text{Sr}) \quad ; \quad \mathcal{W}_p^*(s, \text{Ab}) = \sum_{k=0}^{\infty} Q_{k,p}^+ \mathcal{W}_{k,p-1}^*(s, \text{Ab}), \\ \mathcal{H}_p^*(s, \text{Sr}) &= \sum_{k=0}^{\infty} Q_{k,p}^+ \mathcal{H}_{k,p-1}^*(s, \text{Sr}) \quad ; \quad \mathcal{H}_p^*(s, \text{Ab}) = \sum_{k=0}^{\infty} Q_{k,p}^+ \mathcal{H}_{k,p-1}^*(s, \text{Ab}), \end{aligned}$$

$$\begin{aligned}
\mathcal{T}_p^*(s, \text{Sr}) &= \sum_{k=0}^{\infty} Q_{k,p}^+ \mathcal{T}_{k,p-1}^*(s, \text{Sr}) & ; & \quad \mathcal{T}_p^*(s, \text{Ab}) = \sum_{k=0}^{\infty} Q_{k,p}^+ \mathcal{T}_{k,p-1}^*(s, \text{Ab}), \\
\mathcal{P}_p\{\text{Sr}\} &= \sum_{k=0}^{\infty} Q_{k,p}^+ \mathcal{P}_{k,p-1}\{\text{Sr}\} & ; & \quad \mathcal{P}_p\{\text{Ab}\} = \sum_{k=0}^{\infty} Q_{k,p}^+ \mathcal{P}_{k,p-1}\{\text{Ab}\}, \\
\mathcal{W}_p^*(s) &= \sum_{k=0}^{\infty} Q_{k,p}^+ \mathcal{W}_{k,p-1}^*(s) & ; & \quad \mathcal{H}_p^*(s) = \sum_{k=0}^{\infty} Q_{k,p}^+ \mathcal{H}_{k,p-1}^*(s) & ; & \quad \mathcal{T}_p^*(s) = \sum_{k=0}^{\infty} Q_{k,p}^+ \mathcal{T}_{k,p-1}^*(s),
\end{aligned}$$

where the probabilities $\{Q_{k,p}^+; k \geq 0\}$ are given in Section 2.

6.2. PR-LCFS for customers of the same class

Those customers of class $1, 2, \dots, p-1$ found in the system by an arriving tagged customer of class p are served before him. Those customers of class $1, 2, \dots, p$ who arrive after the tagged customer of class p are served before him. Therefore we obtain

$$\begin{aligned}
\mathcal{W}_p^*(s, \text{Sr}) &= \sum_{k=0}^{\infty} Q_{k,p-1}^+ \mathcal{W}_{k,p}^*(s, \text{Sr}) & ; & \quad \mathcal{W}_p^*(s, \text{Ab}) = \sum_{k=0}^{\infty} Q_{k,p-1}^+ \mathcal{W}_{k,p}^*(s, \text{Ab}), \\
\mathcal{H}_p^*(s, \text{Sr}) &= \sum_{k=0}^{\infty} Q_{k,p-1}^+ \mathcal{H}_{k,p}^*(s, \text{Sr}) & ; & \quad \mathcal{H}_p^*(s, \text{Ab}) = \sum_{k=0}^{\infty} Q_{k,p-1}^+ \mathcal{H}_{k,p}^*(s, \text{Ab}), \\
\mathcal{T}_p^*(s, \text{Sr}) &= \sum_{k=0}^{\infty} Q_{k,p-1}^+ \mathcal{T}_{k,p}^*(s, \text{Sr}) & ; & \quad \mathcal{T}_p^*(s, \text{Ab}) = \sum_{k=0}^{\infty} Q_{k,p-1}^+ \mathcal{T}_{k,p}^*(s, \text{Ab}), \\
\mathcal{P}_p\{\text{Sr}\} &= \sum_{k=0}^{\infty} Q_{k,p-1}^+ \mathcal{P}_{k,p}\{\text{Sr}\} & ; & \quad \mathcal{P}_p\{\text{Ab}\} = \sum_{k=0}^{\infty} Q_{k,p-1}^+ \mathcal{P}_{k,p}\{\text{Ab}\}, \\
\mathcal{W}_p^*(s) &= \sum_{k=0}^{\infty} Q_{k,p-1}^+ \mathcal{W}_{k,p}^*(s) & ; & \quad \mathcal{H}_p^*(s) = \sum_{k=0}^{\infty} Q_{k,p-1}^+ \mathcal{H}_{k,p}^*(s) & ; & \quad \mathcal{T}_p^*(s) = \sum_{k=0}^{\infty} Q_{k,p-1}^+ \mathcal{T}_{k,p}^*(s),
\end{aligned}$$

where the probabilities $\{Q_{k,p-1}^+; k \geq 0\}$ are given by replacing p with $p-1$ in Section 2.

For both disciplines described above, we have the relation

$$E[\mathcal{W}_p] = \frac{1}{\theta} \mathcal{P}_p\{\text{Ab}\} = \frac{E[\mathcal{L}_p]}{\lambda_p} \quad ; \quad E[\mathcal{H}_p] = \frac{1}{\mu} \mathcal{P}_p\{\text{Sr}\} = \frac{E[\mathcal{S}_p]}{\lambda_p}$$

and an instance of Little's theorem

$$E[\mathcal{T}_p] = \frac{1}{\theta} \mathcal{P}_p\{\text{Ab}\} + \frac{1}{\mu} \mathcal{P}_p\{\text{Sr}\} = \frac{E[\mathcal{N}_p]}{\lambda_p},$$

where $E[\mathcal{L}_p]$, $E[\mathcal{S}_p]$, and $E[\mathcal{N}_p]$ are the mean number of customers of class p present in the waiting room, the service facility, and the entire system, respectively, which are given in Section 2.

References

- [1] I. Brosh, Preemptive priority assignment in multichannel systems, *Operations Research*, Vol.17, No.3, pp.526–535, May–June 1969.
- [2] J. P. Buzen and A. B. Bondi, The response times of priority classes under preemptive resume in M/M/m queues, *Operations Research*, Vol.31, No.3, pp.456–465, May–June 1983.
- [3] F. Iravani and B. Balcioglu, On priority queues with impatient customers, *Queueing Systems*, Vol.58, No.4, pp.239–260, May 2008.
- [4] M. Segal, A multiserver system with preemptive priorities, *Operations Research*, Vol.18, No.2, pp.316–323, March–April 1970.
- [5] H. Takagi, Times to service completion and abandonment in the M/M/m preemptive LCFS queue with impatient customers, *QTNA '16, Proceedings of the 11th International Conference on Queueing Theory and Network Applications*, held on December 13–15, 2016, in Wellington, New Zealand (Best paper award). ACM, ISBN 978-1-4503-4842-3/16/12.

Table 1: Unified framework for M/M/m preemptive-resume priority queues with different disciplines (FCFS and PR-LCFS) for customers of the same priority class.

(a) Customers who compete with a tagged customer of class p arriving in state k

	Customers present at arrival	Customers who come afterwards
FCFS	class 1, 2, ..., p for $k \geq 0$	class 1, 2, ..., $p - 1$ for $k \geq 0$
PR-LCFS	class 1, 2, ..., $p - 1$ for $k \geq 0$	class 1, 2, ..., p for $k \geq 0$

(b) Specific notation for the three disciplines

	FCFS	PR-LCFS
Competing customers who are present in the system at an arbitrary time		
Arrival rate and traffic intensity	$\lambda_p^+, \rho_p^+ := \lambda_p^+ / (m\mu)$	$\lambda_{p-1}^+, \rho_{p-1}^+ := \lambda_{p-1}^+ / (m\mu)$
Number of customers	$\mathcal{N}_p^+$	$\mathcal{N}_{p-1}^+$
Probability distribution, $k \geq 0$	$Q_{k,p}^+ := P\{\mathcal{N}_p^+ = k\}$	$Q_{k,p-1}^+ := P\{\mathcal{N}_{p-1}^+ = k\}$
Competing customers who arrive after a tagged customer in state k , $0 \leq k \leq m - 1$		
Arrival rate	λ_{p-1}^+	λ_p^+
State transition rate	$\alpha_{k,p-1}, \beta_{k,p-1}$	$\alpha_{k,p}, \beta_{k,p}$
State residence time	$B_{k,p-1}^*(s)$	$B_{k,p}^*(s)$
Auxiliary function	$h_{k,p-1}^*(s)$	$h_{k,p}^*(s)$
Service time until preemption	$H_{k,p-1}^*(s, \text{Pr})$	$H_{k,p}^*(s, \text{Pr})$
Service time until completion	$H_{k,p-1}^*(s, \text{Sr})$	$H_{k,p}^*(s, \text{Sr})$
Competing customers who arrive after a tagged customer in state k , $k \geq m$		
Arrival rate	λ_{p-1}^+	λ_p^+
State transition rate	$\alpha'_{k,p-1}, \beta'_{k,p-1}$	$\alpha'_{k,p}, \beta'_{k,p}$
State residence time	$B'_{k,p-1}(s)$	$B'_{k,p}(s)$
Busy period duration	$G_{k,p-1}^*(s)$	$G_{k,p}^*(s)$
Waiting time until resumption	$W_{k,p-1}^*(s, \text{Rs})$	$W_{k,p}^*(s, \text{Rs})$
Waiting time until abandonment	$W_{k,p-1}^*(s, \text{Ab})$	$W_{k,p}^*(s, \text{Ab})$

- [6] H. Takagi, Analysis of the response and waiting times in the M/M/m LCFS preemptive-resume priority queue, *International Journal of Pure and Applied Mathematics*, Vol.109, No.2, pp.325–370, September 2016.
- [7] H. Takagi, Distribution of the times to service completion and abandonment in the M/M/m preemptive LCFS queue with impatient customers, 待ち行列シンポジウム「確率モデルとその応用」報文集, pp.54–63, 東京理科大学森戸記念館 (東京都新宿区), 2017 年 1 月 19–21 日.
- [8] H. Takagi, Unified and refined analysis of the response time and waiting time in the M/M/m FCFS preemptive-resume priority queue, *Journal of Industrial and Management Optimization*, Vo.13, No.4, pp.1945–1973, October 2017.
- [9] H. Takagi, Distribution of the waiting and service time in an M/M/m preemptive-resume LCFS queue with impatient customers, *International Journal of Pure and Applied Mathematics*, Vol. 116, No.2, pp.501–545, October 2017.
- [10] H. Takagi, Times until service completion and abandonment in an M/M/m preemptive-resume LCFS queue with impatient customers, to appear in *Journal of Industrial and Management Optimization*.
- [11] A. G. Tatashev, Calculation of the distribution of the waiting time in a multiple-channel queueing system with fixed priorities, *Engineering Cybernetics*, Vol.22, No.1, pp.59–62, 1984 (Originally published in *Tekhnicheskaya Kibernetika*, No.6, pp.163–166, 1983).
- [12] S. Zeltyn, Z. Feldman, and S. Wasserkrug, Waiting and sojourn times in a multi-server queue with mixed priorities, *Queueing Systems*, Vol.61, No.4, pp.305–328, March 2009.

Acknowledgment

This work is supported by the Grant-in-Aid for Scientific Research (C) No. 17K00435 from the Japan Society for the Promotion of Science (JSPS) in 2017.

A Toy Model of Preferentially-Attached Networks with Break-Ups

Hiroshi Toyoizumi

Takahiro Shimomura

Graduate School of Accounting and Department of Applied Mathematics
Waseda University

Abstract A toy model of break-ups in the scale-free network are studied. Break-ups can be found in many social networks, such as political factions, class division in society, interest groups in the social media. Also, the evolution of the fragments draws more interest in the marketing to specific interest groups on social media such as twitter and facebook. We analyze a toy model to give a scale-free network with possible break-ups constructed by the preferential attachment. Our toy model can be easily generated by random-pairing of edges in the network. We study the basic features such as the dynamics of the number of break-ups, and the size of the fragments. We also discuss the extension of our toy model to $M/M/\infty$ stochastic dynamic networks with joins and leaves.

1. Introduction

There have been many analysis on complex networks in different fields such as statistical physics, communication theory, and probability theory (see review on probabilistic treatment [1] and interesting critical phenomena in statistical physics [2]). Standard network models that have specific statistical features which capture real world are important, for example, to study spread dynamics on networks [3, 4, 5, 6, 7]. One such standard network model is the configuration model which pairs each endpoints randomly [8, 9]. The configuration model is a useful tool for generating uncorrelated networks with a specific degree distribution [10, 11], and they were used in studying epidemics [4, 5, 6, 7].

Most of the current network models including configuration models are focused on generating connected graphs [12], which is ideal for studying spread dynamics. However, the focus is shifting to the segregation or how closed communities arise in the network [13, 14]. We propose a simple toy model of network with preferential attachment to study the segregation phenomena. In scale-free networks both in the fluid model and discrete model [15, 1, 16], nodes join the network and they distribute links to the existing nodes preferentially with their degree, and the network evolves progressively. We use a discrete type of scale-free networks [16] so that it may have break-ups. We show that a simple random walk model to study the detail of the break-ups, and derive the basic properties such as the break-up probability and the size of the fragments.

Further, we extend our model to dynamic network with Poisson arrival steam of individuals who leave the network spending independent exponential sojourn times.

2. Preferential Attachment with Possible Break-ups

Consider an evolving social network with arriving individuals. The individuals in the social network are connected with pairwise relationships called links. At the time of its arrival, a new comer will extend new links to the existing individuals based on the current links inside the network. More specifically, the new comer selects and establishes new links with existing individuals stochastically and more preferentially if they have more links. In addition to this preferential attachment, the new comer may select to be independent and start a new fraction in the network. The fraction founded by one individual eventually obtains followers and grows, by the mechanism of the stochastic preferential attachment (see Figure 1).

We describe the above procedure in more detail. In the following, we use the word “node” more neutral than the word “individual”. Suppose one node arrives at the social network at each time instant $t = 1, 2, \dots$. Note that we use t for both time and node index. Let $G(t) = (V(t), E(t))$ be the graph of the network, where $V(t)$ and $E(t)$ are the set of node and links in the network at

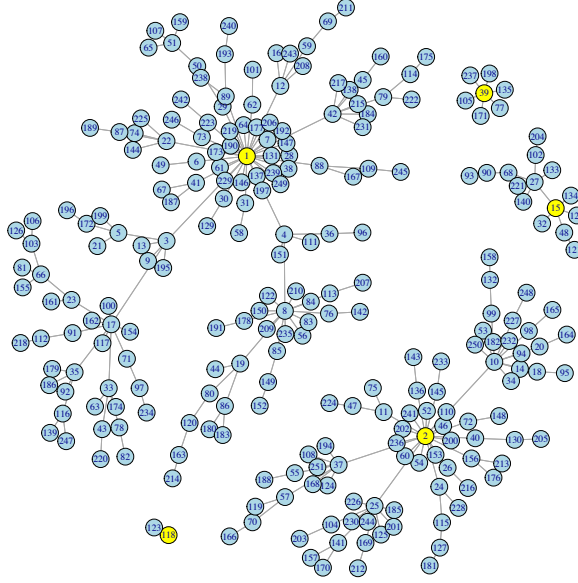


Figure 1: An example of the network $G(250)$ constructed by the preferential attachment with break-ups ($m = 1$). Yellows are the nodes who start break-ups.

the time t . The network starts from $G(1)$ with one node indexed by 1 and m self loops $E(1) = \{(1, 1), \dots, (1, 1)\}$. At time t , a new m -degree node t joins to $V(t-1) = \{1, 2, \dots, t-1\}$, and new m links are added. The destination node of the new links are selected from $V(t)$ with probability proportional to its degree. Here we allow the node t to have self-loops by assuming it has already some outgoing links.

More precisely, we adopt the following preferential attachment procedure to select the destination nodes at the arrival of the node t . Let $k_s(t)$ be the degree of the node s at the time t . Note that each time when a new link is attached to the network, the total degree is increased by 2, and the total degree at time t is given by $\sum_{s \in V(t)} k_s(t) = 2mt$.

Preferential Attachment [16]:

1. At time t , start with selecting the first destination node d_1 . Set $k_t(t) = 1$ at this step as if one new link is already existed at the node t . A node in $V(t)$ is picked as d_1 with the probability $k_s(t)/\{2m(t-1) + 1\}$, where the denominator is the sum of all degree of nodes in $V(t-1)$ and 1 for the node t . Increase the degree of the picked node d_1 by 1. If the new link is self loop, $k_t(t)$ is increased to 2 for the next step.
2. At the selection of the i -th link (t, d_i) , increase $k_t(t)$ by 1. Pick d_i with the probability $k_s(t)/\{2m(t-1) + 2(i-1) + 1\}$. Increase the degree of the picked node by 1.
3. Repeat this procedure up to deciding the last destination node d_m .

When all m new links from the node t are self-loop, a break-up occurs at time t , and its probability is given by

$$\begin{aligned}
 P(\text{a break-up starts at time } t) &= \frac{1}{2m(t-1) + 1} \frac{3}{2m(t-1) + 3} \frac{5}{2m(t-1) + 5} \cdots \frac{2m-1}{2mt-1} \\
 &= \prod_{i=1}^m \frac{2i-1}{2m(t-1) + 2i-1},
 \end{aligned}$$

since the self-loop on the node t at each step will increase $k_t(t)$ by 2. Note that when $m = 1$, we have simply

$$P(\text{a break-up starts at time } t) = \frac{1}{2t-1}. \quad (1)$$

Note that, for $m \geq 2$, the break-ups originated at the arrival of new nodes may be repaired and connected to other fractions later.

3. Equivalence of Random Pairing and Preferential Attachment

The network constructed by the above preferential attachment is indeed random pairing network described below. Let $\tilde{G}(t)$ be the network of the size of t with the random pairing, and let $F(t) = \{1, 2, \dots, 2mt\}$ be the set of endpoint indexes. We consider the smaller index as the left side.

Random Pairing [16]:

1. Connect the left and the right end points randomly to obtain the mt pairs of the endpoints from $F(t)$.
2. Every m -th right-end point from the left is sequentially indexed as $i = 1, 2, \dots, t$, which represents a node in $\tilde{G}(t)$.
3. Designate all the end points in-between $i - 1$ and i for the node i .

Theorem 1 ([16]). *The networks constructed by the preferential attachment $G(t)$ and the random pairing $\tilde{G}(t) = (V(t), \tilde{E}(t))$ are stochastically equivalent, i.e.*

$$P(E(t) = \{e_1, e_2, \dots, e_{mt}\}) = P(\tilde{E}(t) = \{e_1, e_2, \dots, e_{mt}\}),$$

for all possible combinations of mt edges $\{e_1, e_2, \dots, e_{mt}\}$ on $V(t)$.

Proof. For simplicity, we prove the case of $m = 1$ by induction, which can be easily generalized. It is obvious that $G(1) = \tilde{G}(1)$. Suppose $G(t - 1) = \tilde{G}(t - 1)$ stochastically, so $G(t - 1)$ can be constructed by random-pairing of $2(t - 1)$ endpoints at time $t - 1$. At time t , according to the preferential attachment, a node for the new outgoing link from t is selected by its degree, which is equivalent to selecting one from $2(t - 1) + 1$ endpoints including the degree of the arriving node t . Thus, the endpoint $2t$ will be randomly paired with one of $2(t - 1) + 1$ endpoints in $G(t)$. Since the selected endpoint i is a left endpoint, it is connected to the same node with the next endpoint $i + 1$ in $G(t - 1)$. To obtain the full connectivity of $G(t)$, the rest of the endpoints $1, 2, \dots, i - 1, i + 1, \dots, 2(t - 1) + 1$ can be re-wired as $G(t - 1)$ by re-indexing $i + 1, \dots, 2(t - 1)$ as $i, \dots, 2(t - 1)$. Thus, $G(t)$ is constructed by random pairing. $\square$

It can be shown that $G(t)$ has the power law degree distribution for large t .

Theorem 2 ([17]). *Let $N_k(t)$ be the number of node with degree- k in the network $G(t)$. Then, the ratio of degree- k node is converged as*

$$\frac{N_k(t)}{t} \rightarrow \frac{2m(m+1)}{k(k+1)(k+2)} \text{ as } t \rightarrow \infty.$$

Thus, in particular, $G(t)$ has the power law degree-distribution,

$$P\{K = k\} \approx \frac{N_k(t)}{t} \approx \frac{2m(m+1)}{k(k+1)(k+2)} \approx k^{-3}.$$

Proof. For simplicity, we prove the case when $m = 1$ and we know that $N_k(t)/t$ converges to a constant q_k almost surely (more detail can be found at [1]).

Let A_k be the event that a degree- k node is selected at time t . By the observation of the link constructed at time t , we can easily find the dynamics of $N_k(t)$ as

$$\begin{aligned} N_k(t) - N_k(t-1) &= 1_{A_{k-1}} - 1_{A_k} \text{ for } k \geq 2, \\ N_1(t) - N_1(t-1) &= 1 - 1_{A_1}, \end{aligned} \tag{2}$$

where 1_A is the indicator function of the set A . Since there are $N_k(t-1)$ degree- k nodes at the arrival of node t , by taking the expectation we have

$$E[N_k(t) - N_k(t-1)] = \frac{k-1}{2t-1} E[N_{k-1}(t-1)] - \frac{k}{2t-1} E[N_k(t-1)].$$

Then, for large t , neglecting the terms of $1/t$, we have

$$q_k = \frac{k-1}{2}q_{k-1} - \frac{k}{2}q_k.$$

Thus, we have the recursion for $k \geq 2$

$$q_k = \frac{k-1}{k+2}q_{k-1}.$$

Similarly, by (2)

$$q_1 = 1 - \frac{1}{2}q_1,$$

and $q_1 = 2/3$. Thus, using this initial condition, we can conclude that if $N_k(t)/t$ converges, it would converges to

$$q_k = \frac{4}{k(k+1)(k+2)}.$$

□

4. The Number of Break-Ups

In this section we only deal with the case when $m = 1$. The number of break-ups $B(t)$ of the network $G(t)$ can be modeled by a stochastic process on $\mathbb{Z}$ starting from $B(1) = 1$. In the case of $m = 1$, $B(t)$ is given by a simple random walk;

$$B(t) = \sum_{k=1}^t X_k,$$

where $X_1, X_2, \dots, X_t$ are Bernoulli trials with different parameters $p_k = P(X_k = 1) = 1/(2k-1)$ as in (1). Thus, we have the following result indicating the logarithmic growth of $E[B(t)]$.

Theorem 3. *Let $B(t)$ be the number of break-ups at time t of the network $G(t)$ for $m = 1$. Then,*

$$E[z^{B(t)}] = \prod_{k=1}^t E[z^{X_k}] = \prod_{k=1}^t \frac{2k-2+z}{2k-1},$$

and

$$E[B(t)] = \sum_{k=1}^t \frac{1}{2k-1} \approx \frac{1}{2} \log(2t+1) + \gamma.$$

where γ is the Euler's gamma.

Break-ups can occur any time but the growth of its fraction may depend on the time it borns. It turns out that the growth of the size of the fraction is linear and the slope is inversely proportional to the break-up time (see Figure 2).

Note that the degree of a node $d_s(t)$ is known to increase as

$$E[d_s(t)] = \prod_{i=s}^t \frac{2i}{2i-1} \approx O(\sqrt{t/s}),$$

(see [16]).

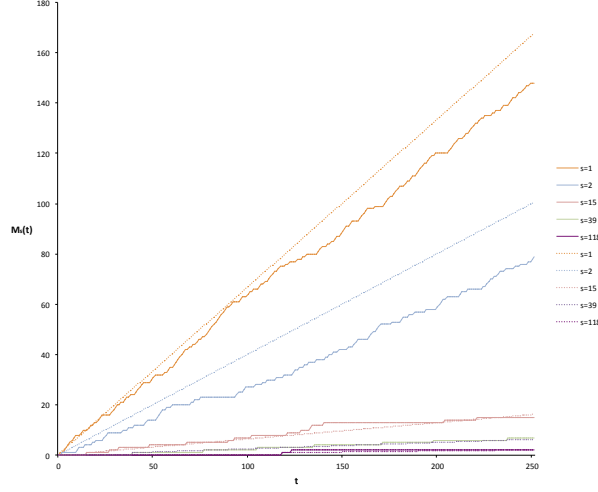


Figure 2: An example of the fraction size evolution $M_s(t)$ with the estimates (dotted lines) obtained in Theorem 4. The resulted fragmented network at $t = 250$ is shown in Figure 1.

Theorem 4. Let $M_s(t)$ be the size of the fraction started by the node s at its arrival of $G(s)$ for $m = 1$. Then,

$$E[M_s(t) | a \text{ break-up occurs at time } s] = \frac{2t + 1}{2s + 1}.$$

Moreover, the conditional probability generating function given by

$$\phi_s(z, t) = E \left[z^{M_s(t)} | a \text{ break-up occurs at time } s \right]$$

is recursively derived by

$$\phi_s(z, t) = \phi_s(z, t-1) + \frac{2z(z-1)}{2t-1} \frac{\partial}{\partial z} \phi_s(z, t-1),$$

starting from $\phi_s(z, s) = z$.

Proof. Let $B_s(t) = \{\text{the node } t \text{ selects the target fraction of the size } M_s(t-1)\}$ and $C = \{\text{a break-up occurs at time } s\}$, then we have

$$M_s(t) = M_s(t-1) + 1_{B_s(t)}, \quad (3)$$

on C . Since $P(B_s(t) | C, M_s(t-1)) = 2M_s(t-1)/(2t-1)$, by taking the expectation on the both sides gives

$$\begin{aligned} E[M_s(t) | C] &= \frac{2t+1}{2t-1} E[M_s(t-1) | C] = \frac{2t+1}{2t-1} \frac{2t-1}{2t-3} E[M_s(t-2) | C] \\ &= \frac{2t+1}{2s+1}, \end{aligned}$$

where we used the fact $M_s(s) = 1$ on C .

For the second statement, we have

$$E \left[z^{1_{B_s(t)}} | M_s(t-1), C \right] = zP(B_s(t) | C) + 1 - P(B_s(t) | C) = 1 + \frac{2(z-1)}{2t-1} M_s(t-1).$$

Substituting this and (3) gives

$$\begin{aligned}
\phi_s(z, t) &= E \left[z^{M_s(t-1)} E \left[z^{1_{B_s(t)}} | M_s(t-1), C \right] \middle| C \right] \\
&= E \left[z^{M_s(t-1)} \middle| C \right] + \frac{2(z-1)}{2t-1} E \left[M_s(t-1) z^{M_s(t-1)} \middle| C \right] \\
&= \phi_s(z, t-1) + \frac{2z(z-1)}{2t-1} \frac{\partial}{\partial z} \phi_s(z, t-1).
\end{aligned}$$

□

Break-ups may be followed by no-one, and then the fraction is consisted only by the starter. Indeed, in our toy model those isolation can occur uniformly regardless of the arrival time. Thus all nodes have the equal chance for isolation.

Theorem 5.

$$P(\text{node } s \text{ is single at time } t) = \frac{1}{2t-1}.$$

Proof. Since single nodes are always degree-2, we have

$$\begin{aligned}
P(s \text{ is single at time } t) &= P(\text{a break-up at time } s, \text{ all nodes } (s+1, \dots, t) \text{ do not select } s) \\
&= \frac{1}{s-1} \prod_{k=s+1}^t \frac{2k-3}{2k-1} = \frac{1}{2t-1}.
\end{aligned}$$

□

5. Break-ups in an $M/M/\infty$ Stochastic Dynamic Network

Consider a dynamic network on continuous time with leaves as well as joins. Let $C_1, C_2, \dots$ be the nodes in the arriving order. Starting from the empty network $H(0)$, let $H(t) = \{V(t), E(t)\}$ be the network at the time t . The nodes in $V(t)$ are determined as the dynamics of $M/M/\infty$ queue, i.e., nodes arrive at the network according to a Poisson process with the rate λ and each node stays in the network for an independent exponential time with the mean $1/\mu$. Thus, $N(t)$, the number of nodes at time t , is Poisson random variable with the mean $\rho = \lambda/\mu$ in the steady state.

Given $H(t-)$ at the time of its arrival, the arriving node selects a node in $V(t)$ to establish the link between them according to the preferential attachment. At the time of leave, the leaving node C withdraws its outgoing link from the links $E(t-)$ and the incoming links to C are relocated to the next node in the arriving order existed in $V(t-)$. Note that C is the last in $V(t)$, then there is no incoming link to C .

Theorem 6. *Conditioning on $N(t) = n$, $H(t)$ is stochastically equivalent to the network $G(n)$ constructed by preferential attachment after n joins.*

Proof. By Theorem 1, we can always replace $G(n)$ with $\tilde{G}(n)$. Suppose $H(t) = \tilde{G}(n)$, and check what happens at the arrivals and leaves. Also, by Theorem 1, the preferential attachment of a new arriving node to $\tilde{G}(n-1)$ gives $\tilde{G}(n)$. The leaves may occur randomly among the existing nodes because of memoryless property among their independent exponential sojourn times in $V(t)$. The withdrawal of the existing outgoing links from a randomly selected node, which corresponds to one of the right endpoints in $F(n)$ of $\tilde{G}(n)$, will give $\tilde{G}(n-1)$.

Thus, both arrivals and leaves of the nodes preserve the random pairing network structure. Since the first arrival gives self-loop which is a random pairing network, after some joins and leaves $H(t)$ is stochastically equivalent to the network $G(n)$ given $N(t) = n$. □

Theorem 6 shows that we can build the network $H(t)$ by neglecting the leave of nodes. By combining this result with Theorem 3, we obtain the following estimation of the break-ups in our toy model as depicted in Figure 3.

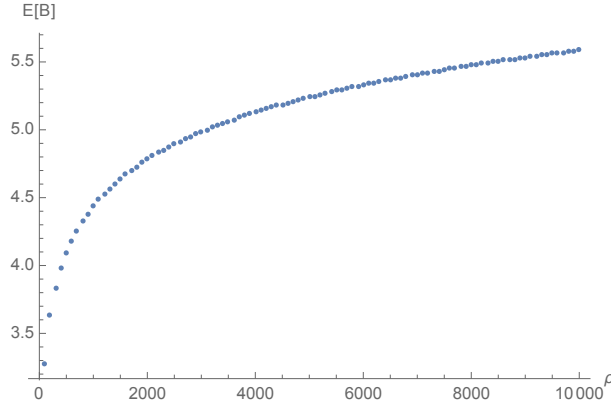


Figure 3: The expected number of break-ups $E[B(t)]$ in the $M/M/\infty$ dynamic stochastic network $H(t)$.

Theorem 7. For $m = 1$, the number of breakups in the $M/M/\infty$ stochastic network $H(t)$ in the steady state with $\rho = \lambda/\mu$ can be given by

$$E[B(t)] = e^{-\rho} \sum_{n=1}^{\infty} \sum_{k=1}^n \frac{1}{2k-1} \frac{\rho^n}{n!}.$$

References

- [1] R. Durrett, *Random graph dynamics*, vol. 20 (Cambridge Univ Pr, 2007).
- [2] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008).
- [3] H. Toyozumi, S. Tani, N. Miyoshi, and Y. Okamoto, *Phys. Rev. E* **86**, 021103 (2012), URL <http://link.aps.org/doi/10.1103/PhysRevE.86.021103>.
- [4] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, *Journal of Theoretical Biology* **235**, 275 (2005), URL <http://dx.doi.org/10.1016/j.jtbi.2005.01.011>.
- [5] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
- [6] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* **65**, 035108 (2002), URL <http://link.aps.org/doi/10.1103/PhysRevE.65.035108>.
- [7] M. Boguñá and R. Pastor-Satorras, *Phys. Rev. E* **66**, 047104 (2002).
- [8] N. Berger, C. Borgs, J. Chayes, and A. Saberi, in *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms* (Society for Industrial and Applied Mathematics, 2005), pp. 301–310.
- [9] M. Molloy and B. Reed, *Combin. Probab. Comput* **7**, 295 (1998).
- [10] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, *Proceedings of the National Academy of Sciences of the United States of America* **99**, 2566 (2002), URL <http://dx.doi.org/10.1073/pnas.012582999>.
- [11] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
- [12] X. F. Wang and G. Chen, *IEEE circuits and systems magazine* **3**, 6 (2003).
- [13] X. Han, S. Cao, Z. Shen, B. Zhang, W.-X. Wang, R. Cressman, and H. E. Stanley, *Proceedings of the National Academy of Sciences* **114**, 2887 (2017), <http://www.pnas.org/content/114/11/2887.full.pdf>, URL <http://www.pnas.org/content/114/11/2887.abstract>.
- [14] P. S. Chodrow, *Proceedings of the National Academy of Sciences* **114**, 11591 (2017), URL <http://www.pnas.org/content/114/44/11591>.
- [15] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999), URL <http://DOI:10.1126/science.286.5439.509>.

- [16] B. Bollobás and O. Riordan, *Combinatorica* **24**, 5 (2004).
- [17] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Physical review letters* **85**, 4633 (2000).

レベル依存する G/G/1 型マルコフ連鎖における 条件付き定常分布の誤差評価付き数値計算法

木村雅俊 滝根哲哉

大阪大学 工学研究科 電気電子情報工学専攻

概要： 本稿では、レベル依存する G/G/1 型連続時間マルコフ連鎖における、レベルが N 以下であるという条件付き定常分布を、誤差評価付きで数値計算する方法について議論する。なお、本稿で扱う条件付き定常分布の数値計算法の元になる理論的結果は、既存研究の成果から直ちに導かれるものである。本稿では、この条件付き定常分布の数値計算法において、数値解がもつ誤差上界が数値的に計算可能であることを示す。また、関連する諸定理へ確率的な解釈を与える。

1. はじめに

本稿では、既約で正再帰的な連続時間 2 変数マルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ を考察する。ただし、レベルと呼ばれる変数 $L(t)$ は非負整数値を取り、相と呼ばれる変数 $J(t)$ は、 $L(t) = k$ ($k = 0, 1, 2, \dots$) が与えられたとき、有限集合 $\mathcal{M}_k = \{1, 2, \dots, M_k\}$ の要素から値を取る。以下では、連続時間マルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ の状態空間 $\Omega = \{(k, j); k = 0, 1, \dots, j \in \mathcal{M}_k\}$ を互いに素な部分集合 $\mathcal{L}_k = \{(k, j); j \in \mathcal{M}_k\}$ ($k = 0, 1, \dots$) に分割し、各部分集合 $\mathcal{L}_k$ をレベル k と呼ぶ。よって、辞書式順序に従って状態を並べると、連続時間マルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ の無限小生成作用素 Q は以下の形をとる。

$$Q = \begin{matrix} & \mathcal{L}_0 & \mathcal{L}_1 & \mathcal{L}_2 & \mathcal{L}_3 & \dots \\ \begin{matrix} \mathcal{L}_0 \\ \mathcal{L}_1 \\ \mathcal{L}_2 \\ \mathcal{L}_3 \\ \vdots \end{matrix} & \begin{pmatrix} Q_{0,0} & Q_{0,1} & Q_{0,2} & Q_{0,3} & \cdots \\ Q_{1,-1} & Q_{1,0} & Q_{1,1} & Q_{1,2} & \cdots \\ Q_{2,-2} & Q_{2,-1} & Q_{2,0} & Q_{2,1} & \cdots \\ Q_{3,-3} & Q_{3,-2} & Q_{3,-1} & Q_{3,0} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{matrix} \quad (1)$$

ただし、 $Q_{k,\ell}$ ($k = 0, 1, \dots, \ell = -k, -k+1, \dots$) は $M_k \times M_{k+\ell}$ 行列であり、レベル k からレベル $k+\ell$ への遷移を表している。マルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ の定常分布を $\mathbf{q} = (q_0 \ q_1 \ \cdots)$ と表す。ただし、 $\mathbf{q}_k$ ($k = 0, 1, \dots$) は $1 \times M_k$ ベクトルであり、その j 番目 ($j \in \mathcal{M}_k$) の要素は定常状態において $(L(t), J(t)) = (k, j)$ となる確率で与えられる。無限小生成作用素が式 (1) で示したブロック構造をもつ 2 変数マルコフ連鎖はレベル依存する G/G/1 型と呼ばれる。以下では、レベル依存する G/G/1 型連続時間マルコフ連鎖を LD-G/G/1 型 MC と略記する。

一般に、可算無限の状態空間をもつ任意の連続時間マルコフ連鎖は、各レベル k の相の数 M_k が全て 1 であるような LD-G/G/1 型 MC として定式化可能であり、ブロック構造を用いた定式化は本質的ではない。しかし、待ち行列モデルの系内客数過程をマルコフ連鎖として定式化する際には、系内客数に対応するレベル変数と補助的な情報を与える相変数の組でモデルの状態を表現することが自然である場合が頻出するため、本稿では、ブロック構造を明示的に導入した LD-G/G/1 型 MC を考察する。なお、可算無限の状態空間をもつ離散時間マルコフ連鎖の定常分布 $\mathbf{q}$ に興味がある場合には、そのマルコフ連鎖の遷移確率行列を $\mathbf{P}$ としたとき、定常分布は $\mathbf{q} = \mathbf{q}\mathbf{P}$ を満たすため、無限小生成作用素 $Q := \mathbf{P} - \mathbf{I}$ をもつ LD-G/G/1 型 MC の定常分布を考察すればよい。

可算無限の状態空間をもつマルコフ連鎖における定常分布の数値計算では、多くの場合、状態空間の有限部分集合に着目した条件付き定常分布しか得ることができない。本稿では、LD-G/G/1 型

MC において、レベルが閾値 N 以下であるという条件付きの定常分布

$$\mathbf{q}(N) = \left(q_0(N) \quad q_1(N) \quad \cdots \quad q_N(N) \right)$$

の数値計算法を議論する。ただし、

$$q_k(N) = q_k / \left(\sum_{\ell=0}^N q_\ell e \right)$$

であり、 e はすべての要素が 1 である適当な次元の列ベクトルである。次節で示すように、本稿で議論する条件付き定常分布の数値計算法の元になる理論的結果は、既存研究の成果から直ちに導かれるものである。

一般に、条件付き定常分布 $\mathbf{q}(N)$ とその数値解

$$\mathbf{q}^{\text{comp}}(N) = \left(q_0^{\text{comp}}(N) \quad q_1^{\text{comp}}(N) \quad \cdots \quad q_N^{\text{comp}}(N) \right)$$

の間には誤差が生じる。よって、数値解 $\mathbf{q}^{\text{comp}}(N)$ がもつ誤差を数値的に評価できることが望ましい。LD-G/G/1 型 MC の特別な場合であるレベル依存する M/G/1 型マルコフ連鎖に対しては、

$$\|\mathbf{q}^{\text{comp}}(N) - \mathbf{q}(N)\| \leq \epsilon$$

を満たす誤差上界 ϵ が評価可能な数値計算アルゴリズムが提案されている [Takine 2016, Kimura 2017]。ただし、 $\|\cdot\|$ は 1 ノルムを表す。しかし、レベル依存する G/M/1 型マルコフ連鎖を含めて、その他の 2 変数マルコフ連鎖における条件付き定常分布を誤差評価付きで数値計算するアルゴリズムは、著者の知る限り報告されていない。本稿の主な目的は、LD-G/G/1 型 MC の条件付き定常分布を誤差上界を伴って計算可能であることを示すことである。

もし N を十分に大きくとることができれば、条件付き定常分布 $\mathbf{q}(N)$ の数値解 $\mathbf{q}^{\text{comp}}(N)$ ならびに全ての要素が 0 である適当な次元をもつ行ベクトル $\mathbf{0}$ を用いて構成された分布

$$\mathbf{q}^{\text{comp}} = \left(\mathbf{q}^{\text{comp}}(N) \quad \mathbf{0} \quad \mathbf{0} \quad \cdots \right)$$

は定常分布の良い近似となる [Bright 1995]。具体的には [Takine 2016, Kimura 2017]

$$\sum_{\ell=N+1}^{\infty} q_\ell e \leq \eta \quad \text{かつ} \quad \|\mathbf{q}^{\text{comp}}(N) - \mathbf{q}(N)\| \leq \epsilon \quad \Rightarrow \quad \|\mathbf{q}^{\text{comp}} - \mathbf{q}\| \leq \epsilon + 2\eta$$

が成立する。上式から明らかなように、 $\mathbf{q}^{\text{comp}}$ が定常分布 $\mathbf{q}$ の良い近似となっているか否かは、条件付き定常分布の数値解 $\mathbf{q}^{\text{comp}}(N)$ がもつ誤差（その上界は ϵ ）のみならず、定常分布の裾確率（その上界は η ）にも強く依存する。よって、 $\mathbf{q}^{\text{comp}}$ が定常分布 $\mathbf{q}$ の良い近似となっているかを数値的に判断することは一般には困難であり、定常分布の近似解 $\mathbf{q}^{\text{comp}}$ が良い近似となるような N の選択に関しては、計算対象となる具体的なモデル毎に裾確率の評価が不可欠である。

本稿の構成は以下の通りである、2 節において関連研究で示されている結果を整理し、LD-G/G/1 型 MC における条件付き定常分布の数値計算法の元となる理論的結果を提示する。3 節では、LD-G/G/1 型 MC における条件付き定常分布の誤差上界に関する定理を示した上、その確率的な解釈を与える。4 節では、本稿で議論する数値計算法の実行に関して幾つかコメントを与える。

以下では、任意のベクトル $\mathbf{x}$ に対して、 $\bar{\mathbf{x}}$ を

$$\bar{\mathbf{x}} = \begin{cases} \frac{\mathbf{x}}{\mathbf{x}e}, & \mathbf{x}e \neq 0 \\ \mathbf{0}, & \mathbf{x}e = 0 \end{cases}$$

と定義する. また, 任意のベクトル $\mathbf{y}$ に対して $\text{diag}^*(\mathbf{y})$ を

$$[\text{diag}^*(\mathbf{y})]_{i,i} = \begin{cases} [\mathbf{y}]_i^{-1}, & [\mathbf{y}]_i \neq 0 \\ 0, & [\mathbf{y}]_i = 0 \end{cases}$$

で定める対角行列と定義し, これを用いて, 任意の行列 $\mathbf{X}$ に対して, $\overline{\mathbf{X}}$ を

$$\overline{\mathbf{X}} = \text{diag}^*(\mathbf{X}\mathbf{e})\mathbf{X}$$

と定義する. すなわち, 行列 $\mathbf{X}$ の i 番目の行ベクトルが $\mathbf{x}(i)$ で与えられるとき, $\overline{\mathbf{X}}$ の i 番目の行ベクトルは $\overline{\mathbf{x}}(i)$ で与えられる.

2. 関連研究

条件付き定常分布 $\mathbf{q}(N)$ の数値解がもつ誤差の評価に関しては, レベル依存する M/G/1 型マルコフ連鎖 (以下, LD-M/G/1 型 MC と略記) を対象に研究されている [Takine 2016, Kimura 2017]. これらの既存研究では, 条件付き定常分布を議論するに当たって, レベル k ($k = 0, 1, \dots$) への再帰時間に着目している. $\mathbf{u}_{k,n}(i)$ ($n, k = 0, 1, \dots, n \neq k, i \in \mathcal{M}_k$) を $1 \times M_n$ ベクトルとし, その j 番目の要素を状態 (k, i) から始めたレベル k への再帰時間内に状態 (n, j) に滞在する平均時間を状態 (k, i) での平均滞在時間で割った値を表すものとする. また $M_k \times M_n$ 行列 $\mathbf{U}_{k,n}$ を

$$\mathbf{U}_{k,n} = \begin{pmatrix} \mathbf{u}_{k,n}(1) \\ \mathbf{u}_{k,n}(2) \\ \vdots \\ \mathbf{u}_{k,n}(M_k) \end{pmatrix}$$

と定める. $\mathbf{U}_{k,n}$ は, いわゆる率行列 (R -行列) であり

$$\mathbf{q}_n = \mathbf{q}_k \mathbf{U}_{k,n}, \quad n, k = 0, 1, \dots \quad (2)$$

を満たす. ただし $\mathbf{U}_{k,k} = \mathbf{I}$ ($k = 0, 1, \dots$). [Takine 2016] では LD-M/G/1 型 MC に対して, $\mathbf{U}_{k,n} = \mathbf{Z}_k \mathbf{Z}_{k-1} \cdots \mathbf{Z}_{n+1}$ ($k > n \geq 0$) となる $\mathbf{Z}_m := \mathbf{U}_{m,m-1}$ ($m = 1, 2, \dots$) の存在, $\mathbf{Z}_m$ ($m = 1, 2, \dots$) を昇順に計算する手順, ならびに全ての k, n ($k > n$) に対して $\mathbf{U}_{k,n} > \mathbf{O}$ であるための必要十分条件である, 全ての $\mathbf{Q}_{k,-1}$ が正則であるという仮定の下で, 以下の極限公式が導かれている.

$$\lim_{k \rightarrow \infty} \overline{\mathbf{U}}_{k,N} = \mathbf{e}\overline{\mathbf{q}}_N = \mathbf{e}\mathbf{q}_N/(\mathbf{q}_N\mathbf{e}) \quad (3)$$

これらの結果から, 全ての $\mathbf{Q}_{k,-1}$ が正則である LD-M/G/1 型 MC に対しては,

- (i) まず $\mathbf{Z}_m$ ($m = 1, 2, \dots$) を $m = 1$ から昇順に計算していき,
- (ii) $\overline{\mathbf{U}}_{k,N}$ から正規化された境界ベクトル $\overline{\mathbf{q}}_N$ を得た後,
- (iii) 式 (2) の特別な場合である $\mathbf{q}_k = \mathbf{q}_{k+1} \mathbf{Z}_{k+1}$ を用いてレベル $N - 1$ 以下に関する定常確率ベクトル $\mathbf{q}_k$ ($k = N - 1, N - 2, \dots, 0$) のスカラ倍 $\mathbf{q}_k/(\mathbf{q}_N\mathbf{e})$ を降順に計算し,
- (iv) 最後に正規化する

ことで, 条件付き定常分布の数値解を得ることができる [Takine 2016].

この手順に従って得られた数値解に含まれる誤差の発生源は, 数値計算の際に式 (3) の極限公式を有限の $k = K$ で停止して利用する点である. すなわち, $\mathbf{U}_{K,N}$ のランクが 1 であるような一部の例外を除いて, 有限の K に対して

$$\overline{\mathbf{U}}_{K,N} \neq \mathbf{e}\overline{\mathbf{q}}_N$$

であるため、 $\bar{\mathbf{q}}_N$ の数値解 $\bar{\mathbf{q}}_N^{\text{comp},K}$ には誤差が含まれている。[Takine 2016] では、真の境界ベクトル $\bar{\mathbf{q}}_N$ が $\{\bar{\mathbf{u}}_{k,N}(i); i \in \mathcal{M}_k\}$ で張られる凸包の相対的内部に含まれることが示されている。

$$\bar{\mathbf{q}}_N \in \{\mathbf{x} \in \mathbb{R}^{M_N} \mid \mathbf{x} = \sum_{i \in \mathcal{M}_k} \alpha(i) \bar{\mathbf{u}}_{k,n}(i), \sum_{i \in \mathcal{M}_k} \alpha(i) = 1, \alpha(i) > 0 (i \in \mathcal{M}_k)\}$$

すなわち真の境界ベクトル $\bar{\mathbf{q}}_N$ が存在する $\mathbb{R}^{M_N}$ 上の有界な部分集合は、数値計算可能なベクトル $\bar{\mathbf{u}}_{k,N}(i)$ を頂点とする凸包として表現できる。よって、この凸包の相対的内部上、例えば重心 $\bar{\mathbf{q}}_N^{\text{grav},K} = \mathbf{e}^T \mathbf{U}_{K,N} / (\mathbf{e}^T \mathbf{U}_{K,N} \mathbf{e})$ を境界ベクトル $\bar{\mathbf{q}}_N$ の近似的数値解 $\bar{\mathbf{q}}_N^{\text{comp},K}$ として用いることにすると、境界ベクトルの近似的数値解 $\bar{\mathbf{q}}_N^{\text{comp},K}$ がもつ誤差の上界を得ることができる。

$$\|\bar{\mathbf{q}}_N^{\text{comp},K} - \bar{\mathbf{q}}_N\| \leq \max_{i \in \mathcal{M}_K} \|\bar{\mathbf{q}}_N^{\text{comp},K} - \bar{\mathbf{u}}_{K,N}(i)\|$$

[Kimura 2017] では、LD-G/G/1 型 MC に対して、以下の二つの命題が示されており、前述のアルゴリズムに課されていた、全ての $\mathbf{Q}_{k,-1}$ が正則であるという条件は不要であることが指摘されている。

命題 1 ([Kimura 2017, Theorem 2.5]). 既約で正再帰的な LD-G/G/1 型 MC において、任意の $\epsilon > 0$ および $n = 0, 1, \dots$ に対して、次式を満たす K ($K > n$) が存在する。

$$\|\bar{\mathbf{u}}_{k,n}(i) - \bar{\mathbf{q}}_n\| < \epsilon \text{ if } \mathbf{u}_{k,n}(i)\mathbf{e} > 0, \quad \forall k = K, K+1, \dots, i \in \mathcal{M}_k$$

命題 2 ([Kimura 2017, Corollary 3.2]). 既約で正再帰的な LD-G/G/1 型 MC において、次式が成立する。

$$\bar{\mathbf{q}}_n \in \left\{ \mathbf{x} \in \mathbb{R}^{M_n} \mid \mathbf{x} = \sum_{\substack{i \in \mathcal{M}_k, \\ \mathbf{u}_{k,n}(i)\mathbf{e} > 0}} \alpha(i) \bar{\mathbf{u}}_{k,n}(i), \sum_{\substack{i \in \mathcal{M}_k, \\ \mathbf{u}_{k,n}(i)\mathbf{e} > 0}} \alpha(i) = 1, \alpha(i) > 0 (i \in \mathcal{M}_k, \mathbf{u}_{k,n}(i)\mathbf{e} > 0) \right\}$$

さらに、[Kimura 2017] では、境界ベクトルの数値解 $\bar{\mathbf{q}}_N^{\text{comp},k}$ の誤差上界を利用して、条件付き定常分布の数値解 $\mathbf{q}^{\text{comp},k}(N)$ の誤差上界が計算可能であることが示されている。

命題 3 ([Kimura 2017, Lemma 3.6, Theorem 3.7]). 既約で正再帰的な LD-G/G/1 型 MC において

$$\|\bar{\mathbf{q}}_N^{\text{comp},k} - \bar{\mathbf{q}}_N\| \leq \max_{\substack{i \in \mathcal{M}_k, \\ \mathbf{u}_{k,N}(i)\mathbf{e} > 0}} \|\bar{\mathbf{q}}_N^{\text{comp},k} - \bar{\mathbf{u}}_{k,N}(i)\|$$

が成立し、これより条件付き定常分布の近似 $\mathbf{q}^{\text{comp},k}(N)$ がもつ誤差に関して、次式が成立する。

$$\|\mathbf{q}^{\text{comp},k}(N) - \mathbf{q}(N)\| \leq \left(\frac{2e \sum_{\ell=0}^N \mathbf{U}_{N,\ell} \mathbf{e}}{\bar{\mathbf{q}}_N^{\text{comp},k} \sum_{\ell=0}^N \mathbf{U}_{N,\ell} \mathbf{e}} \right) \cdot \max_{\substack{i \in \mathcal{M}_k, \\ \mathbf{u}_{k,N}(i)\mathbf{e} > 0}} \|\bar{\mathbf{q}}_N^{\text{comp},k} - \bar{\mathbf{u}}_{k,N}(i)\|$$

これらの命題および式 (2) は既約で正再帰的な LD-G/G/1 型 MC における $\mathbf{U}_{k,n}$ が数値計算可能であれば、条件付き定常分布 $\mathbf{q}(N)$ を誤差評価付きで数値計算できることを示している。しかし、現時点において $\mathbf{U}_{k,n}$ が数値的に計算可能なマルコフ連鎖は LD-M/G/1 型 MC に限られており、結果として、条件付き定常分布の誤差評価付き数値計算法が与えられているクラスも LD-M/G/1 型 MC のみである。

[Masuyama 2017] では可算無限の状態空間をもつ一般のマルコフ連鎖を対象に、初到達時間を通して定常分布が考察されている。LD-G/G/1 型 MC における状態空間のうちのレベル k 以下を表す部分集合を $\mathcal{L}_{\leq k}$ とする。

$$\mathcal{L}_{\leq k} = \bigcup_{\ell=0}^k \mathcal{L}_{\ell} = \{(\ell, i); \ell = 0, 1, \dots, k, i \in \mathcal{M}_{\ell}\}$$

また、 Q の $|\mathcal{L}_{\leq k}| \times |\mathcal{L}_{\leq k}|$ の北西角を $Q_{\leq k}$ とする.

$$Q_{\leq k} = \begin{matrix} & \begin{matrix} \mathcal{L}_0 & \mathcal{L}_1 & \mathcal{L}_2 & \cdots & \mathcal{L}_k \end{matrix} \\ \begin{matrix} \mathcal{L}_0 \\ \mathcal{L}_1 \\ \mathcal{L}_2 \\ \vdots \\ \mathcal{L}_k \end{matrix} & \begin{pmatrix} Q_{0,0} & Q_{0,1} & Q_{0,2} & \cdots & Q_{0,k} \\ Q_{1,-1} & Q_{1,0} & Q_{1,1} & \cdots & Q_{1,k-1} \\ Q_{2,-2} & Q_{2,-1} & Q_{2,0} & \cdots & Q_{2,k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Q_{k,-k} & Q_{k,-k+1} & Q_{k,-k+2} & \cdots & Q_{k,0} \end{pmatrix} \end{matrix}$$

さらに、 $F_{\leq k}$ を次式で定義する.

$$F_{\leq k} := (-Q_{\leq k})^{-1} \quad (4)$$

$F_{\leq k}$ の $((\ell, i), (n, j))$ 要素は状態 (ℓ, i) から始めたレベル $k+1$ 以上への初到達時間内に状態 (n, j) に滞在する平均時間を表す. 条件付き定常分布に関する [Masuyama 2017, Theorem 2.3] の結果から LD-G/G/1 型 MC に対して次の命題が直ちに導かれる.

命題 4. 既約で正再帰的な LD-G/G/1 型 MC $\{(L(t), J(t)); t \geq 0\}$ において、以下の極限公式が成立する.

$$\lim_{k \rightarrow \infty} \overline{F}_{\leq k, N}^{(NW)} = eq(N)$$

ただし、 $F_{\leq k, N}^{(NW)}$ は $F_{\leq k}$ の $|\mathcal{L}_{\leq N}| \times |\mathcal{L}_{\leq N}|$ の北西角を表す.

命題 4 の略証を付録に与える. 命題 4 は任意の既約で正再帰的な LD-G/G/1 型 MC における条件付き定常分布 $q(N)$ が数値的に近似計算可能であることを示している. 次節では、既約で正再帰的な LD-G/G/1 型 MC に対する命題 4 に基づく数値計算において、条件付き定常分布 $q(N)$ が誤差評価 (誤差上界) 付きで数値計算可能であることを示すと共に、関連する定理の確率的な解釈を与える.

3. LD-G/G/1 型 MC における条件付き定常分布の誤差評価付き数値計算

まず始めに、既約で正再帰的な LD-G/G/1 型 MC $\{(L(t), J(t)); t \geq 0\}$ に対する、命題 4 の確率的な解釈を与える. そのため、 $Q_{\leq k}$ と $N < k$ なる N に対して、以下のブロック表記を導入する.

$$Q_{\leq k} = \begin{matrix} & \begin{matrix} \mathcal{L}_{\leq N} & \mathcal{L}_{\leq k} \setminus \mathcal{L}_{\leq N} \end{matrix} \\ \begin{matrix} \mathcal{L}_{\leq N} \\ \mathcal{L}_{\leq k} \setminus \mathcal{L}_{\leq N} \end{matrix} & \begin{pmatrix} Q_{\leq k, N}^{(NW)} & Q_{\leq k, N}^{(NE)} \\ Q_{\leq k, N}^{(SW)} & Q_{\leq k, N}^{(SE)} \end{pmatrix} \end{matrix}$$

定義より、 $Q_{\leq k, N}^{(NW)} = Q_{\leq N}$ である. 次に、LD-G/G/1 型 MC $\{(L(t), J(t)); t \geq 0\}$ に関連して、レベル $k+1$ 以上を吸収状態とした吸収マルコフ連鎖を構築した上で、この吸収マルコフ連鎖において、レベル N 以下を観測状態とするセンサー過程を考える. このセンサー過程も吸収マルコフ連鎖であり、その劣無限小生成作用素 $\tilde{Q}_{\leq k, N}$ は

$$\tilde{Q}_{\leq k, N} = Q_{\leq N} + Q_{\leq k, N}^{(NE)} (-Q_{\leq k, N}^{(SE)})^{-1} Q_{\leq k, N}^{(SW)}$$

で与えられる. さらに、この劣無限小生成作用素 $\tilde{Q}_{\leq k, N}$ に付随して、再生間隔が表現 $(\alpha, \tilde{Q}_{\leq k, N})$ をもつ相型分布に従う再生過程を考える. この相型再生過程を駆動するマルコフ連鎖 $\{(L_{\leq k, N}(t; \alpha), J_{\leq k, N}(t; \alpha)); t \geq 0\}$ の無限小生成作用素は $\tilde{Q}_{\leq k, N} + (-\tilde{Q}_{\leq k, N})e\alpha$ で与えられる. よって、マルコフ連鎖 $\{(L_{\leq k, N}(t; \alpha), J_{\leq k, N}(t; \alpha)); t \geq 0\}$ の定常分布 $\bar{f}_{\leq k, N}(\alpha)$ は次式を満たす.

$$\bar{f}_{\leq k, N}(\alpha) \left(\tilde{Q}_{\leq k, N} + (-\tilde{Q}_{\leq k, N})e\alpha \right) = 0, \quad \bar{f}_{\leq k, N}(\alpha)e = 1 \quad (5)$$

マルコフ連鎖 $\{(L_{\leq k, N}(t; \alpha), J_{\leq k, N}(t; \alpha)); t \geq 0\}$ は元のマルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ に対して

- (i) レベル k 以下の状態からレベル $k+1$ 以上の状態への全ての遷移に対して、遷移先を分布 α に従ってレベル N 以下へ遷移するように変更したマルコフ連鎖を考え、
- (ii) さらに、この変更後のマルコフ連鎖において、レベル N 以下を観測状態としたセンサー過程である。一方、元のマルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ の条件付き定常分布 $\mathbf{q}(N)$ はレベル N 以下を観測状態としたセンサー過程 $\{(L_N(t), J_N(t)); t \geq 0\}$ の定常分布に等しい。すなわち $\mathbf{q}(N)$ と $\bar{\mathbf{f}}_{\leq k, N}(\alpha)$ の違いは、上記の手順 (i) で示された変更によって生じる。

元のマルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ の条件付き定常分布 $\mathbf{q}(N)$ の (ℓ, j) 要素 ($\ell = 0, 1, \dots, N, j \in \mathcal{M}_\ell$) は元のマルコフ連鎖における状態 (ℓ, j) での平均滞在時間 $s_{\ell, j}$ と状態 (ℓ, j) への平均再帰時間 $\nu_{\ell, j}$ の比 $s_{\ell, j}/\nu_{\ell, j}$ に比例する。一方、 $\bar{\mathbf{f}}_{\leq k, N}(\alpha)$ の (ℓ, j) 要素 ($\ell = 0, 1, \dots, N, j \in \mathcal{M}_\ell$) は、上記の手順 (i) の変更を施したマルコフ連鎖における状態 (ℓ, j) での平均滞在時間 $s_{\ell, j}$ と状態 (ℓ, j) への平均再帰時間 $\nu_{\ell, j}(k, \alpha)$ の比 $s_{\ell, j}/\nu_{\ell, j}(k, \alpha)$ に比例する。 k の値が大きくなるに従って、上記の手順 (i) の変更を施したマルコフ連鎖における状態 (ℓ, j) への再帰時間が、変更を施したことによる影響を受ける（すなわち、元のマルコフ連鎖における状態 (ℓ, j) への再帰時間内にレベル $k+1$ 以上の状態を訪れる）確率は単調に減少していき、 $k \rightarrow \infty$ の極限では 0 となる。よって、任意の α に対して

$$\lim_{k \rightarrow \infty} \nu_{\ell, j}(k, \alpha) = \nu_{\ell, j}, \quad \ell = 0, 1, \dots, N, j \in \mathcal{M}_\ell$$

であり、これより任意の初期ベクトル α に対して

$$\lim_{k \rightarrow \infty} \bar{\mathbf{f}}_{\leq k, N}(\alpha) = \mathbf{q}(N) \quad (6)$$

となることが分かる。

一方、 $\mathbf{F}_{\leq k, N}^{(\text{NW})}$ は

$$\mathbf{F}_{\leq k, N}^{(\text{NW})} = (-\tilde{\mathbf{Q}}_{\leq k, N})^{-1} = \left(-(\mathbf{Q}_{\leq N} + \mathbf{Q}_{\leq k, N}^{(\text{NE})}(-\mathbf{Q}_{\leq k, N}^{(\text{SE})})^{-1}\mathbf{Q}_{\leq k, N}^{(\text{SW})}) \right)^{-1} \quad (7)$$

と書くことができ、 $\mathbf{F}_{\leq k, N}^{(\text{NW})}$ の (ℓ, j) 番目 ($\ell = 0, 1, \dots, N, j \in \mathcal{M}_\ell$) の行ベクトルの各要素は、状態 (ℓ, j) から開始された元のマルコフ連鎖が初めてレベル $k+1$ 以上の状態を訪れるまでの初到達時間内にレベル N 以下の各状態で滞在する平均時間である。そこで、 (ℓ, j) 要素が 1 である適当な次元の単位列ベクトル $\mathbf{e}_{(\ell, j)}^T$ を用いて、 $\alpha = \mathbf{e}_{(\ell, j)}^T$ とすると、これは確率 1 で状態 (ℓ, j) から再生するように上記の手順 (i) の変更を施したマルコフ連鎖 $\{(L_{\leq k, N}(t; \mathbf{e}_{(\ell, j)}^T), J_{\leq k, N}(t; \mathbf{e}_{(\ell, j)}^T)); t \geq 0\}$ となるため、マルコフ再生過程に対する再生報酬定理より、 $\bar{\mathbf{F}}_{\leq k, N}^{(\text{NW})}$ の (ℓ, j) 行目は $\bar{\mathbf{f}}_{\leq k, N}(\mathbf{e}_{(\ell, j)}^T)$ と一致する。すなわち、式 (6) を用いると

$$\lim_{k \rightarrow \infty} \bar{\mathbf{F}}_{\leq k, N}^{(\text{NW})} = \lim_{k \rightarrow \infty} \begin{pmatrix} \bar{\mathbf{f}}_{\leq k, N}(\mathbf{e}_{(0,1)}^T) \\ \bar{\mathbf{f}}_{\leq k, N}(\mathbf{e}_{(0,2)}^T) \\ \vdots \\ \bar{\mathbf{f}}_{\leq k, N}(\mathbf{e}_{(N, M_N)}^T) \end{pmatrix} = \begin{pmatrix} \mathbf{q}(N) \\ \mathbf{q}(N) \\ \vdots \\ \mathbf{q}(N) \end{pmatrix}$$

を得る。以上が命題 4 の確率的な解釈である。

次に $\mathbf{q}(N)$ の存在範囲を $\bar{\mathbf{f}}_{\leq k, N}(\mathbf{e}_{(\ell, j)}^T)$ を用いて特徴付ける。そのため、レベル N 以下の状態集合 $\mathcal{L}_{\leq N}$ を二つの互いに素な部分集合に分割する。 $\mathcal{I}_{k, N}^+(\ell, j)$ ($\ell = k+1, k+2, \dots, j \in \mathcal{M}_\ell$) をレベル $k+1$ 以上の状態 $(\ell, j) \in \Omega \setminus \mathcal{L}_{\leq k}$ から $\mathcal{L}_{\leq N}$ への初到達時間の終了時点において取り得る状態のインデックス集合とする。定義より $\mathcal{I}_{k, N}^+(\ell, j) \subseteq \mathcal{L}_{\leq N}$ である。これを用いて $\mathcal{I}_{k, N}^+$ ならびに $\bar{\mathcal{I}}_{k, N}^+$ を次式で定義する。

$$\mathcal{I}_{k, N}^+ = \bigcup_{\ell=k+1}^{\infty} \left(\bigcup_{j \in \mathcal{M}_\ell} \mathcal{I}_{k, N}^+(\ell, j) \right), \quad \bar{\mathcal{I}}_{k, N}^+ = \mathcal{L}_{\leq N} \setminus \mathcal{I}_{k, N}^+$$

$\mathcal{I}_{k,N}^+$ はレベル $k+1$ 以上のいずれかの状態から開始したレベル N 以下への初到達時間の終了時点において取り得る状態のインデックス集合であり、 $\bar{\mathcal{I}}_{k,N}^+$ はレベル N 以下の状態のうち、レベル $k+1$ 以上から開始したマルコフ連鎖がレベル N 以下の他の状態に少なくとも一度遷移した後でなければ到達できない状態のインデックス集合である。

定理 5. 既約で正再帰的な LD-G/G/1 型 MC $\{(L(t), J(t)); t \geq 0\}$ において、次式が成立する。

$$\mathbf{q}(N) \in \text{ri } \mathcal{P}_{k,N}^+$$

ただし、 $\text{ri } \mathcal{P}_{k,N}^+$ は $\{\bar{\mathbf{f}}_{\leq k,N}(\mathbf{e}_{(\ell,j)}^T); (\ell,j) \in \mathcal{I}_{k,N}^+\}$ で張られる凸包 $\mathcal{P}_{k,N}^+$ の相対的内部を表す。

$$\text{ri } \mathcal{P}_{k,N}^+ = \left\{ \mathbf{x} \in \mathbb{R}^{|\mathcal{L}_{\leq N}|} \mid \mathbf{x} = \sum_{(\ell,j) \in \mathcal{I}_{k,N}^+} \alpha(\ell,j) \bar{\mathbf{f}}_{\leq k,N}(\mathbf{e}_{(\ell,j)}^T), \right. \\ \left. \sum_{(\ell,j) \in \mathcal{I}_{k,N}^+} \alpha(\ell,j) = 1, \alpha(\ell,j) > 0 ((\ell,j) \in \mathcal{I}_{k,N}^+) \right\} \quad (8)$$

定理 5 の証明は省略し、以下では、定理 5 の確率的解釈を与える。式 (5) の 1 番目の式の両辺に後ろから $\mathbf{F}_{\leq k,N}^{(\text{NW})}$ を掛け、整理すると

$$\bar{\mathbf{f}}_{\leq k,N}(\boldsymbol{\alpha}) = c \cdot \boldsymbol{\alpha} \mathbf{F}_{\leq k,N}^{(\text{NW})} = c \cdot \boldsymbol{\alpha} \text{diag}(\mathbf{F}_{\leq k,N}^{(\text{NW})} \mathbf{e}) \bar{\mathbf{F}}_{\leq k,N}^{(\text{NW})} \quad (9)$$

を得る。ただし、 c は $\bar{\mathbf{f}}_{\leq k,N}(\boldsymbol{\alpha}) \mathbf{e} = 1$ となるための正規化定数である。式 (9) は $\bar{\mathbf{f}}_{\leq k,N}(\boldsymbol{\alpha})$ が $\{\bar{\mathbf{f}}_{\leq k,N}(\mathbf{e}_{(\ell,j)}^T); (\ell,j) \in \mathcal{L}_{\leq N}\}$ の凸結合で与えられることを示している。センサー過程 $\{(L_N(t), J_N(t)); t \geq 0\}$ において、センサーアウトされている期間に、 $k+1$ 以上へ遷移したとすると、元のマルコフ連鎖の遷移率に従って、やがてレベル N 以下に初到達する。このことから、

$$\bar{\mathbf{f}}_{\leq k,N}(\boldsymbol{\beta}_{k,N}) = \mathbf{q}(N)$$

を満たす $\boldsymbol{\beta}_{k,N}$ が存在することが分かる。すなわち、 $\boldsymbol{\beta}_{k,N}$ は元のマルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ において、レベル k 以下から一度、レベル $k+1$ 以上へ遷移した後、レベル N 以下へ初到達した直後の分布である。そのため $\{(L(t), J(t)); t \geq 0\}$ において、レベル $k+1$ 以上から開始した際、レベル N 以下の他の状態に少なくとも一度遷移した後でなければ到達できない状態 $(\ell,j) \in \bar{\mathcal{I}}_{k,N}^+$ に対応する $\boldsymbol{\beta}_{k,N}$ の (ℓ,j) 番目の要素は 0 となる。さらに、 $\bar{\mathbf{f}}_{\leq k,N}(\boldsymbol{\beta}_{k,N})$ も式 (9) で示されているように $\{\bar{\mathbf{f}}_{\leq k,N}(\mathbf{e}_{(\ell,j)}^T); (\ell,j) \in \mathcal{L}_{\leq N}\}$ の凸結合で表現することができる。以上の議論により、定理 5 が成立することが説明される。

定理 5 より、[Takine 2016, Kimura 2017] と同様に誤差上界の計算も可能となる。

系 6. 既約で正再帰的な LD-G/G/1 型 MC $\{(L(t), J(t)); t \geq 0\}$ において、次式が成立する。

$$\|\mathbf{q}^{\text{comp},k}(N) - \mathbf{q}(N)\| \leq e_F(k, N) \quad (10)$$

ただし、誤差上界 $e_F(k, N)$ は次式で与えられる。

$$e_F(k, N) := \max_{(\ell,j) \in \mathcal{I}_{k,N}^+} \|\mathbf{q}^{\text{comp},k}(N) - \bar{\mathbf{f}}_{\leq k,N}(\mathbf{e}_{(\ell,j)}^T)\| \quad (11)$$

また、 $\mathcal{P}_{k,N}^+$ は k の増加に伴う非拡大性を有し、 $\{\mathbf{q}(N)\}$ に漸近する。

定理 7. 既約で正再帰的な LD-G/G/1 型 MC $\{(L(t), J(t)); t \geq 0\}$ において、次式が成立する。

$$\mathcal{P}_{k+1,N}^+ \subseteq \mathcal{P}_{k,N}^+, \quad k = N, N+1, \dots$$

定理 8. 既約で正再帰的な LD-G/G/1 型 MC $\{(L(t), J(t)); t \geq 0\}$ において、次式が成立する.

$$\bigcap_{k=0}^{\infty} \mathcal{P}_{k,N}^+ = \lim_{k \rightarrow \infty} \mathcal{P}_{k,N}^+ = \{\mathbf{q}(N)\}$$

これら二つの定理から、一般に $\bar{\mathbf{F}}_{\leq k,N}^{(NW)}$ から得られる条件付き定常分布の近似解は、 k の増加に伴って精度が向上し、任意の $\epsilon > 0$ について、 $e_F(k, N) < \epsilon$ を満たす $k = K$ が存在することが分かる.

4. 数値計算手順についてのコメント

命題 4 および系 6 から、条件付き定常分布 $\mathbf{q}(N)$ の近似解 $\mathbf{q}^{\text{comp},k}(N)$ ならびにその誤差上界 $e_F(k, N)$ の数値計算は、 $\mathbf{F}_{\leq k,N}^{(NW)} = (-\mathbf{Q}_{\leq k,N})^{-1}$ の各行を正規化したベクトルを数値計算することに帰着されることが分かる. $\mathbf{Q}_{\leq k,N}$ は任意の k ($k > N$) に対して理論上は正則であるが、 $k \rightarrow \infty$ の極限において、 $\mathbf{Q}_{\leq k,N}$ は元のマルコフ連鎖の（既約な）センサー過程 $\{(L_N(t), J_N(t)); t \geq 0\}$ の無限小生成作用素に収束する. すなわち、 k が大きくなるに従い、正則でない行列へと漸近する. そのため、逆行列 $\mathbf{F}_{\leq k,N}^{(NW)} = (-\mathbf{Q}_{\leq k,N})^{-1}$ を直接、数値計算しようとする、大きな k では数値的に不安定となり、計算精度が保証されない可能性がある.

一方、 $\bar{\mathbf{F}}_{\leq k,N}^{(NW)}$ の各行 $\bar{\mathbf{f}}_{\leq k,N}(\mathbf{e}_{(\ell,j)}^T)$ は、マルコフ連鎖 $\{(L_{\leq k,N}(t; \mathbf{e}_{(\ell,j)}^T), J_{\leq k,N}(t; \mathbf{e}_{(\ell,j)}^T)); t \geq 0\}$ の定常分布と解釈することもできる. すなわち、以下の平衡方程式が成立する.

$$\bar{\mathbf{f}}_{\leq k,N}(\mathbf{e}_{(\ell,j)}^T) \left(\tilde{\mathbf{Q}}_{\leq k,N} + (-\tilde{\mathbf{Q}}_{\leq k,N}) \mathbf{e} \cdot \mathbf{e}_{(\ell,j)}^T \right) = \mathbf{0}, \quad \bar{\mathbf{f}}_{\leq k,N}(\mathbf{e}_{(\ell,j)}^T) \mathbf{e} = 1 \quad (12)$$

初期分布 α が $\mathbf{e}_{(\ell,j)}^T$ で与えられているとき、マルコフ連鎖 $\{(L_{\leq k,N}(t; \mathbf{e}_{(\ell,j)}^T), J_{\leq k,N}(t; \mathbf{e}_{(\ell,j)}^T)); t \geq 0\}$ は状態 (ℓ, j) からレベル $k+1$ 以上の状態を訪れることなしに到達可能な状態からなる $\mathcal{L}_{\leq N}$ の部分集合を唯一の既約クラスにもつ. よって、 $\bar{\mathbf{f}}_{\leq k,N}(\mathbf{e}_{(\ell,j)}^T)$ は式 (12) の解として、一意に定まる.

系 6 では、インデックス集合 $\mathcal{I}_{k,N}^+$ が同定できるという暗黙の仮定が置かれている. しかし、複雑あるいは不規則な構造をもつ一般のマルコフ連鎖では、 $\mathcal{I}_{k,N}^+$ を正確に知ることができない場合があり得る. 真の条件付き分布 $\mathbf{q}(N)$ の存在する領域 $\text{ri } \mathcal{P}_{k,N}^+$ (式 (8) 参照) は全ての $\bar{\mathbf{f}}_{\leq k,N}(\mathbf{e}_{(\ell,j)}^T)$ ($(\ell, j) \in \mathcal{L}_{\leq N}$) で張られる凸包に含まれるため、そのような場合には、

$$\mathcal{I}_{k,N}^+ \subseteq \mathcal{I}_{k,N}^* \subseteq \mathcal{L}_{\leq N}$$

を満たすインデックス集合 $\mathcal{L}_{\leq N}$ の部分集合 $\mathcal{I}_{k,N}^*$ を $\mathcal{I}_{k,N}^+$ の代わりに用いればよい. ただし、 $\mathcal{I}_{k,N}^*$ ($\neq \mathcal{I}_{k,N}^+$) を用いた場合、条件付き定常分布の数値解 $\mathbf{q}^{\text{comp},k}(N)$ がもつ真の誤差と数値的に求められる誤差上界との差は一般に大きくなる.

通常、 $\bar{\mathbf{f}}_{k,N}(\mathbf{e}_{(\ell,j)}^T)$ を得るための計算量は大きい. そのため、 k の値を 1 ずつ増加させる、といった逐次的な手順を実行しようとする、膨大な計算時間が必要となる可能性がある. 一方、本稿の結果を用いれば、有限の k に対して得られた数値解 $\mathbf{q}^{\text{comp},k}(N)$ の誤差が評価できるため、収束判定を目的として逐次的な計算を行う必要はない. 特定の問題に対して数値計算を実行する際、この性質は非常に有用である. 例えば、本稿で議論した数値計算法の利用方法として、以下の二通りの場合を考える.

- (i) 予め興味のある状態空間の部分集合 $\mathcal{L}_{\leq N}$ と数値解がもつ許容誤差が与えられている
- (ii) 数値解がもつ誤差を許容範囲内に押さえつつ、できるだけ大きな N に対して条件付き定常分布 $\mathbf{q}(N)$ を計算したい

(i) の場合、与えられた N に対して $k = k_1$ ($k_1 > N$) を適当に定めてアルゴリズムを実行すれば、数値解 $\mathbf{q}^{\text{comp},k_1}(N)$ と共に誤差上界 $e_F(k_1, N)$ が出力される. もし、誤差上界 $e_F(k_1, N)$ が誤差の許容範囲内に収まっていなければ、新たに $k = k_2$ ($k_2 > k_1$) を定め、再度、アルゴリズムを実行し、誤差が許容範囲に収まっているか判定すればよい. ただし、この手順を繰り返す際に、新しい k の

値をどのように定めれば適切かという問題に関しては、扱うモデル毎に個別に検討する必要がある。なお、 $k = k_2$ に対するアルゴリズムの実行に際して、前段の $k = k_1$ に対する計算結果の一部を流用することが可能であり、この流用によって $k = k_2$ における計算量を大きく減少させることができるが、詳細は省略する。

一方、(ii) の場合には、式 (4) で与えられる $\mathbf{F}_{\leq k}$ の正規化された各行を（計算時間とメモリー容量を勘案して）なるべく大きな $k = K$ に対して求めたのち、これらを用いて、各 N ($N = 0, 1, \dots, K$) に対する $\mathbf{q}^{\text{comp}, K}(N)$ とその誤差上界 $e_F(K, N)$ を求め、誤差上界が誤差の許容範囲内に収まる最も大きな N に対する結果を出力させればよい。

付録. 命題 4 の略証

$\bar{\mathbf{F}}_{\leq k, N}^{(\text{NW})}$ の各行が $k \rightarrow \infty$ の極限において $\mathbf{q}(N)$ に収束することを示す。以下では、任意に固定された N ($N = 0, 1, \dots$) ならびに $(\ell, i) \in \mathcal{L}_{\leq N}$ に対して、 $\bar{\mathbf{F}}_{\leq k, N}^{(\text{NW})}$ の (ℓ, i) 行目を表すベクトルを $\bar{\mathbf{f}}_k$ と略記する。なお、 $\bar{\mathbf{f}}_k$ は 3 節以降で用いられる $\bar{\mathbf{f}}_{\leq k, N}(e_{(\ell, j)}^T)$ と等価である。

まず、定義より

$$\bar{\mathbf{f}}_k = \mathbf{e}_{(\ell, i)}^T \bar{\mathbf{F}}_{\leq k, N}^{(\text{NW})} = c \cdot \mathbf{e}_{(\ell, i)}^T \mathbf{F}_{\leq k, N}^{(\text{NW})}$$

と書くことができる。ただし、 $\mathbf{e}_{(\ell, i)}$ は (ℓ, i) 番目の要素が 1 である単位列ベクトルを表し、 c は $\bar{\mathbf{f}}_k \mathbf{e} = 1$ となるための正規化定数である。ここで $\tilde{\mathbf{Q}}_{\leq k, N} := (-\mathbf{F}_{\leq k, N}^{(\text{NW})})^{-1}$ とする。上式の両辺に $(-\tilde{\mathbf{Q}}_{\leq k, N})\mathbf{e}$ を掛けることにより $c = \bar{\mathbf{f}}_k(-\tilde{\mathbf{Q}}_{\leq k, N})\mathbf{e}$ を得る。よって

$$\bar{\mathbf{f}}_k = \bar{\mathbf{f}}_k(-\tilde{\mathbf{Q}}_{\leq k, N})\mathbf{e} \cdot \mathbf{e}_{(\ell, i)}^T \mathbf{F}_{\leq k, N}^{(\text{NW})} \quad (13)$$

が成立する。さらに、式 (13) の両辺に $(-\tilde{\mathbf{Q}}_{\leq k, N})$ を掛けて整理すると次式を得る。

$$\bar{\mathbf{f}}_k \left(\tilde{\mathbf{Q}}_{\leq k, N} + (-\tilde{\mathbf{Q}}_{\leq k, N})\mathbf{e} \mathbf{e}_{(\ell, i)}^T \right) = \mathbf{0} \quad (14)$$

一方、マルコフ連鎖 $\{(L(t), J(t); t \geq 0)\}$ においてレベル N 以下を観測状態としたセンサー過程 $\{(L_N(t), J_N(t); t \geq 0)\}$ の無限小生成作用素を $\tilde{\mathbf{Q}}_N$ とすると、次式が成立する。

$$\mathbf{q}(N)\tilde{\mathbf{Q}}_N = \mathbf{0} \quad (15)$$

式 (14) および式 (15) から、 $\bar{\mathbf{f}}_k$ は次式を満たすことが分かる。

$$\begin{aligned} (\bar{\mathbf{f}}_k - \mathbf{q}(N)) \left(\mathbf{e}\mathbf{q}(N) - \tilde{\mathbf{Q}}_N \right) &= \bar{\mathbf{f}}_k \left(-\tilde{\mathbf{Q}}_N \right) \\ &= \bar{\mathbf{f}}_k \left(\tilde{\mathbf{Q}}_{\leq k, N} + (-\tilde{\mathbf{Q}}_{\leq k, N})\mathbf{e} \mathbf{e}_{(\ell, i)}^T - \tilde{\mathbf{Q}}_N \right) \end{aligned}$$

定義より、センサー過程 $\{(L_N(t), J_N(t); t \geq 0)\}$ は既約であり、その無限小生成作用素が $\tilde{\mathbf{Q}}_N$ 、定常分布が $\mathbf{q}(N)$ であるため、上式に現れる $\mathbf{e}\mathbf{q}(N) - \tilde{\mathbf{Q}}_N$ は正則である。よって、両辺に $\mathbf{e}\mathbf{q}(N) - \tilde{\mathbf{Q}}_N$ の逆行列を掛けて整理すると

$$\bar{\mathbf{f}}_k = \mathbf{q}(N) + \bar{\mathbf{f}}_k \left(\tilde{\mathbf{Q}}_{\leq k, N} + (-\tilde{\mathbf{Q}}_{\leq k, N})\mathbf{e} \mathbf{e}_{(\ell, i)}^T - \tilde{\mathbf{Q}}_N \right) \left(\mathbf{e}\mathbf{q}(N) - \tilde{\mathbf{Q}}_N \right)^{-1} \quad (16)$$

を得る。

一方、 $\tilde{\mathbf{Q}}_{\leq k, N}$ は $k \rightarrow \infty$ の極限において $\tilde{\mathbf{Q}}_N$ に収束することを示すことができる（証明略）。

$$\lim_{k \rightarrow \infty} \tilde{\mathbf{Q}}_{\leq k, N} = \tilde{\mathbf{Q}}_N$$

さらに、 $\tilde{\mathbf{Q}}_N \mathbf{e} = \mathbf{0}$ より $\lim_{k \rightarrow \infty} (-\tilde{\mathbf{Q}}_{\leq k, N})\mathbf{e} = \mathbf{0}$ も成立する。以上より

$$\lim_{k \rightarrow \infty} \left(\tilde{\mathbf{Q}}_{\leq k, N} + (-\tilde{\mathbf{Q}}_{\leq k, N})\mathbf{e} \mathbf{e}_{(\ell, i)}^T - \tilde{\mathbf{Q}}_N \right) = \mathbf{0}$$

を得る. よって, 式 (16) において $k \rightarrow \infty$ の極限を取ると, 任意に固定された N と (ℓ, i) に対して

$$\lim_{k \rightarrow \infty} \overline{\mathbf{f}}_k = \mathbf{q}(N)$$

が示される. □

謝辞

本研究の一部は JSPS 科研費 16K00034 の助成を受けている.

参考文献

- [Bright 1995] L. Bright and P. G. Taylor, Calculating the equilibrium distribution in level-dependent quasi-birth-and-death process. *Stochastic Models*, 11 (1995) 497–526.
- [Takine 2016] T. Takine, Analysis and computation of the stationary distribution in a special class of level-dependent M/G/1-type and its application to BMAP/M/ ∞ and BMAP/M/c+M queues. *Queueing Systems*, 84 (2016) 49–77.
- [Kimura 2017] M. Kimura and T. Takine, Computing the conditional stationary distribution in Markov chains of level-dependent M/G/1-type, submitted for publications (2017).
- [Masuyama 2017] H. Masuyama, Limit formulas for the normalized fundamental matrix of the northwest-corner truncation of Markov chains: Matrix-infinite-product-form solutions of block-Hessenberg Markov chains. Available online: <https://arxiv.org/abs/1603.07787v5> (2017).

フロー到着とクロストラフィックを考慮した SDN の待ち行列モデルとその解析

牛垣 龍一¹ Phung-Duc Tuan² 河西 憲一³ Yuan-Cheng Lai⁴ Ying-Dar Lin⁵

¹ 東京工業大学 ² 筑波大学 ³ 群馬大学

⁴National Taiwan University of Science and Technology ⁵National Chiao Tung University

概要： Software Defined Network (SDN) は、簡単かつ柔軟にネットワーク環境を構築することができる技術として、近年実用化が進められている。本論文では、1組の SDN スイッチとコントローラをフロー到着とクロストラフィックを有する待ち行列モデルとして表し、解析を行った。送信データはパケットの流れ（フロー）として系に到着し、各フローの先頭パケットのコントロールサーバでの処理が完了したのち、先頭以外のパケットがスイッチキューへの到着を始める。また、コントロールサーバは着目しているスイッチサーバ以外から到着する先頭パケットの処理も行う。性能評価指標として、呼損率や平均滞在時間などを導出した。そして、導出した評価指標とフローの到着率や制御サーバにかかる負荷との関係を考察した。

1. はじめに

近年、ネットワークの仮想化技術として、Software Defined Network (SDN) が注目されている。SDN とは、ネットワークをソフトウェア制御するための技術や、それにより構成されたネットワークを指す。ネットワークをすべてソフトウェア制御することで、設定の変更やネットワークの再構成が容易になったり、ネットワーク全体を見渡したルーティングが可能になったりする利点がある。SDN の中心技術として OpenFlow がある。OpenFlow の特徴は、これまで1つのルータが担っていたデータ転送機能と経路判断機能を分離したことである。ネットワーク内に複数存在する SDN スイッチがデータ転送を行い、SDN コントローラがネットワーク内の複数の SDN スイッチの経路判断を担当する。データは SDN スイッチに到着し、先頭パケットはスイッチ処理完了後に SDN コントローラに向かう。そして、先頭パケットはコントロール処理でデータの転送先が決定された後、もう一度スイッチ処理を受けて転送先に向かう。非先頭パケットは SDN スイッチで処理を受けた後、SDN コントローラが決定した経路にしたがって転送先に向かう。SDN は既に実用化が始まっている。本研究の目的は、ネットワーク設計やシステムの性能保障に役立てるために、SDN を適切にモデル化して有効な評価指標を求めることである。

待ち行列モデルを用いた SDN の解析としては次のようなものが存在する。[1] は、SDN スイッチと SDN コントローラをそれぞれ M/M/1 待ち行列と M/M/1/c 待ち行列を用いてモデル化し、解析を行った。[2] では、SDN が1つのコントローラに対して複数のスイッチを担当することを、ジャクソン・ネットワークを用いてモデル化することで表している。[3] は、送信パケットが独立でなくデータ送信時にまとめて到着することを表すため、 $M^X/M/1$ 待ち行列を用いて SDN スイッチを表した。[4] では、SDN スイッチと SDN コントローラの相互関係を調べるため、3次元待ち行列モデルを用いて1組の SDN スイッチとコントローラを表している。[5] はまとめて到着するパケットを表すために2状態の背後状態を持つ MMPP(Markov Modulated Poisson Process) でパケットの到着過程を表している。[6] は SDN スイッチのサービス時間を SDN コントローラに向かう場合と向かわない場合に分け、解析を行った。ここまで紹介した研究は、系に到着するパケットの中から、確率的に先頭パケットを決定していた。[7] はデータ送信時にまとめて到着するパケットをフロー到着を用いて表現した。すなわち、ある期間続くフローの間に複数個のパケットが到着することで表した。こうすることで、パケットの到着をより現実的に表すと同時に、先頭パケットの判定を正しく行うことに成功している。

本研究では、[7] で提案されたフロー到着を導入した1組の SDN スイッチとコントローラでの TCP 通信を待ち行列を用いてモデル化し、解析を行った。このモデル化では、SDN コントローラの着目するスイッチ以外から到着する先頭パケットの処理も考慮した。これにより、SDN コントローラが複数の SDN スイッチを担当する場合の解析が可能になった。複数の SDN スイッチを担当する場合の SDN コントローラの解析には、[8] などがある。そして、性能評価のための指標として呼損率や平均滞在時間などを計算し、シミュレーションと比較した。また、このモデルは複雑になりやすいため、評価指標を高速に計算できる近似モデルの提案を行い、もとのモデルと比較した。

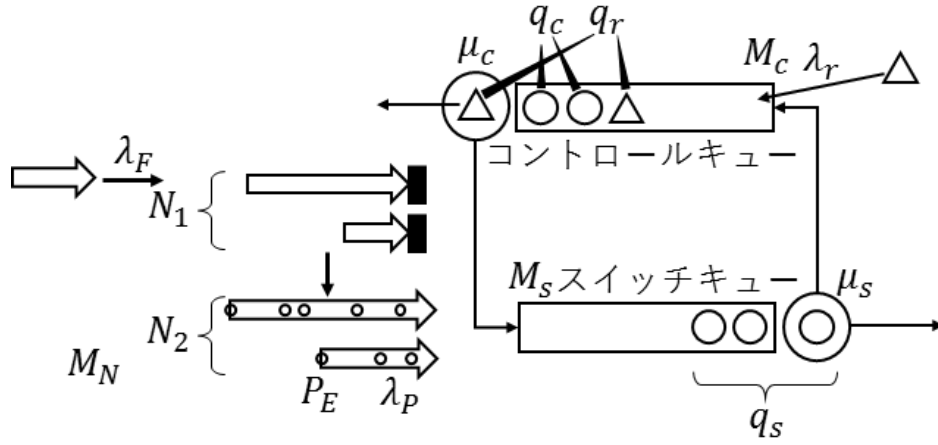


図 1: 5次元モデル (記号は表 1 参照)

表 1: 記号の定義

記号	定義
λ_F	フロー到着率
λ_P	フローあたりのパケット到着率
P_E	パケット到着時フロー終了確率
λ_r	非着目スイッチからの先頭パケット到着率

記号	定義
μ_s	スイッチサービス率
μ_c	コントロールサービス率
M_N	接続フローの最大数
M_s	スイッチキュー容量
M_c	コントロールキュー容量

2. SDN における TCP 通信

本節では、本研究でモデル化する SDN における TCP 通信でのパケットの動きについて説明する。TCP とは、インターネットで標準的に利用されているプロトコルの一つである。TCP を用いてデータを送る際、まず送信者は受信者に通信コネクションを確立するためのパケット (SYN パケット) を送信する。SYN パケットを受け取った受信者は、受信可能であることを送信者に伝えるパケット (ACK パケット) を送り返す。そして、送信者は受信者からの ACK パケットを受け取り次第、データを送り始める。SDN において TCP 通信をする際は、まず SYN パケットがスイッチキューに入り、スイッチ処理を受ける。次に、SYN パケットはコントロールキューに入り、コントロール処理を受ける。そして SYN パケットはもう一度スイッチサービスを受けて転送先へ向かう。その後受信者からの ACK パケットが到着次第、パケットに分割されたデータがスイッチキューに到着する。このデータの転送先は SYN パケットのコントロール処理時に決定されているので、データはスイッチ処理を受けた後、コントロールキューを経由せずに転送先へ向かう。モデル化の簡単のため、本研究では SYN パケットがコントロール処理を終えてから ACK パケットが届くまでの時間が小さいと仮定し、SYN パケットのコントロール処理が終わり次第データの転送が始まるとしている。また、本論文では、SYN パケットを先頭パケット、データを分割したパケットを非先頭パケットと呼ぶ。

3. 5次元待ち行列モデル

本節では 1 組の SDN スイッチとコントローラにおける、TCP 通信時のパケットの動きを、待ち行列を用いてモデル化する。まず、3.1 節でモデル化について説明し、状態遷移を定める。そして、3.2 節で評価指標の計算方法を述べる。モデル化に際しては、[7] で提案されたフロー到着を取り入れ、さらに SDN コントローラの着目するスイッチ以外からの先頭パケットの処理も考慮した。これらについては、3.1.1 節と 3.1.2 節で説明する。

3.1. 待ち行列モデル定義

図 1 は SDN の待ち行列モデルを表している。系には 1 つのスイッチと 1 つのコントローラがあり、それぞれが有限の大きさの待合室を持つ。系にはフローがパラメータ λ_F のポアソン到着をする。モデル化の簡単のため、同時に接続できるフローの数、すなわち同時に転送できるデータの数の最大値を M_N とする。実際には同時に接続するフロー数に制限はないため、 M_N はできるだけ大きな値を設定

表 2: $(N_1, N_2, q_s, q_c, q_r)$ からの遷移

number	to	rate	condition
遷移 1	$(N_1 + 1, N_2, q_s + 1, q_c, q_r)$	λ_F	$N_1 + N_2 < M_N, q_s < M_s$
遷移 2	$(N_1, N_2, q_s + 1, q_c, q_r)$	$N_2 \lambda_P (1 - P_E)$	$q_s < M_s, N_2 > 0$
遷移 3	$(N_1, N_2 - 1, q_s + 1, q_c, q_r)$	$N_2 \lambda_P P_E$	$q_s < M_s, N_2 > 0$
遷移 4	$(N_1, N_2 - 1, q_s, q_c, q_r)$	$N_2 \lambda_P P_E$	$q_s = M_s, N_2 > 0$
遷移 5	$(N_1, N_2, q_s, q_c, q_r + 1)$	λ_r	$q_c + q_r < M_c$
遷移 6	$(N_1, N_2, q_s - 1, q_c + 1, q_r)$	$\mu_s(N_1 - q_c)/q_s$	$N_1 - q_c > 0, q_c + q_r < M_c$
遷移 7	$(N_1 - 1, N_2, q_s - 1, q_c, q_r)$	$\mu_s(N_1 - q_c)/q_s$	$N_1 - q_c > 0, q_c + q_r = M_c$
遷移 8	$(N_1, N_2, q_s - 1, q_c, q_r)$	$\mu_s(q_s - N_1 + q_c)/q_s$	$q_s - N_1 + q_c > 0$
遷移 9	$(N_1 - 1, N_2 + 1, q_s + 1, q_c - 1, q_r)$	$\mu_c q_c / (q_c + q_r)$	$q_c > 0, q_s < M_s$
遷移 10	$(N_1 - 1, N_2 + 1, q_s, q_c - 1, q_r)$	$\mu_c q_c / (q_c + q_r)$	$q_c > 0, q_s = M_s$
遷移 11	$(N_1, N_2, q_s, q_c, q_r - 1)$	$\mu_c q_r / (q_c + q_r)$	$q_r > 0$

することが望ましい。

フローが到着すると、先頭パケットのみがスイッチキューに入る。スイッチサーバでのサービス時間は、それぞれ独立に、パラメータ μ_s の指数分布に従うとする。スイッチサービスが完了した先頭パケットはコントロールキューに入る。コントロールキューには着目しているスイッチ以外からの先頭パケットも到着する。この到着過程は、パラメータ λ_r のポアソン到着とする。コントロールサーバでのサービス時間は、それぞれ独立に、パラメータ μ_c の指数分布に従うとする。

先頭パケットがコントロールサービスを終わると、対応するフローで非先頭パケットの到着が始まる。非先頭パケットはフローごとにパラメータ λ_P のポアソン到着をし、到着するごとに確率 P_E でフローが終了する。到着した非先頭パケットはスイッチキューに入り、先頭パケットと同じサービス率 μ_s でサービスを受ける。サービス完了後は系を離脱する。スイッチキューとコントロールキューの最大容量をそれぞれ M_s, M_c とする。

接続フロー数が上限値のときに到着したフローは呼損する。スイッチキュー、コントロールキューがいっぱいのときにそれぞれのキューに到着したパケットも呼損する。先頭パケットがコントロール処理を終える前に呼損した場合、該当フローも失う。

本来、サーバはサービスを先着順に行うが、これをモデル化するためにはキューの並び順を記憶しておく必要があり、モデルが非常に複雑になってしまう。そこで本論文では、サービスが完了したときにキュー内にいるパケット数の比に応じて確率的にサービス完了パケットを決定するように近似している。この近似の妥当性は、数値実験の際に先着順サービスの場合のシミュレーションと比較を行うことで確かめる。

このモデルで、時刻 t において、先頭パケットのコントロールサービスが完了前のフローの数を $N_1(t)$ 、先頭パケットのコントロールサービスが完了後のフローの数を $N_2(t)$ 、スイッチキュー内のパケットの数を $q_s(t)$ 、コントロールキュー内の着目スイッチから転送された先頭パケットの数を $q_c(t)$ 、コントロールキュー内の非着目スイッチから転送された先頭パケットの数を $q_r(t)$ とする。 $X(t) = \{(N_1(t), N_2(t), q_s(t), q_c(t), q_r(t))\}$ とおくと、確率過程 $\{X(t); t \geq 0\}$ は以下のような有限な離散状態空間 S を持つ 5 次元連続時間マルコフ連鎖である。

$$S = \{(N_1, N_2, q_s, q_c, q_r); N_1 + N_2 \leq M_N, q_s \leq M_s, q_c + q_r \leq M_c, q_c \leq N_1 \leq q_s + q_c, N_1, N_2, q_s, q_c, q_r \geq 0\}.$$

確率過程 $\{X(t)\}$ の状態 $(N_1, N_2, q_s, q_c, q_r)$ からの遷移は、表 2 のようになる。マルコフ連鎖の平衡方程式と正規化条件から、 $\{X(t)\}$ の定常分布を数値的に計算することができる。

以下、本モデルの特徴であるフロー到着とクロストラフィックについて説明する。

3.1.1. フロー到着

[7] では 1 つのデータを送信する際のデータの流れ（フロー）を、指数分布に従うフロー継続時間中にパケットがポアソン到着することで表した。本論文では、フローが継続している間、パケットがポアソ

ン到着し、到着する度に確率的にフローが終了している。以下の定理より、本論文のモデルは [7] のフロー到着を含むことが分かる。

定理 1. フローの長さをパラメータ μ_{F1} の指数分布に従う確率変数とし、その間パラメータ λ_{P1} のポアソン到着を行うパケットの到着を到着過程 1 とする。パラメータ λ_{P2} のポアソン到着を行い、到着する度に確率 P_{E2} でそれ以降の到着が無くなるパケットの到着を到着過程 2 とする。各パラメータが以下の条件を満たすとき、パケットが 1 つ以上到着する条件の下でのフロー内のパケットの個数の分布は一致する。また、フロー内のパケットの間隔の分布も一致する。

$$P_E = \frac{\mu_{F1}}{\lambda_{P1} + \mu_{F1}}, \quad \lambda_{P2} = \lambda_{P1} + \mu_{F1}.$$

3.1.2. クロストラフィック

SDN では、1 つのコントローラが複数のスイッチの経路判断を担い、全体を通したルーティングを行っている。本論文のモデルでは、1 つのコントローラが複数のスイッチからの先頭パケットの処理をすることを表すために、系外からコントロールキューに到着しコントロール処理終了後に系外へ離脱する非着目スイッチからの先頭パケットをモデル化している。これにより、先頭パケットの動きがより現実に近いものになり、さらに SDN コントローラへの負荷が与える影響についての解析が可能になった。

3.2. 評価指標計算

本論文では性能評価指標として、連続時間マルコフ連鎖 $\{X(t)\}$ の定常分布から、フロー呼損率、パケット呼損率、パケット平均滞在時間、フロー平均滞在時間の 4 つを導出した。本節では、それぞれの評価指標についての説明と計算方法を示す。 $\{X(t)\}$ が定常状態において状態 (i, j, k, l, m) をとる確率を $\pi(i, j, k, l, m)$ と記す。

3.2.1. フロー呼損率

フロー呼損率 B_F はフロー到着時に接続フロー数が最大であるか、コントロールサービス完了前に先頭パケットが呼損して全体が失われるフローの割合であり、次のように計算される。

$$\begin{aligned} B_F &= 1 - P(\text{フロー到着時に先頭パケットがスイッチキューに入る}) \\ &\quad \times P(\text{スイッチサービスを終えた先頭パケットがコントロールキューに入る}) \\ &= 1 - P(\text{フロー到着時に } N_1 + N_2 < M_N, q_s < M_s) \\ &\quad \times P(\text{スイッチサービス完了時に } q_c + q_r < M_c) \\ &= 1 - \left(\sum_{\substack{(N_1, N_2, q_s, q_c, q_r) \in S \\ N_1 + N_2 < M_N, q_s < M_s}} \pi(N_1, N_2, q_s, q_c, q_r) \right) \left(\sum_{\substack{(N_1, N_2, q_s, q_c, q_r) \in S \\ q_c + q_r < M_c}} \pi_c(N_1, N_2, q_s, q_c, q_r) \right). \end{aligned}$$

ただし、 π_c は定常状態において先頭パケットが 1 度目のスイッチ処理を終えた時点のパルム分布であり、以下のように与えられる。

$$\pi_c(N_1, N_2, q_s, q_c, q_r) = \frac{\pi(N_1, N_2, q_s, q_c, q_r) \times \mu_s \frac{N_1 - q_c}{q_s}}{\sum_{(i, j, k, l, m) \in S} \pi(i, j, k, l, m) \times \mu_s \frac{i - l}{k}}.$$

3.2.2. パケット呼損率

パケット呼損率 B_P は、呼損しなかったフローの非先頭パケットのうち、スイッチキュー到着時にキューがいっぱいで呼損するパケットの割合であり、次のように計算される。

$$B_P = \sum_{(N_1, N_2, q_s, q_c, q_r) \in S, q_s = M_s} \pi_p(N_1, N_2, q_s, q_c, q_r).$$

ただし、 π_p は定常状態において非先頭パケットが到着した時点のパルム分布で、次式で与えられる。

$$\pi_p(N_1, N_2, q_s, q_c, q_r) = \frac{\pi(N_1, N_2, q_s, q_c, q_r) \times N_2 \lambda_P}{\sum_{(i, j, k, l, m) \in S} \pi(i, j, k, l, m) \times j \lambda_P}.$$

3.2.3. パケット平均滞在時間

パケット平均滞在時間 $E[D_P]$ は、スイッチキューに入った非先頭パケットの滞在時間の平均である。スイッチキュー内に q_s 個のパケットが存在するときに到着したパケットの平均滞在時間 $E[T|q_s]$ は、 Y_i を i 番目のパケットの残りサービス時間とすると、サービス時間の独立性と指数分布の無記憶性より、

$$E[T|q_s] = E\left[\sum_{i=1}^{q_s+1} Y_i\right] = (q_s + 1) \times \frac{1}{\mu_s}$$

と表される。したがって $E[D_P]$ は、次のように計算される。

$$\begin{aligned} E[D_P] &= \sum_{(N_1, N_2, q_s, q_c, q_r) \in S} \tilde{\pi}_{p(N_1, N_2, q_s, q_c, q_r)} \times E[T|q_s] \\ &= \sum_{(N_1, N_2, q_s, q_c, q_r) \in S} \tilde{\pi}_{p(N_1, N_2, q_s, q_c, q_r)} \times (q_s + 1) \times \frac{1}{\mu_s}. \end{aligned}$$

ただし、 $\tilde{\pi}_p$ は定常状態において非先頭パケットがスイッチキューに入った時点でのパルム分布で、次式で与えられる。

$$\tilde{\pi}_{p(N_1, N_2, q_s, q_c, q_r)} = \begin{cases} \frac{\pi(N_1, N_2, q_s, q_c, q_r) \times N_2 \lambda_P}{\sum_{(i, j, k, l, m) \in S, k < M_s} \pi(i, j, k, l, m) \times j \lambda_P}, & \text{if } q_s < M_s, \\ 0, & \text{otherwise.} \end{cases}$$

3.2.4. フロー平均滞在時間

フロー平均滞在時間 $E[D_F]$ は、フローが到着してから、最後のパケットが系を離脱するまでの平均時間である。これは、フローが到着してから先頭パケットがスイッチ処理を終えるまでの時間 T_{D1} 、スイッチ処理を終えてからコントロール処理を終えるまでの時間 T_{D2} 、コントロール処理を終えてから最終パケットが到着するまでの時間 T_{D3} 、最終パケットの系内滞在時間 T_{D4} の和の期待値として表される。 N_P をフロー内非先頭パケット数とすると、 T_{D3} の期待値 $E[T_{D3}]$ は次のように計算される。

$$E[T_{D3}] = E[E[T_{D3}|N_P]] = E\left[N_P \times \frac{1}{\lambda_P}\right] = \frac{1}{P_E \lambda_P}.$$

一方、 T_{D1} 、 T_{D2} 、 T_{D4} の期待値 $E[T_{D1}]$ 、 $E[T_{D2}]$ 、 $E[T_{D4}]$ は、先頭パケットがスイッチキュー内に入った時点でのパルム分布 $\tilde{\pi}_f$ と、先頭パケットがコントロールキューに入った時点でのパルム分布 $\tilde{\pi}_c$ と、非先頭パケットがスイッチキューに入った時点でのパルム分布 $\tilde{\pi}_p$ を用いて次のように計算される。

$$\begin{aligned} E[T_{D1}] &= \sum_{(N_1, N_2, q_s, q_c, q_r) \in S} \tilde{\pi}_f(N_1, N_2, q_s, q_c, q_r) \times (q_s + 1) \times \frac{1}{\mu_s}, \\ E[T_{D2}] &= \sum_{(N_1, N_2, q_s, q_c, q_r) \in S} \tilde{\pi}_c(N_1, N_2, q_s, q_c, q_r) \times (q_c + q_r + 1) \times \frac{1}{\mu_c}, \\ E[T_{D4}] &= \sum_{(N_1, N_2, q_s, q_c, q_r) \in S} \tilde{\pi}_p(N_1, N_2, q_s, q_c, q_r) \times (q_s + 1) \times \frac{1}{\mu_s}. \end{aligned}$$

ただし、

$$\begin{aligned} \tilde{\pi}_f(N_1, N_2, q_s, q_c, q_r) &= \begin{cases} \frac{\pi(N_1, N_2, q_s, q_c, q_r)}{\sum_{\substack{(i, j, k, l, m) \in S \\ k < M_s}} \pi(i, j, k, l, m)}, & \text{if } q_s < M_s, \\ 0, & \text{otherwise.} \end{cases} \\ \tilde{\pi}_c(N_1, N_2, q_s, q_c, q_r) &= \begin{cases} \frac{\pi(N_1, N_2, q_s, q_c, q_r)}{\sum_{\substack{(i, j, k, l, m) \in S \\ l+m < M_c}} \pi(i, j, k, l, m)}, & \text{if } q_c + q_r < M_c, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

4. 高速計算可能な近似モデル

3節ではSDNを5次元の連続時間マルコフ連鎖を用いてモデル化した。このモデルは精密ではあるが状態数が多くなりやすく、特に M_N の値を大きくすると数値計算に時間を要するという欠点があるため、高速計算可能なSDNの近似モデルを考える。具体的には、モデルをそれぞれ独立に動くSDNスイッチとSDNコントローラに分割し、さらにSDNスイッチについては有界な準出生死滅過程に対するアルゴリズム [9] を用いることで計算を高速化する。本節では、この近似モデルと評価指標について述べる。

4.1. SDN スwitchの近似モデル

本節では、SDN スwitchの近似モデルについて説明する。本研究ではこれをMAP/M/1/ M_s 待ち行列で表す。系には、フローがパラメータ λ_F のポアソン到着をする。同時に接続できるフロー数は最大 M_N 本とする。フローが系に到着した際、1つのパケットが先頭パケットとしてキューに到着する。キューの最大容量は M_s とする。フローが到着したときにキューがいっぱいで先頭パケットが呼損した場合、フローも呼損する。フローごとに非先頭パケットがパラメータ λ_P のポアソン到着をし、到着するごとに確率 P_E でフローが終了する。また、フローとは関係なく、コントロールキュー処理を終えた先頭パケットがパラメータ λ_c のポアソン到着をする。サービスは先着順とし、サービス時間はどのパケットも独立にパラメータ μ_s の指数分布に従うとする。キューに入ったパケットはキュー内で区別せず、サービスを終えたパケットはすべて系外へと離脱する。

このモデルで、時刻 t でのキュー内のパケット数を $q(t)$ 、接続フロー数を $N(t)$ とする。 $Y(t) = \{(q(t), N(t))\}$ とおくと、確率過程 $\{Y(t); t \geq 0\}$ は有限な離散状態空間 $\hat{S} = \{(i, j); i \leq M_s, j \leq M_N, i, j \geq 0\}$ を持つ連続時間マルコフ連鎖である。このマルコフ連鎖の無限小生成作用素 Q は次のようになる。ただし、 $A_0, A, A_{M_s}, \Lambda_P, M$ はそれぞれ $(M_N + 1) \times (M_N + 1)$ 行列である。

$$Q = \begin{pmatrix} A_0 & \Lambda_P & & & \\ M & A & \Lambda_P & & O \\ & M & A & \Lambda_P & \\ & & \ddots & \ddots & \ddots \\ & O & & M & A & \Lambda_P \\ & & & & M & A_{M_s} \end{pmatrix},$$

$$\begin{aligned} A_0(i, j) &= \begin{cases} -(\lambda_c + \lambda_F + i\lambda_P), & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} & \Lambda_P(i, j) &= \begin{cases} \lambda_F, & \text{if } i = j - 1, \\ \lambda_c + i\lambda_P(1 - P_E), & \text{if } i = j, \\ i\lambda_PP_E, & \text{if } i = j + 1, \\ 0, & \text{otherwise,} \end{cases} \\ A(i, j) &= \begin{cases} -(\lambda_c + \lambda_F + \mu_s + i\lambda_P), & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \\ A_{M_s}(i, j) &= \begin{cases} -(\mu_s + i\lambda_PP_E), & \text{if } i = j, \\ i\lambda_PP_E, & \text{if } i = j + 1, \\ 0, & \text{otherwise,} \end{cases} & M(i, j) &= \begin{cases} \mu_s, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

無限小生成作用素 Q を見ると、このマルコフ連鎖 $\{Y(t)\}$ が有限な準出生死滅過程になっていることが分かる。有限な準出生死滅過程の定常分布を求めるアルゴリズム [9] に従い、 $\{Y(t)\}$ の定常分布 $p(q, N)$ を求める。

4.2. SDN コントローラの近似モデル

本節では、SDN コントローラの近似モデルについて説明する。本研究ではこれを、SDN スwitchとは独立なM/M/1/ M_c 待ち行列で表す。キューにはパラメータ λ_s で着目スイッチからの先頭パケットがポアソン到着し、それ以外のスイッチからのパケットがパラメータ λ_r のポアソン到着をする。キューの最大容量を M_c とする。サービスは先着順とし、サービス時間はどのパケットも独立にパラメータ μ_c の指数分布に従うとする。キューに入ったパケットはキュー内で区別せず、サービスを終えたパケットはすべて系外へと離脱する。

このモデルで、時刻 t でのキュー内のパケット数を $Z(t)$ とおくと、確率過程 $\{Z(t); t \geq 0\}$ は有限な離散状態空間 $S_c = \{i; 0 \leq i \leq M_c\}$ を持つ連続時間マルコフ連鎖である。トラフィック強度を

$\rho_c = (\lambda_s + \lambda_r)/\mu_c$ とすると, $\rho_c \neq 1$ のとき, このコントロールキューでの呼損率 $\hat{B}_c$ と平均滞在時間 $E[\hat{T}_c]$ はそれぞれ以下のように計算できる.

$$\hat{B}_c = \frac{1 - \rho_c}{1 - \rho_c^{M_c+1}} \rho_c^{M_c},$$

$$E[\hat{T}_c] = \frac{\rho_c}{1 - \rho_c} \frac{1 - \rho_c^{M_c-1}}{1 - \rho_c^{M_c}} - \frac{(M_c - 1)\rho_c^{M_c}}{1 - \rho_c^{M_c}}.$$

4.3. 近似評価指標計算

本節では近似モデルを用いた各評価指標の計算方法について説明する. 3.2 節で求めた各評価指標を, 4.1 節と 4.2 節で求めた値を用いて近似する.

- フロー呼損率 $\hat{B}_F$

$$\hat{B}_F = \sum_{\substack{(q,N) \in \hat{S} \\ q=M_s \cup N=M_N}} p_{(q,N)}.$$

- パケット呼損率 $\hat{B}_P$

$$\hat{B}_P = \sum_{(q,N) \in \hat{S}, q=M_s} p_{p(q,N)}.$$

ただし, p_p は定常状態において非先頭パケットが到着した時点のパルム分布であり, 以下のよう
に表される.

$$p_{p(q,N)} = \frac{p_{(q,N)} \times N\lambda_P}{\sum_{(i,j) \in \hat{S}} p_{(i,j)} \times j\lambda_P}.$$

- パケット平均滞在時間 $E[\hat{D}_P]$

$$E[\hat{D}_P] = \sum_{(q,N) \in \hat{S}} \tilde{p}_{p(q,N)} \times (q+1) \times \frac{1}{\mu_s}.$$

ただし, $\tilde{p}_p$ は定常状態において非先頭パケットがスイッチキューに入った時点でのパルム分布
であり, 以下のよう表される.

$$\tilde{p}_{p(q,N)} = \begin{cases} \frac{p_{(q,N)} \times N\lambda_P}{\sum_{(i,j) \in \hat{S}, q < M_s} p_{(i,j)} \times j\lambda_P}, & \text{if } q < M_s, \\ 0, & \text{otherwise.} \end{cases}$$

- フロー平均滞在時間 $E[\hat{D}_F]$

フロー滞在時間を求めるための 4 つの時間の期待値の近似 $E[\hat{T}_{D1}]$, $E[\hat{T}_{D2}]$, $E[\hat{T}_{D3}]$, $E[\hat{T}_{D4}]$ は
それぞれ次のように計算される.

$$E[\hat{T}_{D1}] = \sum_{(q,N) \in \hat{S}} \tilde{p}_{(q,N)} \times (q+1) \times \frac{1}{\mu_s}, \quad E[\hat{T}_{D2}] = E[\hat{T}_c], \quad E[\hat{T}_{D3}] = \frac{1}{P_E \lambda_P}, \quad E[\hat{T}_{D4}] = \hat{D}_P.$$

ただし, $\tilde{p}$ は定常状態において先頭パケットがスイッチキューに入った時点でのパルム分布であ
り, 以下のよう表される.

$$\tilde{p}_{(q,N)} = \begin{cases} \frac{p_{(q,N)}}{\sum_{(i,j) \in \hat{S}, q < M_s} p_{(i,j)}}, & \text{if } q < M_s, \\ 0, & \text{otherwise.} \end{cases}$$

$E[\hat{D}_F]$ は, これらを用いて次のように計算される.

$$E[\hat{D}_F] = E[\hat{T}_{D1}] + E[\hat{T}_{D2}] + E[\hat{T}_{D3}] + E[\hat{T}_{D4}].$$

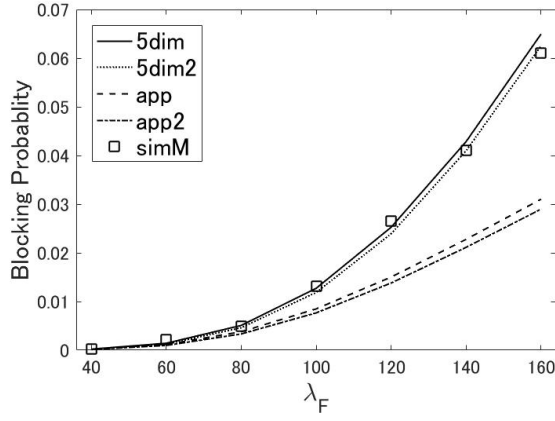


図 2: フロー到着時呼損率 $M_N = 6$

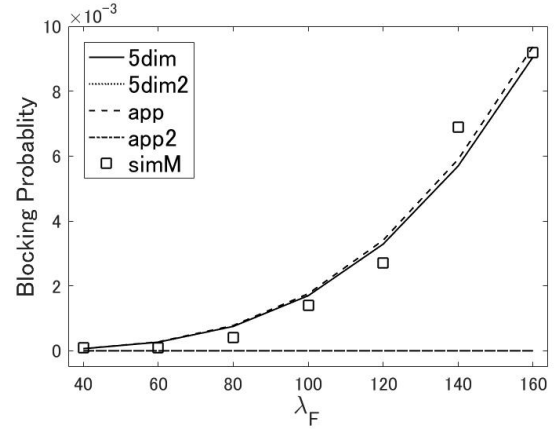


図 3: フロー到着時呼損率 $M_N = 15$

5. 数値実験

本節では、提案モデルから求めた評価指標値とシミュレーション値の比較を行う。3節のモデル化では簡単のため、サービス完了パケットを確率的に決定していたが、シミュレーションは先着順サービスで評価値を求めた。また、実際のスイッチ・コントローラの処理時間は定数に近いことを考慮し、サービス時間が指数分布に従う場合と平均が同一の定数の場合の2通りでシミュレーションを行った。サービス順とサービス時間以外の条件は3.1節と同一である。シミュレーションは系にフロー・パケットが存在しない状態からスタートし、フロー呼損率・フロー平均滞在時間はそれぞれ10000個目のフローまで、パケット呼損率・パケット平均滞在時間はそれぞれ1000000個目のパケットまで発生させて評価値を導出した。

5.1節で近似モデルの精度の評価とモデルの状態数の比較を行い、5.2節と5.3節で性能評価を行う。本節では特に指定のない限り、フロー到着率 $\lambda_F = 100$ 、フローあたりのパケット $\lambda_P = 4000$ 、フロー終了確率 $P_E = 1/80$ 、着目パケット外からの先頭パケット到着率 $\lambda_r = 10000$ 、スイッチサービス率 $\mu_s = 30000$ 、コントロールサービス率 $\mu_c = 20000$ 、最大接続フロー数 $M_N = 15$ 、スイッチキュー容量 $M_s = 8$ 、コントロールキュー容量 $M_c = 6$ として計算を行った。また、今回はコントロールキューでの呼損率は十分小さいとして、近似モデルにおけるコントロールキューからスイッチキューへの到着率 λ_c は、 λ_F と同じ値で計算した。

5.1. 近似モデルの評価

図2・図3はそれぞれ $M_N = 6, 15$ のときの、フロー到着時呼損率と、そのうち接続フロー数不足による呼損率を表している。それぞれ、フロー到着率 λ_F を40から160まで変化させて計算している。5dimが3節の5次元モデルから求めたフロー到着時の呼損率、appが4節で提案した近似モデルから求めた呼損率、simMはサービス時間が指数分布に従う場合のシミュレーションでの呼損率を表す。また、5dim2とapp2はそれぞれ5次元モデルと近似モデルから求めた、フロー到着時に接続可能フロー数の不足が原因で呼損する確率を表す。これらは各モデルの定常分布から求められる。

表3には5次元モデル、近似モデルの $M_N = 6, 15$ のときのマルコフ連鎖の状態数を示す。近似モデルに関しては、SDNコントローラの評価指標は解析的に求められるので、SDNスイッチの近似モデルの状態数を示す。

表 3: モデル状態数比較

	$M_N = 6$	$M_N = 15$
5次元モデル $ S $	3402	14280
近似モデル $ \hat{S} $	63	144

図2は、フローの最大接続数が比較的少ない $M_N = 6$ の場合の結果を表している。このとき、フロー呼損の原因の多くは接続可能フロー数の不足によるものである。図2を見ると、フロー到着率 λ_F が大きくなるほど5次元モデルで求めた値と近似モデルで求めた値の差が大きくなっていることが分かる。これは、5次元モデルでのフローが、先頭パケットがコントロール処理を終えるまでの時間の分だけ長く滞在するためだと考えられる。到着時に呼損するフローの割合が異なると、系に到着する非先頭パ

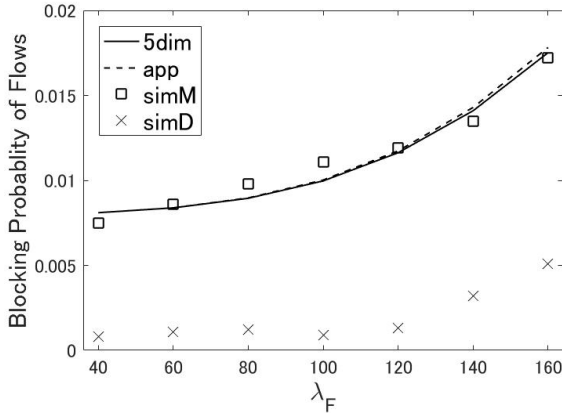


図 4: フロー呼損率 $B_F, \hat{B}_F$

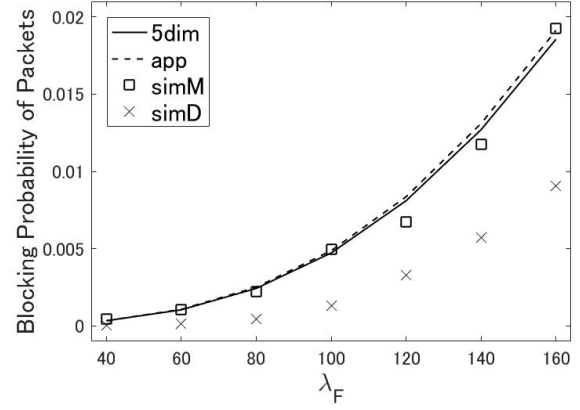


図 5: パケット呼損率 $B_P, \hat{B}_P$

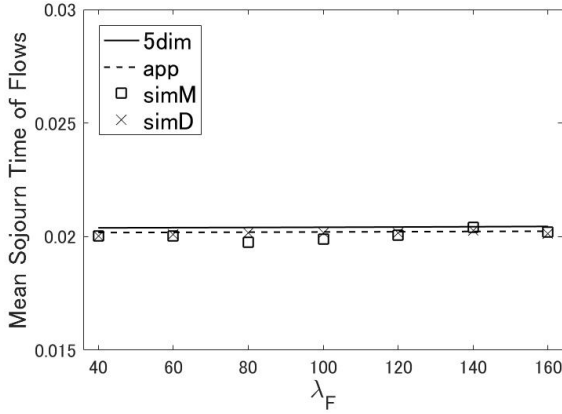


図 6: フロー平均滞在時間 $E[D_F], E[\hat{D}_F]$

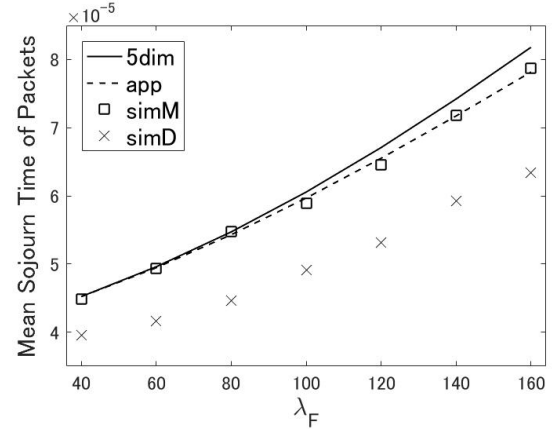


図 7: パケット平均滞在時間 $E[D_P], E[\hat{D}_P]$

ケットの数も大きく変化するので、各評価指標の値も5次元モデルの場合と近似モデルの場合で差異が見られた。また、このとき、シミュレーション値は5次元モデルから求めた値とほぼ等しくなった。

図3はフローの最大接続数が多く、接続可能フロー数の不足がフロー呼損の原因になりにくい $M_N = 15$ の場合の結果を表している。このとき、5次元モデルから求めた呼損率と近似モデルから求めた呼損率はほぼ一致していることが分かる。これは、呼損率が接続可能フロー数の上限の影響を受けにくいいため、フローがとどまる時間の差の影響が小さくなるからだと考えられる。今回、モデル化の簡単のために接続可能フロー数の上限を設定しているが、実際は同時にSDNで送られるデータ数に上限はない。接続可能フロー数が系に影響を与えないほど十分に大きい状態での計算を行ううえで、この近似モデルは有用であることが分かる。

5.2. 評価指標とフロー到着率の関係

本節では、3.2節と4.3節で求めた各評価指標を、フロー到着率 λ_F を40から160まで変化させて計算した。各評価指標に対して、5次元モデル、近似モデル、サービス時間が指数分布に従う場合のシミュレーション、サービス時間が定数の場合のシミュレーションから評価値を求めた。5dimが5次元モデルから求めた評価値、appが近似モデルから求めた評価値、simMがサービス時間が指数分布に従う場合のシミュレーションで求めた評価値、simDがサービス時間が定数の場合のシミュレーションで求めた評価値を表す。図4にフロー呼損率、図5にパケット呼損率、図6にフロー平均滞在時間、図7にパケット平均滞在時間を示す。

図4から図7を見ると、特に呼損率に関して、サービス時間が定数の場合のシミュレーションから求めた値は、他の場合と比較して差が出ている。これは、指数サービスは定数サービスに比べて分散が大きいいため、キュー長がとて長くなるという極端な状態になりやすく、呼損がおりやすいためと考えられる。モデルから求めた値は定数サービスのシミュレーションの上界になっているので、性能保障などに用いることはできると考える。提案モデルから導出した値と指数分布に従うサービス時間を用いたシミュレーション値を比較すると、どの評価指標でも大きな差は見られなかった。これは、

モデル化の際に用いたサービス完了パケットを確率的に決める近似が妥当であることを示している。また、図 6 は、今回の数値実験のパラメータではフロー平均滞在時間がフローの到着率に依存していないことを示している。これは、今回のパラメータではフローの継続時間 $E[T_{D3}]$ が $E[T_{D1}]$, $E[T_{D2}]$, $E[T_{D4}]$ に比べて長すぎることに、フローの平均滞在時間が $E[T_{D3}] = 0.02$ とほとんど同じ値になっているためである。

5.3. 評価指標と着目スイッチ以外からの到着率の関係

本論文では 1 組の SDN スイッチとコントローラのモデル化を行ったが、その中でネットワーク内の着目スイッチ以外のスイッチからコントローラに到着するパケットも組み込んだモデル化をした。これにより、コントロールサーバへの負荷が評価指標にどのような影響を与えるかを調べることができるようになった。これを用いることで、SDN ネットワークを構築する際に、ネットワークの性能を落とさずに、1 つのコントローラで何台までスイッチを担うことができるかなどを評価可能になった。

6. まとめ

本研究では、フロー到着とクロストラフィックを考慮した SDN の待ち行列モデルを考案した。提案したモデルから SDN の性能を測るいくつかの評価指標を計算し、それらを高速に計算可能な近似モデルを作成した。そして 2 つのモデルから導出された評価指標を、シミュレーションから求めた値と比較した。今後の課題としては、UDP などの TCP 以外のプロトコルのモデル化や、待ち行列ネットワークを用いて複数のスイッチを表すことなどが考えられる。

謝辞

本研究を進めるにあたり、クロストラフィックのモデル化をはじめとした、多くの助言を下さった三好直人教授に心から感謝いたします。

参考文献

- [1] Jarschel, M., Oechsner, S., Schlosser, D., Pries, R., Goll, S., Tran-Gia, P., Modeling and performance evaluation of an OpenFlow architecture, 23rd International Teletraffic Congress, (2011), 1-7.
- [2] Mahmood, K., Chilwan, A., Østerbø, O., Jarschel, M., Modelling of OpenFlow-based software-defined networks: the multiple node case, IET Networks, Vol. 4, (2015), 278-284.
- [3] Xiong, B., Yang, K., Zhao, J., Li, W., Li, K., Performance evaluation of OpenFlow-based software-defined networks based on queueing model, Computer Networks, Vol.102, (2016), 172-185.
- [4] Goto, Y., Masuyama, H., Ng, B., Seah, W. K., Takahashi, Y., Queueing analysis of software defined network with realistic OpenFlow-based switch model, IEEE 24th International Symposium, (2016), 301-306.
- [5] Miao, W., Min, G., Wu, Y., Wang, H., Hu, J., Performance modelling and analysis of software-defined networking under bursty multimedia traffic, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), Vol. 12, (2016), 77.
- [6] Muhizi, S., Shamshin, G., Muthanna, A., Kirichek, R., Vladyko, A., Koucheryavy, A., Analysis and performance evaluation of SDN queue model, International Conference on Wired/Wireless Internet Communication, (2017), 26-37.
- [7] Lai, Y. C., Ali, A., Hassan, M. M., Hossain, M. S., Lin, Y. D., Performance modeling and analysis of TCP connections over software defined networks, Proceedings of IEEE GLOBECOM, 4-8 December 2017, Singapore.
- [8] Yao, L., Hong, P., Zhou, W., Evaluating the controller capacity in software defined networking, IEEE 23rd International Conference, (2014), 1-6.
- [9] Gaver, D. P., Jacobs, P. A., Latouche, G., Finite birth-and-death models in randomly changing environments, Advances in Applied Probability, Vol. 16, (1984), 715-731.

着陸待ち行列

石井 宏明 佐藤 真史 豊泉 洋
早稲田大学大学院 基幹理工学研究科

概要： 航空機の着陸時における「着陸待ち行列」に関する考察を行う。この待ち行列は、カフェなどのレジとは異なり、そもそも存在自体が明らかになっていないため、解析することが難しい。しかし、飛行中の民間航空機の現在位置をリアルタイム表示するアプリケーションを用い、各便における最短飛行時間を抽出することで、着陸待ち行列の構成・解析することに成功した。また、優先権付き待ち行列理論を使用することで、現在管制官が指示する着陸順序よりも、1機あたりの待ち時間、燃費の面でより良い方法を提案する。

1. はじめに

私たちの身の回りに、待ち行列は非常に多く存在する。例えば、カフェのレジを待つ列、テーマパークの人気アトラクションの行列、エスカレーターに乗る人々の行列、交差点における信号待ちの車の列など、様々なものが挙げられる [1]。これら待ち行列に共通する性質は、待ち行列が「目に見える」ということである。

夕暮れ時や夜間に空港の展望デッキに行くと、着陸に向けて降下する航空機の光を見たことがある方は多いと思われる。しかし、その行列全体を目視することは不可能である。離陸を待つ航空機は地上でタキシングをするため、待ち行列の存在を確認することができるが、着陸時に関しては、存在するのかということすら明らかになっていない。航空機は渋滞状態で飛行間隔の調整を行っていることも、待ち行列を見えにくくしている原因の一つである。本研究では、航空機の運航情報データベースである「Flightradar24」というアプリケーション [2] と、国土交通省が公開している航空機の大規模な飛行軌跡データ「CARATS Open Data」 [3] を用い、成田国際空港に向けて飛行する航空機のデータを抽出し、シミュレーションをすることで、「着陸待ち行列」の存在を定義した (図 1)。

今後発着枠を拡大していくための議論は、すでに多く交わされている [4]。今日の航空管制では、管制官と交信を開始した航空機から順に着陸させるという方法を採用している。要するに「先着順 (First In First Out, 以下 FIFO と呼ぶ)」である。航空機の燃費は、同じ種類の航空機を使用した場合でも、1km 飛行するのに要する燃料の量は長距離便と中・短距離便で異なる [5]。一般的に長距離便は座席数が多く、貨物重量や搭載燃料も多いため、短距離便よりも総重量が重くなる、よって単位時間あたりに使用する燃料が多くなる、ということが挙げられる。このことより、本研究では「長距離便を優先的に着陸する方が、燃費の面で FIFO よりも効率が良いのではないか」という仮説を検証する。

待ち行列において客をいくつかのパターンに分類して考える際に有効なモデルとして、Accumulating Queue モデル (以下 AQ) がある。このモデルは主に病院で用いられ、病院の患者を「救急患者」と「一般患者」に分け、前者を優先的に診察するというものである [6]。本研究では、2017 年 9 月 30 日の成田国際空港における全着陸便 (貨物便を除く) 236 便を、飛行時間が 8 時間以上の「長距離便 (計 64 便)」, 8 時間未満の「短距離便 (計 172 便)」の 2 種類に分け、分析した。また、FIFO モデル, AQ モデル, 長距離便が飛来した場合は最優先で着陸させる長距離最優先モデル (Priority Queue, 以下 PQ) の 3 種類のモデルを、待ち時間及び燃費の点で比較し、検討した。

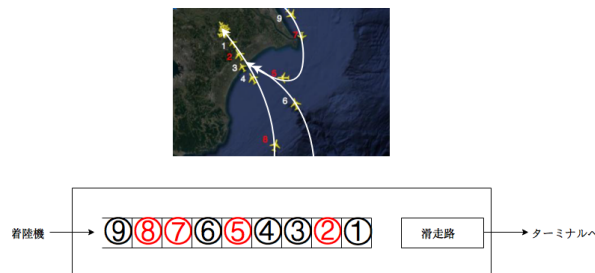


図 1: 成田国際空港に向け降下をする航空機とその待ち行列 (赤字：長距離便, 黒字：短距離便)

2. 着陸待ち行列の定義

「Flightradar24」を用いることで、各航空機の着陸時刻 L 、離陸時刻 D を取得することができるが、待ち時間 W や、待ち行列に到着する時刻 T は分からない。そこで、過去 60 日間のうち最も早い飛行時間 F で飛来したものを、着陸待ち時間が 0 であると定義することにより、 T 及び W を明らかにすることに成功した (図 2)[1]。

実際の成田国際空港には滑走路は 2 本あり、どちらも着陸便及び離陸便を扱っている。しかし本研究では、単純化のために着陸専用、離陸専用の滑走路に分け、着陸時には 1 本の滑走路のみ使用すると仮定した。また、着陸待ち行列への到着時刻を推定するために、全機が最速飛行時間で滑走路に飛来した場合を考え、シミュレーションを行った。着陸に要する時間 S は、①滑走路進入端から 1 マイルの地点（着陸または着陸復行を行うかの決断点）までの所要時間②滑走路進入端を通過し、滑走路縁を通過するまでの所要時間③滑走路を完全に出るまでに要する時間の合計で求められる (図 3)。着陸直前の航空機を 200km/h とすると、①は約 30 秒となる。②を 1 分、③を 30 秒とし、着陸復行の確率を極力 0 にするために 1 分間を加えると、 S は約 3 分となる。よって本研究では、着陸に要する時間を一律で 3 分とする。

次に、過去 60 回の飛行データより、各々の離陸時刻 D_i 、離陸から着陸終了までの飛行時間 A_i を抽出し、待ちのない飛行時間 F を次のように推定する。

$$F = \min_i \{A_i \mid |A_i - E[A]| < 1 \text{ 時間} \}. \quad (1)$$

航空機の離陸時刻及び飛行時間は、その日の空港や天気の状態により毎回異なるため、各便における平均離陸時刻 $E[D]$ 及び平均飛行時間 $E[A]$ を求め、過去 60 日間における飛行時間のうち、最も短い飛行時間で成田国際空港に着陸したものを F とした。しかし、 $E[A]$ と F の差が 1 時間以上のものもあったため、これらは天候やその他何らかの要因により記録された異常値だとみなし、差が 1 時間以内のものの中で最も早い飛行時間を F とした (式 (1))。このとき、待ち行列退去時刻 L 、待ち時間 W 、サービス開始時刻 X は、

$$L = D + A = D + F + W + S, \quad (2)$$

$$W = L - S - T, \quad (3)$$

$$X = L - S \quad (4)$$

で表せる。つまり、時刻 D に出発、時刻 T にシステムに到着、行列に W 時間並び、サービス開始時刻 X から S 時間サービスを受け、 L で待ち行列を出るとすることで、M/D/1 モデルと同じような「仮想待ち行列」(図 2) を考えることができる、実際の運航便にこの待ち行列を適用すると、図 4 のようになる。こうして、一般的には存在すら確認されていない着陸待ち行列を定義することができた。

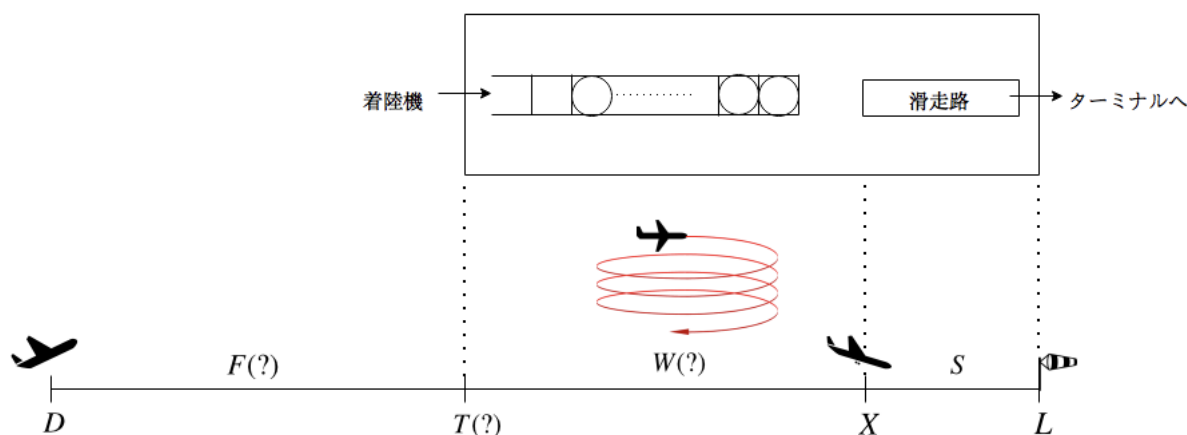


図 2: 仮想待ち行列

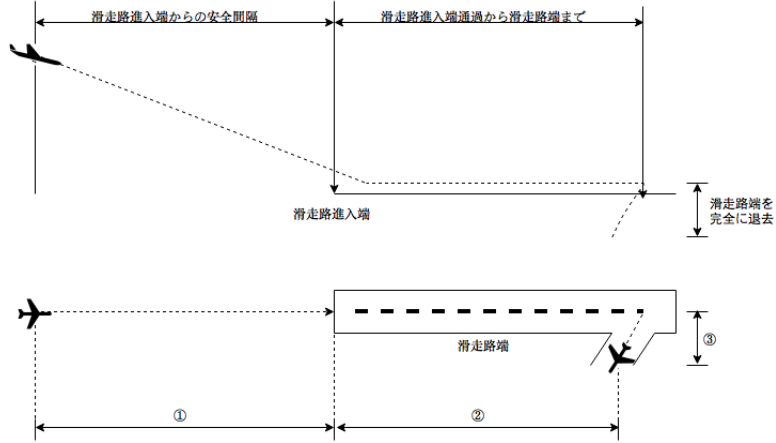


図 3: 着陸に要する時間の考え方

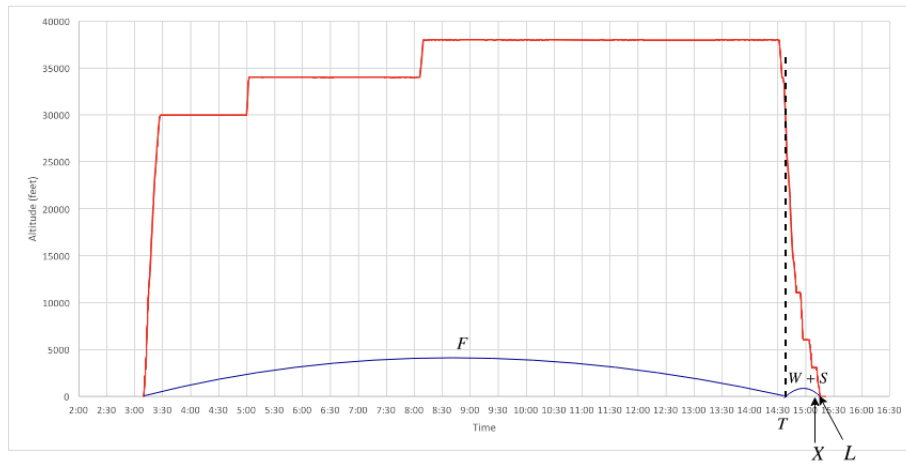


図 4: AA153 の最速ラップと待ち時間 (2017 年 9 月 20 日の高度データ)

3. 着陸順序の変更

現在の世界の航空管制は、着陸機が管制官にコンタクトを開始した順番に着陸をさせるという方法を採用している。しかし、単位距離あたりに使用する燃料は長距離便の方が短距離便よりも多いため、長距離便を優先的に着陸させる方が燃料の面でより効率的である。しかし長距離便を優先し続けると、短距離便の待ち時間が伸び過ぎる可能性がある。そこで、先述の着陸待ち行列に、待ち時間に対して異なる重みを付けてサービス順序をコントロールする Accumulating Queue の適用を考える。

図 5 において、横軸を時間 t 、縦軸を優先度 V とし、待ち行列に到着した順番に A_n とする。また、重み b を長距離便と短距離便それぞれ

$$b_n = \begin{cases} 1 & (F_n \geq 8:00) \\ 0.5 & (F_n < 8:00), \end{cases} \quad (5)$$

とする。このとき優先度は、

$$V_n(t) = b_n(t - T_n), \quad (6)$$

に従う。いま、3 番目に待ち行列に到着した短距離便 A3 が滑走路を使い終わった瞬間 $t = 10$ における他機の優先度を考える。FIFO モデルであれば、4 番目に到着した短距離便 A4 がそのまま着陸するが、AQ モデルにおいては、 $V_5(10) = 10 - 7 = 3 > V_4(10) = 0.5(10 - 6) = 2$ となるため、5 番目に飛来した長距離便 A5 が着陸する。よって、A4 が着陸できる順番は 5 番目となる。このように考えると、AQ モデルの着陸順序は、図 6 (2) のようになる。

また、FIFO モデルは、AQ において長距離便の優先度 b_n を全て 1 としたもの、PQ モデルは ∞ にしたものと同じである。このとき図 6 より、PQ モデルにおいては、短距離便が圧倒的に待たされてしまう場合があることが分かる。

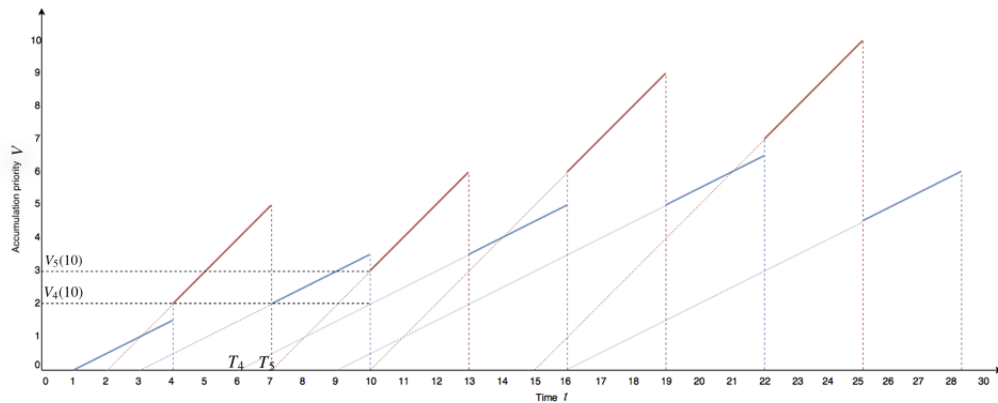


図 5: Accumulating Queue

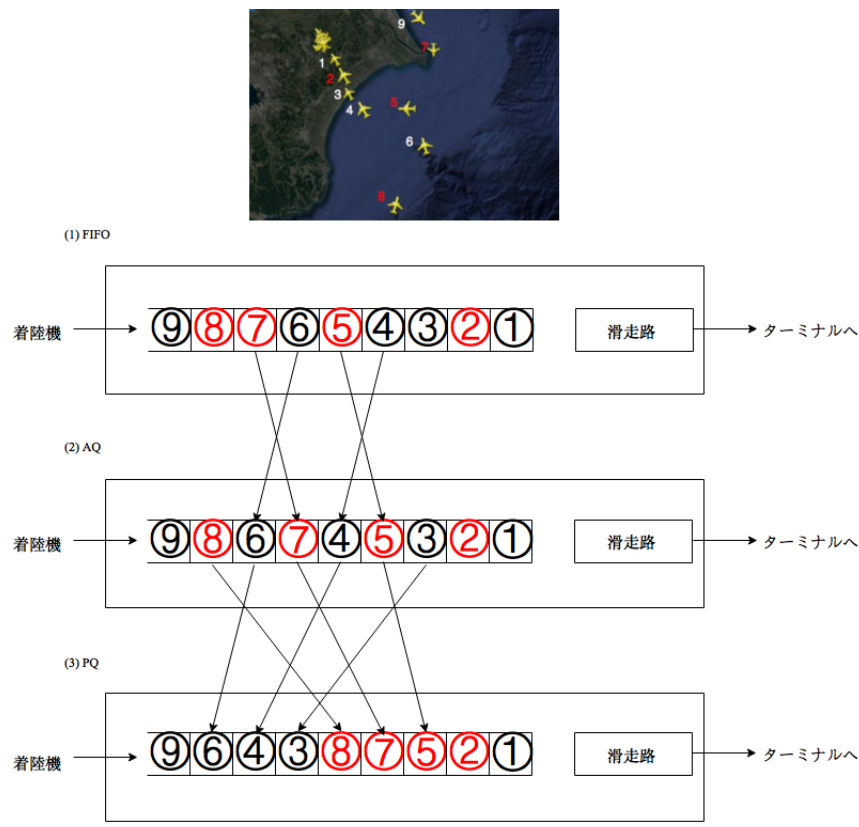


図 6: 各モデルごとの待ち行列 (赤字：長距離便，黒字：短距離便)

4. 現モデルと AQ 及び PQ の比較

4.1. 待ち時間

全着陸便における合計待ち時間は、保存則が成立するため FIFO モデルと変わらないが、短距離便と長距離便の平均値をそれぞれ見ると、AQ モデルでの短距離便は 5 分 19 秒余計に遅れてしまう一方、長距離便は 14 分 18 秒早く着陸できる。また TPLD モデルでは、短距離便は 11 分 17 秒遅れ、長距離便は 30 分 20 秒早く着陸できることが明らかになった (表 1)。しかし、PQ モデルにおける短距離

便では，最大で 1 時間余計に待たなければならない便が生じてしまう一方で，AQ モデルでは最大でも 30 分の待ち時間となった (表 2)．

表 1: 各モデルにおける待ち時間 W_n の比較 (括弧内は FIFO との差)

サービス順	全体	短距離	長距離
FIFO	0:31:44	0:30:18	0:35:36
AQ	0:31:44	0:35:37 (+0:05:19)	0:21:18 (-0:14:18)
PQ	0:31:44	0:41:35 (+0:11:17)	0:05:16 (-0:30:20)

表 2: 各モデルにおける最大待ち時間と FIFO モデルとの差分

	AQ	PQ
短距離	+0:30:00	+1:00:00
長距離	-0:45:00	-1:30:00

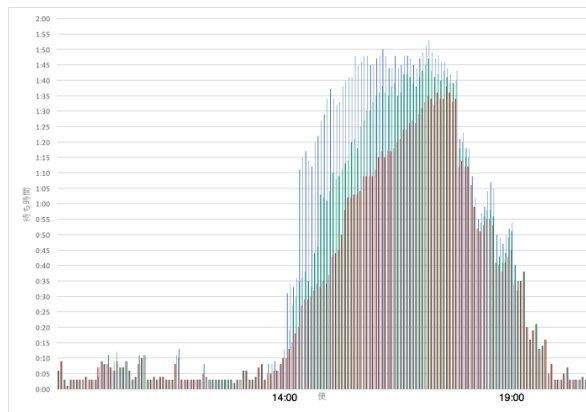


図 7: 各モデルにおける待ち時間 (短距離便)

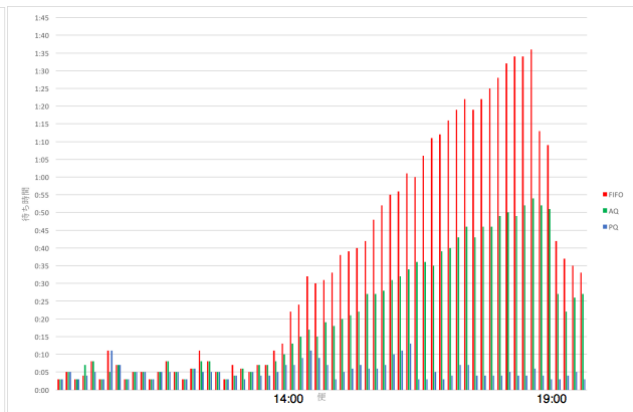


図 8: 各モデルにおける待ち時間 (長距離便)

4.2. 燃費

着陸便 k の巡航速度を α_k (km/h)，燃費を β_k (kg/km) とする．待ち行列に並び，着陸開始を待っている時間 W_k に消費する全 236 機の総燃料は，

$$\sum_{k=1}^{236} W_k \times \alpha_k \times \beta_k \quad (7)$$

によって求められるが，この値は，巡航速度，燃費の各平均値より算出することも可能なので，

$$(\text{全短距離便の消費燃費}) + (\text{全長距離便の消費燃料}) = \bar{W}_k \times \bar{\alpha}_s \times \bar{\beta}_s \times 172 + \bar{W}_k \times \bar{\alpha}_l \times \bar{\beta}_l \times 64 \quad (8)$$

となる．ここで，全 236 機を機材ごとに分類すると，表 3 のようになった [8][9]．このデータより短距離便の平均燃費 $\bar{\alpha}_s$ ，長距離便の平均燃費 $\bar{\alpha}_l$ を求めると，それぞれ 4.91kg/km，7.93kg/km となり，巡航速度は 861km/h，892km/h となった (表 4)．これらを式 (8) に代入し，全消費燃料を計算すると，表 5 の結果が得られた．すなわち，AQ，PQ を用いることによって，現行モデルの FIFO よりも，32553.92kg，64137.76kg もの燃料を削減できることが明らかになった．これらは，待ち時間に消費する全燃料の約 2.67%，5.27% に相当する．

表 3: 全便の機体数, 巡航速度, 燃費

機種	巡航速度 (km/h)	短距離便		長距離便	
		機体数	燃費 (kg/km)	機体数	燃費 (kg/km)
A320neo	833	2	2.79	0	no data
A320	829	24	3.12	0	no data
A321	829	13	3.61	0	no data
A332	871	4	6.00	3	6.40
A333	877	19	6.25	6	6.25
A343	871	1	6.81	3	7.32
A359	903	0	no data	1	7.07
A388	903	1	13.8	1	13.8
B738	838	31	3.17	0	no data
B739	838	2	3.42	0	no data
B744	933	2	10.8	1	11.1
B748	933	2	9.90	0	10.5
B752	854	3	4.60	0	no data
B763	850	27	5.38	3	5.51
B772	892	5	6.83	6	7.44
B773	892	3	7.88	0	8.49
B77L	892	0	no data	2	7.57
B77W	829	4	8.49	13	8.58
B788	903	21	5.26	12	5.38
B789	903	8	5.77	13	7.18

表 4: 平均燃費 $\bar{\alpha}$ 及び平均巡航速度 $\bar{\beta}$

	短距離便	長距離便
燃費 (kg/km)	4.91	7.93
巡航速度 (km/h)	861	892
積 (km/h)	4227.51	7073.56

表 5: 待ち時間に消費する総燃料 (kg)

	AQ	PQ
短距離	64460.49	141627.79
長距離	-97014.42	-205765.55
差分	-32553.92	-64137.76

5. 成果・今後の課題

最短飛行時間を抽出することで、存在すら明らかになっていない着陸待ち行列を新たに定義することに成功した。これにより Accumulating Queue を適用でき、現行の First In First Out モデルとは異なった2つのモデルを構築することができた。

これら3つのモデルを比較すると、FIFO モデルよりも AQ, PQ 両モデルを用いた場合の方が、全体の燃費という点で利点があることが明らかになった。しかし待ち時間という面では、長距離便が立て続けに飛来した場合、短距離便の待ち時間は最大で1時間も伸びてしまう(表2)ため、PQ モデルは良いものとは言えない。従って、燃費も一定の割合で削減でき、短距離便にもそれほどの余分な待ち時間を与えない AQ モデルが、最も良いモデルと考えられる。

今後の課題としては、到着便を2つのパターンだけでなく、飛行距離に応じて細かく分類し、それぞれの飛行距離に応じて優先レートを設定することで、より正確に燃費を計算することができる。さらには、滑走路を2本にし、離陸便も含めることで、より現実に近い形でシミュレーションを行うことができるのではないかと感じた。

謝辞

本研究を進めるにあたり、データを頂いた国土交通省航空局交通管制部交通管制企画課の方々に深く感謝致します。

参考文献

- [1] 塩田 茂雄, 河西 憲一, 豊泉 洋, 会田 雅樹, 待ち行列理論の基礎と応用, 共立出版, (2014).
- [2] Flightradar24.com - Live flight tracker! (2017, September 30). Retrieved from <https://www.flightradar24.com/>
- [3] 国土交通省, CARATS Open Data 2014, (2014)
- [4] 二見康友, 離着陸順序の実態からみた滑走路容量算定手法に関する研究 成田国際空港を対象として, (2014).
- [5] Antonio Filippone: *Advanced Aircraft Flight Performance*, Cambridge University Press, (2012), p.454.
- [6] David A. Stanford, Peter Taylor, Ilze Ziedins: Waiting time distributions in the accumulating priorityqueue, *Queueing Syst*, **77** (2014), p. 297–330.
- [7] 平松健志, 平田輝満, 屋井鉄雄, 空港容量算定シミュレーションの開発と 容量拡大効果に関する研究, 運輸政策研究, (2006), 26–37.
- [8] Boeing.com (2017, September 30). Retrieved from <http://www.boeing.com/commercial/787/>
- [9] Airbus.com (2017, September 30). Retrieved from <http://www.airbus.com>

Cooperation area を考慮した高密度セルラ移動体ネットワークの性能解析

原口 昌太

東京工業大学 情報理工学院

概要：本論文では、ユーザーが移動する場合の高密度ネットワーク (ultra-dense network; UDN) の通信性能の評価を行う。UDN では、基地局が高密度に配置されるため、干渉が大きくなるというデメリットが存在する。したがって、cooperation という戦略をとることにより干渉を軽減させる。この戦略のもとで、ユーザーの移動を random waypoint モデル、基地局の配置を空間点過程を用いてモデル化する。その後、ハンドオフレートとアウトエージ確率 (outage probability) の導出を行い、これらの指標を用いた新しい指標を提案する。

1. はじめに

近年、スマートフォンの普及によりトラヒックの需要が増加している [1]。そこで、次世代の 5G 通信によってさらに拡大するトラヒックの需要をカバーするために、高密度に基地局が配置されている高密度ネットワーク (ultra-dense network; UDN) を考えることが必要である [2]。UDN における通信性能に関する研究は数多く存在する [3, 4, 5, 6]。UDN では、高密度に基地局が配置されているため各ユーザーから最寄りの基地局までの距離が近くなるというメリットがあるが、基地局が増えることによって干渉が大きくなるというデメリットも存在する。そこで、干渉を軽減するために、cooperation という方策がとられることがある。一般的に各基地局は異なる信号を送信しているが、ある領域 (cooperation area) 内の基地局が協力して同じ信号を送ることにより干渉電波を減らすことができる。Cooperation area は、基地局の通信を制御する収容局によりユーザーの移動に基づいて決定される [7, 8]。

本論文では、cooperation area を考慮したうえで、移動端末の通信性能をハンドオフレートとアウトエージ確率 (outage probability) の 2 つの指標を用いて評価する。ハンドオフレートは、移動を伴う場合の性能評価の指標である [9, 10]。ユーザーの移動によって接続する基地局が切り替わることがあるが、この切り替え操作をハンドオフといい、ハンドオフが発生する確率をハンドオフレートという。本論文では、cooperation area 内の基地局は同一の信号を送信しており、通信中に cooperation area 内にある基地局とはハンドオフをしないモデルを考える。ハンドオフレートの導出にあたってはユーザーの移動をモデル化する必要があり、random waypoint モデル (RWP モデル) がよく用いられる [10, 11]。RWP モデルでは、ユーザーが行き先の座標をランダムに選択し、ある速度 v で直線に移動する。そして、目的地に到達するとその場で停止し、次の行き先を決定することを繰り返す。

また、通信性能を評価する指標として被覆確率 (coverage probability) がよく知られている。被覆確率は、端末が受信する信号の強さと干渉の信号の強さの比である SIR がしきい値 θ を越える確率であり、端末が通信をするために十分な強度の電波を受信できる確率を表す。本論文では、被覆確率の補分布であるアウトエージ確率を考える。

本論文では、[11] 内で導出されているハンドオフレートとその導出方法の問題点を解消した提案手法との比較、アウトエージ確率の導出を行い、またハンドオフレートとアウトエージ確率を用いた新しい指標を提案する。

2. システムモデル

本節では、ハンドオフレートとアウトエージ確率を導出するための空間確率モデルを構築する。以下では、 $\mathbb{R}$ を実数の集合、 $\mathbb{N}$ を自然数の集合、 $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ とする。また、 $b(c, R) = \{x \in \mathbb{R}^2 \mid \|x - c\| < R\}$ ($c \in \mathbb{R}^2, R > 0$) とする。

2.1. 基地局

$\Phi = \{X_k\}_{k \in \mathbb{N}_0}$ を 2 次元平面 $\mathbb{R}^2$ 上における強度 λ_r の定常ポアソン点過程とする。この点過程の点がそれぞれ基地局の地点を表すと考え、地点 X_k にある基地局を基地局 B_k とする。基地局の送信電力 p (> 0) は一定とする。距離による電波の減衰はユーザーと基地局との距離 r を用いて、パスロス関数 $l(r) = cr^{-\beta}$ として表される [12]。定数 c (> 0) は伝搬損失係数、定数 β (> 2) は伝搬損失指数と呼ばれ、都市部では β は 3~4 程度であることが知られている [12, 13]。また、地点 x にいるユーザーの基

地局 B_k からの電波に対するフェージング $H_{k,x}$ はそれぞれ独立に平均 1 の指数分布に従い、ノイズは無視できるものとする。

2.2. ユーザーの移動

基地局は定常ポアソン点過程であるから、注目するユーザー u_0 の位置を原点 O としても一般性を失わない。また、ユーザー u_0 から最寄りの基地局 B_0 までの距離 $|X_0|$ が R であるとき、同様の理由から $X_0 = (-R, 0)$ となるように座標平面をとっても一般性を失わない。以降は、この仮定のもとで考える。

本論文では、ユーザーの移動を時間 n ごとに観測し、その間は直線的に移動していると近似する。これを再現するには RWP モデルが最適である。以降は、ユーザーが RWP モデルに従って 2 回の移動を行う場合を考える。すなわち、ユーザーは原点 O から M_1 へ 1 回目の移動を行った後、 M_1 から M_2 へ 2 回目の移動を行う。今回のモデルでは、1 回目と 2 回目の移動距離 L_1, L_2 がそれぞれパラメータ $(2\pi\lambda_l)^{-\frac{1}{2}}$ のレイリー分布に従うものとする [10]。移動距離 $L_i (i = 1, 2)$ の密度関数は、

$$f_{L_i}(l_i) = 2\pi\lambda_l l_i \exp(-\pi\lambda_l l_i^2), \quad (i = 1, 2)$$

となる。移動する方向は、 x 軸から反時計回りに角度を取り、1 回目と 2 回目の移動角をそれぞれ ω_1, ω_2 と固定する。 $\omega_1, \omega_2 \in (\pi, 2\pi)$ の場合も対称性から同様に考えることができるため、 $\omega_1, \omega_2 \in [0, \pi]$ を考える。このとき、1 回目、2 回目の移動後の点をそれぞれ $M_1 = (M_1x, M_1y)$, $M_2 = (M_2x, M_2y)$ とすると、

$$\begin{aligned} M_1x &= L_1 \cos(\omega_1), \\ M_1y &= L_1 \sin(\omega_1), \\ M_2x &= L_1 \cos(\omega_1) + L_2 \cos(\omega_2), \\ M_2y &= L_1 \sin(\omega_1) + L_2 \sin(\omega_2) \end{aligned}$$

となるので、基地局 B_0 から M_1, M_2 までの距離 R_1, R_2 は、それぞれ、

$$\begin{aligned} R_1 &= \sqrt{(M_1x - (-R))^2 + (M_1y - 0)^2} \\ &= \sqrt{(L_1 \cos(\omega_1) + R)^2 + (L_1 \sin(\omega_1))^2}, \\ R_2 &= \sqrt{(M_2x - (-R))^2 + (M_2y - 0)^2} \\ &= \sqrt{(L_1 \cos(\omega_1) + L_2 \cos(\omega_2) + R)^2 + (L_1 \sin(\omega_1) + L_2 \sin(\omega_2))^2} \end{aligned}$$

となる。

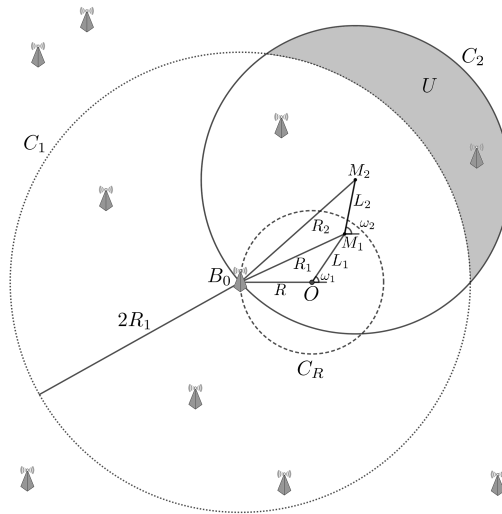


図 1: システムモデル

2.3. Cooperation area

本論文では、ユーザーが1回目の移動を終えるまでにハンドオフをしないと仮定する。また, cooperation area を中心が基地局 B_0 で半径 $2R_1$ の円の領域と設定する。ここで R_1 は1回目の移動後の基地局 B_0 との距離である。これは, 1回目の移動を終えるまでハンドオフをしないために, ユーザーが M_1 に到達し停止したときに収容局が cooperation area を設定すると仮定しているためである。

2.4. ハンドオフレート

ユーザーが移動をする場合, それに伴い各基地局との距離が変動するため, 最寄りの基地局が B_0 でなくなる場合がある。このとき, cooperation area が考慮されていなければ新たに最も近くなる基地局に通信先を変更する。この操作をハンドオフといい, 本論文では2回目の移動中にハンドオフが発生する確率をハンドオフレートと呼ぶ。

図1を用いてユーザーが M_1 から M_2 に移動する際にハンドオフが起こるケースを説明する。Cooperation area が考慮されていない場合, 中心が M_2 で半径 R_2 の円の領域 C_2 内に基地局が存在するとき, M_2 における最寄りの基地局は B_0 でなくなる。このような場合にハンドオフが発生する。一方, cooperation area が考慮されている場合, 領域 U (C_2 のうち cooperation area に含まれない領域) に基地局が存在する場合にハンドオフが発生する。したがって, 本論文では領域 U に基地局が1つ以上存在する確率をハンドオフレートとする。

2.5. アウテージ確率

ユーザーが基地局 B_0 と接続していると仮定すると, 地点 x にいるユーザーの SIR は,

$$\text{SIR}(x) = \frac{pH_{0,x}l(|X_0 - x|)}{\sum_{k=1}^{\infty} pH_{k,x}l(|X_k - x|)} \quad (1)$$

によって与えられる。本論文では, ユーザーが原点 O から M_1 まで移動するまでの間に1度でも SIR がしきい値 θ を下回る確率をアウテージ確率として定義する。本来であれば, 原点 O から M_1 までの移動を連続的に捉えて計算することが理想的である。しかし, デジタル通信では連続的に観測することができないため, 移動を離散の点に分割して考える。ユーザーの移動を単位時間ごとに分け, 1回目の移動経路を n 等分した点を x_i ($i = 0, 1, \dots, n$) とする。すなわち,

$$x_i = \frac{i}{n} OM_1$$

である。これを用いて, アウテージ確率は

$$p_{out}(\theta) = P\left(\min_{0 \leq i \leq n} \text{SIR}(x_i) < \theta\right) \quad (2)$$

によって計算できる。本来は, M_1 から M_2 まで移動する間の SIR も考慮する必要があるが, 本論文では簡単のため原点 O から M_1 までの移動のみを考えている。

3. ハンドオフレートの導出

本章では, [11] でのハンドオフレートの導出方法を紹介した後, この手法が抱える問題点を修正した提案手法を紹介する。

3.1. 既存手法

領域 C_1, C_2 をそれぞれ,

$$\begin{aligned} C_1 &= b(-R, 2R_1), \\ C_2 &= b(M_2, R_2) \end{aligned}$$

として, $U_0 = C_2 \setminus C_1$ とする。このとき, この領域 U_0 の中に基地局があれば cooperation area 外でかつ M_2 から距離が B_0 よりも近い場合ハンドオフをする。領域 U_0 は $R, L_1, L_2, \omega_1, \omega_2$ によって決定さ

れ, ハンドオフレート H_0 は,

$$H_0 = \iiint (1 - \exp(-\lambda_l S_{U_0})) f_R(r) f_{L_1}(l_1) f_{L_2}(l_2) dr dl_1 dl_2,$$

$$S_{U_0} = 2R_1 \sqrt{R_2^2 - R_1^2} + (2R_1^2 - 4R_2^2) \cos^{-1} \left(\frac{R_1}{R_2} \right)$$

となる. ただし, S_{U_0} は領域 U_0 の面積を表す.

6 節の数値実験で示すとおり, 既存手法 (図 8) では ω_1 が π に近づくとシミュレーションとの誤差が大きくなってしまふ. その原因は, 基地局 B_0 が原点 O から最も近い基地局なため $C_R = b(0, R)$ 内には基地局が存在しないという条件を考慮していないためである. ω_1 が π に近い場合, 基地局 B_0 に近づいていく方向に進むため R_1 が小さくなりやすい. R_1 が R よりも小さくなると cooperation area から C_R がはみ出してしまう. しかし, 既存手法ではその領域に基地局の存在を許しているため正確なハンドオフレートの導出ができていない.

3.2. 提案手法

本節では, C_R の影響を加味したうえでハンドオフレートを導出する. $U = C_2 \setminus (C_R \cup C_1)$ とする. $C_R \subset C_1$ の場合 (図 2, 3) は, $C_R \cup C_1 = C_1$ となるので既存手法と同様になる. $C_R \not\subset C_1$ の場合 (図 4~6) は, 基地局が存在しうる領域という制約を加えたうえで, この領域の中に基地局があれば cooperation area 外でかつ M_2 から距離が B_0 よりも近い場合ハンドオフをする. このとき, ハンドオフレート $H(\omega_1, \omega_2)$ は,

$$H(\omega_1, \omega_2) = \iiint (1 - \exp(-\lambda_l S_U)) f_R(r) f_{L_1}(l_1) f_{L_2}(l_2) dr dl_1 dl_2$$

$$S_U = \begin{cases} 0, & (\text{case1} = \{R_1 > R_2\}), \\ S_{C_2 \setminus C_1}, & (\text{case2} = \{R_1 \leq R_2, R \leq R_1\}), \\ S_{C_2 \setminus \{C_R \cup C_1\}}, & (\text{case3} = \{R_1 \leq R_2, R > R_1, |M_2 D| \leq R_2\}), \\ S_{C_2 \setminus C_1}, & (\text{case4} = \{R_1 \leq R_2, R > R_1, |M_2 D| > R_2, \angle OB_0 D < \angle OB_0 M_2\}), \\ 0, & (\text{case5} = \{R_1 \leq R_2, R > R_1, |M_2 D| > R_2, \angle OB_0 D \geq \angle OB_0 M_2\}) \end{cases}$$

となる. ただし, 点 D は C_R と C_1 の共有点のうち y 座標が正のものとする. S_U の場合分けは, case1 (図 2) は, $C_2 \subseteq C_1$ の場合, case2 (図 3) は, $C_2 \not\subset C_1$ かつ $C_R \subset C_1$ の場合, case3 (図 4) は, $C_2 \not\subset C_1$ かつ $C_R \not\subset C_1$ のうち C_R が領域 U に隣接している場合, case4 (図 5) は, $C_2 \not\subset C_1$ かつ $C_R \not\subset C_1$ のうち C_R が領域 U に隣接していない場合, case5 (図 6) は, $C_2 \subset C_1 \cup C_R$ かつ $C_R \subseteq C_1$ の場合である. 各領域の面積は,

$$S_{C_2 \setminus C_1} = S_{C_2} - S_{C_1 \cap C_2},$$

$$S_{C_2 \setminus \{C_R \cup C_1\}} = S_{C_2} - S_{C_1 \cap C_2} - S_{C_R \cap C_2} + S_{C_R \cap C_1 \cap C_2},$$

$$S_{C_2} = \pi R_2^2,$$

$$S_{C_1 \cap C_2} = \pi R_2^2 + (4R_1^2 - 2R_2^2) \cos^{-1} \left(\frac{R_1}{R_2} \right) - 2R_1 \sqrt{R_2^2 - R_1^2},$$

$$S_{C_R \cap C_2} = R^2 \cos^{-1} \left(\frac{R^2 + |OM_2|^2 - R_2^2}{2R|OM_2|} \right) + R_2^2 \cos^{-1} \left(\frac{R_2^2 + |OM_2|^2 - R^2}{2R_2|OM_2|} \right)$$

$$- 2\sqrt{\left(\frac{R + R_2 + |OM_2|}{2} \right) \left(\frac{-R + R_2 + |OM_2|}{2} \right) \left(\frac{R - R_2 + |OM_2|}{2} \right) \left(\frac{R + R_2 - |OM_2|}{2} \right)},$$

$$S_{C_R \cap C_1 \cap C_2} \approx R_2^2 \sin^{-1} \left(\frac{R_1}{R_2} \right) - R_1 \sqrt{R_2^2 - R_1^2} + R^2 \sin^{-1} \left(\frac{R_1}{R} \right) - R_1 \sqrt{R^2 - R_1^2}$$

$$+ 2R_1^2 \left(\cos^{-1} \left(\frac{R_1}{R_2} \right) + \cos^{-1} \left(\frac{R_1}{R} \right) + \tan^{-1} \left(\frac{M_2 x + R}{M_2 y} \right) \right)$$

で与えられる. ここで, $S_{C_R \cap C_1 \cap C_2}$ の面積が近似になっているのは, case3 において, 図 7 のような例外があるためである. この場合, 領域 $C_R \cap C_1 \cap C_2$ には図 7 中の斜線部は含まれないが, 右辺の計算

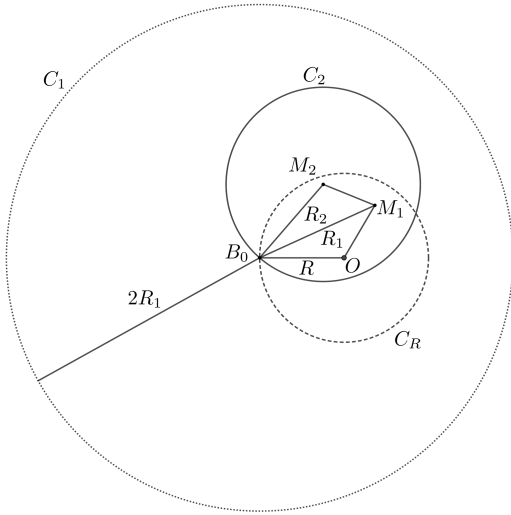


図 2: case1

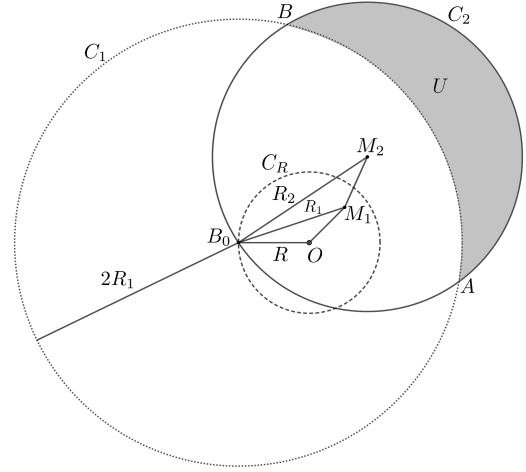


図 3: case2

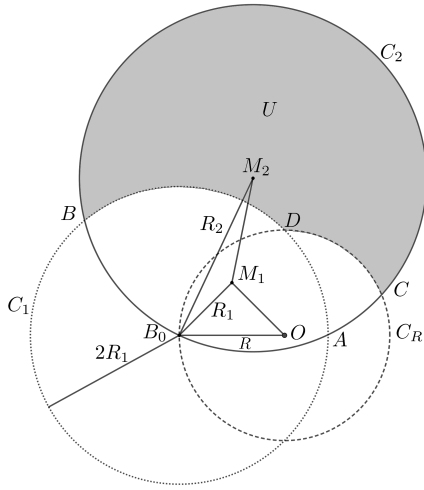


図 4: case3

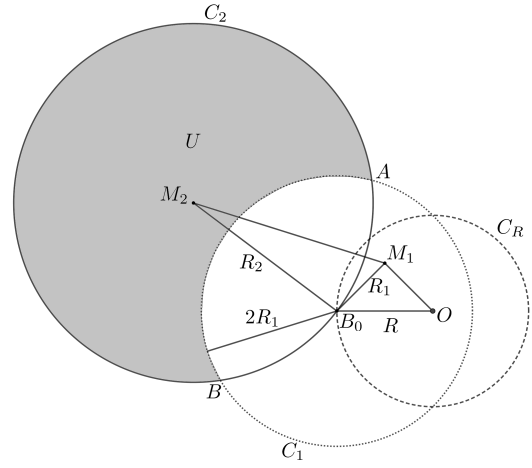


図 5: case4

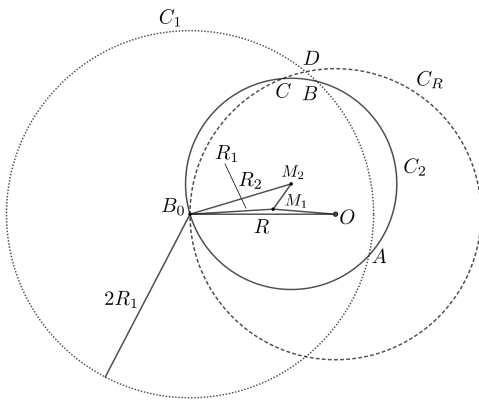


図 6: case5

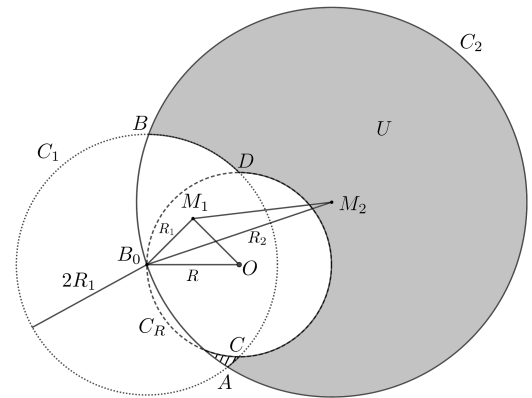


図 7: case3 の例外

では斜線部も加えた面積となっている。したがって、領域 U と斜線部の面積の合計を計算することになるが、斜線部の面積は領域 U に対して十分小さいため近似できる。

4. アウテージ確率の導出

本章では、(2) 式のアウテージ確率を数値計算可能な形に変換する。まず、被覆確率を考える。各 x_i における最小の SIR を考えるので、

$$P\left(\min_{0 \leq i \leq n} \text{SIR}(x_i) \geq \theta\right) = P(\text{SIR}(x_0) \geq \theta, \text{SIR}(x_1) \geq \theta, \dots, \text{SIR}(x_n) \geq \theta) \quad (3)$$

となる。基地局の配置 Φ 、ユーザーの移動距離 L_1 、原点 O と基地局 B_0 の距離 R で条件付けすると、各 x_i における SIR は互いに条件付き独立なので、(3) 式は、

$$P\left(\bigcap_{i=0}^n \{\text{SIR}(x_i) \geq \theta\} \mid \Phi, L_1, R\right) = \prod_{i=0}^n P(\text{SIR}(x_i) \geq \theta \mid \Phi, L_1, R) \quad (4)$$

となる。ここで、(1) より、

$$\begin{aligned} \{\text{SIR}(x_i) \geq \theta\} &= \left\{ \frac{H_{0,x_i} l(|X_0 - x_i|)}{\sum_{k=1}^{\infty} H_{k,x_i} l(|X_k - x_i|)} \geq \theta \right\} \\ &= \left\{ H_{0,x_i} \geq \theta \frac{\sum_{k=1}^{\infty} H_{k,x_i} l(|X_k - x_i|)}{l(|X_0 - x_i|)} \right\} \end{aligned}$$

であるから、 H_{0,x_i} 以外のフェージング (以降、 $H = \{H_{k,x_i} \mid k = 1, 2, \dots, i = 0, 1, \dots, n\}$ と表記する) を条件付けすることによって、(4) 式は、

$$\begin{aligned} P(\text{SIR}(x_i) \geq \theta \mid \Phi, L_1, R, H) &= \exp\left(-\theta \frac{\sum_{k=1}^{\infty} H_{k,x_i} l(|X_k - x_i|)}{l(|X_0 - x_i|)}\right) \\ &= \prod_{k=1}^{\infty} \exp\left(-\theta \frac{H_{k,x_i} l(|X_k - x_i|)}{l(|X_0 - x_i|)}\right) \end{aligned}$$

となる。さらに $H_{k,x_i} (k = 1, 2, \dots)$ の独立性から、平均 1 の指数分布のラプラス変換 $\mathcal{L}_{H_{k,x_i}}(s) = (1 + s)^{-1}$ を用いれば、

$$P(\text{SIR}(x_i) \geq \theta \mid \Phi, L_1, R) = \prod_{k=1}^{\infty} \left(1 + \theta \frac{l(|X_k - x_i|)}{l(|X_0 - x_i|)}\right)^{-1}$$

となる。 x_i から各基地局までの距離は、

$$\begin{aligned} |X_0 - x_i| &= \sqrt{R^2 + \left(\frac{i}{n} L_1\right)^2 + 2R \left(\frac{i}{n} L_1\right) \cos(\omega_1)} \\ |X_k - x_i| &= \sqrt{|X_k|^2 + \left(\frac{i}{n} L_1\right)^2 + 2|X_k| \left(\frac{i}{n} L_1\right) \cos(\angle X_k O M_1)} \end{aligned}$$

であり、 $\mathbb{R}^2 \setminus b(0, R)$ 上の強度 λ_r の定常ポアソン点過程の点配置を Y_k とすれば、

$$E \left[\prod_{i=0}^n P(\text{SIR}(x_i) \geq \theta \mid \Phi, L_1, R) \right] = E \left[\prod_{k=1}^{\infty} \prod_{i=0}^n \left(1 + \theta \frac{l\left(\sqrt{|Y_k|^2 + \left(\frac{i}{n} L_1\right)^2 - 2|Y_k| \left(\frac{i}{n} L_1\right) \cos \angle Y_k O M_1}\right)}{l\left(\sqrt{R^2 + \left(\frac{i}{n} L_1\right)^2 + 2R \left(\frac{i}{n} L_1\right) \cos \omega_1}\right)} \right)^{-1} \right]$$

となる。強度 λ_r の定常ポアソン点過程のラプラス汎関数は、 $\mathcal{L}_\Phi(h) = \exp(-\lambda_r \int_{\mathbb{R}^2} (1 - e^{-h(z)}) dz)$ で与えられるため、

$$P \left(\bigcap_{i=0}^n \{SIR(x_i) \geq \theta\} \mid L_1, R \right) \\ = \exp \left(-\lambda_r \int_{\mathbb{R}^2 \setminus b(0,r)} 1 - \prod_{i=0}^n \left(1 + \theta \frac{l \left(\sqrt{y^2 + (\frac{i}{n} l_1)^2} - 2y(\frac{i}{n} l_1) \cos(\angle yOM_1) \right)}{l \left(\sqrt{r^2 + (\frac{i}{n} l_1)^2} + 2r(\frac{i}{n} l_1) \cos \omega_1 \right)} \right)^{-1} dy \right)$$

となる。極座標を用いて変数変換を行うと、

$$P \left(\bigcap_{i=0}^n \{SIR(x_i) \geq \theta\} \mid L_1, R \right) \\ = \exp \left(-\lambda_r \int_r^\infty \int_0^{2\pi} \left(1 - \prod_{i=0}^n \left(1 + \theta \frac{l \left(\sqrt{s^2 + (\frac{i}{n} l_1)^2} - 2s(\frac{i}{n} l_1) \cos \phi \right)}{l \left(\sqrt{r^2 + (\frac{i}{n} l_1)^2} + 2r(\frac{i}{n} l_1) \cos \omega_1 \right)} \right)^{-1} \right) sd\phi ds \right)$$

となる。最後に、 L_1 と R についての期待値をとり全体を 1 から引く。最も近い基地局までの距離 R の確率密度関数は $f_R(r) = 2\pi\lambda_r r \exp(-\pi\lambda_r r^2)$ であるから、

$$p_{out}(\theta) = 1 - \int_0^\infty \int_0^\infty f_R(r) f_{L_1}(l_1) \\ \times \exp \left(-\lambda_r \int_r^\infty \int_0^{2\pi} \left(1 - \prod_{i=0}^n \left(1 + \theta \frac{l \left(\sqrt{s^2 + (\frac{i}{n} l_1)^2} - 2s(\frac{i}{n} l_1) \cos \phi \right)}{l \left(\sqrt{r^2 + (\frac{i}{n} l_1)^2} + 2r(\frac{i}{n} l_1) \cos \omega_1 \right)} \right)^{-1} \right) sd\phi ds \right) dr dl_1$$

となる。

5. 新しい評価指標

本節では、ハンドオフレートとアウトージ確率を用いて新しい評価指標を提案する。Cooperation area を広くすることでハンドオフレートを抑えることができる一方、基地局 B_0 から遠くに移動した場合でもハンドオフをしないためアウトージ確率は高くなる。この 2 つの指標はトレードオフの関係にあるため両方を加味した指標が必要になる。ユーザーにとって障害となるのは、端末のネットワーク接続が途切れることである。接続が途切れるのは、ハンドオフして切り替えに失敗した場合と受信電力の強度が十分でない場合である。よって、ハンドオフ先の基地局が交信の許容量を超えているなどの理由からハンドオフをしたときに接続に失敗する確率 m が既知であるとして、接続失敗率 CF を

$$CF(\theta) = \max_{\omega_1, \omega_2 \in [0, \pi]} (p_{out}(\theta) + mH(\omega_1, \omega_2)) \quad (5)$$

として定義すれば、ハンドオフレートとアウトージ確率を加味した指標が得られる。接続失敗率では、アウトージ確率と m 倍したハンドオフレートの和が最大になる場合を考えている。したがって、接続が途切れる確率が最も高い場合を評価している。

6. 数値実験

本章では、既存手法と提案手法でのハンドオフレート、アウトージ確率の解析結果とシミュレーションの比較を行う。考える領域は、 $100\sqrt{5}\text{m}$ 四方の正方形で、基地局の密度が 0.001 個/ m^2 、すなわち $\lambda_r = 0.001$ 、ユーザーの移動距離のパラメータを $\lambda_l = 0.001$ とする。

6.1. ハンドオフレート

本節では、ハンドオフレートの比較を行う。図8は既存手法によって求めたハンドオフレートとシミュレーション結果との比較、図9は提案手法によって求めたハンドオフレートとシミュレーション結果との比較である。ユーザーの移動方向 ω_1 は 0.25π , 0.5π , 0.75π , 0.95π の4つの値を取り、それぞれの場合において $[0, \pi]$ を12等分した13個の ω_2 でハンドオフレートの計算を行った。

既存手法では、 ω_1 が0に近くなると解析結果とシミュレーションで差異が見られる。しかし、提案手法では、 ω_1 が0に近い場合でも正確な結果が得られている。これは、 ω_1 が小さい場合に C_R の内部に基地局が B_0 以外存在しないことを考慮しないことによる影響が大きいためである。また、提案手法においても面積を導出する際に近似を行ったが、シミュレーションと合致していることからこの近似の影響が小さく、妥当な近似であることがわかる。

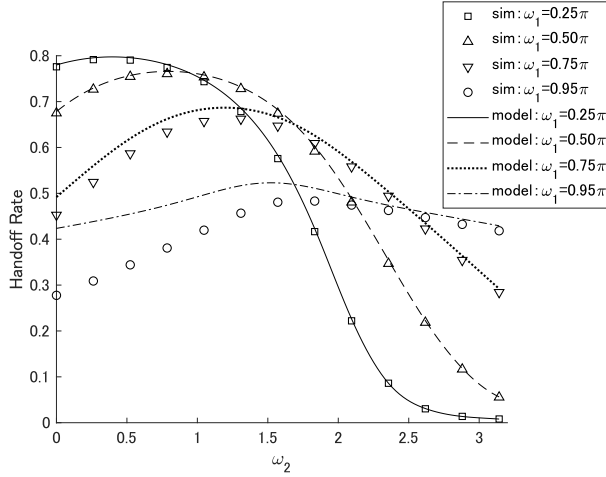


図 8: 既存手法

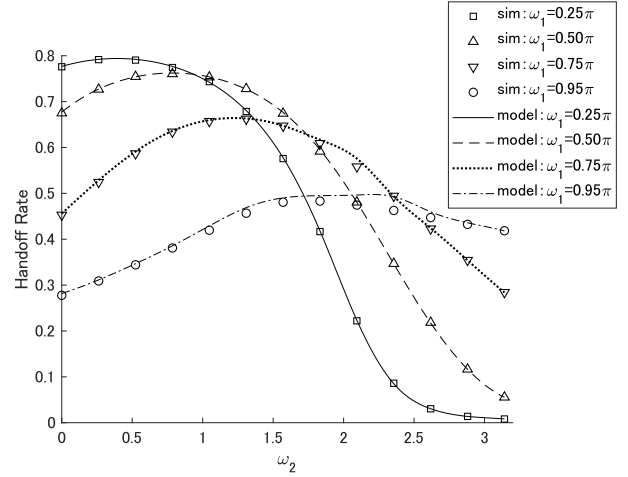


図 9: 提案手法

6.2. アウテージ確率

本節では、4章で求めたアウテージ確率とシミュレーション結果との比較を行った。パスロス関数の伝搬損失係数 $c = 1$ 、伝搬損失指数 $\beta = 4$ 、すなわち $l(r) = r^{-4}$ とする。しきい値 θ の単位はデシベル(dB); $x[\text{dB}] = 10^{x/10}$ とする。

今回は、経路の分割数 $n = 2$ と $n = 10$ の場合で数値実験を行った。図10では、 $n = 2$ の場合のアウテージ確率をシミュレーションと比較し、図11では、 $n = 2$ と $n = 10$ の場合のアウテージ確率の比較を行う。まず、しきい値の変動について考察する。しきい値が高くなるほど、接続するためには高いSIRが必要であるため、アウテージ確率が1に近づいていく。逆にしきい値が小さくなるほど、アウテージ確率も0に近づいていく。次に、 ω_1 について考察する。 ω_1 が π に近づくほどアウテージ確率は低くなっている。これは、 ω_1 が π に近づくときユーザーは基地局 B_0 に近づいていくため受信電力の強度が強くなるためと考えられる。最後に、分割数 n について考察する。 $n = 2$ の場合よりも $n = 10$ の場合のほうがアウテージ確率が高くなっている。 $n = 10$ のときは、11個の点の1つでもSIRがしきい値を下回れば接続が途切れたと判断されるので、 $n = 2$ のときよりもアウテージ確率が高くなりやすい。

6.3. 接続失敗率

本節では、ハンドオフレートとアウテージ確率から接続失敗率を計算し、図12に示す。ハンドオフに失敗する確率を $m = 0.1$ とする[14, 15]。ハンドオフレートの導出は6.1節、アウテージ確率の導出では6.2節と同様の設定で経路の分割数 $n = 2$ として実験を行った。

接続失敗率は、(5)式のようにハンドオフレートとアウテージ確率の和の最大値である。まず、ハンドオフレートに着目する。 ω_1 が0に近づくにつれてハンドオフレートが最大値をとるときの ω_2 は0に近づいていき、 $H(\omega_1, \omega_2)$ は、 $(\omega_1, \omega_2) = (0, 0)$ において最大値をとる。次に、アウテージ確率に着目する。アウテージ確率は、 ω_1 が小さくなるにつれて大きくなり、 $\omega_1 = 0$ において最大値をとる。以上

から本論文でのモデルでは, $(\omega_1, \omega_2) = (0, 0)$ でハンドオフレート, アウテージ確率はともに最大値をとる. したがって, 接続失敗率では, $(\omega_1, \omega_2) = (0, 0)$ の場合を考えている.

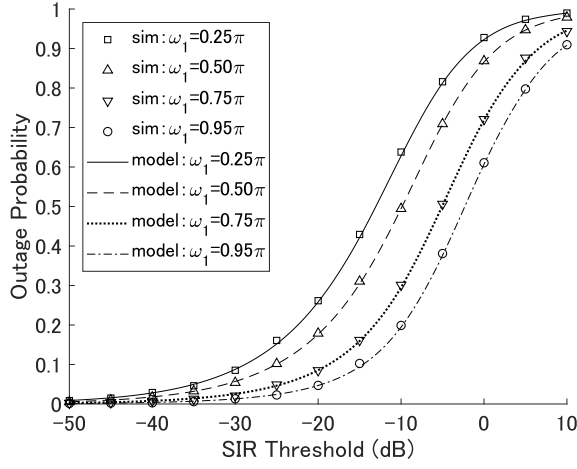


図 10: アウテージ確率 ($n = 2$)

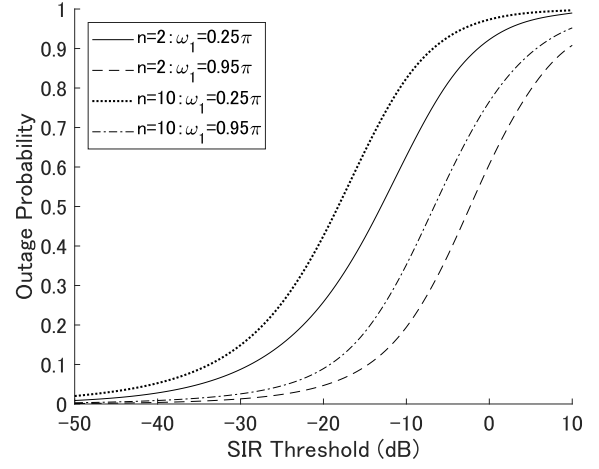


図 11: アウテージ確率 ($n = 2$ と $n = 10$ の比較)

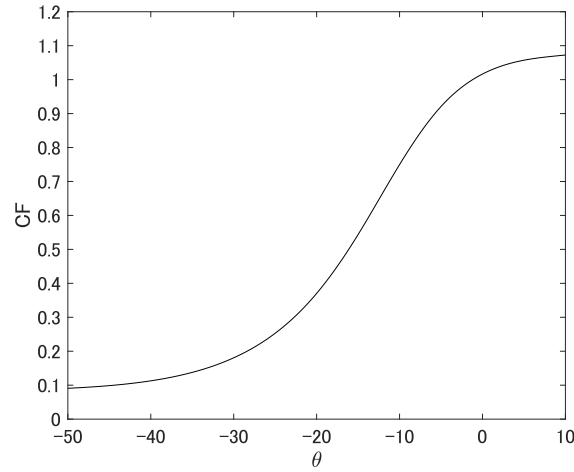


図 12: 接続失敗率 ($m = 0.1$)

7. おわりに

本論文では, cooperation area を考慮した高密度セルラネットワークにおける移動するユーザーの通信性能の評価を行う手法を提案した. まず, [11] で与えられたハンドオフレートを修正した. また, 原点 O から M_1 まで移動する場合のアウテージ確率を導出した. そして, この2つの指標を用いて新たな指標を提案した.

今後の課題としては, 2点考えられる. まず1つ目は, 原点 O から M_2 まで移動する場合のアウテージ確率を導出することである. また, ハンドオフ後の cooperation 戦略を定義する必要がある. 2つ目は, 今回提案した指標を用いて, 異なる cooperation 戦略を取った場合との比較を行うことである. 今回提案した指標がより低くなるような cooperation 戦略を取ることによって移動時においてもクオリティの高い通信を提供することができるだろう.

参考文献

- [1] Cisco, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016?2021 White Paper, Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html
- [2] Galinina, O., Pyattaev, A., Andreev, S., Dohler, M., Koucheryavy, Y., 5G multi-RAT LTE-WiFi ultra-dense small cells: Performance dynamics, architecture, and trends, *IEEE Journal on Selected Areas in Communications*, Vol.33 (2015), 1224-1240.
- [3] Liu, L., Garcia, V., Tian, L., Pan, Z., Shi, J., Joint clustering and inter-cell resource allocation for CoMP in ultra dense cellular networks, 2015 IEEE International Conference, 2560-2564.
- [4] Morosa, H., Harada, H., Morimoto, A., Nasata, S., Ishii, H., Okumura, Y., Cell identification performance based on hierarchical synchronization channels in dense small cell environment, *Signal Processing Advances in Wireless Communications (SPAWC)*(2013), 734-738.
- [5] Hao, P., Yan, X., Yu-Ngok, R., Yuan, Y., Ultra dense network: challenges, enabling technologies and new trends, *China Commun* Vol.13 (2016), 30-40.
- [6] Qu, D., Zhou, Y., Tian, L., Shi, J., User-Centric QoS-Aware Interference Coordination for Ultra Dense Cellular Networks, *Global Communications Conference (GLOBECOM)*(2016), 1-6.
- [7] Yang, S. R., Kao, C. C., Kan, W. C., Shih, T. C., Handoff minimization through a relay station grouping algorithm with efficient radio-resource scheduling policies for IEEE 802.16 j multihop relay networks, *IEEE transactions on vehicular technology* Vol.59 (2010), 2185-2197.
- [8] Dai, B., Yu, W., Sparse beamforming and user-centric clustering for downlink cloud radio access network, *IEEE Access*, Vol.2 (2014), 1326-1339.
- [9] Sadr, S., Adve, R. S., Handoff rate and coverage analysis in multi-tier heterogeneous networks, *IEEE Transactions on Wireless Communications*, Vol.14 (2015), 2626-2638.
- [10] Lin, X., Ganti, R. K., Fleming, P. J., Andrews, J. G., Towards understanding the fundamentals of mobility in cellular networks, *IEEE Transactions on Wireless Communications*, Vol.12 (2013), 1686-1698.
- [11] Chen, C., Zhao, X., Performance of handoff rate for two transitions under base stations cooperation in ultra-dense network, *IET Communications*, Vol.11 (2017), 218-224.
- [12] 三好直人, セルラネットワークの空間点過程モデル, オペレーションズ・リサーチ Vol.59 (2014), 678-683.
- [13] 電気情報通信学会, 知識ベース 知識の森 4 群 モバイル・無線, Available:http://www.ieice-hbkb.org/portal/doc_195.html
- [14] Gervasi, O., Murgante, B., Pardede, E., Apduhan, B. O., Computational Science and Its Applications-ICCSA 2010, Springer-Verlag Berlin Heidelberg (2010).
Gervasi, O., Gavrilova, M. L., Kumar, V., Laganà, A., Lee, H. P., Mun, Y., Tan, C. J. K., Computational Science and Its Applications-ICCSA 2005.
- [15] Zhao, H., Huang, R., Zhang, J., Fang, Y., Handoff for wireless networks with mobile relay stations, *Wireless Communications and Networking Conference (WCNC)*(2011), 826-831.

エスカレーターは歩くべきか、止まるべきか

鶴田 裕太郎 豊泉 洋
早稲田大学基幹理工学部応用数理学科

概要： エスカレーターでは止まるべきであるかを乗客の待ち時間の観点から分析する．ロンドンのホルボーン駅でエスカレーターの歩行に関する社会実験が行なわれ、片側を歩行者優先レーンとした場合よりも、両側止まった場合の方がより多くの乗客を運べて効率が良くなったと報告されている．今回は待ち行列理論を利用してエスカレーターの効率を待ち時間の観点で分析した．M/D/1 待ち行列を応用することで期待待ち時間を導出し、エスカレーターの混雑度により望ましい使用方法が異なることを明らかにした．

1. はじめに

エスカレーターの利用方法は、片側を歩行者優先レーンとして空ける習慣がある [1]．エスカレーターの歩行にはメリットとデメリットが存在する．歩くことによって移動の速度が上がる．しかし、歩行することによって間隔が空き、少ない人数しか運ぶことができない．エスカレーターの利用方法に関してロンドンのホルボーン駅で社会実験が行われた [2]．実験ではエスカレーターを歩かないように呼びかけを行い、乗客にエスカレーターの両側に立ち止まってもらった場合の乗客の動きを観測したところ、1日あたりの駅のエスカレーターを利用できた人数が増加した．ホルボーン駅での実験では、エスカレーターの利用者数の理論値は簡単な計算 [3] によって評価されている．一方、エスカレーターの消費電力については、M/D/1 待ち行列 [4] による数理モデルがある [5]．今回は M/D/1 待ち行列による期待待ち時間の数理モデル化を行った．M/D/1 待ち行列を利用することにより混雑度も考慮したより精確な評価が行なえる．

2. エスカレーターモデル

エスカレーターの速さを v ，エスカレーターの長さを l ，次の利用者が乗るまでの乗車間隔を d ，一度にエスカレーターに乗ることのできる最大人数を $n = l/d$ ，客の到着レートを λ とする．[図 1 参照]．利用者の到着はポアソン過程に従い、サービス時間は一定となる M/D/1 モデルを用いて考える．また使用率は $\rho (= \lambda / \mu) < 1$ で考える．一人当たりの期待待ち時間 $W(\lambda, \mu)$ は到着レート λ とサービスレート μ で表すことができる．ここでサービスレートは $\mu = v/d$ となる．エスカレーターに乗っている時間 (stop) を S とする．この時、利用時間は $S = S(l, v) = l/v$ となる．よってエスカレーター内での利用時間の期待値 h はエスカレーター内の期待待ち時間 W とエスカレーターに乗っている時間 S の和で表すことができる．

$$h(\lambda, v, l, d) = W_{M/D/1}(\lambda, v/d) + S(l, v) \quad (1)$$

次にエスカレーターを右側、左側の 2 列で利用する方法として次の 2 パターンを考える．

パターン 1[止まって利用するエスカレーター]

両レーン立ち止まって利用する．乗車間隔 d が半分となり、一人当たりの平均利用時間 $h_{2,stop}$ は以下のように表すことができる．

$$h_{2, stop}(\lambda, v, l, d) = W_{M/D/1}(\lambda, 2v/d) + S(l, v) \quad (2)$$

パターン 2[歩いて利用するエスカレーター]

パターン 2 では、容量が異なる独立な 2 本の列にそれぞれ利用者が来ると仮定する．また利用客は同じ割合でそれぞれのレーンに来ると仮定する．止まって乗る左側は $h_{1, stop}(\lambda/2, v, l, d) = W_{M/D/1}(\lambda/2, v/d) + S(l, v)$ となる．

エスカレーター上を歩く速さを v_W とする．エスカレーターに乗っている時間を S_W ，歩く際の間隔を d_W とする．この時、エスカレーターに乗っている時間 S_W は $S_W = l/(v + v_W)$ と表すことが

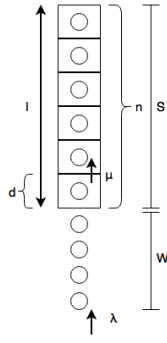


図 1: エスカレーターモデル図

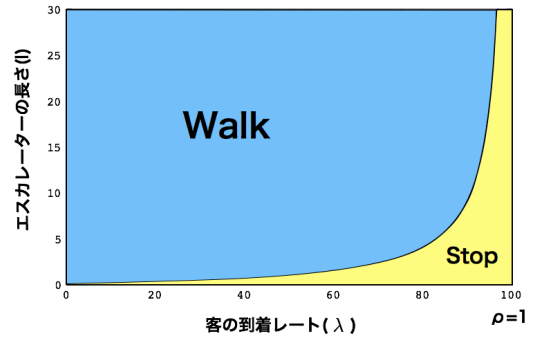


図 2: エスカレーターの混雑度と長さによっての利用時間効率境界図

できる．また，サービスレートは $\mu = (v + v_W)/d$ となる．右側のレーンを歩く一人当たりの期待利用時間は $W_{M/D/1}(\lambda/2, (v + v_W)/d) + S(l, v + v_W)$ で表すことができる．よって全体の一人当たりの期待利用時間は以下の式で表すことができる．

$$h_{2, walk}(\lambda, v, l, d) = 1/2 \left(W_{M/D/1}(\lambda/2, v/d) + S(l, v) + W_{M/D/1}(\lambda/2, (v + v_W)/d_W) + S(l, v + v_W) \right) \quad (3)$$

3. エスカレーターの例からみた結果

$v_W = v$, $d_W = 2d$, $d = l/30$, $v = 10l/3$ の場合を考える．利用時間の効率境界を図 2 に示す．エスカレーターが混雑していない場合はエスカレーターは歩いた方が良い場合が多くなることがわかる．しかし，エスカレーターが混み合っている時は止まって利用する方が良い場合が多くなる．

次に，エスカレーターの長さに着目する．短いエスカレーターの時の境界条件を見ると，あまり混んでいない時でも止まった方が良いという結果となった．

エスカレーターが空いている時は，エスカレーターに乗るまでの待ち時間があまり発生せず，利用時間はエスカレーターに乗っている時間のみに関わってくるので歩く場合の方が利用時間の効率が高くなっていると考えられる．一方，エスカレーターが混雑している時はエスカレーターに乗るまでの時間が大幅に伸びるため，全体の利用時間が伸びると考えられる．エスカレーターの長さが短い時に止まって利用する方が良い場合が多い理由としては，一度にエスカレーターに乗ることのできる最大人数が少ないためあまり混んでいなくても待ち時間が多く発生することが考えられる．

4. 今後の展望

実際のエスカレーターの条件で社会実験を行い，エスカレーターのより望ましい利用方法を明らかにする．

参考文献

- [1] 斗鬼 正一，エスカレーター片側空けという異文化と日本人のアイデンティティ，「情報と社会」第 25 号，(2015)．
- [2] Archie Bland, The tube at a standstill: why TfL stopped people walking up the escalators, The Guardian, (2016), 16 January.
- [3] Paul Davis, Goutam Dutta, Estimation of Capacity of Escalators in London Underground, (2002)．
- [4] 塩田 茂雄, 河西 憲一, 豊泉 洋, 会田 雅樹, 待ち行列理論の基礎と応用, 共立出版, (2014)．
- [5] Wei-wu Ma, Xiao-yan, Liu, Liqing, Xiangnan Shi, Chenn Q. Zhou, Research on the waiting time of passengers and escalator energy consumption at the railway station, *Journal of the Energy and Buildings*, (2009)．

雀荘における待ち行列に関する考察

井手 広康 奥田 隆史
愛知県立大学大学院 情報科学研究科

1. はじめに

雀荘とは料金を支払い麻雀を遊戯する場所のことを指し、客にはフリーとセットの2種類のタイプが存在する。フリーの客は1人で来店し、他のフリーの客あるいはメンバー（店員）を組み合わせた4人1組でゲームを行う。一方、セットの客は4人で来店し、麻雀卓を時間単位で貸し切る形態を指す。本研究では、雀荘における客はフリーのみを対象とし、セットは考慮しないものとする。

ここで雀荘における待ち行列は「客（ジョブ）のゲーム回数やプレイヤとしてのレベルが多様である」ことや、「卓（サーバ）における1回のゲーム時間が毎回異なり、客あるいはメンバーの卓への配分を最適化する必要がある」という観点から、VCHS 型待ち行列システム [1] として考えることができる。そこで本研究では、雀荘の待ち行列を VCHS 型待ち行列システムにおける量的 VCHS 問題としてモデル化し、多様な客に対する最適なメンバー数と卓数の組み合わせについて考察する。

2. 雀荘の待ち行列のモデル化

客は雀荘へ到着すると待ち席で自身の順番が来るまで待ち、メンバーは客が4人揃ったところで卓へと案内しゲームを開始する（4人にはメンバーを含むこともある）。また客ごとにゲーム回数は任意であるため、客がゲームを終えて卓の人数が欠けると、待ち状態の客あるいはメンバーを補充しゲームを継続させる（4人に満たない場合は待ち席へと戻る）。ここで図1に雀荘の待ち行列モデルを示す。以下の節では、雀荘の待ち行列モデルにおける客、メンバー、卓についてそれぞれ説明する。

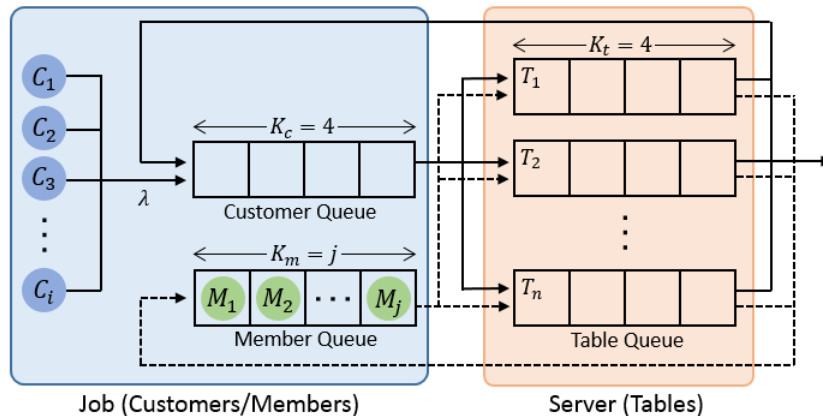


図 1: 雀荘の待ち行列モデル

2.1. 客

客 C_i ($i = 1, 2, \dots$) は1時間当たりの平均到着率が $\lambda = \{2, 4, 6\}$ のポアソン分布に従って到着し、図1のように待ち席である客キュー ($K_c = 4$) へと入る。なお客が客キューにいる間は待ち時間 W [分] が発生する。また客は残りゲーム回数 g を属性として持ち、 $g = 0$ となればゲームを終えて退出する。ここで客の到着時における残りゲーム回数 g は平均6回の指数分布に従うものとする。この客の残りゲーム回数 g についてメンバーは把握することができないが、ゲーム回数が残り1回 ($g = 1$) となった場合は申告し、メンバーはこれを知ることができる（以下これを「ラス半コール」と呼ぶ）。

2.2. メンバー

メンバー M_j ($j = 1, 2, \dots, 10$) は初期状態において図1のようにメンバーキュー ($K_m = j$) に入っている。雀荘においてメンバーは客に対してさまざまなサービスを提供するが、本研究では客数が4人に満たない場合において「次の客が到着するまでの“つなぎ”として卓に入る機能」のみ提供することとする（客が到着しなかった場合は最後までゲームを続ける場合もある）。なおメンバーは客とは異なり、図1のように卓キューから退出する場合には再度メンバーキューへと戻る。

2.3. 卓

雀荘に設置されている卓数を $n = \{1, 2, \dots, 10\}$ と置き、それぞれの卓キュー ($K_t = 4$) には客あるいはメンバーの4人が入りゲームを開始する (4人全員をメンバーとすることはできない)。ただしゲーム開始後 t 分までに新たな客が到着し、かつ1人以上のメンバーが卓に入っていれば、メンバーは客を自分の席へと案内することができる (ゲーム途中での交代)。一方、ゲーム開始後 t 分以上経過していれば、該当のゲーム終了時に客を卓へ案内する (ゲーム終了時での交代)。ここで交代可能時間 t は平均5分、1回のゲーム時間 s は平均30分の指数分布に従うものとする。

ゲーム終了までに待ち状態の4人が揃い、かつ使用可能な卓があった場合、卓キューに4人が移動しゲームを開始する。ただしこの段階でラス半コールが1以上発生し、かつ待ち状態の客数がラス半コール数以下であれば、卓キューへ移動せずにラス半コールが発生している卓のゲームが終了するまで各キューで待機する。ゲームが終了したら客は残りゲーム回数 g を1減らし次のゲームを開始する。ここで1人以上の客が $g = 0$ を満たし次のゲームを開始できないとき、つまり客が退出したことで4人に満たなくなった場合 (待ち状態の客やメンバーを含む)、図1のように各キューへと戻る。

3. シミュレーション結果

本章ではシミュレーション結果から到着率 $\lambda = \{2, 4, 6\}$ における客の平均待ち時間 $\bar{W}$ を求め、各到着率での最適なメンバー数と卓数の組み合わせについて考察する。なお本研究における最適な平均待ち時間は $\bar{W} < 10$ を満たす場合とする。1回のシミュレーションは一般的な雀荘の営業時間である15時間 (900分) を設定した。ここで10回のシミュレーション結果の平均値を図2(a)-(c)に示す。

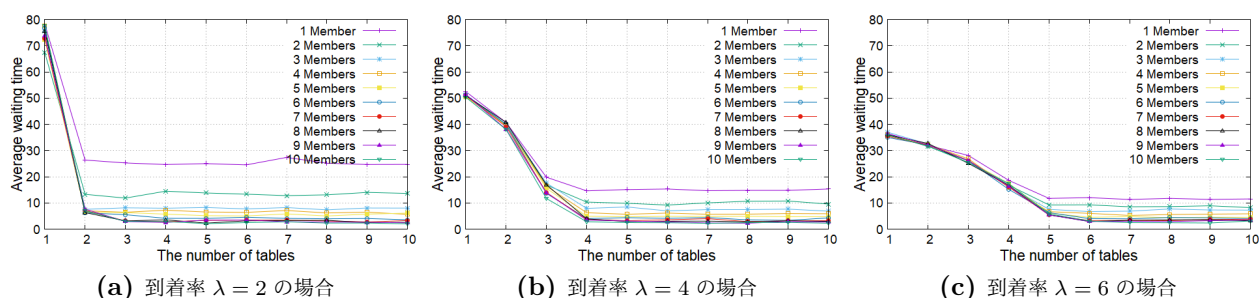


図 2: 到着率 λ の違いによる客の平均待ち時間 $\bar{W}$ の比較

まず図2(a)の到着率 $\lambda = 2$ の場合 (到着率: 小) について考察する。卓数 n が1のときに平均待ち時間は $\bar{W} \simeq 75$ と非常に高い数値を示しているが、 $n = 2$ では急激に減少し、これ以降はほぼ横ばいとなっている。また $n \geq 2$ においてメンバー数が3以上のとき $\bar{W} < 10$ を満たしていることから、 $\lambda = 2$ の場合では {メンバー数: 3, 卓数: 2} の組み合わせが最適解であるといえる。

次に図2(b)の到着率 $\lambda = 4$ の場合 (到着率: 中) について考察する。卓数 n が1のときに平均待ち時間は $\bar{W} \simeq 50$ と高い数値を示しているが、 $n = 4$ までは減少していき、これ以降はほぼ横ばいとなっている。また $n \geq 4$ においてメンバー数が3以上のとき $\bar{W} < 10$ を満たしていることから、 $\lambda = 4$ の場合では {メンバー数: 3, 卓数: 4} の組み合わせが最適解であるといえる。

最後に図2(c)の到着率 $\lambda = 6$ の場合 (到着率: 大) について考察する。卓数 n が1のときに平均待ち時間は $\bar{W} \simeq 35$ と高い数値を示しているが、 $n = 5$ までは緩やかに減少していき、これ以降はほぼ横ばいとなっている。また $n \geq 5$ においてメンバー数が2以上のとき $\bar{W} < 10$ を満たしていることから、 $\lambda = 6$ の場合では {メンバー数: 2, 卓数: 5} の組み合わせが最適解であるといえる。

4. おわりに

本研究では雀荘の待ち行列をモデル化し、多様な客に対する最適なメンバー数と卓数の組み合わせについて考察した。シミュレーション結果から、メンバー数よりも卓数が客の平均待ち時間に大きな影響を及ぼすことが明らかとなった。しかし現実にはメンバーは客に対してさまざまなサービスを提供する必要があるため、本研究のシミュレーション結果が一概に最適解とはいえない。そのため今後はサービスに対する客の満足度を考慮した最適なメンバー数についても分析していきたい。

参考文献

- [1] 清水哲, 能上慎也, 量的・質的 VCHS 問題のシミュレーション評価, 信学技報, vol.110, IN2010-108, pp.63-68, 2010.

マルコフモデルを用いたテニスプレイヤーの分類

大久保 幸美 豊泉 洋
早稲田大学 基幹理工学部

概要： テニスのスコアの動きをマルコフ連鎖としてモデル化し，ゲーム取得率の理論値と世界ランキングのプレイヤーの実測値の差を調べることで，プレイヤーを分類した．サーブでは，イーブンスコアでポイントを取得するプレイヤーが，効率的にゲームを取得し，リターンでは，自分に有利なポイントに力を入れているプレイヤーが効率的にゲームを取れていることがわかった．

1. はじめに

近年，あらゆるスポーツでデータ解析が進んでおり，監督はiPadを手放さない時代へと変わった．例えば，野球界でもマルコフモデル [1] を利用し，どのフェーズにおいてバント，敬遠，盗塁をすべきかといった戦略が，研究されている [2]．同様にテニス界においてもデータ解析は進んでおり，ポイント取得率から，ポイント取得率が一樣である単純なマルコフチェーンによりゲームの勝率が算出できる [3]．しかし，野球とは異なり，マルコフモデルを用いた戦略の研究はまだなされていない．そこで本研究では，マルコフモデルを利用し，世界ランキング Top10（2017 年 9 月 11 日時点）の各プレイヤーの 2014 年から 2017 年の 50 試合の試合データ [4] を用い，マルコフモデルから得られる理論値と比較した．サーブ，リターンそれぞれにおいて，効率的にポイントを取得しているプレイヤーを「強いプレイヤー」，非効率的にポイントを取得しているプレイヤーを「弱いプレイヤー」とし，抽出した．さらに，強いプレイヤーと弱いプレイヤーの 50 試合における各ポイントカウントごとの条件付きのポイント取得率から，各プレイヤーの特徴を抽出した．

2. ベルヌーイモデル

テニスの 1 ゲームは，相手より先に 4 ポイント取れば勝ち，0, 15, 30, 40, GAME とゲームのポイントを数える．(40-40) になった場合は，1 ポイント取ればアドバンテージとなり，先に 2 点差をつけてポイントを取ったプレイヤーの勝利となる．自分のカウントを x ，相手のカウントを y とし，ゲーム中のカウントを $i = (x - y)$ と表す．これにより，ゲームカウントごとに 20 状態 $\{i = (x - y) \mid x, y = 0, 15, 30, 40, \text{GAME}\}$ に分類できる (図 2 参照)．ベルヌーイモデルでは， p_{ij} をゲームカウント i から j への遷移確率とし， $\mathbf{P}_0 = (p_{ij})$ を吸収状態である SW(サーバー勝ち) と RW(リターン勝ち) を抜いた 18×18 の行列とし，ゲーム内でのポイントカウントに依らず，独立に確率 p でポイントを取るとした (ベルヌーイ試行) ときの状態遷移行列とした．ゲームは (0-0) からスタートし，初期状態確率ベクトルを $\pi_0 = (1, 0, \dots, 0)$ とおく． $\mathbf{q}$ はゲーム取得への遷移をあらわし，(40-0), (40-15), (40-30), (AD-40) の状態のとき p でその他は 0 である 18 次元ベクトルとし， $\mathbf{I}$ を単位行列とする．自分がゲームを取得する確率 $w(p)$ は次式で表せる．

$$w(p) = \sum_{n=0}^{\infty} \pi_0 \mathbf{P}_0^n \mathbf{q} = \pi_0 (\mathbf{I} - \mathbf{P}_0)^{-1} \mathbf{q}. \quad (1)$$

3. テニスプレイヤーの分類

トップ 10 のプレイヤーに関し，サーブ，リターンに分け，各プレイヤーの 50 試合の平均ポイント取得率 $(p_k)_{k=1, \dots, 10}$ を算出した． (p_k) を (1) に代入することで，各プレイヤーのゲーム取得率の理論値 $w(p_k)$ を算出した．また，各プレイヤーの 50 試合の平均ゲーム取得率 w_k を算出し，理論値 $w(p_k)$ と実測値 w_k をサーブリターンごとに比較した (図 1)．サーブのポイント取得率 p_k は 70 パーセント辺りに集中しており，10 人のプレイヤーの全体の傾向として，理論値 $w(p)$ の曲線より下回る．一方リターンは，40 パーセント辺りに集中し，理論値 $w(p)$ の曲線より上回った．この結果より， $w_k - w(p_k)$ の値でプレイヤーの分類を行った．サービスゲームの強い 3 人のプレイヤー (Murray, Djokovic, Wawrinka : ●) とサービスゲームの弱い 3 人のプレイヤー (Cilic, Dimitrov, Thiem : ▲)，またリターンゲーム

の強い3人のプレーヤー (Zverev, Thiem, Dimitrov : ●) とリターンゲームの弱い3人のプレーヤー (Nadal, Djokovic, Busta : ▲) に分類した。

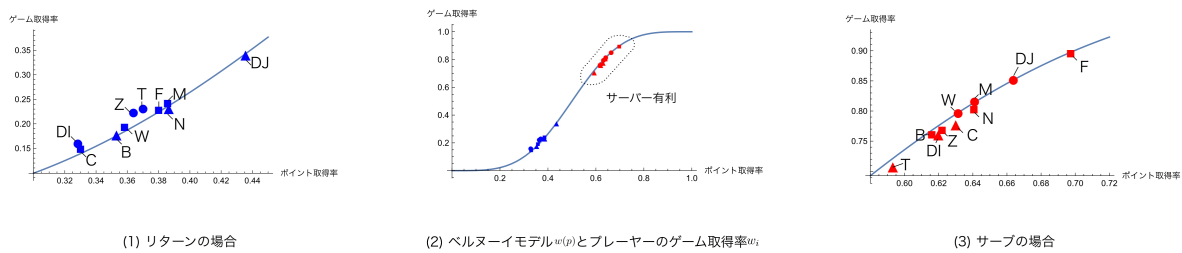


図1 トッププレーヤー10人 (Federer(F), Nadal(N), Murray(M), Zverev(Z), Cilic(C), Djokovic(DJ), Thiem(T), Dimitrov(DI), Wawrinka(W), Busta(B)) の分析

4. 分類したプレーヤーごとの特徴

図1の様に理論値 $w(p_k)$ と実測値 w_k が異なっている。この原因は、独立に同じ確率でポイントを取る単純なベルヌーイモデルと仮定していたが、実際はポイントカウントによってポイント取得率が変化しているためであると考えられる。ポイントカウントごとのポイント取得率 p_{ij} を実データから分析し、強いプレーヤーと弱いプレーヤーの特徴を抽出した。図2は、サーブ、リターン共に、強い3人のプレーヤーと弱い3人のプレーヤーのマルコフチェーン上の流れを表している。各ポイントカウントごとのポイント取得率 p_{ij} と p を比較し、それぞれのポイントカウントにおいて p_{ij} の方が高いプレーヤーが多いほど太い矢印で表している。

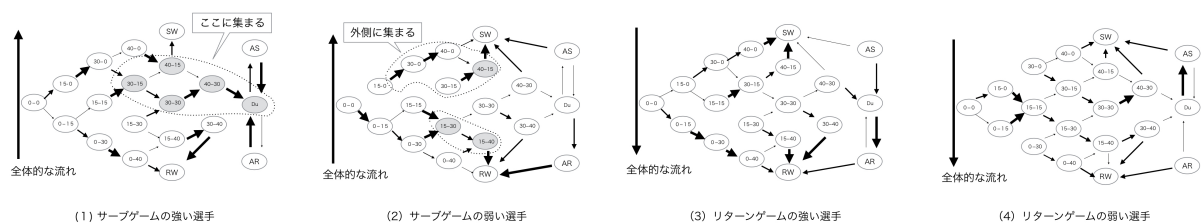


図2 強いプレーヤーと弱いプレーヤーの特徴

サーブの強いプレーヤーは、平衡カウントの (15-15), (30-30), (40-40) のときに力を入れているため、図2(1)の囲まれた部分にポイントが集まっている。一方リターンの強いプレーヤーは相手 (サーバー) からみて、(0-15), (0-30), (0-40), (15-40), (30-40) など自分が有利なときに力を入れていることがわかる。

5. 試合ごとのポイント取得率の変化

ポイント取得率の低いプレーヤーほど理論曲線 $w(p)$ からずれている理由として、ポイント取得率が低いプレーヤーは試合ごとにポイント取得率がばらついていることも考えられ、今後はこの検証をプロのプレーヤーのみならず、アマチュアのプレーヤーについても行っていきたい。

参考文献

- [1] Rick Durrett, Essentials of Stochastic Processes, Almost Final Version of the 2nd Edition, (2011).
- [2] 大井 一輝, マルコフ連鎖を用いた野球における状況別勝率計算とその応用, 第31回日本OR学会, (2013).
- [3] T. J. Barnett and S. R. Clarke, Combining player statistics to predict outcomes of tennis matches, IMA Journal of Management Mathematics, (2005).
- [4] tennisabstract.com

セルラネットワークにおける移動体のデータ量解析とハンドオフの機会制限モデルの考察

徳山 喜一

東京工業大学大学院 情報理工学研究科

概要：本論文では、セルラネットワークにおいて任意の軌道上を移動する移動体に対し、他の基地局からの障害電波強度を踏まえた形での平均受信データ量を計算する手法を提案する。さらに、これらの結果を用いて移動体のハンドオフに係るコストと受信データ量からなる評価関数の解析を行う。特に、ハンドオフに係るコストの低減を目的として、移動体通信においてハンドオフの対象となる基地局の割合を制限するモデルを考察し、移動体の移動速度に応じて制限の割合を調整する手法を提案した。その結果、移動体が速度を変化させながら移動する場合において、一定のハンドオフの制限が通信の評価関数を増大させる効果があることが分かった。

1. はじめに

今日の無線通信セルラネットワークにおいては、通信端末の普及も相まって通信量が膨大化を続けている。特に、IoT 技術の発展によって現代社会における様々な機器や装置がセルラ通信に対応し、それに伴う、移動体通信の更なる普及と通信量の増大が予測されている。2020 年に導入開始予定のセルラ通信の新方式 (5G) においても、移動体通信をより低コストかつ高パフォーマンスに実現することが喫緊の課題とされている [1]。また以上の背景を受けて、セルラネットワークのモデル解析の分野においても、移動体通信に対応したモデル解析の必要性が高まっている。

移動体通信に対応したセルラネットワークのモデル解析の分野として代表的なものに、移動体が通信先の基地局を変更する操作 (ハンドオフ) の平均回数、すなわちハンドオフレートの解析と、移動体の平均受信データ量の解析とがある。さらに、これら 2 つの解析量を用いて、移動体通信におけるハンドオフに係るコストと平均受信データ量とのトレードオフを、評価関数によって解析する研究が Bao と Linag によってなされている [2]。また、これら評価関数におけるハンドオフのコスト低減に関連して、密集した複数の基地局の間を通過する移動体に対して、ハンドオフを行う対象の基地局を制限する戦略 (CoMP) についても、多くの関連研究が存在する [3][4]。しかし、これら基地局の制限によるハンドオフレート低減の効果が通信の評価関数に及ぼす影響について、分析を行った研究は少ない。一方、これらの分析のためには、移動体のハンドオフレートに加えて、平均受信データ量の解析手法の確立が肝要となる。

移動体の平均受信データ量の解析においては、SN 比 (SNR) によって導出した受信電波強度から平均受信データ量を導出する方法が Madadi らによって提案されている [5]。しかし本結果では、平均受信データ量の導出において他の基地局からの障害電波 (インターフィアレンス) が考慮されておらず、また、通信端末の移動軌道として直線のみが想定されており、直線以外の軌道上を移動する場合に本結果は対応していない。セルラネットワークのモデル解析ではインターフィアレンスに基づくデータ量の導出が重要であること、また任意の軌道上の移動体に対する通信性能解析の必要性から、これらの課題の解決が望まれる。

本研究では、セルラネットワークにおいて任意の軌道上を任意の移動速度によって移動する移動体に対し、インターフィアレンスを基にした SI 比 (SIR) による平均受信データ量の計算方法を提案する。さらに、ハンドオフに係るコストの低減効果が評価関数に及ぼす影響を確認するため、移動体の移動速度に応じてハンドオフの回数に制限を加えるモデルを定義し、ハンドオフの制限と通信の評価関数の推移の関係について分析した。その結果、移動体が速度を変化させながら移動する場合において、単位区間における移動体のハンドオフ回数に一定の制限を加えた場合に評価関数が最大値を示すことが分かった。以上の結果により、速度変化を伴って移動する移動体に対して、ハンドオフを行う基地局の数を一定以下に制限し、ハンドオフに係るコストを低減するアプローチが有効であることが分かった。

2. システムモデル

2.1. ネットワークモデル

単一周波数帯の通信のみを扱うダウンリンク型の無線通信ネットワーク（ホモジニアスネットワーク）を考える。ここで、通信を供給する基地局は配置強度を λ に持つ 2 次元定常ポワソン点過程 Φ に従って配置されるとする。セルラネットワーク上の通信端末は最も近い位置に存在する基地局と通信することを仮定した場合、基地局は、それらの配置がなすポワソンボロノイセルの内部に含まれる通信端末に対し無線通信を行うこととなる。

通信端末が基地局から電波を受信する際、その電波の強度は $\alpha \geq 2$ の指数係数によって減衰するとする。また、受信電波に係るフェーディング係数はレイリーフェーディング、すなわち平均 1 の指数分布を仮定し、時刻 $t (t \geq 0)$ において通信先の基地局 b_0 の発信電波に発生するフェーディングを h_t 、それ以外からの基地局 $i \in \Phi/b_0$ からの発信電波に発生するフェーディングを $g_{t,i}$ と表す。このとき、任意の異なる値を持つ $s, t \geq 0$ に対して h_s と h_t は独立であり、また $g_{s,i}$ と $g_{t,i}$ は独立である。また、基地局が発信する電波強度は一律に P であるとする。

平面上の任意の点に配置された通信端末に対し、最短距離に位置する基地局 $b_0 \in \Phi$ との距離を r と表すとき、時刻 t における通信端末の基地局 $b_0 \in \Phi$ からの受信電波強度は $Ph_t r^{-\alpha}$ と表せられ、通信端末に発生するインターフィアレンス I_{b_0} は、

$$I_{b_0} = \sum_{i \in \Phi/\{b_0\}} P g_{t,i} R_i \quad (1)$$

と表される。これより、通信端末に対する SIR は

$$SIR = \frac{Ph_t r^{-\alpha}}{I_{b_0}} = \frac{h_t r^{-\alpha}}{\sum_{i \in \Phi/\{b_0\}} g_{t,i} R_i} \quad (2)$$

となる。ここで、 R_i は通信端末と基地局 $i \in \Phi/\{b_0\}$ との距離を表す。

通信端末とその通信先である基地局との間の距離 R_{b_0} が従う分布関数 $F(r)$ は

$$F(r) = P\{R_{b_0} \leq r\} = 1 - e^{-\lambda \pi r^2} \quad (3)$$

である。よって、その密度関数 $f(r)$ は

$$f(r) = \frac{dF(r)}{dr} = e^{-\lambda \pi r^2} 2\pi \lambda r \quad (4)$$

と表せられる。

2.2. 通信端末の移動モデルと通信の評価関数

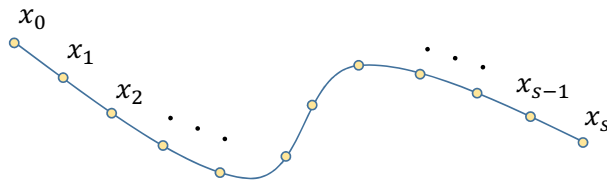


図 1: 移動体の移動軌道 $T(l)$ の模式図

長さを l とした任意の軌道 $T(l)$ に対し、その始点と終点をそれぞれ x_0, x_s として、 $s+1$ 個の点 $x_0, \dots, x_s$ を等間隔に配置する。そして、通信端末が軌道 $T(l)$ に沿って、 x_0 から x_s までの区間を s 秒間で等速で移動する。このとき、通信端末は軌道 $T(l)$ 上を速度 $v = l/s$ によって移動する。以上のモデルにより、任意の軌道を任意の速度で移動するモデルを考えることができる。

軌道 $T(l)$ 上を移動する通信端末に対し、通信の評価関数 $Q_1(l, s, \lambda)$ を以下のように定義する。

$$Q_1(l, s, \lambda) = D_1(s, \alpha) - C \cdot \mathcal{H}(l, \lambda) \quad (5)$$

ここで $D_1(s, \alpha)$ と $\mathcal{H}(l, \lambda)$ はそれぞれ、軌道 $T(l)$ 上を通信端末が移動する間の平均受信データ量とハンドオフレートであり、 $\mathcal{C}$ はハンドオフレートのデータ量への換算係数である。なお、ハンドオフレート $\mathcal{H}(l, \lambda)$ は以下のように求まることが知られている [2]。

$$\mathcal{H}(l, \lambda) = \frac{4\sqrt{\lambda}}{\pi} l \quad (6)$$

$D_1(s, \alpha)$ の解析については本論文の第 4 章で行う。

3. ハンドオフの機会制限のモデル

移動体の通信端末に対するハンドオフレートは、単位時間における通信端末の移動距離 (移動速度)、また基地局の配置密度に応じて増大する。ゆえに、移動体通信端末のハンドオフレートを小さくすることを考えるとき、移動速度に応じてハンドオフの対象となる基地局の配置密度を調整する必要がある。以上の考えに基づき、ここでは、移動体の移動速度に応じてハンドオフの回数を制限するハンドオフ制限レートを導入する。

3.1. ハンドオフ制限レート

配置強度 λ の 2 次元ポワソン点過程 Φ に対して、2 つの独立な 2 次元ポワソン点過程 Φ_1, Φ_2 を考え、それぞれの配置強度 λ_1, λ_2 が $\lambda = \lambda_1 + \lambda_2$ をみたすとする。このとき、 Φ の定常性から、 $\Phi = \Phi_1 + \Phi_2$ がほとんどいたるところで成立する。以上のことから、通信を供給する基地局の全体は、2 つの独立な 2 次元ポワソン点過程 Φ_1, Φ_2 によって配置される 2 つの基地局の集合に分割することができる。ここで、移動体の通信端末が Φ_1 によって配置される基地局とのみ通信を行うと仮定することで、ハンドオフの対象の基地局を制限することができる。特に、ハンドオフ制限レート η を

$$\eta = \frac{\lambda_1}{\lambda} \quad (7)$$

と表す ($0 \leq \eta \leq 1$)。

軌道 $T(l)$ 上を移動する通信端末に対し、ハンドオフ制限レートを η としてハンドオフの機会制限を行う場合の通信の評価関数 $Q_2(l, s, \eta, \lambda)$ は以下となる。

$$Q_2(l, s, \eta, \lambda) = D_2(s, \eta, \alpha) - \mathcal{C} \cdot \mathcal{H}(l, \eta\lambda) \quad (8)$$

$D_2(s, \eta, \alpha)$ の解析結果については、本論文の第 4 章で与える。

3.2. ハンドオフ制限レートの決定方法

上記で定めたハンドオフ制限レートの値を決定するためには、ハンドオフの対象となる基地局の配置が従う分布 Φ の配置強度 λ_1 を決定する必要がある。ここでは、単位時間において移動体通信端末がハンドオフを行う回数に上限を設けることで、 λ_1 を決定するための方法を与える。

proposition 1 移動体通信端末が距離 l を移動する間に、通信端末がハンドオフを行う回数の上限を k 回と定める ($k \geq 1$)。このとき、移動距離 l とハンドオフ上限回数 k に対する Φ の配置強度 $\lambda_1(k, l, \lambda)$ は以下のように定まる

$$\lambda_1(k, l, \lambda) = \frac{\pi^2}{16v^2} \left(k - k \sum_{i=0}^k Po(3\lambda | B |, i) - 3\lambda | B | \sum_{i=0}^{k-1} Po(3\lambda | B |, i) \right) \quad (9)$$

ここで、 $Po(\nu, i)$ と $|B|$ はそれぞれ以下で与えられる。

$$Po(\nu, i) = \frac{\nu^i}{i!} e^{-\nu} \quad (10)$$

$$|B| = \frac{4l}{3\sqrt{\lambda\pi}} \quad (11)$$

証明 定常ポワソン点過程 Φ によって配置された基地局の生成する通信区画は、 Φ により生成されるポワソンボロノイセルに等しいことに注意する。移動体が軌道 $T(l)$ 上を移動するときのハンドオ

プレートは、ポワソンボロノイセルを構成する辺の集合と軌道 $T(l)$ との交点の数に等しい。よって、 Φ の生成するポワソンボロノイセルを構成する辺の、中心点の配置分布を Ψ と表せば、ハンドオフレートは

$$\sum_{i=0}^{\infty} iP(\Psi(|B|) = i) = 3\lambda |B| = \frac{4\sqrt{\lambda}}{\pi} l \quad (12)$$

である。ここで、 Ψ の配置強度が 3λ であることと、軌道 $T(l)$ とポアソンボロノイセルの辺との交点数が、平均面積が式 (11) をみたす領域 B と Ψ との交点数に等しいことを用いた [6]。次に、移動体が軌道 $T(l)$ 上を移動するときに k 回を超えるハンドオフを行わないと仮定する。このとき、ハンドオフの対象となる基地局の配置密度 $\lambda_1(k, l, \lambda)$ は以下の式をみたす。

$$\begin{aligned} \frac{4\sqrt{\lambda_1(k, l, \lambda)}}{\pi} &= \sum_{i=0}^k iP(\Psi(|B|) = i) + \sum_{i=k+1}^{\infty} kP(\Psi(|B|) = i) \\ &= \sum_{i=0}^k iP(\Psi(|B|) = i) + k\left(1 - \sum_{i=1}^k P(\Psi(|B|) = i)\right) \\ &= k - e^{-3\lambda|B|} \left\{ k \sum_{i=0}^k \frac{(3\lambda|B|)^i}{i!} - 3\lambda|B| \sum_{i=0}^{k-1} \frac{(3\lambda|B|)^i}{i!} \right\} \end{aligned} \quad (13)$$

以上より、 $\lambda_1(k, l, \lambda)$ の式 (9) が従う。 □

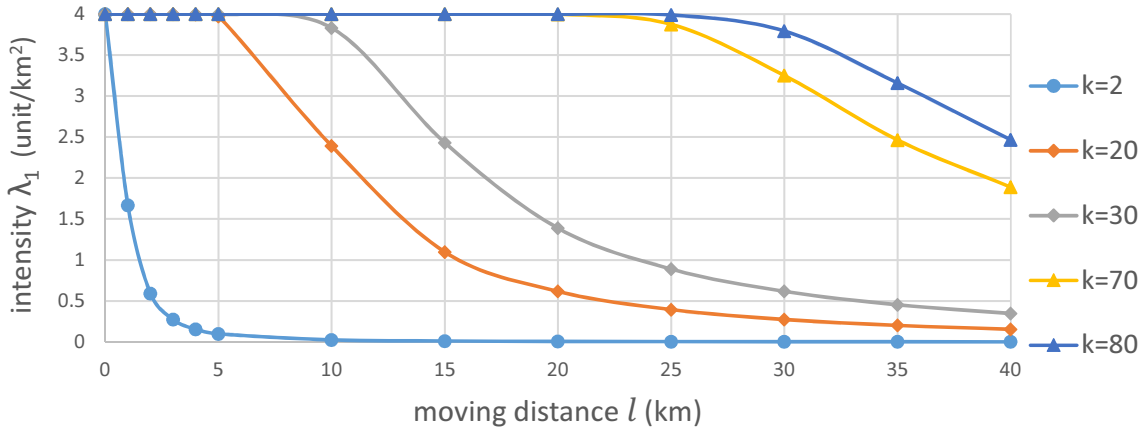


図 2: $\lambda_1(k, l, \lambda)$ の数値例

図 2 は、ハンドオフ制限回数をそれぞれ $k = 2, 20, 30, 70, 80$ として、 $\lambda_1(k, l, \lambda)$ を通信端末の移動距離 l について推移させた結果である。なお、基地局の配置密度は $\lambda = 4$ としている。このように、通信端末が一定の距離を移動する際のハンドオフ制限レートの調整を、 k の値によって行うことが可能である。なお、単位時間における移動距離を考えることで、通信端末の速度によって $\lambda_1(k, l, \lambda)$ を推移させることも可能である。

4. 任意の移動軌道におけるデータ量の解析

ここでは、第 2 章で定義した通信端末の移動モデルにおいて、距離 l の移動軌道 $T(l)$ 上を s 秒で移動する通信端末に対する平均受信データ量 $D_1(s, \alpha)$ の導出方法を与える。さらに、第 3 章で定めたハンドオフの機会制限のモデルに基づき、軌道 $T(l)$ 上をハンドオフ制限レートを η として移動する通信端末に対する平均受信データ量 $D_2(s, \eta, \alpha)$ の導出方法も与える。これにより、第 2 章の式 (5) や第 3 章の式 (8) で定めた通信の評価関数の解析が可能となる。

proposition 2 距離 l の軌道 $T(l)$ 上を s 秒かけて通信端末が移動し、かつ通信端末は常に最短距離に配置されている基地局と通信を行うものとする。このとき、通信端末が s 秒間に受信するデータ量

の平均 $D_1(s, \alpha)$ は以下で与えられる。

$$D_1(s, \alpha) = s \cdot \xi_1(\lambda, \alpha) \quad (14)$$

ここで、 $\xi_1(\lambda, \alpha)$ は以下の式で与えられる。

$$\xi_1(\lambda, \alpha) = 2\pi\lambda \int_{r>0} r e^{-\pi\lambda r^2} \int_{t>0} e^{-\sigma^2 r^\alpha (e^t - 1)} \cdot \mathcal{L}_{I_r}(r^\alpha (e^t - 1)) dt dr, \quad (15)$$

$$\mathcal{L}_{I_r}(r^\alpha (e^t - 1)) = \exp \left(-\pi\lambda r^2 (e^t - 1)^{2/\alpha} \int_{(e^t - 1)^{-2/\alpha}}^{\infty} \frac{1}{1 + x^{\alpha/2}} dx \right). \quad (16)$$

証明 通信端末の移動軌道 $T(l)$ 上に等間隔に配置された $s + 1$ 個の点 $x_0, \dots, x_s$ について、定常ポワソン点過程 Φ により配置された最短距離にある基地局と $x_i (0 \leq i \leq s)$ との距離を R_i とする。このとき、 x_i におけるインターフィアレンスを I_{x_i} と表せば、地点 x_i における通信端末の平均受信データ量 $\tau(\alpha, i)$ は

$$\tau_1(\alpha, i) = \log \left(1 + \frac{h_i R_i^{-\alpha}}{I_{x_i}} \right), \quad (17)$$

と表される。ゆえに、軌道 $T(l)$ 全体における受信データ量 $D_1(s, \alpha)$ は

$$D_1(s, \alpha) = \mathbf{E} \left[\sum_{i=1}^s \log \left(1 + \frac{h_i R_i^{-\alpha}}{I_{x_i}} \right) \right], \quad (18)$$

である。ここで、 Φ の定常性から、 x_i における最短距離の基地局との距離 R_i は式 (3) と同一の分布関数を持つ。ゆえに $D_1(s, \alpha)$ は以下のように計算される。

$$\begin{aligned} D_1(s, \lambda, \alpha) &= \sum_{t=1}^s \mathbf{E} \left[\log \left(1 + \frac{h r^{-\alpha}}{I_{b_0}} \right) \right] \\ &= s \cdot \mathbf{E} \left[\log \left(1 + \frac{h r^{-\alpha}}{I_{b_0}} \right) \right]. \end{aligned} \quad (19)$$

式 (19) の最終行における平均はホモジニアスネットワークにおけるある一点での平均受信データ量に等しい。よって式 (19) には [7], 定理 1 を適用でき、その結果は式 (15), (16) に合致する。 $\square$

proposition 3 距離 l の軌道 $T(l)$ 上を s 秒かけて通信端末が移動し、かつ通信端末は、ハンドオフ制限レートを η とした配置密度により配置される基地局とのみ通信を行うものとする。このとき、通信端末が s 秒間に受信するデータ量の平均 $D_2(s, \eta, \alpha)$ は以下で与えられる。

$$D_2(s, \eta, \alpha) = s \cdot \int_0^\infty \frac{1}{1+z} \cdot \frac{1}{1 + 2^{\frac{1-\eta}{\eta}} \mathcal{J}(\alpha) z^{2/\alpha} + z^{2/\alpha} \int_{z^{-2/\alpha}}^\infty \frac{1}{1+v^{\alpha/2}} dv} dz \quad (20)$$

ここで、 $\mathcal{J}(\alpha)$ は以下をみたす量である。

$$\mathcal{J}(\alpha) = \int_0^\infty \frac{v}{1+v^\alpha} dv \quad (21)$$

証明 通信端末の移動軌道 $T(l)$ 上に等間隔に配置された $s + 1$ 個の点 $x_0, \dots, x_s$ のそれぞれについて、定常ポワソン点過程 Φ_1 により配置された最短距離にある基地局 b_0 と $x_i (0 \leq i \leq s)$ との距離を R_{i,b_0} とする。このとき、 x_i におけるインターフィアレンス I_{i,b_0} は、 Φ_1 から b_0 を除いた他の基地局の集合と、 Φ_2 全体の基地局からの発信電波の合計である。*proposition 2* の式 (19) までの議論と同様に、地点 x_i における通信端末の平均受信データ量 $D_2(s, \eta, \alpha)$ は

$$\begin{aligned} D_2(s, \eta, \alpha) &= \mathbf{E} \left[\sum_{i=1}^s \log \left(1 + \frac{h_i R_{i,b_0}^{-\alpha}}{I_{i,b_0}} \right) \right] \\ &= s \cdot \mathbf{E} \left[\log \left(1 + \frac{h r^{-\alpha}}{I_{i,b_0}} \right) \right] \end{aligned} \quad (22)$$

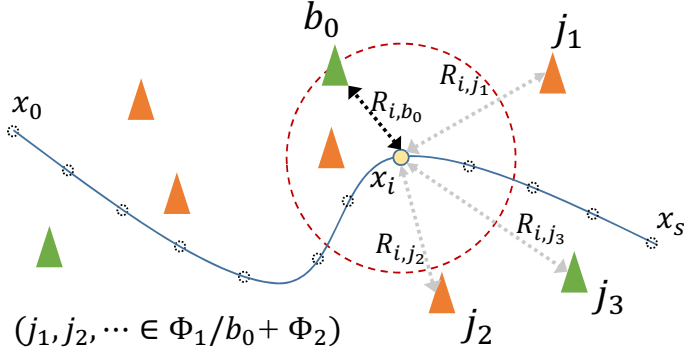


図 3: 軌道上の通信端末と基地局の配置の模式図

と表される。ここで、式 (22) の最終行における r は Φ_1 における最短距離の点との距離分布の密度関数 (4) のパラメータであり、すなわち

$$f_1(r) = e^{-\lambda_1 \pi r^2} 2\pi \lambda_1 r \quad (23)$$

をみだす。式 (22) の最終行における平均は、[8], Lemma 1 によって以下のように計算可能である。

$$\begin{aligned} \mathbf{E} \left[\log \left(1 + \frac{hr^{-\alpha}}{I_{i,b_0}} \right) \right] &= \int_0^\infty f_1(r) \mathbf{E}_{h, I_{i,b_0}} \left[\log \left(1 + \frac{hr^{-\alpha}}{I_{i,b_0}} \right) \mid r \right] dr \\ &= \int_0^\infty \int_0^\infty 2\pi \lambda_1 r e^{-\lambda_1 \pi r^2} \frac{1}{z} (1 - \mathbf{E}_h[e^{-zh}]) \mathbf{E}_{I_{i,b_0}} [e^{-zr^\alpha I_{i,b_0}} \mid r] dz dr \\ &= \int_0^\infty \int_0^\infty 2\pi \lambda_1 r e^{-\lambda_1 \pi r^2} \frac{1}{1+z} \mathcal{L}_{I_{i,b_0}}(zr^\alpha) dz dr \end{aligned} \quad (24)$$

ここで、式 (24) の最終行における $\mathcal{L}_{I_{i,b_0}}(w)$ はインターフィアレンス I_{i,b_0} の Laplace 変換である。 $\mathcal{L}_{I_{i,b_0}}(w)$ は以下のように計算される。

$$\begin{aligned} &\mathcal{L}_{I_{i,b_0}}(w) \\ &= \mathbf{E}_{\Phi_1/b_0 + \Phi_2, g_{t,i}} \left[\exp(-w \sum_{j \in \Phi_1/b_0 + \Phi_2} g_{t,i} R_{i,j}^{-\alpha}) \right] \\ &= \mathbf{E}_{\Phi_1/b_0} \left[\prod_{j \in \Phi_1/b_0} \mathbf{E}_g [\exp(-wg R_{i,j}^{-\alpha})] \right] \cdot \mathbf{E}_{\Phi_2} \left[\prod_{j \in \Phi_2} \mathbf{E}_g [\exp(-wg R_{i,j}^{-\alpha})] \right] \\ &= \exp \left(-2\pi \lambda_1 \int_r^\infty (1 - \mathbf{E}_g [\exp(-wgv^{-\alpha})]) v dv \right) \exp \left(-2\pi \lambda_2 \int_0^\infty (1 - \mathbf{E}_g [\exp(-wgv^{-\alpha})]) v dv \right) \\ &= \exp \left(-2\pi \lambda_1 \int_r^\infty \frac{v}{1 + w^{-1}v^\alpha} dv \right) \cdot \exp \left(-2\pi \lambda_2 \int_0^\infty \frac{v}{1 + w^{-1}v^\alpha} dv \right) \\ &= \exp \left(-2\pi \lambda_1 \int_r^\infty \frac{v}{1 + w^{-1}v^\alpha} dv \right) e^{-2\pi \lambda_2 \mathcal{J}(\alpha) \cdot w^{2/\alpha}} \end{aligned} \quad (25)$$

ゆえに、式 (24), (25) を組み合わせれば、

$$\begin{aligned} &\mathbf{E} \left[\log \left(1 + \frac{hr^{-\alpha}}{I_{i,b_0}} \right) \right] \\ &= \int_0^\infty \int_0^\infty 2\pi \lambda_1 r e^{-\lambda_1 \pi r^2} \frac{1}{1+z} \exp \left(-2\pi \lambda_1 \int_r^\infty \frac{v}{1 + (zr^\alpha)^{-1}v^\alpha} dv \right) e^{-2\pi \lambda_2 \mathcal{J}(\alpha) \cdot (zr^\alpha)^{2/\alpha}} dz dr \\ &= \int_0^\infty \int_0^\infty \frac{2\pi \lambda_1}{1+z} r e^{-\pi \lambda_1 r^2 \left\{ 1 + 2 \frac{\lambda_2}{\lambda_1} \mathcal{J}(\alpha) z^{2/\alpha} + z^{2/\alpha} \int_{z^{-2/\alpha}}^\infty \frac{1}{1+v^{2/\alpha}} dv \right\}} dr dz \\ &= \int_0^\infty \frac{1}{1+z} \cdot \frac{1}{1 + 2 \frac{1-\eta}{\eta} \mathcal{J}(\alpha) z^{2/\alpha} + z^{2/\alpha} \int_{z^{-2/\alpha}}^\infty \frac{1}{1+v^{2/\alpha}} dv} dz \end{aligned} \quad (26)$$

以上、式 (22) と式 (26) により、題意が従う。 $\square$

5. ハンドオフの機会制限を伴った通信の評価関数の推移

これまでの解析結果により、任意の移動軌道を移動する通信端末に対する平均受信データ量とハンドオフレートが得られた。これらの結果を用いて、距離 l の軌道 $T(l)$ 上を s 秒で移動する通信端末に対する通信の評価関数の推移を観察する。特に、第3章で定めたハンドオフの機会制限のモデルに基づき、距離 l を移動する間のハンドオフ制限回数 k ($k = 1, 2, \dots$) についての評価関数の推移を観察し、その特徴を述べる。評価関数の推移は、通信端末が軌道上を一定の速度で移動する場合と、ある確率分布に基づき不規則に速度を変化させて移動する場合の評価関数の平均によって与える。

5.1. 等速の移動体に対する通信の評価関数

図4は、軌道 $T(l)$ 上を移動する通信端末の移動速度を定数として、評価関数 $Q_2(l, s, \eta, \lambda)$ を k について推移させた結果である ($\lambda = 3, \alpha = 4$)。

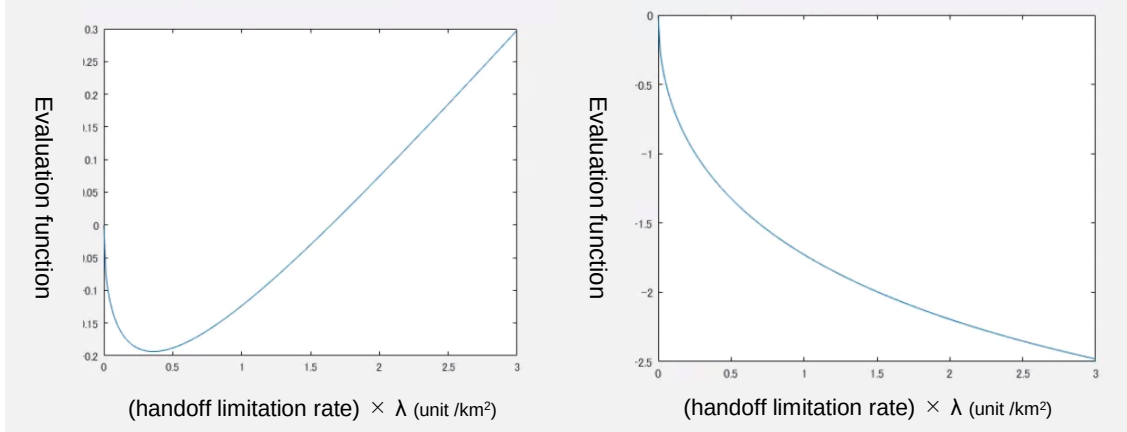


図 4: $Q_2(l, s, \eta, \lambda)$ の数値例 (左図: $v = l/s = 0.01$ (km/s)、右図: $v = l/s = 0.06$ (km/s))

以上の結果から、評価関数は k の推移について上方向の凸性を示さず、 $k = 0$ もしくは $k \rightarrow \infty$ (ハンドオフ制限なし) の場合に最大値を示すことが分かる。すなわち、軌道上を等速で移動する通信端末に対しては、ハンドオフ機会制限を行わず、常に最短距離の基地局と通信を行うようにハンドオフを行うか、移動速度が一定以上大きい場合はハンドオフを全く行わない場合が、もっとも評価関数の値が高いということである。

5.2. 速度変化の移動体に対する通信の評価関数の平均

図5は、軌道 $T(l)$ 上を移動する通信端末の移動速度 v を、区間 $[0, 2\mu]$ の一様分布に従う確率変数 (平均 μ) として、評価関数 $\mathbf{E}_v[Q_2(l, s, \eta, \lambda)]$ (ただし $v = l/s$) を k について推移させた結果である。 ($\mu = l/s = 0.012$ (km/s), $\lambda = 3, \alpha = 4$) 以上の結果から、評価関数は k の推移について上方向の凸性

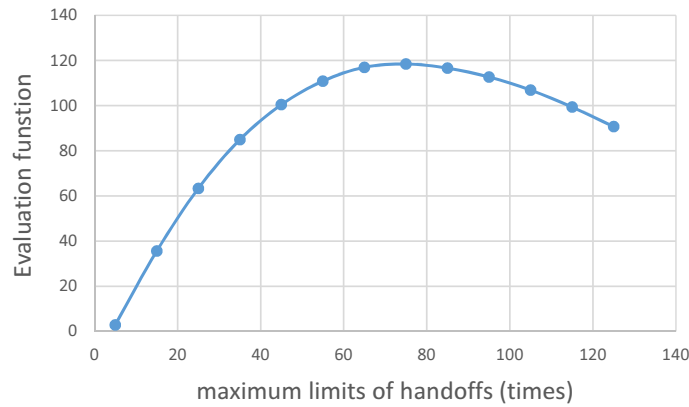


図 5: $\mathbf{E}_v[Q_2(l, s, \eta, \lambda)]$ の数値例

を示すことが分かる。すなわち、軌道上を不規則な等速で移動する通信端末に対しては、ハンドオフ

機会制限によるハンドオフのコスト低減のアプローチが有効であることが分かる。なお、速度が不規則に変化する場合に評価関数の平均が上方向の凸性を示す理由としては以下が推測される：

- 通信端末の移動速度が大きい場合には、式 (9) で定められたハンドオフ対象の基地局 Φ_1 の配置強度 $\lambda_1(k, v, \lambda)$ が小さい値を示すため、ハンドオフ制限レートも小さくなるが、このとき $Q_2(l, s, \eta, \lambda)$ はハンドオフ制限レートに関して単調減少の傾向を示すため、結果 $Q_2(l, s, \eta, \lambda)$ は k が比較的小さい場合に大きな値をとることが期待される。
- 通信端末の移動速度が小さい場合には、式 (9) で定められたハンドオフ対象の基地局 Φ_1 の配置強度 $\lambda_1(k, v, \lambda)$ が大きい値を示すため、ハンドオフ制限レートも大きくなるが、このとき $Q_2(l, s, \eta, \lambda)$ はハンドオフ制限レートに関して単調増加の傾向を示すため、結果 $Q_2(l, s, \eta, \lambda)$ は k が比較的大きい場合に大きな値をとることが期待される。
- 上記の2つの相互作用により、物体の移動速度が一定の条件をみたして変動する場合に、 $Q_2(l, s, \eta, \lambda)$ は k の値が中間の範囲で最大値をとり、 k の推移は上方向の凸性を示す。

6. 結論

本研究では、セルラネットワークにおいて任意の軌道上を移動する移動体に対し、SIR による平均受信データ量の計算方法を提案した。さらにこれを用いて、ハンドオフレートと平均受信データ量からなる通信の評価関数の推移分析を行った。特に、ハンドオフのコストの低減を目的として、移動体通信においてハンドオフの対象となる基地局の割合を制限するモデルを考察し、移動体の移動速度に応じて制限の割合を調整する手法を提案した。その結果、移動体が速度を変化させながら移動する場合において、一定のハンドオフレートの制限が通信の評価関数を増大させる効果があることが分かった。移動体が不規則な速度で移動する例は現実においても十分あり得ることから、ハンドオフの機会制限の有効性を確認できた。なお、そのほかに通信の評価関数を増大させる手段として、受信電波の周波数帯を変更するハンドオフ（垂直ハンドオフ）を行うというものがある。よって、周波数帯の変更が発生しうるヘテロジニアスネットワークへと本モデルを拡張し、移動体の移動速度に応じて垂直ハンドオフを行う場合と、ハンドオフの機会制限を行う場合との間で有効性の程度を比較することが必要である。そのために、ヘテロジニアスネットワークにおける通信の評価関数の解析、ひいては平均受信データ量の解析を行うことが、今後の課題となる。

参考文献

- [1] Osseiran.A., Boccardi.F., Braun.,V., Kusume.K., Marsch.P., Maternia.M., Queseth.O., Schellmann.M., Schotten.H., Scenarios for 5G mobile and wireless communications: the vision of the METIS project, IEEE Communications Magazine, **52** No.5(2014), 26–35.
- [2] W. Bao, B. Liang, Stochastic Geometric Analysis of User Mobility in Heterogeneous Wireless Networks, IEEE Journal. Selected Areas in Communications, **33** No.10(2015), 2212–2225.
- [3] C. Chen, X.Zhao, Performance of handoff rate for two transition under base station cooperation in ultra-dense network, IET Journals. IET Communications, **11** No.2(2017), 218–224.
- [4] M.Liu, Y.Teng, M.song, CoMP Handover Probability Analysis with Different Handover Schemes in Ultra-Dense Networks, Wireless Communications and Networking Conference (WCNC), IEEE 2017.
- [5] P.Madadi, F.Baccelli, and G.Veciana, On Temporal Variation in Mobile User SNR with Applications to Perceived QoS, 14th International Symposium on Modeling and Optimization in Mobile, Ad-Hoc and Wireless Networks (WiOpt), 2016.
- [6] K.Tokuyama, N.Miyoshi, Data Rate and Handoff Rate Analysis for User Mobility in Cellular Networks, Wireless Communications and Networking Conference (WCNC), IEEE 2018.
- [7] J. Andrews, F. Baccelli, and R. Ganti, A Tractable Approach to Coverage and Rate in Cellular Networks, IEEE Journal. Transactions on Communications, **59** No.11 (2011), 3122–3134.
- [8] K. Hamdi, A Useful Lemma for Capacity Analysis of Fading Interference Channels, IEEE Journal. Transactions on Communications, **58** No.2(2010), 411–416.

リードタイムが一般の分布にしたがう (r, Q) 方式在庫管理システムの解析

高木秀

東京工業大学 情報理工学院 数理・計算科学系

概要： リードタイム（調達期間）が一般の分布にしたがう (r, Q) 方式在庫管理システムの変動を再生型確率過程 (regenerative process) として捉え、定常状態における在庫量の分布を導出した。それを用いて、システム全体にかかるコストを定式化し、リードタイムが様々な分布にしたがう場合について数値計算例を行った。

1. はじめに

生産システムやサービス施設に現れる在庫量の変動を確率モデルを用いて表現し、（コストという面から見て）適切な在庫管理を行うための方式やシステムパラメータを決定する研究が長年に渡り行われている。この分野の研究では、リードタイム（調達時間）が指数分布にしたがうことを仮定して、進めている研究が多く ([5][6])、この仮定を外すことができれば、より一般的な結果を得られることは明らかである [1]。リードタイムが指数分布以外のモデルを考えている先行研究として、以下のようなものがある。Sahin [2] は、 (s, S) 方式在庫システム（在庫量が s 個以下になると、それが S 個になる分だけ調達する管理方式）において、需要の到着が一般分布にしたがい、リードタイムが定数の場合を考えている。Srinivasan [3] は、 (s, S) 方式在庫管理システムを再生型確率過程 (regenerative process) を用いて表現し、リードタイムが一般の分布にしたがう場合を考えている。Mohebbi and Hao [4] は、需要が複合ポアソン過程にしたがい、リードタイムがアーラン分布や超指数分布にしたがう (r, Q) 方式在庫管理システム（在庫量が r 個以下になると、 Q 個の在庫を補給する管理方式）を考えている。Saffari ら [1] は、リードタイムが一般の分布にしたがい、 (r, Q) 方式で、需要が在庫を 1 つずつ持っていく場合について考えており、定常分布の解析解を導出している。

そこで、在庫を 1 つずつ持っていく場合のモデルでは表現出来ないような現象も存在するため (ex. 印刷機やコピー機用の紙補充問題)、在庫を複数個持って行く場合も考慮したい。本研究では、リードタイムが一般の分布にしたがい、 (r, Q) 方式で各需要が在庫を幾何分布にしたがう数だけ持っていく (r, Q) 方式在庫管理システムを考える。具体的には、在庫量の変動を再生型確率過程として捉える。そこで、定常状態における各状態で平均滞在時間を導出することにより、各状態の定常状態確率を求める。

本稿の構成は以下の通りである。第 2 節で、どのような問題を考えているのかを確認し、モデルの説明を行う。次に第 3 節で、定常分布の解析解を求める。第 4 節では、第 3 節で得られた結果を用いて、コスト関数を定義し、数値計算を行う。第 5 節では、まとめと今後の課題を述べる。

2. モデルの説明

2.1. 問題設定

- 需要はシステムに到着率 λ のポアソン到着をする。
- 需要がシステム到着後に持っていく在庫の個数は、平均 $\frac{1}{p}$ の幾何分布にしたがう。
- (r, Q) 方式にしたがって在庫を補給し、リードタイムは (密度関数をもつ) 一般の分布にしたがう ($r < Q$)。
- 持っていきたい量がそのときの在庫量を越える需要が到着した場合は、そのときにある在庫だけ持っていく。
- 在庫切れが起きている間に到着した需要は失われる (lost-sales)。

2.2. 再生型確率過程

時刻 t における在庫量を表す確率変数を $I(t)$ とする。確率過程 $\{I(t)\}$ は幾何分布の無記憶性より、「 $I(t)$ が r 以下になり、発注を行う時刻（リードタイムの開始時刻）を再生点とする再生型確率過程」となる。また、確率過程 $\{I(t)\}$ の状態空間を X とすると、在庫量は最大でも $Q + r$ までしかとらないので、状態空間 X は以下のように定義することが出来る。 $X := \{0, 1, \dots, Q + r\}$ 。

2.3. 記号一覧

- L リードタイムを表す確率変数
 $f(l)$ リードタイムがしたがう確率分布の密度関数
 τ 再生型確率過程のサイクルの長さを表す確率変数
 t_k 1 サイクル辺りの各状態 $\{I(t) = k\}$ ($k = 0, 1, \dots, Q + r$) での滞在時間を表す確率変数
 P_k 定常状態において状態 $\{I(t) = k\}$ である確率

2.4. 各状態における定常状態確率について

本研究では、 P_k の定常分布を計算したい。

$$\begin{aligned}
 P_k &= \frac{\int_0^\infty E(t_k|L=l)f(l)dl}{\int_L E(\tau|L=l)f(l)dl} \\
 &= \frac{\int_0^\infty E(t_k|L=l)f(l)dl}{\sum_{k=0}^{Q+r} \int_L E(t_k|L=l)f(l)dl} \quad (1)
 \end{aligned}$$

上式より、各 k について $E(t_k|L=l)$ を計算すれば良い。次節では、そのための準備となる量の値の計算を行う。以下では、サイクルの開始時刻を $t=0$ として、1 サイクル $[0, \tau)$ に着目する。

2.5. 準備

この節では、以下の値の計算を行う。

- $M(n, q)$ q 回の需要で、 n 個の在庫を持っていく確率
 $Z(n, t)$ 時間 t の間に、 n 個の在庫を持っていく確率
 $W(=r - I(0))$ サイクルが始まったときの在庫量が r からどれ程下回っているかを表す確率変数
 $I(L-)$ リードタイムの終了直前の在庫量を表す確率変数

2.5.1. $M(n, q)$ について

「 q 回の需要で n 個の在庫が持って行かれる ($q = 1, 2, \dots, n$)」とき、 k 回目の需要が持っていった在庫の個数を x_k とすると、「 $x_1 + x_2 + \dots + x_q = n$ 」が成立する。 $x_1, x_2, \dots, x_q$ は 1 以上の整数なので、その組み合わせは ${}_{n-1}C_{q-1}$ 通りである。また、任意の $(x_1, x_2, \dots, x_q)$ の組に対して、

$$\begin{aligned}
 (\text{n 個持って行く確率}) &= [p(1-p)^{x_1-1}] \times [p(1-p)^{x_2-1}] \times \dots \times [p(1-p)^{x_q-1}] \\
 &= p^q (1-p)^{x_1+x_2+\dots+x_q-q} \\
 &= p^q (1-p)^{n-q} \quad (2)
 \end{aligned}$$

となるので、

$$M(n, q) = {}_{n-1}C_{q-1} p^q (1-p)^{n-q} \quad (q = 1, 2, \dots, n)$$

となり、 $(x_1, x_2, \dots, x_q)$ の組に依存しない。

2.5.2. $Z(n, t)$ について

前節で $M(n, q)$ を求めたので、時刻 t の間に n 個の在庫を持っていく確率 $Z(n, t)$ は、

$$Z(n, t) = \begin{cases} \sum_{q=1}^n \frac{(\lambda t)^q e^{-\lambda t}}{q!} M(n, q) & (n = 1, 2, \dots, \text{のとき}) \\ e^{-\lambda t} & (n = 0 \text{ のとき}) \end{cases}$$

2.5.3. W の計算

1 回の需要が持っていく個数は幾何分布にしたがうので、1 回の需要が要求する量を表す確率変数を Z とすると、

$$P(Z = k) = p(1-p)^{k-1}, \quad k = 1, 2, \dots \quad (3)$$

状態 $I(0-) = m$ のときに来た需要が $Z(\geq m - r)$ 個の在庫を持って行き、状態 $I(0) = r - w$ になったとすると、 $w = 0, 1, 2, \dots, r - 1$ のとき、

$$\begin{aligned}
 P(W = w) &= P(Z - (m - r) = w | Z \geq m - r) \\
 &= \frac{P(Z = w + m + r \wedge Z \geq m - r)}{P(Z \geq m - r)} \\
 &= \frac{p(1 - p)^{w+m-r-1}}{(1 - p)^{m-r-1}} \\
 &= p(1 - p)^w
 \end{aligned} \tag{4}$$

となり、 m に依存しない (幾何分布の無記憶性). W がとり得る値は $0, 1, \dots, r$ なので、

$$P(W = w) = \begin{cases} p(1 - p)^w & (w = 0, 1, \dots, r - 1 \text{ のとき}) \\ 1 - (1 - p)^r & (w = r \text{ のとき}) \end{cases}$$

2.5.4. $I(L-)$ の条件付き分布

サイクルの初期値 $I(0) = r - w$ のとき

$$P(I(L-) = j | L = l, W = w) = \begin{cases} Z(r - w - j, l) & (j = 1, 2, \dots, r - w - 1 \text{ のとき}) \\ 1 - \sum_{j=0}^{r-w-1} Z(j, l) & (j = 0 \text{ のとき}) \end{cases}$$

したがって、 $L = l$ が与えられたときの $I(L-)$ の条件付き確率は、

$$P(I(L-) = j | L = l) = \begin{cases} p \sum_{w=0}^{r-j} (1 - p)^w Z(r - w - j, l) & (j = 1, 2, \dots, r \text{ のとき}) \\ 1 - p [\sum_{j=1}^r \{ \sum_{w=0}^{r-j} (1 - p)^w Z(r - w - j, l) \}] & (j = 0 \text{ のとき}) \end{cases}$$

3. 定常分布の計算

$E[t_k | L = l]$ が $k = 0, 1, \dots, Q + r$ のすべての k で計算することができれば、式 (1) から求めたい定常分布を計算することができる. 再生サイクルを以下の 2 つに分けて定義する.

第 1 ピリオド：リードタイム開始から終了までの期間

第 2 ピリオド：リードタイムの終了から再び発注を行い、リードタイムが開始するまでの期間

そこで、 $E[t_k | L = l]$ を以下に場合分けして考える.

- $Q \leq k \leq Q + r$ のとき
- $r + 1 \leq k \leq Q - 1$ のとき
- $k = r$ のとき
- $1 \leq k \leq r - 1$ のとき
- $k = 0$ のとき.

ただし、 $Q = r + 1$ のときは k が $r + 1 \leq k \leq Q - 1$ をとることはなく $r = 1$ のときは k が $1 \leq k \leq r - 1$ をとることはない.

3.1. $Q \leq k \leq Q + r$ のとき

この状態は第 2 ピリオドでしか現れない. そこで、1. 「補給後に在庫量が k 個よりも多い状態となり、そこから状態 $\{I(t) = k\}$ に達する場合」と 2. 「補給後にちょうど、状態 $\{I(t) = k\}$ になる場合」の 2 パターンに場合分けして考える.

1. 補給後に在庫量が k 個よりも多い状態となり, そこから状態 $\{I(t) = k\}$ に達する場合.
 $I(L-) = j (j = k - Q + 1, k - Q + 2, \dots, r)$ のとき, 状態 $\{I(L) = j + Q\}$ から状態 $\{I(t) = k\}$ に達する確率は,

$$\begin{aligned} P(\{I(L) = j + Q\} \rightarrow \{I(t) = k\}) &= \sum_{q=1}^{j+Q-k} M(j + Q - k, q) \\ &= p \times \sum_{q=1}^{j+Q-k} j+Q-k-1 C_{q-1} p^{q-1} (1-p)^{j+Q-k} \\ &= p \end{aligned} \quad (5)$$

である.

2. 補給後にちょうど, 状態 $\{I(L) = k\}$ になる場合.
 $I(L-) = k - Q$ のときで, このとき状態 $\{I(t) = k\}$ には確率 1 で到達する.

$$P(\{I(L) = k\} \rightarrow \{I(t) = k\}) = 1 \quad (6)$$

したがって, 上の 1, 2 より求める平均滞在時間 $E(t_k | L = l)$ は,

$$E(t_k | L = l) = \frac{1}{\lambda} [p \times \{ \sum_{j=k-Q+1}^r P(I(L-) = j | L = l) \} + P(I(L-) = k - Q | L = l)] \quad (7)$$

ここでは, $\sum_{j=a}^b f(j) = 0$ ($b < a$ のとき) と定義している.

3.2. $r + 1 \leq k \leq Q - 1$ のとき

この状態は第 2 ピリオドしか現れない. k がこの値をとる際にはすべて, 「補給後に在庫量が k 個よりも多い状態となり, そこから状態 $\{I(t) = k\}$ に達する場合」なので, 以下のように計算する.

$I(L-) = j$ ($j = 0, 1, 2, \dots, r$) がのとき, 状態 $\{I(L) = j + Q\}$ から状態 $\{I(t) = k\}$ に達する確率は, (6) 式と同様に

$$P(\{I(L) = j + Q\} \rightarrow \{I(t) = k\}) = p$$

である. したがって, 求める平均滞在時間 $E(t_k | L = l)$ は,

$$\begin{aligned} E(t_k | L = l) &= \frac{1}{\lambda} \times \sum_{j=0}^r P(I(L-) = j | L = l) \times p \\ &= \frac{p}{\lambda} \times \sum_{j=0}^r P(I(L-) = j | L = l) \\ &= \frac{p}{\lambda} \end{aligned} \quad (8)$$

3.3. $k = r$ のとき

この状態は第 1 ピリオドで $W = 0$ のときのみ現れる. 最初に来る需要がリードタイムの終了よりも「後」か「先」かで分けて考えて計算する.

1. リードタイムの終了よりも後に最初の需要が到着する場合, このときの平均滞在時間は,

$$\begin{aligned} E[t_k 1_{\{\text{需要がリードタイムより後}\}} | L = l] &= \int_l^\infty l \times ((e^{-\lambda y})(\lambda dy)) \\ &= l e^{-\lambda l} \end{aligned} \quad (9)$$

である.

2. リードタイムの終了よりも先に最初の需要が到着する場合このときの平均滞在時間は,

$$\begin{aligned} E[t_k 1_{\{\text{需要がリードタイムより先}\}} | L = l] &= \int_0^l y \times ((e^{-\lambda y})(\lambda dy)) \\ &= \frac{1 - e^{-\lambda l}}{\lambda} - l e^{-\lambda l} \end{aligned} \quad (10)$$

である.

したがって, 求める平均滞在時間 $E(t_k | L = l)$ は,

$$\begin{aligned} E(t_k | L = l) &= P(W = 0) \times [E[t_k 1_{\{\text{需要がリードタイムより後}\}} | L = l] + E[t_k 1_{\{\text{需要がリードタイムより先}\}} | L = l]] \\ &= \frac{p(1 - e^{-\lambda l})}{\lambda} \end{aligned} \quad (11)$$

3.4. $1 \leq k \leq r - 1$ のとき

まずは, $W = r - k$ のときとそれ以外で場合分けする. そのうえで, 「時間 $[0, t)$ で $(r - w - k - d)$ 個」かつ「(微小) 時間 $[t, t + dt)$ で d 個」の在庫が持って行かれて $\{I(t) = k\}$ の状態に達する確率を考えている. 時刻 t 以降は, 前節の結果を用いている.

1. $W = r - k$ のとき, このときの平均滞在時間は前節の結果を用いると,

$$E[t_k 1_{\{W=r-k\}} | L = l] = \frac{1 - e^{-\lambda l}}{\lambda}$$

である.

2. $W = w (w = 0, 1, \dots, r - k - 1)$ のとき, このときの平均滞在時間は,
 $d = 1, 2, \dots, r - w - k - 1, r - w - k$ のとき

$$E[t_k 1_{\{W < r-k\}} | L = l] = \sum_{d=1}^{r-w-k} \int_0^l Z(r - w - k - d, t) \times p(1 - p)^{d-1} \times \frac{1 - e^{-\lambda(l-t)}}{\lambda} \times (\lambda dt) \quad (12)$$

したがって, 求める平均滞在時間 $E(t_k | L = l)$ は,

$$\begin{aligned} E(t_k | L = l) &= P(W = r - k) \times \frac{1 - e^{-\lambda l}}{\lambda} + \sum_{w=0}^{r-k-1} P(W = w) \\ &\quad \times \left\{ \int_0^l \left(\sum_{d=1}^{r-w-k-1} \left(\sum_{q=1}^{r-w-k-d} \left[\frac{(\lambda t)^q e^{-\lambda t}}{q!} M(r - w - k - d, q) \times p(1 - p)^{d-1} \right] \right) \right. \right. \\ &\quad \left. \left. + e^{-\lambda t} (1 - p)^{r-w-k-1} p \right) \times [1 - e^{-\lambda(l-t)}] dt \right\} \\ &= p(1 - p)^{r-k} \frac{1 - e^{-\lambda l}}{\lambda} + \sum_{w=0}^{r-k-1} p(1 - p)^w \\ &\quad \times \int_0^l \left(\left[\sum_{q=0}^{r-w-k-1} \frac{\lambda t}{q!} M(r - w - k, q) \right] \times [e^{-\lambda t} - e^{-\lambda l}] dt \right) \end{aligned} \quad (13)$$

ここでは, d と q の Σ 計算の順序の入れ替えと,

$$\sum_{d=1}^{n-q} n-d-1 C_{q-1} = n-1 C_q$$

および,

$$M(r - w - k - 1, q) \frac{r - w - k - 1}{q} p = M(r - w - k, q + 1)$$

を用いて変形を行っている.

3.5. $k = 0$ のとき

まずは, $W = r$ のときとそれ以外で場合分けする. そのうえで, 「時間 $[0, t)$ で $(r - w - d)$ 個」あるいは「(微小) 時間 $[t, t + dt)$ で d 個以上」の在庫が持って行かれて $\{I(t) = k\}$ の状態に達する確率を考えている. 時刻 t 以降は, リードタイムまでの時間 $(l - t)$ だけ $\{I(t) = k\}$ に滞在することになる.

1. $W = r$ のとき, このときの平均滞在時間は, リードタイムの l である.

2. $W = w$ ($0, 1, \dots, r - 1$) のとき, このときの平均滞在時間は,

$$E[t_k 1_{\{W < r\}} | L = l] = \sum_{d=1}^{r-w-d} \int_0^l [Z(r-w-d, t) \times (1-p)^{d-1} \times (l-t)(\lambda dt)] \quad (14)$$

したがって,

$$\begin{aligned} E[t_k | L = l] &= P(W = r) \times l + \left[\sum_{w=0}^{r-1} P(W = w) \right. \\ &\quad \times \left\{ \int_0^l \left[\sum_{d=1}^{r-w-1} \left[\sum_{q=1}^{r-w-d} \frac{(\lambda t)^q e^{-\lambda t}}{q!} M(r-w-d, q) \times (1-p)^{d-1} \right] \right. \right. \\ &\quad \left. \left. + e^{-\lambda t} (1-p)^{r-w-1} \right] \times \lambda \times (l-t) dt \right\} \\ &= [1 - (1-p)^r] + \\ &\quad \left. \sum_{w=0}^{r-1} p(1-p)^w \times \int_0^l \left[\left\{ \frac{1}{p} \times \left\{ \sum_{q=1}^{r-w-1} \left(\frac{(\lambda t)^q e^{-\lambda t}}{q!} M(r-w, q+1) \right) \right\} + (1-p)^{r-w-1} \right\} \times (l-t)(\lambda dt) \right] \right] \right\} \quad (15) \end{aligned}$$

ここで, d と q の Σ 計算の順序の入れ替えと,

$$\sum_{d=1}^{r-w-q} r-w-d-1 C_{q-1} = r-w-1 C_q$$

および

$$M(r-w-1, q) \times \frac{r-w-1}{q} = M(r-w, q+1) \times \frac{1}{p}$$

を用いて変形を行っている.

4. 数値計算例

4.1. コストの定式化

HC : 在庫を 1 個保有するのにかかるコスト

VC : 1 回当たりの補給をするのにかかるコスト

LC : 在庫量が 0 個の状態のときに, 1 つ当たりの需要が失われるコスト

とくと, 各コストは以下のように表される.

保有コスト: 在庫を保有するのにかかるコストは, $HC \times \{\sum_{k=0}^{Q+r} k \times (P_k)\}$

発注コスト: 補給にかかるコストは, $VC \times \frac{1}{E[\tau]}$

lost-sales コスト: 在庫が 0 個の状態のときに来た需要が失われるコストは, $LC \times \lambda \{P_0(0)\}$

以上より, 定常分布と平均不足量から, システム全体にかかるコスト $TC(r, Q)$ を以下のように定義出来る [1].

$$\begin{aligned} TC(r, Q) &= HC \times \left\{ \sum_{k=0}^{Q+r} k \times (P_k) \right\} + VC \times \frac{1}{E[\tau]} + LC \times \lambda \{P_0\} \quad (16) \end{aligned}$$

前章までの結果を用いて, L が様々な分布にしたがうとき ($f(l)$ が変化したとき) に, この $TC(r, Q)$ を様々な (r, Q) で計算する.

4.2. 数値計算例

本研究の数値計算では, $p = 1/2, \lambda = 20, HC = 1, VC = 200, LC = 50, L$ を平均 $\frac{5}{2}$ の確率分布にしたがう確率変数として計算を行い, その結果については発表日に詳しく説明する.

5. まとめ

需要が(幾何分布にしたがう)複数個の在庫を持っていく場合について, リードタイムが一般の分布にしたがう (r, Q) 方式在庫管理システムを再生型確率過程として捉え, その各状態の条件付平均滞在時間を計算することで, 各状態の定常分布を導出した. そこから, システム全体にかかるコストを定式化し, それをもとに数値計算を行った.

今後の課題としては, システムパラメータを様々な値にしたときに総コストの変化のを考察していきたい. また, 本研究では在庫切れが起きた際に, lost-sales 方式で考えていた. 今後の課題としてはそのときの backorder を考慮したモデルを提案し, その安定性等についても議論していきたい.

参考文献

- [1] M. Saffari, S. Asmussen, R. Haji, The M/M/1 queue with inventory, lost sale, and general lead times. *Queueing Systems*. Vol.75 (2013), 65–77.
- [2] I. Sahin, Regenerative Inventory Systems. Springer. NewYork (1990).
- [3] S. K. Srinivasan, Analysis of s,S inventory systems with general lead time and demand distribution and adjustable reorder size. *Mathematical Programming and Operations Research*. Vol.19 (1988), 557–576.
- [4] E. Mohebbi, D. Hao, When supplier's availability affects the replenishment lead time-An extension of the supply-interruption problem. *Eur. J. Oper. Res.* Vol. 175 (2006) 992-1008.
- [5] A. Marand, H. Li, A. Thorstenson, Joint inventory control and pricing in a service-inventory system. *International Journal of Production Economics*. Vol. 10 (2017), 1-14.
- [6] J.Park,J.Baek,Y.Kim,simulation study on the lost sales inventory systems with attached service queue. *International Journal of Industrial Engineering*. Vol. 24 (2017), 171-181.
- [7] E. Porteus, Foundations of Stochastic Inventory Theory. Stanford University Press (2002).

GI/M^b/1 モデルによるビットコイン・トランザクション承認時間解析

河瀬 良亮 笠原 正治
奈良先端科学技術大学院大学 情報科学研究科

概要：ビットコインは暗号化技術と P2P ネットワーク技術により構築された仮想通貨である。ビットコインではトランザクションに送金の情報が含まれており、そのためトランザクションの承認時間はユーザがビットコインを使用する上で重要な指標となる。先行研究では、トランザクションの承認プロセスを M/G^b/1 待ち行列でモデル化し、平均承認時間について解析を行っている。また、サービス時間に相当するブロック生成間隔が指数分布で近似できることも合わせて報告されている。しかしながら、実際のビットコインにおけるトランザクションの到着間隔は指数分布よりも変動が大きいたことが知られている。そこで本研究では、トランザクションの到着は独立同一一般分布に従って到着し、マイニングの終了時点で複数のトランザクションが離脱する GI/M^b/1 モデルを考える。トランザクション到着直前の待ち行列内トランザクション数をシステムの状態として定義し、行列解析法により定常分布および平均トランザクション滞在時間を導出する。数値例において、実データに基づいてパラメータを決定した理論解析結果とシミュレーションを比較して妥当性を検証する。

1. はじめに

ビットコインは暗号化技術と P2P ネットワーク技術により構築された自律分散型の仮想通貨である [1]。ビットコインの取引件数は日々増加しており、2016 年の一日当たりの平均取引件数は 226,000 件で、その前年の 125,000 件の約 2 倍にまで増加している [5]。日本国内においては家電量販店で使用できるようになるなど普及し始めている [3]。

ユーザがビットコインを用いて送金するとき、まず送金者、受信者、送金金額などの情報が含められたトランザクションが発行される。そしてマイナーと呼ばれるノードがトランザクションを複数個集めてブロックに格納し、パズル的問題を解くことでそのトランザクションに記載された取引が有効になる [2]。この取引の承認過程はマイニングと呼ばれる。

ビットコインでは最大ブロックサイズが 1 Mbyte に制限されており、またマイニングが平均 10 分に 1 回行われようように調節されているため、トランザクションの処理数には上限がある。トランザクションのデータから電子署名を切り離すことで、トランザクションのデータ量を削減する segwit [6] の導入によりブロックに格納できるトランザクション数は増加したが依然としてスケーラビリティの問題は残っている。

また、トランザクションは送金金額、前回の承認からの経過時間、データサイズによって優先権の値が決められている [2]。マイナーはトランザクションの優先権の値と送金手数料を基にそのトランザクションをブロックに含めるかどうか決定する。そのため、低優先権のトランザクションは承認に時間がかかる傾向がある。文献 [9] では、トランザクションを低優先権と高優先権の 2 種類に分けてトランザクション承認時間を解析し、低優先権のトランザクションはマイナーがブロックに意図的に含めていない可能性を指摘している。このようなことからマイクロペイメントのような低優先権になりやすいトランザクションや高負荷時のトランザクションの処理には問題がある。

ビットコインのトランザクション承認プロセスは、集団サービスによる待ち行列を用いてモデル化できることが文献 [9] で報告されている。集団サービスを持つ待ち行列の解析は、文献 [7, 8] において M/G^b/1 に対し、経過サービス時間と待ち行列内客数の結合分布関数について解析が行われている。文献 [9] では、それをビットコインのトランザクション承認時間の解析に応用し、さらに優先権付き M/G^b/1 待ち行列を考え、高優先と低優先の二種類のトランザクションに対する平均承認時間の解析が行われている。

また文献 [9] において、極値統計理論を用いてマイニングにかかる時間 (ブロック生成時間) が指数分布に従っていること、トランザクションが到着してから承認されるまでの時間がサービス時間の約二倍近くかかっていること、が報告されている。また新たに到着したトランザクションは現在マイニング中のブロックには含められず、早くてもその次のブロックに含められることを指摘している。

我々の過去の研究 [10] では、集団サービスが行われているときに到着したトランザクションはそのサービスに含められず、待ち行列で待機させられる M/G^b/1 待ち行列モデルを検討した。具体的には、

サービス中のトランザクション数と待ち行列内のトランザクション数の結合確率母関数を考え、平均系内トランザクション数からリトルの公式より平均トランザクション承認時間を導出した。しかし、トランザクションの到着時間間隔分布の変動係数は指数分布よりも大きいため、数値例において解析とシミュレーションの結果で乖離が見られた。

文献 [11] では文献 [10] のモデルを発展させ、トランザクション承認処理に優先権を考慮したモデルについて解析を行った。数値例において、トランザクションを送金金額に基づいて高優先権と低優先権の2種類に分類し、解析結果と実データのトランザクション承認時間を比較したところ、低優先権トランザクションの承認時間が理論値よりも大きい値を取ることが判明した。この結果は、マイナーは低優先権のトランザクションをブロックに含めない傾向にあることを示唆している。

本研究では、我々の過去の研究 [10] で判明したトランザクション到着時間間隔の変動を考慮するため、トランザクションの到着は独立同一な一般分布に従い、マイニング終了時点で複数のトランザクションが離脱する $GI/M^b/1$ モデルを考える。トランザクション到着直前の待ち行列内トランザクション数をシステムの状態として定義し、行列解析法により定常分布を導出する。それにより、平均トランザクション滞在時間を求めることができる。数値例において、EM アルゴリズムを用いてトランザクションの到着時間間隔分布を推定し、解析結果とシミュレーションを比較することで解析モデルの妥当性を検証する。

本論文の構成は以下の通りである。2章ではトランザクション処理の待ち行列モデルを示す。3章では、トランザクションの到着直前の待ち行列内トランザクション数をシステムの状態として定義し、行列解析法により定常分布を導出する。4章で数値例を示し、5章で結論を述べる。

2. $GI/M^b/1$ 待ち行列モデル

本稿ではビットコインシステムを $GI/M^b/1$ 待ち行列としてモデル化する。

トランザクションの到着時間間隔 A は独立同一な一般分布 $G(x)$ に従う。平均到着時間間隔は次式で表されるものとする。

$$E[A] = \int_0^{\infty} x dG(x) \equiv \frac{1}{\lambda}.$$

システムに到着したトランザクションは複数個まとめられて一つのブロックに収容される。マイニングによってブロックが生成されるため、以下ではブロック生成間隔をサービス時間と呼ぶ。文献 [9] より、ブロック生成間隔が独立同一な指数分布に従うことが報告されている。ここではブロック生成間隔 S が率 μ の指数分布に従うことを仮定し、

$$F(x) = \Pr\{S \leq x\} = 1 - e^{-\mu x},$$

とおく。サーバは複数のトランザクションを集団サービスし、その最大サイズを b とする。

ビットコインでは到着したトランザクションは現在マイニング中のブロックには含まれず、早くても次のブロックに含まれる傾向にあることが文献 [9] で報告されている。ここでは、トランザクションが到着した時点でシステムが稼働中のとき、そのトランザクションは待ち行列で待機し、次以降のブロックに到着順で含められるものとする。システム内にトランザクションが存在しない場合でも、マイニングによるサービスは継続的に行われる。そのため、新たに到着したトランザクションは次のサービス開始時点まで待機させられるものと仮定する。

3. 解析

X_n ($n = 1, 2, \dots$) を n 番目のトランザクション到着直前の系内トランザクション数、 Y_n を n 番目と $n+1$ 番目のトランザクション到着時点の間に完了するマイニング回数とする。また、 $D_k^{(n)}$ を n 番目のトランザクション到着時点から数えて k 回目のマイニング終了時点で系から離脱するトランザクション数とする。最大の集団離脱サイズが b より、 $Y_n > 1$ のときは $D_k^{(n)} = b$ ($k = 1, \dots, Y_n - 1$) となることに注意すると、 X_n は次式で表現される。

$$X_{n+1} = \max \{X_n - (Y_n - 1)b - D_{Y_n}^{(n)} + 1, 0\}, \quad n = 1, 2, \dots \quad (1)$$

マイニングは系内トランザクション数によらず行われること、およびマイニング間隔が独立同一な指数分布であることから、 $\{Y_n : n = 1, 2, \dots\}$ は独立同一なポアソン分布に従う。よって、(1) 式を満足する $\{X_n : n = 1, 2, \dots\}$ は以下の推移確率をもつ離散時間マルコフ連鎖となる。

$$p_{i,j} = P(X_{n+1} = j | X_n = i) = \begin{cases} a_{(j-i-1)/b}, & (j-i-1)/b \text{ が自然数のとき,} \\ \bar{a}_{[i/b]+1}, & j = 0, \\ 0, & \text{それ以外,} \end{cases}$$

ただし

$$a_k = \int_0^\infty e^{-\mu x} \frac{(\mu x)^k}{k!} dG(x), \quad (2)$$

$$\bar{a}_k = \sum_{i=k}^\infty a_i,$$

であり、 $[x]$ は x 以下の最大の整数を表す。

推移確率行列 $P = (p_{i,j})$ は以下のように書ける。

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & \cdots & b & b+1 & b+2 & \cdots \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ \vdots \\ b-1 \\ b \\ b+1 \\ b+2 \\ \vdots \end{matrix} & \begin{pmatrix} \bar{a}_1 & a_0 & 0 & 0 & \cdots & & & & \\ \bar{a}_1 & 0 & a_0 & 0 & \ddots & \ddots & & & O \\ \bar{a}_1 & 0 & 0 & a_0 & \ddots & \ddots & \ddots & & \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ \bar{a}_1 & 0 & 0 & 0 & \ddots & a_0 & 0 & 0 & \ddots \\ \bar{a}_2 & a_1 & 0 & 0 & \ddots & 0 & a_0 & 0 & \ddots \\ \bar{a}_2 & 0 & a_1 & 0 & \ddots & 0 & 0 & a_0 & \ddots \\ \bar{a}_2 & 0 & 0 & a_1 & \ddots & 0 & 0 & 0 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \end{matrix},$$

$\{X_n : n = 1, 2, \dots\}$ は既約かつ非周期的であり、 $\lambda/\mu < b$ のときに系は安定で定常分布

$$\pi_k = \lim_{n \rightarrow \infty} \Pr\{X_n = k\}, \quad k = 0, 1, \dots,$$

が存在する。 $\pi = (\pi_k)$ なる確率ベクトルを定義すると、 π は次式を満足する。

$$\pi = \pi P, \quad \pi \mathbf{1} = 1.$$

ここで $\mathbf{1} = (1, 1, \dots)^\top$ ($\top$ は転置) である。

P は次のようなブロック行列表現に書き直すことができる。

$$P = \begin{pmatrix} \mathbf{B}_0 & \mathbf{C}_0 & \mathbf{O} & \mathbf{O} & \cdots \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{A}_0 & \mathbf{O} & \cdots \\ \mathbf{B}_2 & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \cdots \\ \mathbf{B}_3 & \mathbf{A}_3 & \mathbf{A}_2 & \mathbf{A}_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

ここで $\mathbf{C}_0$ は $1 \times b$, $\mathbf{B}_0$ は 1×1 , $\mathbf{B}_k$ ($k = 1, 2, \dots$) は $b \times 1$, $\mathbf{A}_l$ ($l = 0, 1, \dots$) は $b \times b$ の行列であり、次式で表現される。

$$\mathbf{C}_0 = (a_0, 0, \dots, 0), \quad \mathbf{B}_0 = (\bar{a}_1), \quad \mathbf{B}_k = (\bar{a}_k, \dots, \bar{a}_k, \bar{a}_{k+1}),$$

$$A_l = \begin{pmatrix} 0 & a_l & & & \\ 0 & 0 & a_l & & O \\ 0 & 0 & 0 & & \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & & 0 & a_l \\ a_{l+1} & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

また

$$\pi_0 = (\pi_0), \quad \pi_n = (\pi_{(n-1)b+1}, \pi_{(n-1)b+2}, \dots, \pi_{nb}), \quad n = 1, 2, \dots,$$

とおくことで、定常分布ベクトル π は $\pi = (\pi_0, \pi_1, \pi_2, \dots)$ と表せる. GI/M/1 型マルコフ連鎖の場合は推移確率行列の構造から,

$$\pi_{n+1} = \pi_1 R^n, \quad n = 0, 1, \dots,$$

が成り立つ [13]. R については、初期値 $R(0)$ を零行列として、 $R(n)$ を

$$R(n) = \left(A_0 + \sum_{k=2}^{\infty} [R(n-1)]^k A_k \right) (I - A_1)^{-1},$$

により反復計算することで、 R を求めることができる.

π_0 と π_1 は次式より求めることができる.

$$\pi_0 = \pi_0 B_0 + \pi_1 \left(\sum_{l=1}^{\infty} R^{l-1} B_l \right),$$

$$\pi_1 = \pi_0 C_0 + \pi_1 \left(\sum_{l=1}^{\infty} R^{l-1} A_l \right),$$

$$\pi_0 \mathbf{1}^\top + \pi_1 (I - R)^{-1} \mathbf{1}^\top = 1.$$

トランザクションの平均待ち時間 $E[W]$ と平均系内滞在時間 $E[T]$ は次のように書ける.

$$E[W] = \sum_{k=0}^{\infty} \pi_k \left(\left\lfloor \frac{k}{b} \right\rfloor + 1 \right) E[S],$$

$$E[T] = E[W] + E[S].$$

以下では $E[T]$ を平均トランザクション承認時間と呼ぶことにする.

4. 数値例

4.1. トランザクション到着時間間隔分布の推定

トランザクションの到着時間間隔の変動係数は指数分布の場合よりも大きいため、超指数分布に適合すると仮定し、文献 [12] に基づいて、トランザクション到着時間間隔分布の推定を行う. 超指数分布は滞在時間が指数分布にしたがう相を並列につないだもので、相の数を m として、

$$G(x) = 1 - \sum_{i=1}^m c_i e^{-\lambda_i x}, \quad x \geq 0,$$

と書ける. ただし $\sum_{i=1}^m c_i = 1$ である. その確率密度関数 $g(x)$ は

$$g(x) = \sum_{i=1}^m c_i \lambda_i e^{-\lambda_i x}, \quad x \geq 0,$$

で与えられる.

トランザクション到着時間間隔の元データとして、[5] の web サイトより 2013 年 10 月 1 日から

表 1: EM アルゴリズムによるトランザクション到着時間間隔分布の推定 (2013/10/1-2014/9/30)

i	1	2	3	4	5
c_i	0.446648	0.327843	0.126944	0.098488	7.41350×10^{-5}
λ_i	0.866223	0.866223	0.866223	0.331555	0.00293430

表 2: EM アルゴリズムによるトランザクション到着時間間隔分布の推定 (2014/10/1-2015/9/30)

i	1	2	3	4	5
c_i	0.566151	0.259492	0.161661	0.012449	2.45808×10^{-4}
λ_i	1.119559	1085.472	1.119559	0.326394	0.001765193

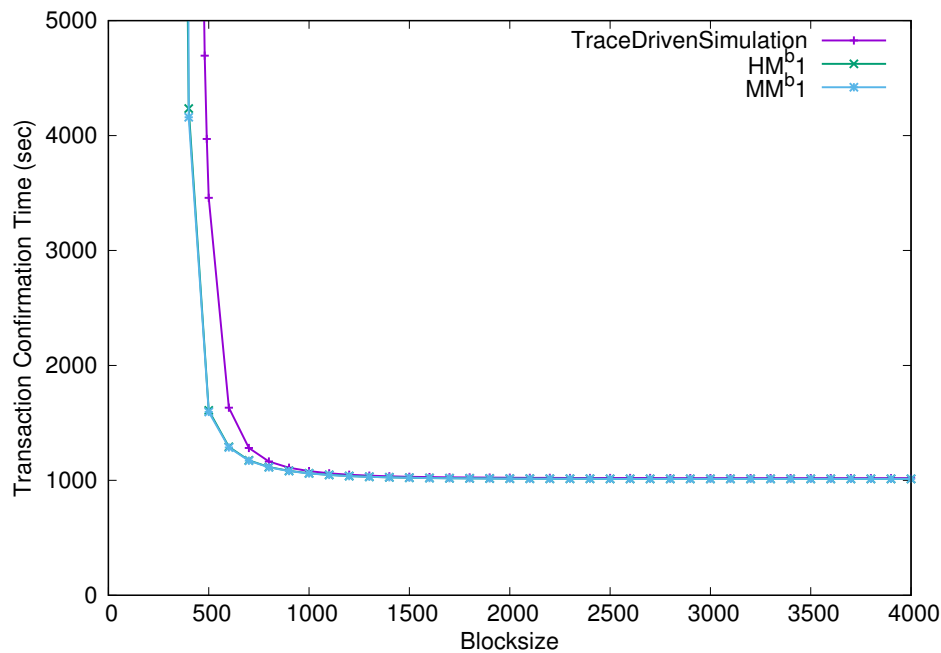


図 1: トレース駆動型シミュレーションと解析モデルの比較 (2013/10/1-2014/9/30)

2015 年 9 月 30 日の 2 年間のブロックに含まれるトランザクションの到着時刻を抽出し, 2013 年 10 月 1 日から 2014 年 9 月 30 日の 1 年間と, 2014 年 10 月 1 日から 2015 年 9 月 30 日までの 1 年間に對してそれぞれ [12] の EM アルゴリズムを適用した. EM アルゴリズムを適用するにあたり, 相の数を $m = 5$, 収束基準を $\epsilon = 10^{-10}$ とした. 表 1 は 2013/10/1 から 2014/9/30 の 1 年間のデータを用いて EM アルゴリズムを適用した結果であり, 表 2 は 2014/10/1 から 2015/9/30 の 1 年間のデータを用いた結果である.

表 3 は実データにおけるトランザクション到着時間間隔に関する統計量を示しており, 表 4 では EM アルゴリズムにより推定したトランザクション到着時間間隔の統計量を示している. 表 3 と表 4 により, EM アルゴリズムにより推定された超指数分布の平均は実データのものと一致し, 標準偏差と変動係数は概ね一致していることがわかる. なお, 表 3 で 1 年目の統計量と 2 年目の統計量を比べると, 2 年目の方が平均到着間隔が小さく, かつ変動も大きいことが観察される. これは 2 年目のトランザクション到着過程の方が 1 年目のものよりも変動が大きいことを示している.

4.2. 実データを用いたトレース駆動型シミュレーションと解析の比較

本節では, [5] のデータよりトランザクションの到着とブロック生成をトレースしたシミュレーションと解析結果を比較することで, 解析モデルの妥当性を検証する.

図 1 は, ブロックサイズを変化させたときの平均トランザクション承認時間について, 解析とシ

表 3: トランザクション到着時間間隔の実データ統計量

期間	平均	標準偏差	変動係数
2013/10/1-2014/09/30	1.36296	5.07571	3.72401
2014/10/1-2015/09/30	0.82772	12.68489	15.32505

表 4: 超指数分布で推定したトランザクション到着時間間隔の統計量

期間	平均	標準偏差	変動係数
2013/10/1-2014/09/30	1.36296	4.42236	3.24465
2014/10/1-2015/09/30	0.82772	12.58913	15.20936

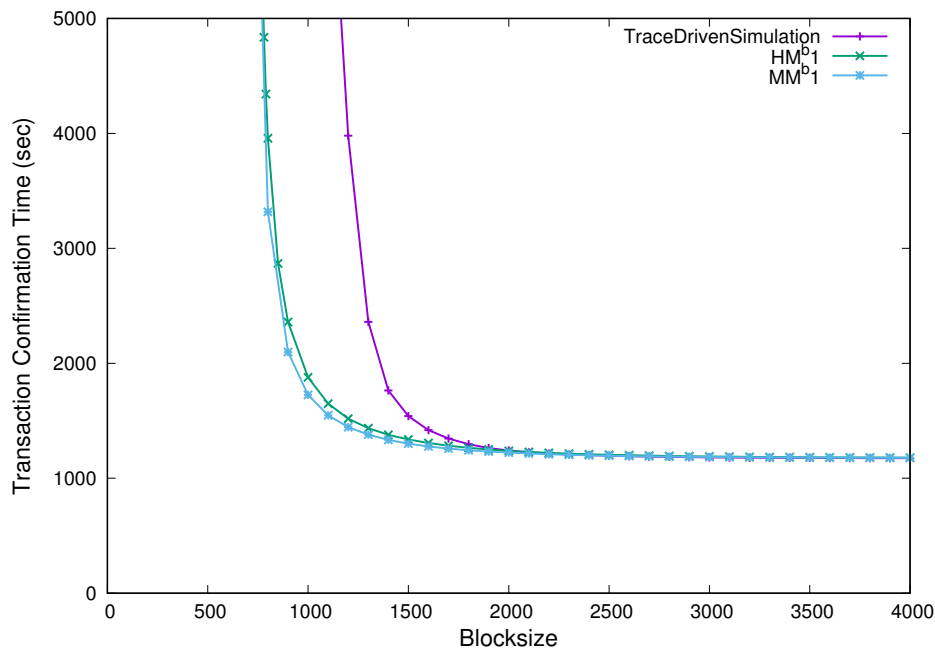


図 2: トレース駆動型シミュレーションと解析モデルの比較 (2014/10/1 から 2015/9/30)

シミュレーションの結果を比較したものである。到着過程の超指数分布のパラメータは 2013/10/1 から 2014/9/30 の 1 年間に対して推定された表 1 の値を用い、指数分布のパラメータはその期間の平均ブロック生成間隔より算出した。また、参考のために平均到着率と平均サービス率を対象期間のデータより算出した $M/M^b/1$ も図に掲載している。

図 1 より、 b が小さいところではシミュレーション結果と二つの解析モデルに差異が観察されるが、 $b = 700$ を超えた辺りからすべての結果が同じ値を取っており、解析モデルはよい近似を与えていることがわかる。また、この図において、 $GI/M^b/1$ と $M/M^b/1$ の結果がほとんど同じであることも観察される。表 1 において、2013/10/1 から 2014/9/30 の期間では到着率 $\lambda = 0.866223$ のトランザクションが約 9 割を占めていることに注意すると、この期間のトランザクション到着過程は単一の指数分布で近似しやすいことが原因と考えられる。

図 2 は 2014/10/1 から 2015/9/30 までのデータを用いたシミュレーションと解析の結果を示したものである。この図より、 $b = 1500$ を超えた辺りまで二つの解析モデルとシミュレーション結果との間に大きな差があることが観察される。図 1 と比較すると、解析結果とシミュレーション結果との差異が図 2 の方が大きくなっている。これは、2014/10/1 から 2015/9/30 の期間のトランザクション到着過程の方がその前の期間のものよりも変動が大きいためである。

また、図 2 より、ブロックサイズが小さい領域では、 $M/M^b/1$ よりも $H/M^b/1$ の方が平均トランザク

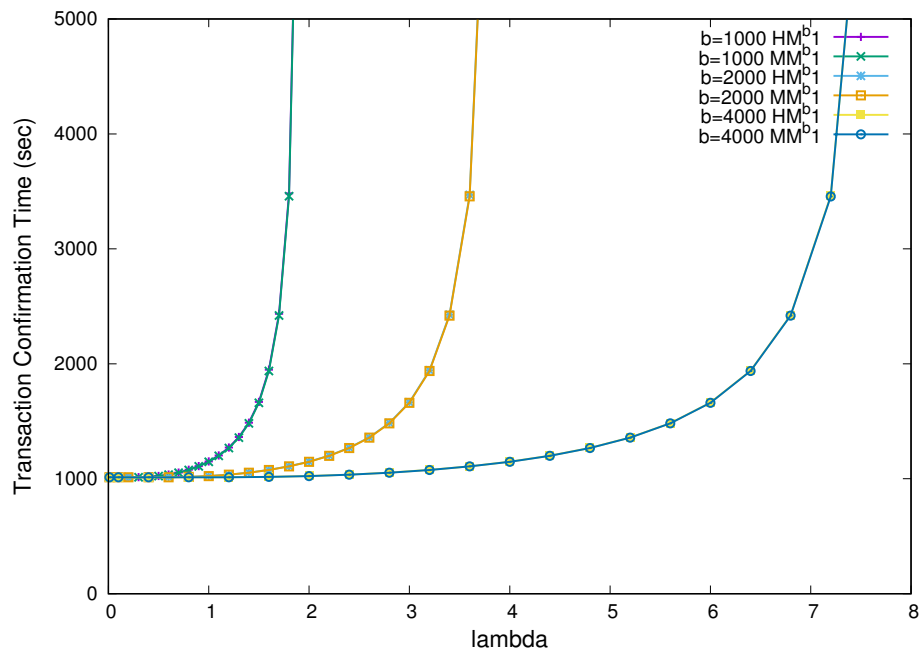


図 3: ブロックサイズと到着率のトランザクション承認時間への影響 (2013/10/1-2014/9/30)

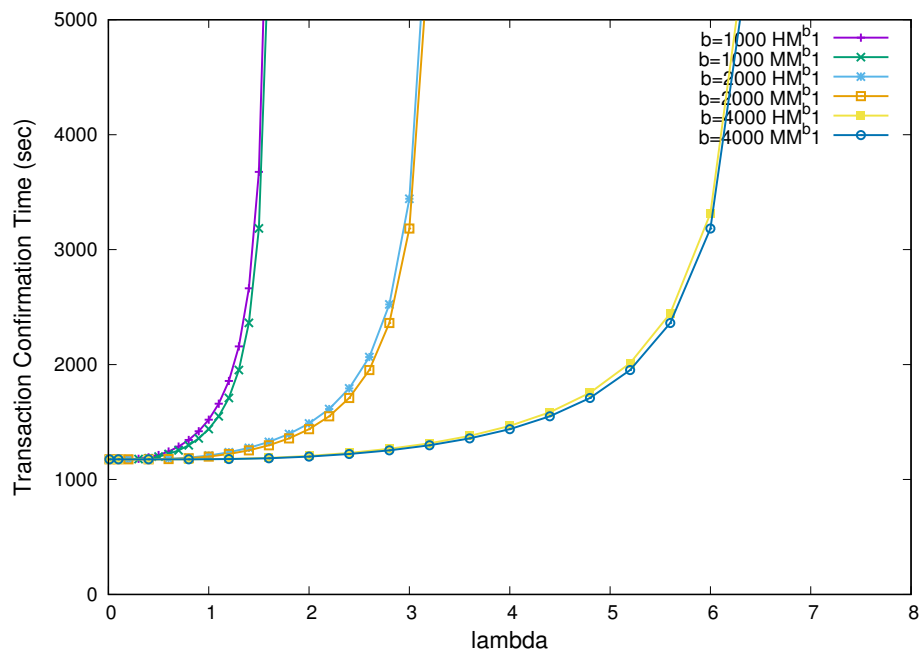


図 4: ブロックサイズと到着率のトランザクション承認時間への影響 (2014/10/1-2015/9/30)

ション承認時間が多少大きくなっているが、ほとんど同一の結果となっていることが観察される。表 2 より、EM アルゴリズムで推定された超指数分布は平均、標準偏差、変動係数を高い制度で近似できているが、平均トランザクション承認時間はポアソン到着過程とほとんど変わらない結果となっており、独立同一な超指数分布の仮定では精度を向上させることが困難であることを示唆している。

4.3. ブロックサイズと到着率によるトランザクション承認時間への影響

本節では、平均到着率の変化が平均トランザクション承認時間に与える影響について検討を行う。トランザクション全体の到着率 λ については

$$\lambda = \frac{\alpha}{\sum_{i=1}^m c_i / \lambda_i}, \quad (3)$$

とし、 α を増加させる形で λ を変化させ、平均トランザクション承認時間を計算する。

図 3 は平均到着率を増大させたときの平均トランザクション承認時間を表している。ここでは 2013/10/1 から 2014/09/30 における実データを基に平均ブロック生成率を $\mu = 0.0019749$ とし、到着率については表 1 の結果を (3) 式に用いた。ブロックサイズについては $b = 1000, 2000, 4000$ の三通りを考え、それぞれに対して平均トランザクション承認時間を計算した。また参考のために、 $M/M^b/1$ の結果も同図に記した。

図 3 より、平均到着率が増加するにつれて平均トランザクション承認時間が増大する様子が観察される。また、 $b = 1000$ の平均トランザクション承認時間が最も大きく、 $b = 4000$ が他の二つの b の場合と比べて緩やかに増大している。文献 [9] において、1 ブロック当たりの平均トランザクション数は 1750 であることが報告されており、 $b = 2000$ がそれに近い値となっていることに注意する。このとき、 $b = 4000$ と 2 倍のブロックサイズにした場合でも到着率が $\lambda = 7$ を超えたあたりから平均トランザクション承認時間が急激に増加しており、ブロックサイズを増加させる方策がトランザクション処理をそれほど改善しないことがわかる。

図 4 は 2014/10/1 から 2015/9/30 までの実データに基づいてパラメータを決定したときの平均トランザクション承認時間の変化を示している。図 3 と同様に、2 番目の期間のブロック生成時間の統計より $\mu = 0.0017009$ と設定し、平均到着率 λ の計算には表 2 の結果を (3) 式に用いている。図 4 において、各ブロックサイズにおける平均トランザクション承認時間は、図 3 と同様の増加傾向を示している。図 4 を図 3 と比較すると、図 4 の平均トランザクション承認時間の方がいずれのブロックサイズの場合でも早く増大することが観察される。これは、2 番目の期間のトランザクション到着過程の変動が大きいこと、及び平均ブロック生成率が期間 1 よりも小さくなっていることが原因として考えられる。

図 3、図 4 とともに、 $M/M^b/1$ の結果と $H/M^b/1$ の結果がほとんど一致していることが観察される。これは前節の結果と同じ傾向であり、独立同一な超指数分布による $GI/M^b/1$ モデルでは平均トランザクション承認時間の精度を向上させることが困難であることを示している。

4.4. ブロックサイズと変動係数によるトランザクション承認時間への影響

本節では、トランザクションの到着時間間隔の変動係数がトランザクション承認時間に与える影響を評価する。ここでは、2 相の超指数分布で以下の確率密度関数 $g(x)$ を持つものを考える。

$$g(x) = (1 - p)\lambda_1 e^{-\lambda_1 x} + p\lambda_2 e^{-\lambda_2 x}.$$

到着時間間隔 A の変動係数を $\alpha = \sqrt{V[A]}/E[A]$ で表現すると、次式を得る。

$$E[A^2] = (1 + \alpha^2)(E[A])^2 \quad (4)$$

また、 A の 1 次と 2 次のモーメントは次式で与えられる。

$$E[A] = \frac{1-p}{\lambda_1} + \frac{p}{\lambda_2} = \frac{1}{\lambda}, \quad (5)$$

$$E[A^2] = \frac{2(1-p)}{\lambda_1^2} + \frac{2p}{\lambda_2^2}. \quad (6)$$

ここで (5) は平均到着間隔 $1/\lambda$ と等しいことに注意する。(5) と (6) を式 (4) に代入すると次式を得る。

$$\frac{2(1-p)}{\lambda_1^2} + \frac{2p}{\lambda_2^2} = \frac{1 + \alpha^2}{\lambda^2}. \quad (7)$$

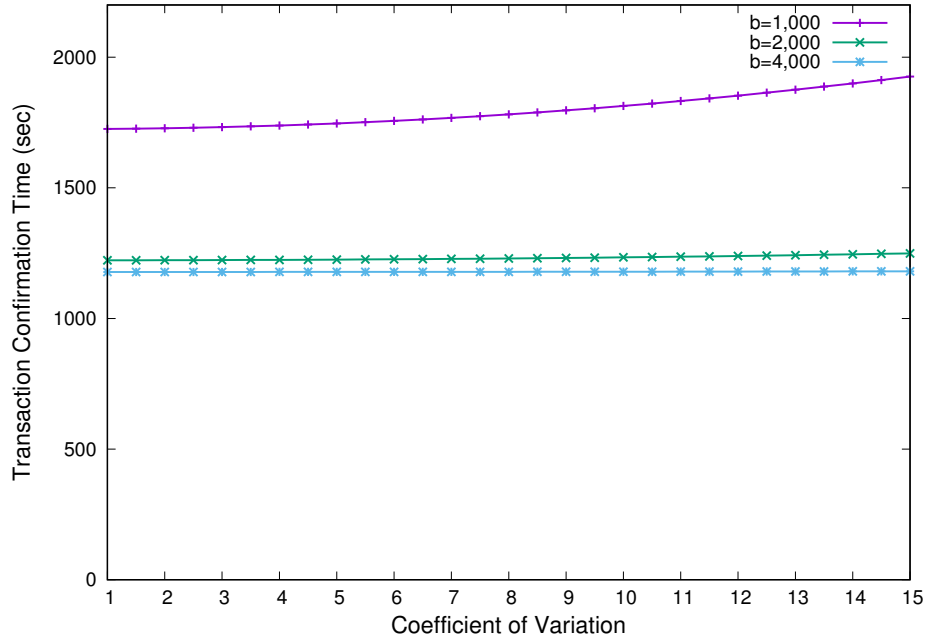


図 5: ブロックサイズと変動係数によるトランザクション承認時間への影響 (2014/10/1-2015/9/30)

ここで, $\lambda_2 = \gamma\lambda_1$ とすると, 式 (5) と式 (7) より, 次の p に関する 2 次方程式を得る.

$$\left(1 - \frac{1}{\gamma}\right)^2 p^2 - 2\left(\left(1 - \frac{1}{\gamma}\right) - \frac{1}{1 + \alpha^2} \left(1 - \frac{1}{\gamma^2}\right)\right)p + \frac{\alpha^2 - 1}{\alpha^2 + 1} = 0 \quad (8)$$

この 2 次方程式により, 変動係数 α と到着率の比例定数 γ を所与として $0 < p < 1$ を満たす p を決定し, トランザクション承認時間を計算する.

図 5 はトランザクション到着時間間隔の変動係数を増大させたときのトランザクション承認時間を図示したものである. ここでは 2 番目の期間の平均到着率 $\lambda = 1.20813$ とサービス率 $\mu = 0.0017009$ を用い, $\gamma = 1000$ と固定してブロックサイズ b を 1000, 2000, 4000 の 3 通りについて平均トランザクション承認時間を計算した.

図 5 より, $b = 1000$ では変動係数とともにトランザクション承認時間が増大していることが分かる. 一方, $b = 2000$ と 4000 では, 平均トランザクション承認時間がほぼ一定であることが観察される. これは平均到着率に対してブロックサイズが十分に大きいからであると考えられる.

5. 結論と今後の課題

本論文ではビットコインのトランザクション承認時間を $GI/M^b/1$ 待ち行列モデルを用いて解析した. この待ち行列モデルでは, 到着したトランザクションは一旦メモリープールに蓄えられ, 到着時点でマイニング中のブロックに空きがあってもそこには含められず, 次以降のブロックに到着順で格納される. 本待ち行列モデルに対し, トランザクション到着直前の待ち行列内トランザクション数をシステムの状態として定義し, 行列解析法により定常分布および平均トランザクション滞在時間を導出した. 数値実験では, トランザクション到着時間間隔の変動係数が指数分布よりも大きいことを考慮して到着時間間隔分布に超指数分布を仮定し, EM アルゴリズムを用いて分布のパラメータを決定した. 数値例においては, トランザクション到着時間間隔の平均値は実データと推定した超指数分布で一致していたが, トレース駆動シミュレーションでは解析の結果と乖離が見られた. 本結果はトランザクション到着時間間隔が独立同一な超指数分布でモデル化できないことを意味している.

数値例ではさらに到着率とブロックサイズの承認時間に与える影響を評価した. ブロックサイズが $b = 1000, 2000, 4000$ の三通りに対して, 到着率を増加させたときのトランザクション承認時間を図示した. 承認時間は到着率の増加とともに増大し, システムの安定条件付近で発散する傾向が見られた.

ブロックサイズを大きくすることでトランザクションの処理量のある程度増加させることはできるが、改善効果はそれほど大きくなく、スケーラビリティの問題を根本的に解決することはできないことが改めて確認された。

また数値例において、2 相から成る超指数分布に対する変動係数の承認時間に与える影響について評価を行った。ブロックサイズが $b = 1000, 2000, 4000$ の三通りに対して、変動係数を増加させたときの承認時間を図示した。 $b = 1000$ では変動係数の増大とともに承認時間は増大するが、 $b = 2000, 4000$ では平均到着率に対してブロックサイズが十分に大きいため、その効果はほとんど認められなかった。

本稿で検討した $GI/M^b/1$ モデルの数値例では、実データを基に到着間隔の平均や変動係数が合うように超指数分布のパラメータを EM アルゴリズムより決定したが、ブロックサイズが小さいところでは解析結果に基づく数値はデータ駆動型シミュレーション結果と大きく乖離することが観察された。これは、独立同一な到着時間間隔の仮定の下では平均トランザクション処理時間を高精度に予測できないことを意味している。今後の課題として、独立同一な到着時間間隔でバッチ到着をもつ到着過程やマルコフ到着過程など、拡張した到着過程をもつ集団サービス待ち行列モデルによる検討が挙げられる。

謝辞

本研究の一部は、科研費基盤 (B) 15H04008 及び SCAT 研究助成による支援を受けて実施している。

参考文献

- [1] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System,” <https://bitcoin.org/bitcoin.pdf>, 2008.
- [2] A. M. Antonopoulos, *Mastering Bitcoin*, O’Reilly, 2014.
- [3] <http://www.biccamera.co.jp/shopguide/campaign/bitcoin/index.html>
- [4] <http://www.coindesk.com/1mb-block-size-today-bitcoin/>
- [5] <https://blockchain.info/>
- [6] <http://www.coindesk.com/segregated-witness-bitcoin-block-size-debate/>
- [7] M. L. Chaudhry and J. G. C. Templeton, “The Queing System $M/G^B/1$ and its Ramifications,” *European Journal of Operational Research*, vol. 6, pp. 56–60, 1981.
- [8] M. L. Chaudhry and J. G. C. Templeton, *A First Course in Bulk Queues*, John Wiley & Sons, 1983.
- [9] S. Kasahara and J. Kawahara, “Effect of Bitcoin Fee on Transaction-Confirmation Process,” To appear in *Journal of Industrial and Management Optimization*.
- [10] Y. Kawase and S. Kasahara, “Transaction-Confirmation Time for Bitcoin: A Queueing Analytical Approach to Blockchain Mechanism,” *Queueing Theory and Network Applications (12th International Conference, QTNA2017, Qinhuangdao, China, August 21–23, 2017, Proceedings)*, W. Yue, Q.-L. Lin, S. Jin, and Z. Ma Eds., LNCS, vol. 10591, pp. 75–88, 2017.
- [11] Y. Kawase and S. Kasahara, “Priority Queueing Analysis of Transaction-Confirmation Time for Bitcoin Blockchain,” Submitted for review.
- [12] R. E. A. Khayari, R. Sadre and B. R. Haverkort, “Fitting World-Wide Web Request Traces with the EM-Algorithm,” *Performance Evaluation*, vol. 52, pp. 175–191, 2003.
- [13] R. Nelson, *Probability, Stochastic Processes, and Queueing Theory*, Springer-Verlag, 1995.

Probabilistic Model Checking for Biochemical Reaction Systems

Ratana Ty

Ken-etsu Fujita

Ken'ichi Kawanishi

Graduated School of Science and Technology
Gunma University

Abstract Probabilistic model checking is an automated verification technique for verifying the correctness and performance of probabilistic models. This paper shows how to use this verification technique to apply to biochemical reaction systems where the numbers of molecules are small and the reactions are governed by stochastic rates with various scales. We work on a case study analysing Fibroblast Growth Factor (FGF) molecular signaling modeled as continuous-time Markov chains (CTMCs). We verify this model by using model checker PRISM and quasi-stationary distributions for CTMCs.

1. Introduction

Probabilistic model checking [1] is an automated verification technique for verifying the correctness and performance of probabilistic models. It involves the systematic analysis of a probabilistic model, typically a variant of a Markov chain. This technique is an important component in the design and analysis of software and hardware systems. It has been widely used to study about systems including biological systems such as biochemical reaction networks and signaling pathways.

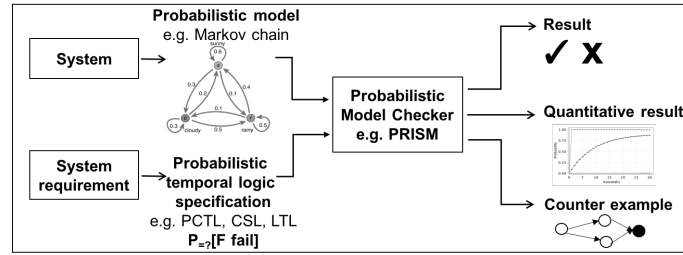


Figure 1: Probabilistic model checking

1.1. Modelling molecular networks

There are two established frameworks for modelling molecular reactions, the continuous deterministic approach and the discrete stochastic approach [2].

- In the continuous deterministic approach, one approximates the number of molecules using a continuous function that represents the change in molecule concentration using differential equations (ODEs) based on mass action kinetics. This approach is suitable for modelling large numbers of molecules and the solutions give the (average) concentration of each molecular species over time.
- The discrete stochastic approach models the stochastic evolution of populations of molecules, where reactions are discrete events, governed by stochastic rates typically assumed to be constant and dependent on the number of molecules, which admits their modelling in terms of continuous-time Markov chains. This approach is more accurate in cases where the numbers of molecules are small.

1.2. Aim

This paper shows how probabilistic model checking can be used to apply to biochemical reaction systems where the numbers of molecules are small and the reactions are governed by stochastic rates with various scales. We observe the mechanisms of the biochemical reaction before reaching the absorbing state by analysing the numbers and time of visiting as well as staying at any states by both model checking tool PRISM and quasi-stationary distributions.

2. Probabilistic Model Checking

We refer to [8] as the standard text of model checking including probabilistic systems.

2.1. Probabilistic model checking for continuous-time Markov Chains

Markov chains (MCs) are viewed as operational models for probabilistic systems and at the same time structures for Continuous Stochastic Logic (CSL) [9]. MCs are handled as a transition system with probabilities rather than a sequence of random variables, and logic serves for expressing properties of probabilistic systems in a systematic method.

A continuous-time Markov chain (CTMC) [1] is defined by the following tuple

$$\mathcal{C} = (\mathbf{S}, \bar{s}, \mathbf{P}, \mathbf{E}, \mathbf{AP}, \mathbf{L}),$$

where $\mathbf{S}$ is a finite set of states, $\bar{s} \in \mathbf{S}$ is a distinguished initial state, $\mathbf{P}$ is a transition probability matrix such that $\sum_{s' \in \mathbf{S}} \mathbf{P}(s, s') = 1$, $\mathbf{E}: \mathbf{S} \rightarrow \mathbb{R}_{\geq 0}$ is the exit rate, $\mathbf{AP}$ is a set of atomic propositions, and $\mathbf{L}: \mathbf{S} \rightarrow 2^{\mathbf{AP}}$ is a labelling function.

In a CTMC $\mathcal{C}$ [1], the residence time of a state $s \in \mathbf{S}$ is a random variable governed by an exponential distribution with exit rate parameter $\mathbf{E}(s)$. The probability to exit state s in t time units is given by $\int_0^t \mathbf{E}(s) \cdot e^{-\mathbf{E}(s)\tau} d\tau$. The probability of transition from state s to another state s' in t time units is equal to $\mathbf{P}(s, s') \cdot \int_0^t \mathbf{E}(s) \cdot e^{-\mathbf{E}(s)\tau} d\tau$.

Alternatively, a CTMC can be defined by specifying the rates matrix $\mathbf{R}: \mathbf{S} \times \mathbf{S} \rightarrow \mathbb{R}_{\geq 0}$ where $\mathbf{R}(s, s')$ is the rate of transitioning from state s to s' . A transition can only occur between states s and s' if $\mathbf{R}(s, s') > 0$. The probability of this transition from s within t time-units is $1 - e^{-\mathbf{R}(s, s') \cdot t}$. If there is more than one state s' for which $\mathbf{R}(s, s') > 0$, the first transition to be taken place determines the next state of the CTMC. The exit rate $\mathbf{E}(s)$ is then equal to $\sum_{s' \neq s} \mathbf{R}(s, s')$ and the embedded discrete-time Markov chain (DTMC) is given by

$$\mathbf{P}(s, s') = \begin{cases} \frac{\mathbf{R}(s, s')}{\mathbf{E}(s)} & \text{if } \mathbf{E}(s) \neq 0 \text{ and } s \neq s', \\ 1 & \text{if } \mathbf{E}(s) = 0 \text{ and } s = s', \\ 0 & \text{otherwise.} \end{cases}$$

2.2. Continuous Stochastic Logic (CSL)

To specify quantitative properties of CTMCs, the logic CSL [2] has been proposed. CSL distinguishes between state formulas (Φ) and path formulas (Ψ) and includes $\mathbf{P}_{\sim p}[\Psi]$ (probabilistic operator), $\mathbf{S}_{\sim p}$ (steady state), path operators $\mathbf{X}\Phi$ (next state), as well as $\Phi \mathbf{U}^{[t, t']}\Phi$ (time bounded until). The syntax of CSL is given by

$$\begin{aligned} \Phi &::= \text{true} \mid \mathbf{a} \mid \neg\Phi \mid \Phi \vee \Phi \mid \Phi \wedge \Phi \mid \mathbf{P}_{\sim p}[\Psi] \mid \mathbf{S}_{\sim p}[\Phi] \\ \Psi &::= \mathbf{X}\Phi \mid \Phi \mathbf{U}^{[t, t']}\Phi \mid \Phi \mathbf{U}\Phi \end{aligned}$$

where $\mathbf{a}$ is an atomic proposition, $\sim \in \{<, \leq, >, \geq\}$, $p \in [0, 1]$, and $t, t' \in \mathbb{R}_{\geq 0}$.

2.3. Model checking for CSL over CTMCs

The infinitesimal generator matrix [2] for the CTMC $\mathcal{C} = (\mathbf{S}, \bar{s}, \mathbf{P}, \mathbf{E}, \mathbf{AP}, \mathbf{L})$ is the matrix $\mathbf{Q}: \mathbf{S} \times \mathbf{S} \rightarrow \mathbb{R}$ defined as follows:

$$\mathbf{Q}(s, s') = \begin{cases} \mathbf{R}(s, s') & \text{if } s \neq s', \\ -\sum_{s' \neq s} \mathbf{R}(s, s') & \text{otherwise.} \end{cases}$$

2.4. Probabilistic model checker PRISM

PRISM [4] is a tool for formal modelling and analysis of systems that exhibit random or probabilistic behaviour. It is an open source application developed at the University of Birmingham and University of Oxford. PRISM can build and analyse some types of probabilistic models: discrete-time Markov chains (DTMCs), continuous-time Markov chains (CTMCs), Markov decision processes (MDPs), probabilistic automata (PAs), probabilistic timed automata (PTAs). The property specification language of PRISM is based on temporal logics: probabilistic computation tree logic

(PCTL), continuous stochastic logic (CSL), linear temporal logic (LTL), probabilistic computation tree logic* (PTCL*)¹ and extensions for quantitative specifications, costs and rewards.

2.4.1. Rewards

PRISM can be used to compute properties such as how long the system has spent in a given state, or how often a certain state transition has taken place based on “rewards”.

3. Biological Case Study

3.1. Fibroblast Growth Factor (FGF)

FGF [1] are family of proteins which play an important role in cell signaling. FGF have been involved with e.g. skeletal development and wound healing, etc. FGF function in many organisms from worm to humans. The mechanisms of FGF signaling are not well understood, and several hypotheses exist.

3.2. Fibroblast Growth Factor Receptor (FGFR)

FGFR are extracellular protein. They receive the signal from the growth signal to promote fibroblast to grow in normal cell. FGFR can only interact with FGF at the cell surface, but relocation causes FGFR to move inside the cell.

3.3. A simple set of reactions

The simplified set of reactions based on the role of FGF in receptor is given below. An FGF (molecule) can bind to an FGF receptor (FGFR) to form the compound FGF:FGFR. When the compound FGF:FGFR is formed, FGFR can become phosphorylated. Binding and phosphorylation, both reactions are also reversible. Finally, when FGFR is phosphorylated, it can be relocated. We assume that FGF disappears if it is bound to FGFR when relocation occurs.

FGF binds FGFR:



FGF releases FGFR:



Phosphorylation² of FGFR (whilst FGF:FGF):



Dephosphorylation³ of FGFR:



Relocation of FGFR (whilst FGF:FGFRP):



Each reaction has an associated kinetic rate ($k_1, \dots, k_5$). The kinetic rate k_1 has units $M^{-1}s^{-1}$, where M refers to the molar concentration, i.e., the number of moles per litre.

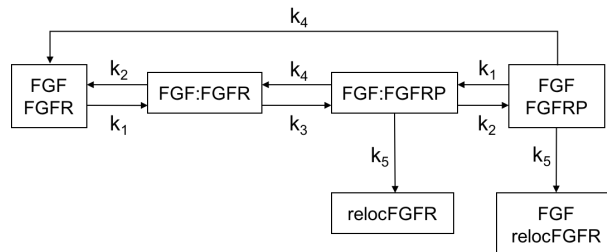


Figure 2: System of reactions of FGF

¹PCTL* can be used to specify properties of DTMCs and MDPs (or untimed properties of CTMCs).

²The addition of a phosphoryl (PO_3^{2-}) group to a molecule

³The process by which phosphate (PO_4^{3-}) group are removed from a molecule by a phosphatase.

4. Model Analysis by PRISM

4.1. Modelling a single number of molecule

We give two alternative approaches [2] to model the reaction system of one of each FGF and FGFR molecule. Both correspond to the model CTMCs and a set of modules. In our first approach, there are two main modules, FGF and FGFR, each with variables representing their current state. Reactions involving more than one species are modelled using synchronisation. In our second approach, we use a single module with one variable representing the six possible states of the whole system. This approach is simple and easy to understand. However, when the system becomes more complex the use of separate variables and synchronisation as our first approach becomes more desirable since we can easily modify and express properties.

4.2. Model analysis

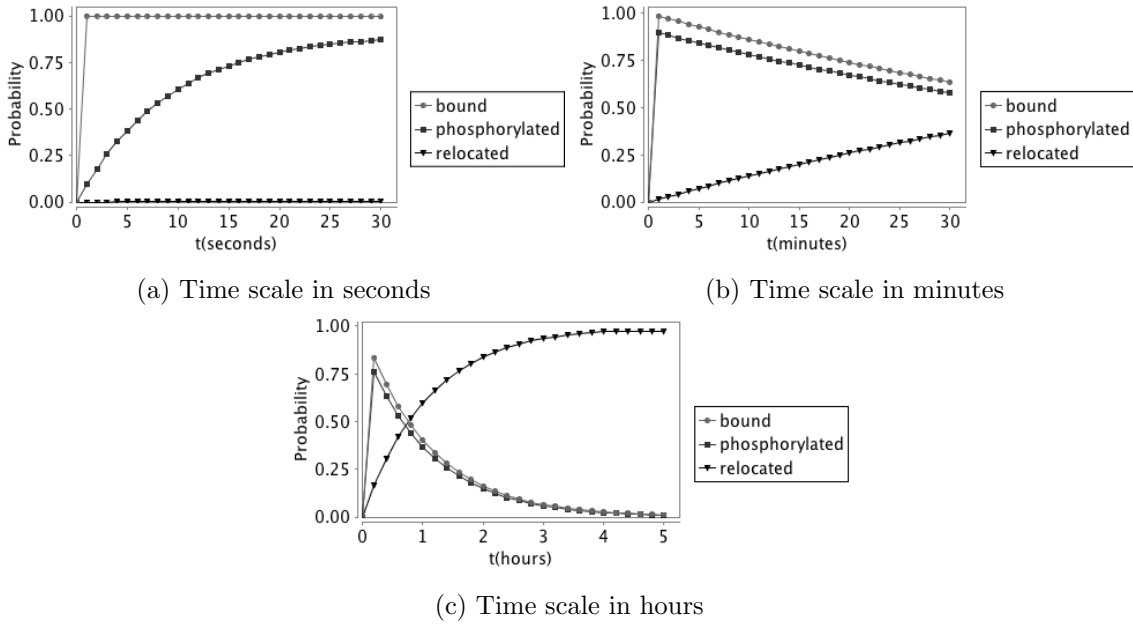


Figure 3: Probabilities over three time scales for the FGF system to be in one of three states

Figure 3 shows results analysed with PRISM for the probability that at time instant t , FGFR is: bound to FGF; phosphorylated; relocated. Results are plotted for ranges of t over 3 different time scales (a) seconds, (b) minutes and (c) hours. Figure 3 (a) shows that in the initial evolution of the system FGF and FGFR bind very quickly after that FGFR becomes phosphorylated while there is almost no chance of FGFR relocating. Figures 3 (b) and 3 (c) show that as time passed the chance that FGF and FGFR are bound diminishes, and the chance that FGFR is phosphorylated diminishes faster, eventually FGFR become relocated.

4.3. Modelling larger number of molecules

In this section we extend our model in previous section by increasing more than one of each FGF and FGFR molecule to occur in the system. We give two possible approaches.

- **Individual-based model:** each individual molecule is modeled separately.
- **Population-based model:** each type of molecule or species are modeled separately rather than the state of each individual component.

4.3.1. The state-space explosion problem

Figure 4 lists the numbers of states and transitions along with their graphs, where X-axis measures the number of FGF and FGFR molecules and the Y-axis measures the base 10 logarithm of the numbers of states and transitions. As can be seen, there is a rapid increase in the number of states and transitions for the individual-based model and quickly make model checking infeasible.

This model will suffer from the well-known state-space explosion problem as the complexity of the system increases where there is an exponential growth in the state space. However, there is also an increase in the number of states for the population-based model but in the way far more gradual increase in the state space, which allow us to analyse much larger models.

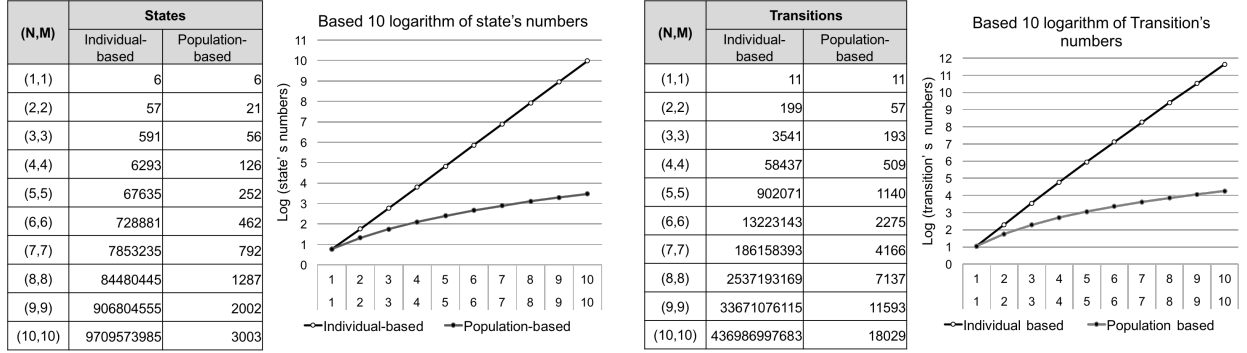


Figure 4: States and transitions of individual-based and population-based model

4.3.2. Population-based model analysis by PRISM

In order to compute the numbers of transitions to any states before relocation occurs, we give labels denoted by $[X] \in \{[\text{unbind1}], [\text{bind1}], [\text{unbind2}], [\text{bind2}], [\text{phos1}], [\text{phos2}], [\text{unbindp}]]\}$ to the transitions in the system as in Figure 5. Table 1 shows the numbers of each transition by simulation and verification.

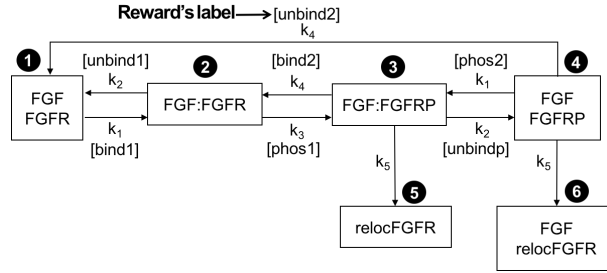


Figure 5: Labels of transitions

Table 1: Expected numbers of transition before relocation occurs by PRISM

State	Label	Number of transitions					
		Simulation (Number of samples)				Verification	Total
		(10)	(100)	(1000)	(10000)		
1	[unbind1]	0.3000	0.7300	0.7490	0.7504	0.7400	0.7400
	[unbind2]	0.0000	0.0000	0.0000	0.0000	< 0.0001	
2	[bind1]	1.4000	1.8400	1.8160	1.7304	1.7400	37.7400
	[bind2]	27.7000	33.4400	34.7100	35.8794	36.0000	
3	[phos1]	45.2000	42.0900	34.5710	36.3347	37.0000	44.2000
	[phos2]	7.2000	7.3900	7.1140	7.2736	7.2000	
4	[unbindp]	9.9000	7.1600	7.0290	7.1517	7.2000	7.2000
Total numbers of transition							89.8000

Table 2 below lists the time spent at states 1–4 before relocation occurs. Table 3 lists the

probability of reaching relocating states (absorbing states) 5 and 6. As can be seen, the probability of reaching state 5 is really high whereas there is almost no chance that it reaches state 6.

Table 2: Expected time staying at any states before relocation occurs by PRISM

State	1	2	3	4	Total
Time of staying	3.4800×10^{-4}	3.7000×10^2	3.6000×10^3	1.4000×10^{-3}	$\approx 3.9700 \times 10^3$

Table 3: Probability of reaching relocating (absorbing) state by PRISM

State	5	6
Probability	9.9999960×10^{-1}	3.9999902×10^{-7}

5. Analysis by Absorbing Markov chain

The transition rate matrix R and the corresponding infinitesimal generator Q of the continuous-time Markov chain are given by

$$R = \begin{pmatrix} 0 & k_1 & 0 & 0 & 0 & 0 \\ k_2 & 0 & k_3 & 0 & 0 & 0 \\ 0 & k_4 & 0 & k_2 & k_5 & 0 \\ k_4 & 0 & k_1 & 0 & 0 & k_5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad Q = \begin{pmatrix} -q_1 & k_1 & 0 & 0 & 0 & 0 \\ k_2 & -q_2 & k_3 & 0 & 0 & 0 \\ 0 & k_4 & -q_3 & k_2 & k_5 & 0 \\ k_4 & 0 & k_1 & -q_4 & 0 & k_5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where $q_1 = k_1, q_2 = k_2 + k_3, q_3 = k_2 + k_4 + k_5$ and $q_4 = k_1 + k_4 + k_5$. The transition probability matrix P for the embedded discrete-time Markov chain is given by

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{k_2}{k_2+k_3} & 0 & \frac{k_3}{k_2+k_3} & 0 & 0 & 0 \\ 0 & \frac{k_4}{k_2+k_4+k_5} & 0 & \frac{k_2}{k_2+k_4+k_5} & \frac{k_5}{k_2+k_4+k_5} & 0 \\ \frac{k_4}{k_1+k_4+k_5} & 0 & \frac{k_1}{k_1+k_4+k_5} & 0 & 0 & \frac{k_5}{k_1+k_4+k_5} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Our P matrix is so-called absorbing Markov chain [5]. The transition probability matrix of the absorbing Markov chain is of the general form written by

$$P = \begin{pmatrix} P_T & P_{T,A} \\ 0 & I \end{pmatrix},$$

where I is the identity matrix of appropriate dimension. In our case, we have two absorbing states 5 and 6 and hence the dimension is two. The sub-stochastic matrix P_T corresponding to the transient states and $P_{T,A}$ are given by

$$P_T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{k_2}{k_2+k_3} & 0 & \frac{k_3}{k_2+k_3} & 0 \\ 0 & \frac{k_4}{k_2+k_4+k_5} & 0 & \frac{k_2}{k_2+k_4+k_5} \\ \frac{k_4}{k_1+k_4+k_5} & 0 & \frac{k_1}{k_1+k_4+k_5} & 0 \end{pmatrix}, \quad P_{T,A} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \frac{k_5}{k_2+k_4+k_5} & 0 \\ 0 & \frac{k_5}{k_1+k_4+k_5} \end{pmatrix}.$$

To compute the average number of times a specific transient state is visited before the absorbing states are entered, we define the fundamental matrix N by

$$N = (I - P_T)^{-1}.$$

The (i, j) entry $N_{i,j}$ of $\mathbf{N}$ is the average number of times the absorbing Markov chain will be in state j before the absorbing states are entered, given that it started in state i . In our case, we can numerically compute $\mathbf{N}$ as follow:

$$\mathbf{N} = \begin{pmatrix} 1.7400 & 37.7400 & 44.2000 & 7.2000 \\ 0.7400 & 37.7400 & 44.2000 & 7.2000 \\ 0.7200 & 36.7200 & 44.2000 & 7.2000 \\ 0.7200 & 36.7200 & 44.2000 & 8.2000 \end{pmatrix}.$$

This means that the biochemical reaction starting from state 1 will, on average, be expected to be in state 2, where the FGF molecule bound with FGFR molecule, 37.74 times before reaching absorbing states.

Let n_i be the mean time before reaching the absorbing states starting from a specific transient state i , and let $\mathbf{n} = (n_i)$ be the column vector composed of n_i for $i=1,2,3$ and 4. In discrete-time Markov chains, the mean time is the mean number of steps. We can obtain $\mathbf{n}$ by multiplying the fundamental matrix $\mathbf{N}$ with a column vector $\mathbf{1}$ that contains only 1's as

$$\mathbf{n} = \mathbf{N}\mathbf{1} = \begin{pmatrix} 90.8800 \\ 89.8800 \\ 88.8400 \\ 89.8400 \end{pmatrix}.$$

This tells us that it takes on average 90.88 time steps before reaching relocation states with the absorbing states starting from state 1. It should be noted that the total number of transitions in Table 1 is equal to $n_1 - 1$.

We can obtain the parallel continuous-time analogues. Let $\mathbf{Q}_T$ be the infinitesimal sub-generator of the continuous-time absorbing Markov chain. In our case, it is given by

$$\mathbf{Q}_T = \begin{pmatrix} -q_1 & k_1 & 0 & 0 \\ k_2 & -q_2 & k_3 & 0 \\ 0 & k_4 & -q_3 & k_2 \\ k_4 & 0 & k_1 & -q_4 \end{pmatrix}.$$

We define a matrix $\mathbf{T}$ whose (i, j) entry $T_{i,j}$ records the mean time that the continuous-time absorbing Markov chain will be in a transient state j before the absorbing states are entered, given that it started in a transient state i . Then it holds that $\mathbf{T} = (-\mathbf{Q}_T)^{-1}$. In our case, we can numerically compute $\mathbf{T}$ as follow:

$$\mathbf{T} = \begin{pmatrix} 3.4800 \times 10^{-4} & 3.7000 \times 10^2 & 3.6000 \times 10^3 & 1.4400 \times 10^{-3} \\ 1.4800 \times 10^{-4} & 3.7000 \times 10^2 & 3.6000 \times 10^3 & 1.4400 \times 10^{-3} \\ 1.4400 \times 10^{-4} & 3.6000 \times 10^2 & 3.6000 \times 10^3 & 1.4400 \times 10^{-3} \\ 1.4400 \times 10^{-4} & 3.6000 \times 10^2 & 3.6000 \times 10^3 & 1.6400 \times 10^{-3} \end{pmatrix}.$$

Let t_i be the mean time of the continuous-time Markov chain before reaching the absorbing states, given that it started from a transient state i . If we denote by $\mathbf{t} = (t_i)$ the column vector composed of t_i for $i=1,2,3$ and 4, then we have $\mathbf{t} = \mathbf{T}\mathbf{1}$ and our example gives numerical values of $\mathbf{t}$ as

$$\mathbf{t} = \begin{pmatrix} 3.9700 \times 10^3 \\ 3.9700 \times 10^3 \\ 3.9600 \times 10^3 \\ 3.9600 \times 10^3 \end{pmatrix}.$$

It should be noted that t_1 is equal to the total expected time by PRISM in Table 2.

6. Probability Distributions Before Absorption

We analyze the behavior of absorbing Markov chains before absorption occurs. For this, we consider two types of probability distributions. First, we define a ratio between the mean time the Markov chain spends at each transient state and the mean time to absorption. The ratio gives us probability distributions depending on each initial transient state. Second, we consider so-called quasi-stationary distribution of absorbing Markov chains. In our case, the existence of the quasi-stationary distribution is guaranteed because the set of transient states is finite and irreducible.

6.1. Probability distributions by ratio of mean times

For transient states i and j of the discrete-time absorbing Markov chain, we define $p(j | i)$ by

$$p(j | i) = \frac{N_{i,j}}{n_i}.$$

It should be noted that n_i is finite and $n_i > 0$ for each transient state $i \in \{1, 2, 3, 4\}$. Because $\mathbf{n} = \mathbf{N}\mathbf{1}$, we can interpret $\mathbf{p}(\cdot | i) = (p(j | i))$ as a probability mass function for each transient state $i \in \{1, 2, 3, 4\}$.

For transient states i and j of the continuous-time absorbing Markov chain, we define $q(j | i)$ by

$$q(j | i) = \frac{T_{i,j}}{t_i}.$$

As is the case of the discrete-time Markov chain, we can interpret $\mathbf{q}(\cdot | i) = (q(j | i))$ as a probability mass function for each transient state $i \in \{1, 2, 3, 4\}$.

6.2. Quasi-stationary distributions

We briefly introduce the quasi-stationary distributions of absorbing Markov chains. For the quasi-stationary distributions, we refer the reader to the seminal work [6, 7].

Let $\mathbf{x}$ be the quasi-stationary distribution of $\mathbf{P}_T$ on the transient states for the discrete-time Markov chain, under the condition that absorption has not yet taken place. It is known that $\mathbf{x}$ is given by the left eigenvector of $\mathbf{P}_T$ for the eigenvalue ρ with maximal real part, that is,

$$\mathbf{x}\mathbf{P}_T = \rho\mathbf{x}.$$

Because $\mathbf{P}_T$ is a nonnegative and irreducible matrix in our case, the eigenvalue ρ is the Perron–Frobenius eigenvalue of $\mathbf{P}_T$. Since $\mathbf{P}_T$ is sub-stochastic, we must have $\rho < 1$. We can obtain the left eigenvector $\mathbf{x}$ as a unique strictly positive vector such that $\mathbf{x}\mathbf{1} = 1$.

Similarly, let $\mathbf{y}$ be the quasi-stationary distribution of $\mathbf{Q}_T$ on the transient states for the continuous-time Markov chain, under the condition that absorption has not yet taken place. It is also known that $\mathbf{y}$ is given by the left eigenvector of $\mathbf{Q}_T$ for the eigenvalue $-\alpha$ with maximal real part, that is,

$$\mathbf{y}\mathbf{Q}_T = -\alpha\mathbf{y}.$$

Because $\mathbf{Q}_T$ is irreducible, $-\alpha$ is unique and simple, and we must have $-\alpha < 0$. It also holds that $\mathbf{y}$ can be chosen strictly positive vector such that $\mathbf{y}\mathbf{1} = 1$.

7. Comparison of Probabilities Computed by PRISM and Those of Quasi-stationary Distributions

Table 4: Comparing probabilities of visiting states 1–4 computed by PRISM with quasi-stationary distributions

State	PRISM			Qusasi-stationary distributions
	Transitions	Visits	Probability	Probability
1	0.7400	1.7400	0.0191	0.0082
2	37.7400	37.7400	0.4153	0.4157
3	44.2000	44.2000	0.4864	0.4946
4	7.2000	7.2000	0.0792	0.0815
Total	89.8800	90.8800	1.0000	1.0000

Table 5: Comparing probabilities of staying at states 1–4 computed by PRISM with quasi-stationary distributions

State	PRISM		Qusasi-stationary distributions
	Time	Probability	Probability
1	3.4800×10^{-4}	0.0000	0.0000
2	3.7000×10^2	0.0932	0.0911
3	3.6000×10^3	0.9068	0.9089
4	1.4000×10^{-3}	0.0000	0.0000
Total	3.9700×10^3	1.0000	1.0000

As can be seen in Tables 4 and 5, the probability of visiting and staying at any states before relocation occurs obtained by model checker PRISM and quasi-stationary distributions are approximately the same.

8. Concluding Remarks and Further Work

In this research, we studied probabilistic model checking [3] for continuous-time Markov chains, and how to analyse this model with model checker PRISM [4]. We worked with a model which contains one molecule each of FGF and FGFR of which signalling mechanisms are not well understood and several hypotheses exist. We also showed how to model a molecular interactions with different stochastic rate using model checker PRISM. We considered two alternative approaches [2]: individual-based model and population-based model.

We analysed the population-based model when there are one of each FGF and FGFR molecule. By comparing the results of PRISM with those of quasi-stationary distribution over absorbing Markov chains, we obtained the probability of visiting and staying at any states before relocation occurs in both techniques are approximately the same. In PRISM, we can specify the initial state and compute properties such as the time spent in a given state or how often a certain state transition has taken by using a feature named “reward”, however we need to put labels to every transitions and states, which makes the model become more complex and easily meet the state-space explosion problem as large numbers of reactions in the system occur. In contrast, in quasi-stationary distributions, we cannot specify the initial state of the system, however we can define numbers of visits or time of staying and the approximated probability in all states in the system at a time. For justifying this approximation method, it would be worthwhile to investigate that the quasi-stationary distributions can provide an upper bound for the probability mass function in terms of stochastic ordering.

As our future work, we want to work on more further and detail about Probabilistic model

checking of larger numbers of molecules of biological system and work on the combination of quasi-stationary distribution and PRISM to compute the system properties without giving labels in each transition.

References

- [1] M. Kwiatkowska and C. Thachuk: Probabilistic Model Checking for Biology. In *Software Safety and Security*, IOS Press, pp. 165–190, 2014.
- [2] M. Kwiatkowska, G. Norman and D. Parker: Probabilistic Model Checking for Systems Biology. In M. Sriram Iyengar (editor), *Symbolic System Biology*, pp. 31–59, Jones and Bartlett, 2010.
- [3] M. Kwiatkowska, G. Norman and D. Parker: Stochastic Model Checking. In M. Bernardo and J. Hillston (editors), *Formal Methods for the Design of Computer, Communication and Software Systems: Performance Evaluation (SFM’07)*, LNCS Vol. 4486, pp. 220–270, 2007.
- [4] PRISM website: <http://www.prismmodelchecker.org>
- [5] D. J. Barnes and D. Chu: *Introduction to Modeling for Biosciences*. Springer, 2010.
- [6] J. N. Darroch and E. Seneta: On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *Journal of Applied Probability*, Vol. 2, pp. 88–100, 1965.
- [7] J. N. Darroch and E. Seneta: On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *Journal of Applied Probability*, Vol. 4, pp. 192–196, 1967.
- [8] C. Baier and J-P. Katoen: *Principle of Model Checking*. The MIT Press, 2008.
- [9] K. Fujita: Theorem Proving and Model Checking. *Proc. of the 2016 IEICE General Conference*, AI-3-1, SS-30–31, 2016.

複合ポアソン過程を用いた企業価値のモデル化とリスク評価

井田 倫宏

東京工業大学情報理工学院数理・計算科学系 三好研究室

概要：近年、日本国内を含む国際社会において、企業の合併や買収といった M&A 事例が数多く報告されている。それに伴い、合併や買収を行う際に行われる企業価値評価の重要性も高まってきている。通常、企業価値の評価を行う際には、対象企業の将来のフリー・キャッシュフローを予測し、それを用いて企業資産および負債の現在価値の算出を行うが、フリー・キャッシュフローの予測は定性的で、希望的観測によるものも多い。本論文では、企業の売上や損失の累積を複合ポアソン過程を用いて表現し、企業価値をモデル化することで、対象企業のリスクを評価する手法を提案する。

1. はじめに

M&A すなわち企業の買収・売却や合併において、対象となる企業の価値を評価することは必要不可欠である。企業価値評価の方法は、大きく分けて3つある。1つ目は、将来期待される経済的な収益やキャッシュフローを、見込まれるリスク等を考慮した割引率で割り引くことにより評価を行う「インカムアプローチ (Income Approach)」である。2つ目は会社の保有している資産の価値に着目する「コストアプローチ (Cost Approach)」である。3つ目は、市場において成立する相場の価格をもとに企業価値を算定する手法である「マーケットアプローチ (Market Approach)」である。通常は、これらのアプローチのいくつかを使用し、最終的な価格幅を求めることとなる。

上場企業の場合、公開された市場において株価の時価総額を知ることができ、これが1つの価値指標となるが、非上場企業の場合、基準とすべき市場価値が存在しないので、企業価値評価の難度は上がる。本研究では、いかなる企業に対しても適用可能な「インカムアプローチ」について考察を行う。

「インカムアプローチ」に分類される手法の中でも特によく用いられているものが「エンタプライズ DCF 法 (Enterprise DCF Method)」である。この手法では、対象企業が将来生み出すフリー・キャッシュフロー (FCF) によって企業の現在価値をもとめる。メリットとしては、経営改善効果やシナジー効果といった、M&A 後に生じるプラスのキャッシュフローを企業価値に反映させることができ、M&A の効果を検証するには最適であるという点が挙げられる。一方でデメリットとしては、将来生じるリスク等をどのように設定するかで結果が大きく変わり、客観性の確保が困難であるという点が挙げられる。

本研究では、「エンタプライズ DCF 法」のデメリットである、評価に関する客観性の確保を解決すべく、マイナスのキャッシュフローや不確実性を伴う損害を、複合ポアソン過程を応用した離散時間型複合ポアソンを用いて表現する。

2. 複合ポアソン過程とその応用

定義 2.1 (複合ポアソン過程 (compound Poisson process)) $\{N_t\}$ を強度 λ の一様ポアソン過程^{*1}、 $\{X_n\}_{n=1}^{\infty}$ を独立同分布かつ $\{N_t\}$ と独立となるような確率変数列であるとする。このとき、次で定義される確率過程 $\{S_t\}$ を考える。:

$$S_t = \sum_{n=1}^{N_t} X_n, \quad t \geq 0 \quad (1)$$

ただし、 $\sum_{n=0}^0 = 0$ とする。このとき、 $\{S_t\}_{t \geq 0}$ は複合ポアソン過程とよばれる。

複合ポアソン過程は、保険会社の持つサープラス (準備金) のモデル化に用いられる。

例 2.1 (保険会社のサープラス) 保険会社の持つサープラス R_t を次のようにモデル化する。

$$R_t = u + \beta t - C_t, \quad C_t = \sum_{n=1}^{N_t} U_n \quad (2)$$

- u : 初期サープラス

^{*1} $N_t \sim Po(\lambda t)$

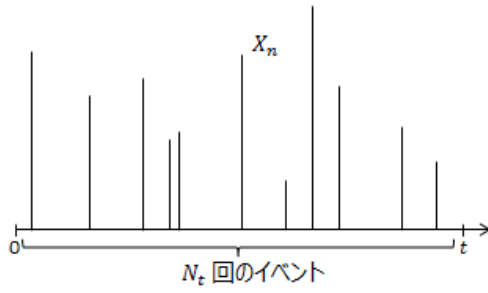


図1 複合ポアソン過程

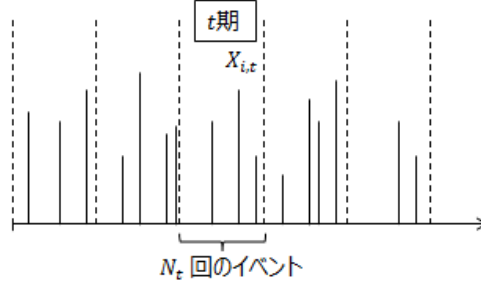


図2 離散時間型複合ポアソン過程

- β : 保険料を表す正の定数
- $\{N_t\}$: 強度 λ の一様ポアソン過程でクレームの頻度に対応する
- $\{U_n\}_{n=1}^{\infty}$: 分布 F_U に従う独立同分布かつ $\{N_t\}$ と独立な正值確率変数列

このモデルは、リスク過程の中でも代表的なモデルで、古典的リスクモデルや Cramér-Lundberg モデルとよばれる。

定義 2.2 (離散時間型複合ポアソン過程) $\{N_t\}$ は強度 λ のポアソン分布に従い^{*2}、 $\{X_{i,t}\}_{\{i,t\} \in \mathbb{N}^2}$ を独立同分布かつ $\{N_t\}$ と独立となるような確率変数列であるとする。このとき、次で定義される確率過程 $\{S_t\}$ を考える。:

$$S_t = \sum_{i=1}^{N_t} X_{i,t}, \quad t \in \{1, 2, 3, \dots\} \quad (3)$$

ただし、 $\sum_{i=0}^0 = 0$ とする。このとき、 $\{\sum_{k=1}^t S_k\}_{t \in \{1, 2, 3, \dots\}}$ を離散時間型複合ポアソン過程とよぶ。

ここで、 t は連続時間のある一定の間隔で区切ったときの期間を表している。一定の間隔で区切っているため、 $\{N_t\}$ は強度 λ のポアソン分布に従うと仮定している。

3. モデル

3.1. フリー・キャッシュフローのモデル化

本研究では、将来予測される売上高などプラスのキャッシュフロー (CF)、生産にかかったコストや株主への配当金支出などによるマイナスのキャッシュフロー、それとは別に自然災害や大規模な不況等による損害を不確実性を伴う損害といった 3 つの要素が企業のフリー・キャッシュフローに影響を及ぼすものと考え、 $FCF(F_t)$ を次のようにモデル化した。

$$F_t = a_t - \left(\sum_{i=1}^{N_t} X_{i,t} + \sum_{i=1}^{M_t} Y_{i,t} + \sigma W_t \right) \quad (4)$$

- a_t : t 期におけるプラスの CF。おもに企業の売上高目標に基づき、与えられる。
- $X_{i,t}$: t 期における i 番目のマイナスの CF。独立同一な分布 F_X に従う。
- $Y_{i,t}$: t 期における i 番目の不確実性を伴う損害。独立同一な分布 F_Y に従う。
- $N_t \stackrel{\text{iid}}{\sim} Po(\lambda), M_t \stackrel{\text{iid}}{\sim} Po(\mu)$: t 期における CF および不確実性を伴う損害の発生回数。
- $W_t \sim N(0, t), \sigma > 0$

3.2. 企業価値のモデル化

加重平均資本コスト (Weighted Average Cost of Capital: WACC) を用いて FCF を割り引くことで企業価値を算出する。

$$\begin{aligned} PV &= \sum_{t=1}^{\infty} \frac{F_t}{(1+r)^t} \\ &= \sum_{t=1}^{\infty} \frac{a_t}{(1+r)^t} - \sum_{t=1}^{\infty} \frac{1}{(1+r)^t} \left(\sum_{i=1}^{N_t} X_{i,t} + \sum_{i=1}^{M_t} Y_{i,t} + \sigma W_t \right) \end{aligned} \quad (5)$$

^{*2} $N_t \sim Po(\lambda)$

ここで、(2) 式の第 2 項を確率変数 V で表す。すなわち、

$$V = \sum_{t=1}^{\infty} \frac{1}{(1+r)^t} \left(\sum_{i=1}^{N_t} X_{i,t} + \sum_{i=1}^{M_t} Y_{i,t} + \sigma W_t \right) \quad (6)$$

- r : WACC

4. Laplace 変換

確率変数 V に関して Laplace 変換を計算する。

$$\begin{aligned} \mathcal{L}_V(s) &:= \mathbb{E} [e^{-sV}] \\ &= \mathbb{E} \left[\exp \left(-s \sum_{t=1}^{\infty} \frac{1}{(1+r)^t} \left(\sum_{i=1}^{N_t} X_{i,t} + \sum_{i=1}^{M_t} Y_{i,t} + \sigma W_t \right) \right) \right] \\ &= \prod_{t=1}^{\infty} \mathbb{E} \left[\exp \left(-\frac{s}{(1+r)^t} \left(\sum_{i=1}^{N_t} X_{i,t} + \sum_{i=1}^{M_t} Y_{i,t} + \sigma W_t \right) \right) \right] \\ &\quad \left(c_t = \frac{s}{(1+r)^t} \right) \\ &= \prod_{t=1}^{\infty} \mathbb{E} \left[\exp \left(-c_t \sum_{i=1}^{N_t} X_{i,t} \right) \right] \mathbb{E} \left[\exp \left(-c_t \sum_{i=1}^{M_t} Y_{i,t} \right) \right] \mathbb{E} \left[\exp \left(-c_t \sigma W_t \right) \right] \end{aligned} \quad (7)$$

ここで、(4) 式の無限積に含まれるそれぞれの期待値について計算していくと、

$$\begin{aligned} \mathbb{E} \left[\exp \left(-c_t \sum_{i=1}^{N_t} X_{i,t} \right) \right] &= \mathbb{E} \left[\mathbb{E} \left[\exp \left(-c_t \sum_{i=1}^{N_t} X_{i,t} \right) \mid N_t \right] \right] \\ &= \mathbb{E} \left[\prod_{i=1}^{N_t} \mathbb{E} [\exp(-c_t X_{i,t})] \right] \\ &= \mathbb{E} \left[(\mathbb{E} [\exp(-c_t X)])^{N_t} \right] \\ &= \sum_{n=0}^{\infty} (\mathbb{E} [\exp(-c_t X)])^n \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \exp(-\lambda(1 - \mathcal{L}_X(c_t))) \end{aligned} \quad (8)$$

$$\mathbb{E} \left[\exp \left(c_t \sum_{i=1}^{M_t} Y_{i,t} \right) \right] = \exp(-\mu(1 - \mathcal{L}_Y(c_t))) \quad (9)$$

$$\mathbb{E} [\exp(-c_t \sigma W_t)] = \exp \left(\frac{\sigma^2 t c_t^2}{2} \right) \quad (10)$$

(5)(6)(7) 式より、

$$\mathcal{L}_V(s) = \prod_{t=1}^{\infty} \exp \left(-\lambda(1 - \mathcal{L}_X(c_t)) - \mu(1 - \mathcal{L}_Y(c_t)) + \frac{\sigma^2 t c_t^2}{2} \right) \quad (11)$$

5. リスク評価の方法

本研究では、マルコフの不等式を利用することで企業価値のリスク評価を行う。

$$\begin{aligned}
P(PV \leq c) &= P\left(\sum_{t=1}^{\infty} \frac{a_t}{(1+r)^t} - \sum_{t=1}^{\infty} \frac{1}{(1+r)^t} \left(\sum_{i=1}^{N_t} X_{i,t} + \sum_{i=1}^{M_t} Y_{i,t} + \sigma W_t\right) \leq c\right) \\
&= P\left(\sum_{t=1}^{\infty} \frac{a_t}{(1+r)^t} - c \leq \sum_{t=1}^{\infty} \frac{1}{(1+r)^t} \left(\sum_{i=1}^{N_t} X_{i,t} + \sum_{i=1}^{M_t} Y_{i,t} + \sigma W_t\right)\right) \\
&\text{ここで、} \sum_{t=1}^{\infty} \frac{a_t}{(1+r)^t} - c = d \text{ とおくと} \\
&= P\left(d \leq \sum_{t=1}^{\infty} \frac{1}{(1+r)^t} \left(\sum_{i=1}^{N_t} X_{i,t} + \sum_{i=1}^{M_t} Y_{i,t} + \sigma W_t\right)\right) \\
&= P\left(\sum_{t=1}^{\infty} \frac{1}{(1+r)^t} \left(\sum_{i=1}^{N_t} X_{i,t} + \sum_{i=1}^{M_t} Y_{i,t} + \sigma W_t\right) \geq d\right) \\
&= P(d \leq V) \\
&= P(e^{sd} \leq e^{sV}) \leq \frac{\mathbb{E}[e^{sV}]}{e^{sd}} \tag{12}
\end{aligned}$$

ただし、 c, d は定数 $s > 0$

6. Tauber 型定理を用いたリスク評価方法 (参考)

確率変数 V の分布 F_V の裾評価のために、Tauber 型定理を用いる手法も検討した。

6.1. Tauber 型定理

原点の近傍の Laplace 変換の漸近的な性質と分布の $x \rightarrow \infty$ の漸近的性質が対応するという関係を述べたものが Tauber 型定理 (Tauberian Theorem) である。

定理 6.1 (Tauber 型定理) $f(x)$ を $\mathbb{R}$ 上の確率密度関数、 $l(x)$ を緩変動関数、 $0 \leq \alpha \leq 1$ とするとき、次は同値である。

$$\begin{aligned}
1 - \int_{-\infty}^{\infty} e^{-sx} f(x) dx &\sim s^{\alpha} l\left(\frac{1}{s}\right) \quad (s \downarrow 0) \\
1 - \int_{-\infty}^x f(y) dy &\sim \frac{l(x)}{x^{\alpha} \Gamma(1-\alpha)} \quad (x \rightarrow \infty)
\end{aligned}$$

ここで、 $\Gamma(z)$ はガンマ関数である。

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$$

注意 6.1 (緩変動関数) $[a, \infty)$ ($a \in \mathbb{R}$) 上で定義された正値可測関数 $l(x)$ が緩変動関数であるとは、任意の $\lambda > 0$ に対して、

$$\lim_{x \rightarrow \infty} \frac{l(\lambda x)}{l(x)} = 1$$

を満たすときを言う。

例 6.1 以下は緩変動関数の例である。

- $l(x) = c + o(1)$ ここで、 c は正の定数、 $\lim_{x \rightarrow \infty} o(1) = 0$
- $l(x) = \log x$
- $l(x) = (\log x)^2$

6.2. Laplace 変換の原点近傍における近似

原点近傍において、

$$1 - \mathcal{L}_V(s) \sim L_{\alpha}(s) := s^{\alpha} l\left(\frac{1}{s}\right) \tag{13}$$

緩変動関数として、 $l(x) = \log x$ を用いると、 $L(s)$ の原点近傍における挙動は図 3 のようになる。

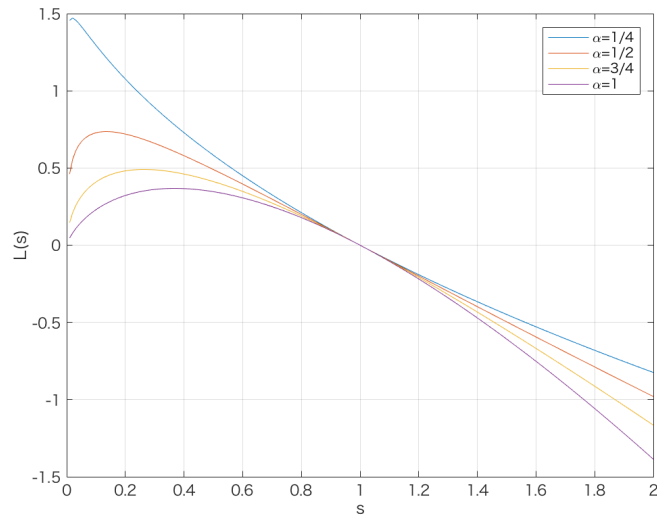


図 3 $L(s)$

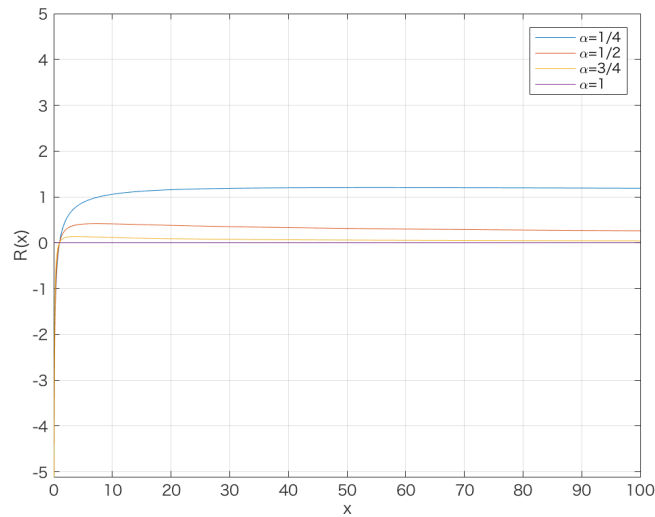


図 4 $R(x)$

6.3. 分布関数の裾評価

定理 4.1 より、原点近傍における Laplace 変換の近似が正しい場合、以下のように分布関数の裾を評価できる。

$$R_{\alpha}(x) := 1 - \int_{-\infty}^x f_V(y) dy \sim \frac{\log x}{x^{\alpha} \Gamma(1 - \alpha)} \quad (x \rightarrow \infty) \quad (14)$$

この $R(x)$ のグラフは図 4 のようになる。

7. ラプラス逆変換による分布関数の導出

確率変数 V のラプラス変換が分かっているので、ラプラス逆変換を用いることで、 V の分布関数を導出することもできると考えている。

$$\mathcal{L}_V^{-1} = \frac{1}{2\pi i} \lim_{T \rightarrow \infty} \int_{\gamma - iT}^{\gamma + iT} e^{st} \mathcal{L}_V(s) ds \quad (15)$$

8. 結論と今後の課題

これまでの研究で、企業価値のモデル化とリスク評価手法の提案に成功した。

モデルに関しては、マイナスの CF および不確実性を伴う損害を表す確率変数間に独立性を仮定しない等、現実をより反映したモデルの構築を試みる余地がありそうである。

リスク評価に関しては、パラメータを適当に設定した場合のリスクについて、数値計算により評価することが必要である。また、一度試みた Tauber 型定理を利用したリスク評価手法については、Laplace 変換の原点近傍の近似が示せなかったため断念したが、再検討してみたいと考えている。

参考文献

- [1] Tomasz Rolski, Hanspeter Schmidli, Volker Schmidt, Jozef Teugels(1998).”Stochastic processes for insurance and finance”.JOHN WILEY SONS.147-204
- [2] Dembo,A and Ofer,Z(1998). ”Large Deviation Techniques and Applications”. Springer-Verlag. Applications of Mathematics. 396.
- [3] Yasutaka Shimizu(2010).”Gerber-Shiu Function in Risk Theory and Statistical Inference”.Proceedings of the Institute of Statistical Mathematics Vol. 59, No. 1, 105-124
- [4] Cohen, A. M. (2007). ”Inversion Formulae and Practical Results”. Numerical Methods for Laplace Transform Inversion. Numerical Methods and Algorithms. 5. p. 23.
- [5] N.H.Bingham, C.M.Goldie and J.L.Teugels(1987).”Regular Variation”.Cambridge University Press

サーバブロックの起動時間を持つデータセンターの待ち行列モデルによる性能解析

日出山 慎人
筑波大学 理工学群 社会工学類

Phung-Duc Tuan
筑波大学 システム情報系

概要：本稿では、有限バッファを持つデータセンターにおける利用客のジョブの到着・サービスの様子を待ち行列モデル化し、その平均待ち時間・サーバ稼働コストに焦点を当てて考察を進める。データセンターは所有するサーバのうち一部を常に稼働させておき、それ以外は利用客の滞在状況に合わせて起動またはシャットダウンを行なうことで稼働数を調整する。サーバの起動・シャットダウンはブロック単位で行い、サーバブロックを起動するためには起動時間が必要とする。1 ブロックに含まれるサーバ数、そして1 ブロックを起動・シャットダウンするための条件となる待ち人数を変数化し、定常分布からコストを導出し、性能解析を行う。

1. はじめに

データセンターは一般に多くのサーバを保有し、利用客の発生に備えて稼働させておく。全てのサーバが常に利用されている訳ではなく、稼働している殆どのサーバが誰にも利用されていない状況（以下、利用されていない状態を idle 状態と呼ぶ）となることも起こり得る。また、idle 状態のサーバは誰にも利用されていない状況であるのも関わらず電力等のコストを消費し続ける。先行研究 [1] では、idle 状態のサーバが消費するコストはピーク時の約 60% にも及ぶ先例があった。したがって、データセンターは所有する全てのサーバを常に稼働させておくことは得策とは言えず、所有するサーバのうち一部を常に稼働させておき、それ以外のサーバを利用客の滞在状況に合わせて起動・シャットダウンを繰り返して利用することが好ましいと考えられる。この時、常に稼働させておくサーバが多い状態は多くの利用客への対応が可能になる利点に対して消費コストが増大となる欠点を持ち、起動・シャットダウンを繰り返すサーバが多い状態はコストを抑えることができる利点に対して多くの利用客に対応しづらくなる欠点を持つ。

以上より、データセンターにおいて常に稼働させるサーバと起動・シャットダウンを繰り返すサーバの比率決定は重要な課題といえる。先行研究より、コールセンターやデータセンターにおける待ち行列モデルはマルコフ連鎖で定式化することで定常状態確率やコストを計算できる [2–4]。特にデータセンターのシステムに関して、[3] では待ち人数に合わせてサーバを 1 つずつ起動・シャットダウンを繰り返すことでそれぞれ利用客の待ち時間・サーバ稼働コストが変動することが示された。この中では、待ち人数が n 人 ($n \in \mathbb{N}$) のときサーバを n 個起動し、idle 状態のサーバが発生するとすぐにシャットダウンしている。また、[4] では残りのサーバを全てまとめてブロック単位で起動・シャットダウンを行うことにより、一定条件下で稼働コストが減少することが示された。この中では、待ち人数が一定以上になるとサーバブロックを起動し、一定以下になるとサーバブロックをシャットダウンしている。

このように、所有するサーバのうち一部を常に稼働させておくサーバ、それ以外を起動・シャットダウンを繰り返すサーバに分けて使用することでコストを抑えることができる。本研究では起動・シャットダウンを繰り返すサーバをいくつかのブロックに分け、待ち人数が mn 人 ($n \in \mathbb{N}$) のときサーバブロックを n 個起動するような変数 m 、および 1 ブロックに含まれるサーバ数 k を変数としておき、起動条件の拡張及び本研究の中で定義されたコストを数値計算し、考察することを目的とする。

本論文の構成は次の通りである。第 2 章ではデータセンターにおける利用客のジョブの到着およびサービスのシステムを設定し、モデル化する。第 3 章では平衡方程式から定常状態確率を計算し、コストを定義する。第 4 章では計算機を用いてコストの数値計算を行う。第 5 章では結論を述べる。

2. データセンターの待ち行列モデル化

本節では、データセンターを待ち行列モデル化するために必要な内容をまとめる。

2.1. システムの設定

待ち行列モデル化するにあたり、利用客・サーバに関するルールを定める。データセンターは全部で $N(= n_0 + c)$ 個のサーバを保有し、そのうち n_0 個を常に稼働させ、 c 個をいくつかのブロックに分

け、利用客の滞在状況に合わせてブロック単位で起動・シャットダウンを繰り返す。また、サービスにおいて利用客1人につき稼働しているサーバ1個を要するものとする。

2.1.1. 利用客の到着とサービス

利用客は独立にデータセンターへアクセスし、idle 状態のサーバがある場合はサービスを受け、idle 状態のサーバがない場合は待ち行列に並び、サービスが受けられるまで待つ。この時、一度並んだ利用客はサービスを受けるまで離れないものとする。また、サービスの提供ルールには FCFS(First Come, First Served) を適用する。また、系内が満員の状態で利用客が到着する場合は列に並ぶことなく離れていくものとする。サービスを受け終えた利用客はそのままシステムから離れる。

利用客はパラメータ λ のポアソン過程に従って到着し、サーバ1個あたりのサービス時間はパラメータ μ の指数分布に従うものとする。

2.1.2. サーバの種類と状態

常に稼働させておくサーバを ON サーバ、起動・シャットダウンを繰り返すサーバを ON-OFF サーバと呼ぶ。ON-OFF サーバが k 個または l 個で1つのセットとしてまとめられているものを ON-OFF サーバブロック（または単にサーバブロック）と呼ぶ。また、サーバが稼働・起動していない状態を off 状態、起動中の状態を setup 状態、起動済みの状態を on 状態、利用されている状態を used 状態、利用されていない状態を idle 状態と呼ぶ。on 状態と used 状態、on 状態と idle 状態は重複できるものとする。

2.1.3. サーバブロックの起動とシャットダウン

待ち行列に並ぶ人数が mn 人 ($n \in \mathbb{N}$) に達した時、新たに n 番目のサーバブロックの起動を開始する。サーバブロックの起動には時間がかかるものとする。また、サーバブロック起動中に待ち人数が mn 人を下回った場合は n 番目のサーバブロックの起動を中断する。

n 個のサーバブロックが起動済みの状態で n 番目のサーバブロックに含まれるサーバが全て idle 状態になった時、 n 番目のサーバブロックをシャットダウンする。シャットダウンに要する時間はここでは無視できるものとする。

サーバブロック1つあたりの起動時間はパラメータ α の指数分布に従うものとする。サーバブロックの起動時間が終わるとそのブロックに含まれる k 個（または l 個）のサーバが全て on 状態になる。

2.2. モデル図の作成

以上の設定を用いてモデル図を作成する。

2.2.1. 待ち行列モデル図

先述したシステムより、今回のデータセンターの待ち行列モデルは図1のようになる。

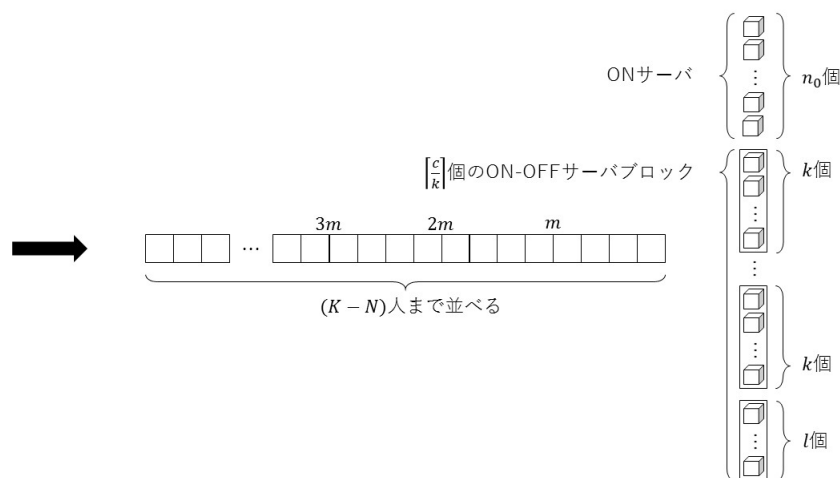


図 1: 待ち行列モデル図.

2.2.2. 状態遷移図

状態 (i, j) を, on 状態のサーバ数が i 個, 滞在中の利用客が j 人の状態とすると, 状態遷移図の例として図 2 のような状態空間になることがわかる.

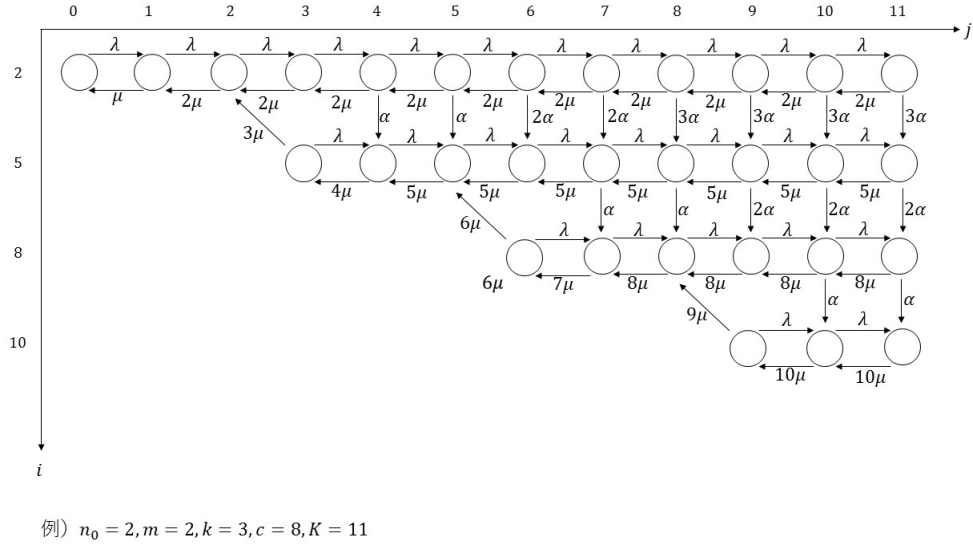


図 2: 状態遷移図の例.

このとき, i, j はそれぞれ

$$i \in I = \{n_0, n_0 + k, \dots, n_0 + \left\lceil \frac{c}{k} - 2 \right\rceil k, n_0 + \left\lceil \frac{c}{k} - 1 \right\rceil k, n_0 + c\},$$

$$j \in J = \{0, 1, \dots, K\},$$

の範囲で値をとり, 状態空間 S を

$$S = \{(i, j) | i \in I, j \in J\},$$

で定義する.

3. 値の計算

本節では, 先程設定した待ち行列モデルにおける平衡方程式から定常状態確率を導出し, 平均待ち時間, 稼働サーバ数からなるコスト式を定義する.

3.1. パラメータの説明

まず, 今回用いる変数を次のようにまとめる.

3.1.1. 状態に依存しない変数

表 1 における変数を状態 (i, j) に依存しない変数として定義する.

ON-OFF サーバが c 個存在し, 1 ブロックに含まれるサーバ数を k 個とすると, 全部で $\left\lceil \frac{c}{k} \right\rceil$ 個のサーバブロックが作成されるものとする. このとき, $\left\lceil \frac{c}{k} - 1 \right\rceil$ 番目までのサーバブロックはそれぞれ k 個のサーバを有し, $\left\lceil \frac{c}{k} \right\rceil$ 番目のサーバブロックは $l (= c - \left\lceil \frac{c}{k} - 1 \right\rceil k)$ 個のサーバを有する.

また, 余分なサーバは準備されないものとするため, $n_0 + \left\lceil \frac{c}{k} - 1 \right\rceil k + m \leq K$ を満たすような c を条件とする.

3.1.2. 状態に依存する変数.

表 2 における変数を状態 (i, j) に依存する変数として定義する.

表 1: 状態に依存しない変数.

変数	説明
λ	利用客の到着パラメータ
μ	サーバ 1 個あたりのサービスパラメータ
α	1 ブロックあたりの起動パラメータ
n_0	常に稼働させておくサーバ数 (ON サーバ数)
c	起動・シャットダウンを繰り返すサーバ数 (ON-OFF サーバ数)
N	全サーバ数 ($= n_0 + c$)
K	系内サイズ
m	新たに 1 ブロックを起動するための待ち行列人数
k	1 ブロックに含まれるサーバ数
l	最後の 1 ブロックに含まれるサーバ数 ($= c - \lceil \frac{c}{k} - 1 \rceil k$)

表 2: 状態に依存する変数

変数	説明
$\mu_{i,j}$	状態 (i, j) におけるサービス率
$\alpha_{i,j}$	状態 (i, j) における起動率

それぞれ

$$\mu_{i,j} = \min(i\mu, j\mu)$$

$$\alpha_{i,j} = \begin{cases} 0, & i \geq j, \\ \min\left(\left\lfloor \frac{j-i}{m} \right\rfloor \alpha, \left\lceil \frac{N-i}{k} \right\rceil \alpha\right), & i < j. \end{cases}$$

で表すことができる.

3.2. 定常状態確率の計算

続いて定常状態確率を計算する. 状態 (i, j) の推移はマルコフ連鎖となっており, マルコフ連鎖の状態空間が有限かつ既約であることから定常分布が一意に定まり, 平衡方程式を満たす定常状態確率を導出する.

以下, 状態 (i, j) における定常状態確率を $\pi_{i,j}$ で表す.

3.2.1. $i = n_0$ の時

$\pi_{n_0,j}$ が定常状態確率のとき, 以下の平衡方程式が満たされる.

$$\lambda\pi_{n_0,0} = \mu\pi_{n_0,1}, \quad j = 0, \quad (1)$$

$$(\lambda + \mu_{n_0,j})\pi_{n_0,j} = j\mu\pi_{n_0,j+1} + \lambda\pi_{n_0,j-1}, \quad 0 < j < n_0, \quad (2)$$

$$(\lambda + n_0\mu)\pi_{n_0,n_0} = n_0\mu\pi_{n_0,n_0+1} + \lambda\pi_{n_0,n_0-1} + (n_0 + 1)\mu\pi_{\min(n_0+k,N),n_0+1}, \quad j = n_0, \quad (3)$$

$$(\lambda + n_0\mu + \alpha_{n_0,j})\pi_{n_0,j} = n_0\mu\pi_{n_0,j+1} + \lambda\pi_{n_0,j-1}, \quad n_0 < j < K, \quad (4)$$

$$(n_0\mu + \alpha_{n_0,K})\pi_{n_0,K} = \lambda\pi_{n_0,K-1}, \quad j = K. \quad (5)$$

ただし, $n_0 = 0$ の場合は (1), (2) 及び (3) の代わりに

$$\lambda\pi_{0,0} = \mu\pi_{\min(k,N),1}, \quad j = 0,$$

を満たす.

さらに, (1), (2), (4) 及び (5) から次の数式を得る.

$$\pi_{n_0,j} = \frac{1}{j!} \left(\frac{\lambda}{\mu}\right)^j \pi_{n_0,0}, \quad 0 < j \leq n_0,$$

$$\pi_{n_0,j} = b_{n_0,j} \pi_{n_0,j-1}, \quad n_0 < j \leq K,$$

ただし,

$$b_{n_0,j} = \begin{cases} \frac{\lambda}{\lambda + \mu_{n_0,j} + \alpha_{n_0,j} - n_0 \mu b_{n_0,j+1}}, & n_0 < j < K, \\ \frac{\lambda}{n_0 \mu + \alpha_{n_0,K}}, & j = K. \end{cases}$$

3.2.2. $i = n_0 + nk$ ($n \in \{1, 2, \dots, \lceil \frac{c}{k} - 1 \rceil\}$) の時

on 状態のサーバが i 個未満の状態集合の中へ入る量と外へ出る量はつり合うため, 次の方程式を得る.

$$(i - k + 1)\mu\pi_{i,i-k+1} = \sum_{j=i-k+1}^K \alpha_{i-k,j}\pi_{i-k,j}.$$

さらに $\pi_{i,j}$ が定常状態確率のとき, 以下の平衡方程式が満たされる.

$$(\lambda + (i - k + 1)\mu)\pi_{i,i-k+1} = (i - k + 2)\mu\pi_{i,i-k+2}, \quad j = i - k + 1, \quad (6)$$

$$(\lambda + j\mu)\pi_{i,j} = (j + 1)\mu\pi_{i,j+1} + \lambda\pi_{i,j-1} + \alpha_{i-k,j}\pi_{i-k,j}, \quad i - k + 1 < j < i, \quad (7)$$

$$(\lambda + i\mu)\pi_{i,i} = (i + 1)\mu\pi_{\min(i+k,N),i+1} + i\mu\pi_{i,i+1} + \lambda\pi_{i,i-1} + \alpha_{i-k,i}\pi_{i-k,i}, \quad j = i, \quad (8)$$

$$(\lambda + i\mu + \alpha_{i,j})\pi_{i,j} = i\mu\pi_{i,j+1} + \lambda\pi_{i,j-1} + \alpha_{i-k,j}\pi_{i-k,j}, \quad i < j < K, \quad (9)$$

$$(i\mu + \alpha_{i,K})\pi_{i,K} = \lambda\pi_{i,K-1} + \alpha_{i-k,K}\pi_{i-k,K}, \quad j = K. \quad (10)$$

ただし, $k = 1$ の場合は (6), (7) 及び (8) の代わりに

$$(\lambda + i\mu)\pi_{i,i} = (i + 1)\mu\pi_{\min(i+k,N),i+1} + i\mu\pi_{i,i+1} + \alpha_{i-k,i}\pi_{i-k,i}, \quad j = i,$$

を満たす.

さらに, (6), (7), (9) 及び (10) から次の数式を得る.

$$\begin{aligned} \pi_{i,j} &= a_{i,j} + b_{i,j}\pi_{i,j-1} + c_{i,j}\pi_{i,j-2}, & i - k + 2 \leq j \leq i, \\ \pi_{i,j} &= a_{i,j} + b_{i,j}\pi_{i,j-1}, & i < j \leq K, \end{aligned}$$

ただし,

$$\begin{aligned} a_{i,j} &= \begin{cases} -\frac{\alpha_{i-k,j-1}\pi_{i-k,j-1}}{\mu_{i,j}}, & i - k + 2 \leq j \leq i, \\ \frac{\mu_{i,j+1}a_{i,j+1} + \alpha_{i-k,j}\pi_{i-k,j}}{\lambda + \mu_{i,j} + \alpha_{i,j} - \mu_{i,j+1}b_{i,j+1}}, & i < j < K, \\ \frac{\alpha_{i-k,K}\pi_{i-k,K}}{\mu_{i,K} + \alpha_{i,K}}, & j = K, \end{cases} \\ b_{i,j} &= \begin{cases} \frac{\lambda + \mu_{i,j-1}}{\mu_{i,j}}, & i - k + 2 \leq j \leq i, \\ \frac{\lambda}{\lambda + \mu_{i,j} + \alpha_{i,j} - \mu_{i,j+1}b_{i,j+1}}, & i < j < K, \\ \frac{\lambda}{\mu_{i,K} + \alpha_{i,K}}, & j = K, \end{cases} \\ c_{i,j} &= \begin{cases} 0, & j = i - k + 2, \\ -\frac{\lambda}{\mu_{i,j}}, & i - k + 2 < j \leq i. \end{cases} \end{aligned}$$

3.2.3. $i = N$ ($= n_o + c$) の時

on 状態のサーバが N 個未満の状態集合の中へ入る量と外へ出る量はつり合うため, 次の方程式を得る.

$$(N - l + 1)\mu\pi_{N,N-l+1} = \sum_{j=N-l+1}^K \alpha_{N-l,j}\pi_{N-l,j}.$$

さらに $\pi_{i,j}$ が定常状態確率のとき, 以下の平衡方程式が満たされる.

$$(\lambda + (N - l + 1)\mu)\pi_{N,N-l+1} = \mu_{N,N-l+2}\pi_{N,N-l+2} + \alpha_{N-l,N-l+1}\pi_{N-l,N-l+1}, \quad j = N - l + 1,$$

$$(\lambda + \mu_{N,j})\pi_{N,j} = \mu_{N,j+1}\pi_{N,j+1} + \lambda\pi_{N,j-1} + \alpha_{N-l,j}\pi_{N-l,j}, \quad N - l + 1 < j < K, \quad (11)$$

$$\mu_{N,K}\pi_{N,K} = \lambda\pi_{N,K-1} + \alpha_{N-l,K}\pi_{N-l,K}, \quad j = K. \quad (12)$$

さらに, (11), (12) から次の数式を得る.

$$\pi_{N,j} = a_{N,j} + b_{N,j}\pi_{N,j-1}, \quad N-l+2 \leq j \leq K,$$

ただし,

$$a_{N,j} = \begin{cases} \frac{\mu_{N,j+1}a_{N,j+1} + \alpha_{N-l,j}\pi_{N-l,j}}{\lambda + \mu_{N,j} - \mu_{N,j+1}b_{N,j+1}}, & N-l+2 \leq j < K, \\ \frac{\alpha_{N-l,K}\pi_{N-l,K}}{\mu_{N,K}}, & j = K, \end{cases}$$

$$b_{N,j} = \begin{cases} \frac{\lambda}{\lambda + \mu_{N,j} - \mu_{N,j+1}b_{N,j+1}}, & N-l+2 \leq j < K, \\ \frac{\lambda}{\mu_{N,K}}, & j = K. \end{cases}$$

3.3. コストの計算

先ほど計算した定常状態確率を用いてコストを定義, 導出する.

3.3.1. 平均待ち時間

まず, 平均待ち時間 W_q を導出する. 系内人数を L とし, 期待値を $E[\cdot]$ で表すものとする, $E[L]$ は次のように計算できる.

$$E[L] = \sum_{(i,j) \in S} j\pi_{i,j}.$$

また, 系内が満員となる定常状態確率を P_b とすると, P_b は次のように計算できる.

$$P_b = \sum_{i \in I} \pi_{i,K}.$$

ここで系内における利用客の滞在時間を W とすると, リトルの公式を用いて $E[W]$ は次のように計算できる.

$$E[W] = \frac{E[L]}{\lambda(1 - P_b)}.$$

サービス時間は平均 $\frac{1}{\mu}$ なので, 系内における利用客の平均待ち時間 W_q は次のように計算できる.

$$W_q = E[W] - \frac{1}{\mu}.$$

3.3.2. サーバ稼働コスト

続いてサーバ稼働コスト C を導出する. used 状態のサーバ数を C_u , idle 状態のサーバ数を C_i , setup 状態のサーバ数を C_s とすると, それぞれ次のようになる.

$$C_u = \sum_{(i,j) \in S} \min(i, j)\pi_{i,j},$$

$$C_i = \sum_{(i,j) \in S} \max(i - j, 0)\pi_{i,j},$$

$$C_s = \sum_{i \in I \setminus \{N\}} \sum_{j=i}^K \min\left(\left\lfloor \frac{j-i}{m} \right\rfloor k, \left(\left\lceil \frac{N-i}{k} \right\rceil - 1\right)k + l\right) \pi_{i,j}.$$

今回は used 状態, setup 状態のサーバ消費コストを 1.0, idle 状態のサーバ消費コストを 0.6 とし, サーバ消費コスト C を以下のように定義する.

$$C = C_u + 0.6C_i + C_s.$$

3.3.3. 統合コスト

先ほど導出した W_q , C に重みづけをして統合したコスト P を以下のように定義する.

$$P = w_1 W_q + w_2 C.$$

今回は W_q, C, P の数値計算を行い, 考察を行う.

4. 数値計算と考察

本節では数値計算の結果および考察内容をまとめる．今回の数値計算には Python を利用した．今回は λ, k, α の値を変化させたときに各コストがどのような変化をするか観察する．また，各変数に関して指定がない限りは $m = 1$, $k = 1$, $\lambda = 150$, $\mu = 1.0$, $\alpha = 0.5$, $n_0 = 100$, $c = 100$, $N = 200$, $K = 250$ として検証する．

4.1. 数値計算

4.1.1. W_q の計算

λ, k, α を変化させて W_q を計算した結果，図 3 のようになった．

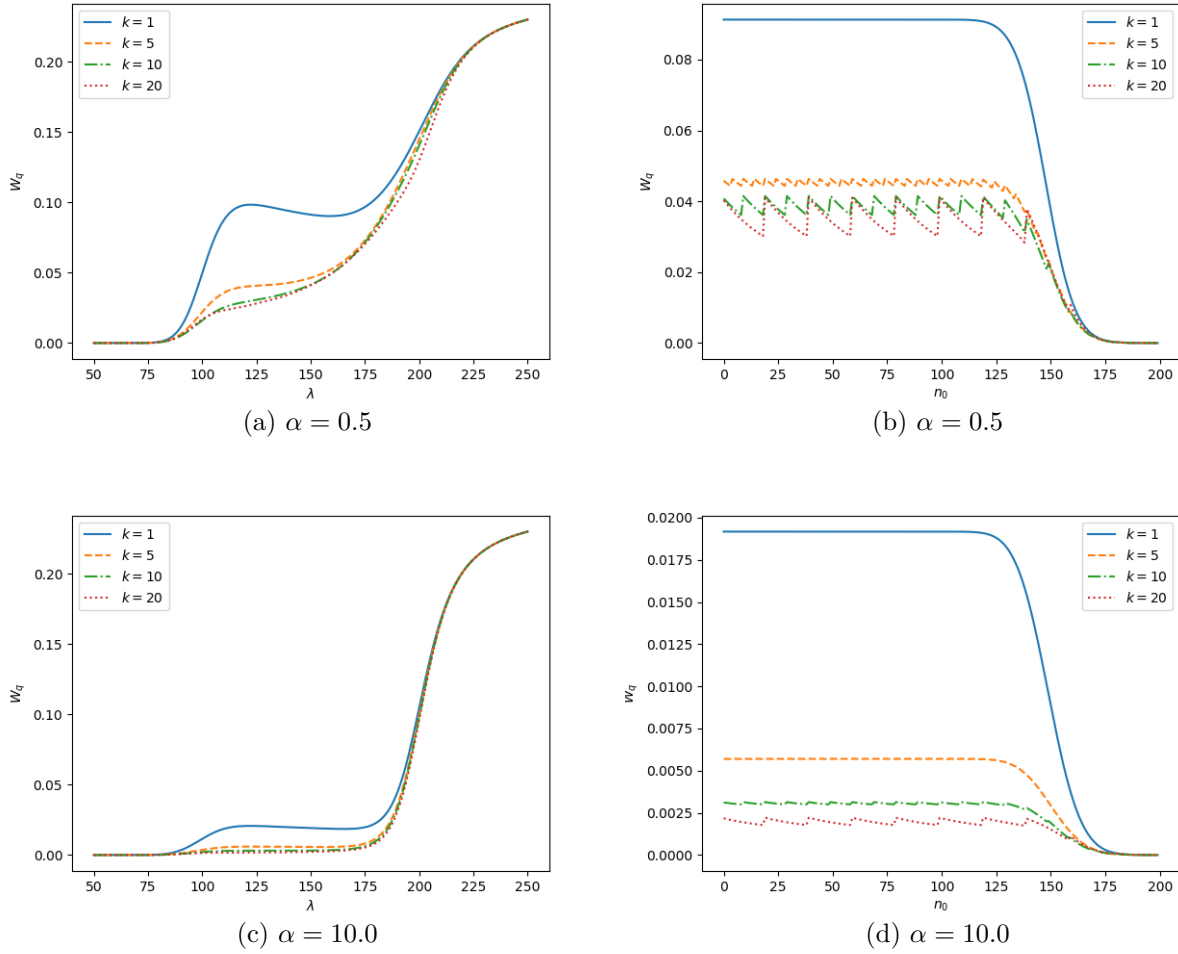


図 3: W_q の数値計算結果．

この結果から以下のことが分かった．

- λ の変化に対して，サーバブロックの起動率 α が小さい場合， λ が $n_0\mu$ に近づいた段階で待ち時間が発生していく．
- λ の変化に対して，サーバブロックの起動率 α が大きい場合， k が大きいほど $n_0\mu < \lambda < N$ における待ち時間 W_q を 0 に近づけることができる．
- λ の変化に対して，サーバブロックの起動率 α によらず， λ が一定以下・一定以上になると待ち時間 W_q が k に依存しなくなる．
- n_0 の変化に対して， n_0 が k の倍数になったタイミングで待ち時間 W_q のジャンプが発生する．
- n_0 の変化に対して，サーバブロックの起動率 α が大きい場合，待ち時間 W_q のジャンプによる変化量が小さくなる．

- n_0 の変化に対して、サーバブロックの起動率 α によらず、 n_0 が一定以上になると待ち時間 W_q が k に依存しなくなる。

λ が一定以下の時は n_0 個の ON サーバで利用客をすべて対応できている、一定以上の時はすべてのサーバブロックが常に on 状態になっている状態を示すため、 k による待ち時間の変化がないものと考えられる。ジャンプが複数回起きる事態は、 n_0 の増加によって起動が必要となるサーバブロック数に変化が現れたタイミングで発生しているものと思われる。また、起動率が大きくなるとサーバブロックの起動がスムーズに行われるようになり、待ち時間が短縮されているものと考えられる。

4.1.2. C の計算

λ, k, α を変化させて C を計算した結果、図 4 のようになった。

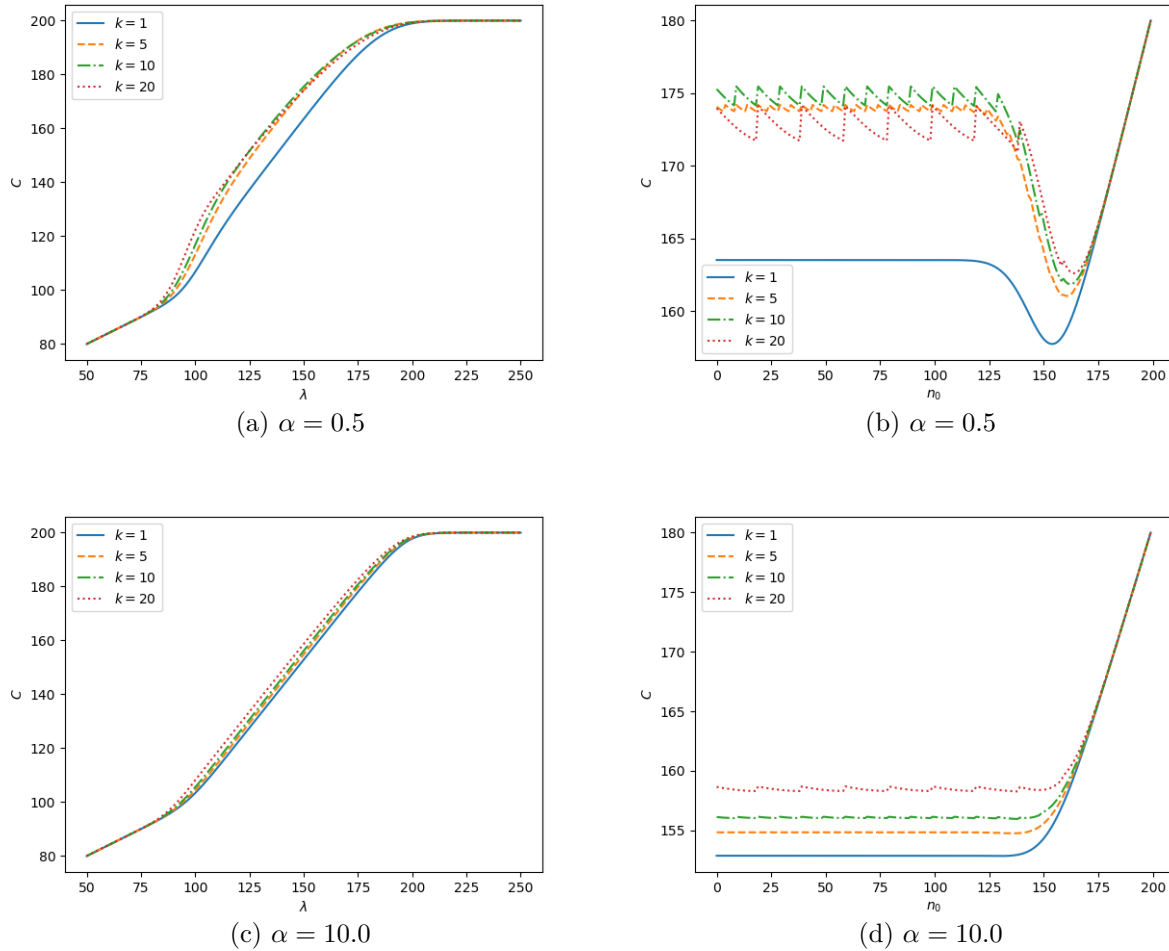


図 4: C の数値計算結果.

この結果から以下のことが分かった。

- λ の変化に対して、サーバブロックの起動率 α が小さい場合、 k が小さいほど消費コスト C は少なく、 k の違いによって消費コストにも明らかな差がある。
- λ の変化に対して、サーバブロックの起動率 α が大きい場合、 k が小さいほど消費コスト C は少ないが、 k の違いによる消費コストの差が小さくなっている。
- λ の変化に対して、サーバブロックの起動率 α によらず、 λ が一定以下・一定以上になると消費コスト C が k に依存しなくなる。
- n_0 の変化に対して、 n_0 が k の倍数になったタイミングで消費コスト C のジャンプが発生する。
- n_0 の変化に対して、サーバブロックの起動率 α が大きい場合、消費コスト C が小さくなる。また、 C のジャンプによる変化量が小さくなる。

サーバ稼働コスト C に関しても、先程の待ち時間 W_q に関する考察と同じ理由で説明できると考えられる。

4.1.3. P の計算

例として $w_1 = 100.0, w_2 = 0.5$ とし, λ, k, α を変化させて P を計算した結果, 図 5 のようになった。

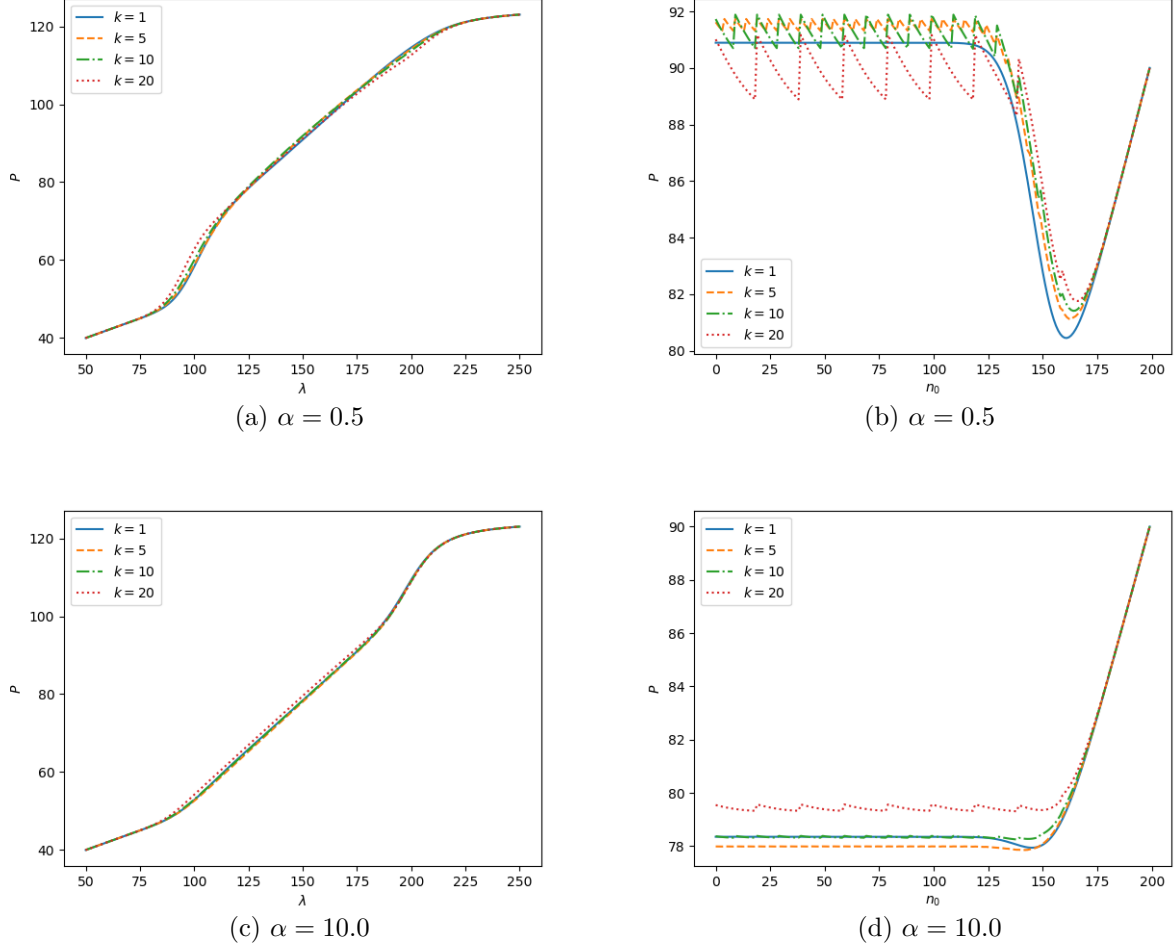


図 5: P の数値計算結果.

この結果から以下のことが分かった。

- λ の変化に対して, サーバブロックの起動率 α によらず, k による P の変化量が小さい。
- λ の変化に対して, サーバブロックの起動率 α が小さい場合に対して α が大きい場合のほうが $n_0 \leq \lambda \leq N$ のあたりで P が小さくなる。
- n_0 の変化に対して, n_0 が k の倍数になったタイミングで P のジャンプが発生する。
- n_0 の変化に対して, サーバブロックの起動率 α が小さい場合, n_0 の変化とともに P が最小となる k が変化している。

待ち時間, サーバ稼働コストに重みづけを行い足し合わせたものが P であるため, 今回もジャンプが複数回発生している。また, w_1, w_2 の値によって待ち時間重視型, 稼働コスト重視型も考えることができる。

4.2. 考察

先程の数値計算を踏まえて考察を行う。

まず, ブロック内サーバ数 k , サーバブロックの起動率 α によってコストが大きく変動した。その中でサーバブロックの起動率 α が k によるコスト変化量に影響を与えていることが分かった。これにより, 複数のサーバブロックを利用するシステムにおいてブロック内のサーバ数等を検討する際には

ブロック起動率にも重点を置くべきと考えられる．今回設定した統合コストでは，起動率が小さい場合にはサーバブロックに含まれるサーバ数を増やして一気に起動することでコストを抑えることができ，起動率が大きい場合にはサーバブロックに含まれるサーバ数を少数にしてこまめに起動することでコストを抑えることができた．しかし今回用いた w_1, w_2 の値はあくまで一例であり，データセンターに応用する場合には待ち時間あたりの損失及びサーバ稼働コストを考慮し，慎重に設定しなければならない．

5. 終わりに

数値計算の結果より，パラメータを設定することでコストの数値計算が可能になることがわかった．また，設定する変数値によってコスト W_q, C, P のグラフが変化し，状況によって1つずつ起動するよりも最適なコスト値をとるようなブロック内サーバ数 k が存在することがわかった．すなわち，データセンターにおける利用客の到着率，サーバのサービス率，サーバブロックの起動率等の数値を測定・適用することによって平均待ち時間・サーバ稼働コストを数値計算し，所有するサーバの運用方法を決める手がかりの一つとして参照することができる．

参考文献

- [1] Barroso, L. A., & Holzle, U. (2007). The case for energy-proportional computing. *Computer*, 40(12).
- [2] 川合洋輔, 高木英明, オペレータの後処理と客の待合せ放棄があるコールセンターの簡易待ち行列モデル, 2011 年度待ち行列シンポジウム「確率モデルとその応用」報文集, 188-197.
- [3] Ren, Y., Phung-Duc, T., Yu, Z. W., & Chen, J. C. (2016). Design and analysis of Dynamic Auto Scaling Algorithm (DASA) for 5G mobile networks. *arXiv preprint arXiv:1604.05803*.
- [4] Mitrani, I. (2013). Managing performance and power consumption in a server farm. *Annals of Operations Research*, 202(1), 121-134.

マルコフ型集団到着過程と裾の軽いサービス分布を持つ 無限サーバ待ち行列の安定性の十分条件

矢島 萌子¹ フンドック トゥアン² 増山 博之³

¹ 東京工業大学 ² 筑波大学 ³ 京都大学

概要：本論文では、裾の軽い分布サービス時間分布を持つ $\text{BMAP/G}/\infty$ 待ち行列の安定性について議論する。裾の軽いサービス分布を持つ $\text{BMAP/G}/\infty$ 待ち行列は、集団の大きさの対数期待値が有界であるとき、安定であることを示す。

1. はじめに

本論文では、客のサービス時間が裾の軽い分布 (light-tailed distribution) [5] に従う $\text{BMAP/G}/\infty$ 待ち行列を考える。 $\text{BMAP/G}/\infty$ は、マルコフ型集団到着過程 (batch Markovian arrival process: BMAP) [9] に従って客が到着し、客のサービス時間は独立で同一の一般分布に従う。BMAP は、集団ポアソン到着過程、相型再生到着過程 [8]、MAP (Markovian arrival process) [10] など多くの到着過程を特殊なケースとして持つ。特に、MAP は BMAP の客が単一で到着するような特殊ケースであり、任意の単純な点過程はある MAP の列の弱収束先と確率的に等価であることが知られている [1]。さらに、相型分布の集合は $[0, \infty)$ 上の分布の集合において稠密であり [1]、任意の相型分布は裾の軽い分布である。よって、本論文が考える裾の軽いサービス分布を持つ $\text{BMAP/G}/\infty$ 待ち行列は、非常に一般的な集団到着型無限サーバ待ち行列である。

無限サーバ待ち行列は、在庫管理 [2]、交通流 [17]、情報通信 [11] などの様々な分野において、システムをモデル化する際に用いられる。そのため、定常状態や過渡状態における無限サーバ待ち行列の解析は、多くの研究者によって盛んに行われている [3, 4, 6, 7, 12–15]。

本論文では、裾の軽いサービス分布を持つ $\text{BMAP/G}/\infty$ 待ち行列の安定条件の導出に向け、安定であるための十分条件が LBSM 条件であることを示す。ここで、待ち行列が安定であるとは、客数過程が唯一つの定常分布を持つことをいう。無限サーバ待ち行列は、客が一人ずつ到着する場合には常に安定であるが、集団で到着する場合には安定とは限らない。そのため、集団到着無限サーバ待ち行列の定常解析を行う場合には、安定条件 (待ち行列が安定であるための必要十分条件) に関して議論を行う必要がある。

しかしながら、集団到着無限サーバ待ち行列の定常解析に関するほとんどの先行研究では、安定条件に関する厳密な議論は行われていない。例えば、Holman ら [6] は、 $\text{M}^X/\text{G}/\infty$ 待ち行列の集団の大きさの 2 次モーメントが有限であることを仮定して、定常系内客数分布の平均と分散を導いた。また、Masuyama と Takine [12] は、多重クラス $\text{BMAP/G}/\infty$ 待ち行列の定常系内客数分布の存在を仮定して、定常系内客数のモーメントに関するいくつかの公式を導いた。

集団到着無限サーバ待ち行列の安定条件について議論を行っている先行研究は、我々の知る限りでは [4, 19, 20] のみである。Cong [4] は、 $\text{M}^X/\text{M}/\infty$ 待ち行列の安定条件が、集団の大きさの対数期待値が有界であることだと示した。これ以降、集団の大きさの対数期待値が有界であることを、LBSM 条件 (logarithmic batch-size moment condition) と呼ぶ。Yajima ら [19] は、[4] の結果を到着過程が BMAP の場合に拡張し、 $\text{BMAP/M}/\infty$ 待ち行列の安定条件も、LBSM 条件であると示した。さらに、Yajima ら [20] により、 $\text{BMAP/PH}/\infty$ 待ち行列の安定条件も LBSM 条件であることが示された。

本論文では、裾の軽いサービス分布を持つ $\text{BMAP/G}/\infty$ 待ち行列が安定であるための十分条件も、LBSM 条件であることを示す。証明の際には、客数過程が再生型確率過程 (regenerative process) [18] の構造を持つことに着目する。再生理論における諸定理を利用する。

本論文の構成は以下の通りである。第 2 節では、本論文で考える裾の軽いサービス分布を持つ $\text{BMAP/G}/\infty$ 待ち行列を詳細に定義し、この待ち行列の客数過程が再生型確率過程の構造を持つことを説明する。第 3 節では、第 2.1 節で定義した待ち行列の安定性を議論するために新たに直列型待ち行列を構築し、元の待ち行列と新たに構築した直列型待ち行列の安定性の関係を示す。第 4 節では、

裾の軽いサービス分布を持つ **BMAP/G/∞** 待ち行列の安定性の十分条件が **LBSM** 条件であることを示す。最後に第 5 節にて、本論文のまとめと今後の課題を述べる。

2. 裾の軽いサービス分布を持つ **BMAP/G/∞** 待ち行列

本節ではまず、第 2.1 節にて、裾の軽いサービス分布を持つ **BMAP/G/∞** 待ち行列について詳細に定義する。その後、第 2.2 節にて、このモデルの客数過程が再生型確率過程の構造を持つことを述べる。

2.1. モデルの説明

システムには無限個のサーバが存在し、客は **BMAP** に従ってシステムに到着する。**BMAP** は既約な連続時間マルコフ連鎖 $\{J(t); t \geq 0\}$ に支配される集団到着過程である。マルコフ連鎖 $\{J(t)\}$ は有界な状態空間 $\mathbb{D} = \{1, 2, \dots, d\}$ を持つ。これ以降、 $\{J(t)\}$ を背後マルコフ連鎖と呼ぶ。 $N(t), t \geq 0$, を時間 $(0, t]$ に **BMAP** から到着する客数とする。ただし、 $N(t) = 0$ 。任意の $k \in \mathbb{Z}_+ := \{0, 1, \dots\}$ と $i, j \in \mathbb{D}$ に対して、次式が成立する。

$$\begin{aligned} & P(N(t + \Delta t) - N(t) = k, J(t + \Delta t) = j \mid J(t) = i) \\ &= \begin{cases} 1 + D_{i,i}(0)\Delta t + o(\Delta t), & k = 0, j = i, \\ D_{i,j}(k)\Delta t + o(\Delta t), & \text{その他.} \end{cases} \end{aligned} \quad (2.1)$$

ただし、 $f(x) = o(g(x))$ は $\lim_{x \downarrow 0} |f(x)|/|g(x)| = 0$ を意味する。(2.1) 式にて、 $\mathbf{D}(k) := (D_{i,j}(k))_{i,j \in \mathbb{D}}$, $k \in \mathbb{N} := \{1, 2, \dots\}$, は非負行列, $\mathbf{D}(0) := (D_{i,j}(0))_{i,j \in \mathbb{D}}$ は負の対角成分と非負の非対角成分をもつ優対角行列である。さらに、 $\mathbf{D} := \sum_{k \in \mathbb{Z}_+} \mathbf{D}(k)$ は背後マルコフ連鎖 $\{J(t)\}$ の既約な無限小生成行列である。ここで、以下の式が成立することを仮定する。

$$\sum_{n=1}^{\infty} \mathbf{D}(n) \mathbf{e} \neq \mathbf{0}.$$

ただし、 $\mathbf{e}$ と $\mathbf{0}$ はそれぞれ、全ての要素が 1 と 0 の列ベクトルとする。

(2.1) 式より、結合確率過程 $\{(N(t), J(t)); t \geq 0\}$ は状態空間 $\mathbb{Z}_+ \times \mathbb{D}$ 上の連続時間マルコフ連鎖であり、無限小生成行列は以下の式で与えられる。

$$\begin{array}{c} \mathbb{L}(0) \quad \mathbb{L}(1) \quad \mathbb{L}(2) \quad \mathbb{L}(3) \quad \dots \\ \begin{array}{c} \mathbb{L}(0) \\ \mathbb{L}(1) \\ \mathbb{L}(2) \\ \mathbb{L}(3) \\ \vdots \end{array} \begin{pmatrix} \mathbf{D}(0) & \mathbf{D}(1) & \mathbf{D}(2) & \mathbf{D}(3) & \dots \\ \mathbf{0} & \mathbf{D}(0) & \mathbf{D}(1) & \mathbf{D}(2) & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{D}(0) & \mathbf{D}(1) & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}(0) & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \end{array}$$

ただし、 $\mathbf{0}$ はゼロ行列で、 $\mathbb{L}(k) = \{k\} \times \mathbb{D}$, $k \in \mathbb{Z}_+$ とする。以上より、**BMAP** は行列の集合 $\{\mathbf{D}(k); k \in \mathbb{Z}_+\}$ によって特徴付けられることが分かる。これ以降、上述で定義した **BMAP** を **BMAP** $\{\mathbf{D}(k); k \in \mathbb{Z}_+\}$ と呼ぶ。

システムに到着した客は、直ちに 1 つのサーバを専有し、サービスが完了後すぐにシステムから退去する。それぞれの客のサービス時間は独立で、同一の分布 G に従う。 G は裾の軽い分布であることを仮定する。裾の軽い分布とは、分布の裾の減衰速度が指数分布より速い確率分布のことであり、以下の式を満たす [5]。

$$\int_0^{\infty} e^{\lambda x} G(dx) < \infty, \quad \text{for all } \lambda \geq 0.$$

上式より、以下の不等式を満たす $\mu \in (0, \infty)$ と $\beta \in (1, \infty)$ が存在する。

$$1 - G(x) \leq \beta e^{-\mu x}, \quad \text{for all } x \geq 0. \quad (2.2)$$

2.2. 客数過程の再生性

本節では、第2.1節で定義した待ち行列の客数過程と背後マルコフ連鎖の結合過程が、再生型確率過程 (regenerative process) [18] であることを述べる。 $L(t), t \geq 0$, を時刻 t でシステムに存在する客数とする。さらに、背後マルコフ連鎖の状態 $i_0 \in \mathbb{D}$ を一つ固定する。

$\tau_n, n \in \mathbb{N}$, を確率過程 $\{(L(t), J(t))\}$ が状態 $(0, i_0)$ を訪問する時刻とする。ただし、 $0 < \tau_1 < \tau_2 < \dots$ 。BMAP のマルコフ性より、各 $n \in \mathbb{N}$ で確率過程 $\{L(t + \tau_n); t \geq 0\}$ と $\{L(t + \tau_1); t \geq 0\}$ は同じ分布を持つ。さらに、各 $n \in \mathbb{N}$ で確率過程 $\{(L(t), J(t)); 0 \leq t < \tau_n\}$ と $\{(L(t), J(t)); t \geq \tau_n\}$ は独立な確率過程である。よって、確率過程 $\{(L(t), J(t)); t \geq 0\}$ は再生点 $\{\tau_n\}_{n \in \mathbb{N}}$ を持つ再生型確率過程である。再生型確率過程の定常性について、以下の定理が存在する。

補題 2.1 ([16, Theorem 3]) 確率過程 $\{X(t); t \geq 0\}$ を再生点 $\{X_n\}_{n \in \mathbb{N}}$ を持つ再生型確率過程とする。 $\{X(t)\}$ が唯一つの定常分布を持つことは、平均再生間隔が有限であること、すなわち以下の式が成立することと同値である。

$$E[X_2 - X_1] < \infty.$$

3. 安定性の証明の準備

第2.2節で述べた通り、確率過程 $\{(L(t), J(t))\}$ は再生型確率過程である。本論文では、この事実に着目して、安定性の十分条件を導く。補題 2.1 より、LBSM 条件のもとで平均再生間隔が有限であることを示せば、LBSM 条件が安定性の十分条件であることが分かる。しかしながら、 $\{(L(t), J(t))\}$ の平均再生間隔 $E[\tau_2 - \tau_1]$ の有限性を直接議論することは難しい。

そこで本論文ではまず、新たに再生型確率過程を、平均再生間隔が $E[\tau_2 - \tau_1]$ よりも大きく、かつ平均再生間隔の有限性が容易に議論できるように構築したい。上述のような再生型確率過程を構築するために、次小節にて直列型待ち行列 T を導入する。

3.1. 直列型待ち行列の構築

直列型待ち行列 T を2つのノードから成る直列型待ち行列 (tandem queue) とする。一つ目の待ち行列は BMAP/M/ ∞ 待ち行列である。到着過程は元の待ち行列と共通の BMAP $\{D(n); n \in \mathbb{Z}_+\}$ であり、サービス時間は率 μ の指数分布に従う。また、二つ目の待ち行列は、一つ目の待ち行列からの退去過程を入力とし、サービス時間は以下の不等式を満たす定数 α の固定長である。

$$\mu^{-1} \log \beta \leq \alpha. \quad (3.1)$$

ただし、一つ目の待ち行列のサービス率 μ と (3.1) 式の β の組は、(2.2) 式を満たす。

直列型待ち行列 T は2つノードから構成されるが、両ノードが無限サーバ待ち行列であるため、ノード間で待ち時間が発生しない。よって、直列型待ち行列 T は、到着過程が元の待ち行列と共通の BMAP $\{D(n); n \in \mathbb{Z}_+\}$ であり、サービス時間は2つのノードのサービス時間の和の分布 F_α に従う無限サーバ待ち行列である解釈できる。分布関数 F_α は率 μ の指数分布と定数 α の畳み込みとなるため、以下の式のように得られる。

$$F_\alpha(x) = \begin{cases} 0, & x < \alpha, \\ 1 - e^{-\mu(x-\alpha)}, & x \geq \alpha. \end{cases} \quad (3.2)$$

(2.2), (3.1), (3.2) 式より、サービス分布 G と F_α の間には以下の関係式が成立する。

$$1 - G(x) \leq 1 - F_\alpha(x), \quad x \geq 0. \quad (3.3)$$

$\hat{L}(t), t \geq 0$, を時刻 t に直列型待ち行列 T に滞在する客数とする。 $\hat{\tau}_n, n \in \mathbb{N}$, を確率過程 $\{(\hat{L}(t), J(t))\}$ が状態 $(0, i_0)$ を訪問する時刻とすると、 $\{(\hat{L}(t), J(t)); t \geq 0\}$ は再生点 $\{\hat{\tau}_n\}_{n \in \mathbb{N}}$ を持つ再生型確率過程である。ただし、 $0 < \hat{\tau}_1 < \hat{\tau}_2 < \dots$ 。次小節にて、 $\{(\hat{L}(t), J(t))\}$ の平均再生間隔 $E[\hat{\tau}_2 - \hat{\tau}_1]$ が $\{(L(t), J(t))\}$ の平均再生間隔 $E[\tau_2 - \tau_1]$ より大きくなることを示す。

3.2. 元の待ち行列と直列型待ち行列 T の安定性の関係

第 2.2 節と第 3.1 節で定義した 2 つの再生型確率過程 $\{(L(t), J(t))\}$ と $\{(\widehat{L}(t), J(t))\}$ の平均再生間隔に関して、以下の補題が成立する。

補題 3.1 以下の不等式が成立する。

$$E[\widehat{\tau}_2 - \widehat{\tau}_1] \geq E[\tau_2 - \tau_1].$$

補題 3.1 より、 $\{(\widehat{L}(t), J(t))\}$ の平均再生間隔 $E[\widehat{\tau}_2 - \widehat{\tau}_1]$ が有限ならば、 $\{(L(t), J(t))\}$ の平均再生間隔 $E[\tau_2 - \tau_1]$ もまた有限であることが示される。すなわち、定理 3.4 より、新たに構築した直行型待ち行列 T が安定ならば、元の待ち行列も安定であることが分かる。

補題 3.1 の証明. 元の待ち行列と直列型待ち行列 T は共通の到着過程から客が到着していたので、(3.3) 式より以下の不等式を満たす元の待ち行列の客数過程 $\{L(t)\}$ と直列型待ち行列 T の客数過程 $\{\widehat{L}(t)\}$ を構築できる。

$$L(t) \leq \widehat{L}(t), \quad t \geq 0. \quad (3.4)$$

(3.4) 式より、 $\{(\widehat{L}(t), J(t))\}$ の平均再生間隔 $E[\widehat{\tau}_2 - \widehat{\tau}_1]$ が有限ならば、 $\{(L(t), J(t))\}$ の平均再生間隔 $E[\tau_2 - \tau_1]$ もまた有限であることは明らかである。以降、(3.4) 式を満たす $\{L(t)\}$, $\{\widehat{L}(t)\}$ を構築できることを示す。

元の待ち行列と直列型待ち行列 T に $\text{BMAP}\{\mathbf{D}(n); n \in \mathbb{Z}_+\}$ から集団が到着する時刻を $T_n, n \in \mathbb{N}$, と記す。ただし、 $0 < T_1 < T_2 < \dots$ 。さらに、 $\{U_m; m \in \mathbb{N}\}$ を $(0, 1)$ 上の連続時間一様分布に従う独立な確率変数列として、 $S_m, \widehat{S}_m$ を各 $m \in \mathbb{N}$ に対して、以下のように定義する。

$$S_m = \inf\{x \geq 0; G(x) \in [U_m, 1]\}, \quad m \in \mathbb{N}, \quad (3.5)$$

$$\widehat{S}_m = -\frac{1}{\mu} \log(U_m) + \alpha, \quad m \in \mathbb{N}. \quad (3.6)$$

S_m と $\widehat{S}_m$ の分布は、それぞれ以下の式のように得られる。

$$P(S_m \leq x) = G(x), \quad x \geq 0, \quad (3.7)$$

$$P(\widehat{S}_m \leq x) = F_\alpha(x), \quad x \geq 0. \quad (3.8)$$

よって、(3.3), (3.5), (3.6) 式より、以下の不等式が確率 1 で成立する。

$$S_m \leq \widehat{S}_m, \quad m \in \mathbb{N}. \quad (3.9)$$

各 $n \in \mathbb{N}$ に対して、時刻 T_n に到着する集団の大きさを B_n と記す。さらに、 $A_n := \sum_{k=1}^n B_k, n \in \mathbb{N}$, とする。(3.7), (3.8) 式より、 $\{S_m; A_{n-1} + 1 \leq m \leq A_n\}$ と $\{\widehat{S}_m; A_{n-1} + 1 \leq m \leq A_n\}$ をそれぞれ、元の待ち行列と直列型待ち行列 T へ時刻 T_n に到着する客のサービス時間と仮定できる。このようにサービス時間を仮定したとき、 $L(t)$ と $\widehat{L}(t)$ は以下の式で与えられる。

$$L(t) = \sum_{n=1}^{\infty} \sum_{m=A_{n-1}+1}^{A_n} \mathbb{1}(T_n \leq t < T_n + S_m), \quad t \geq 0, \quad (3.10)$$

$$\widehat{L}(t) = \sum_{n=1}^{\infty} \sum_{m=A_{n-1}+1}^{A_n} \mathbb{1}(T_n \leq t < T_n + \widehat{S}_m), \quad t \geq 0. \quad (3.11)$$

ただし、 $\mathbb{1}(\chi)$ は事象 χ の指示関数とする。(3.9) 式より、(3.10), (3.11) 式で与えられる確率過程 $\{L(t)\}$, $\{\widehat{L}(t)\}$ は (3.4) 式を満たす。 $\square$

3.3. 直列型待ち行列 $\mathbf{T}$ の一つ目の待ち行列の安定性

直列型待ち行列 $\mathbf{T}$ の一つ目、二つ目の待ち行列の時刻 t での系内客数を、それぞれ $\hat{L}_1(t), \hat{L}_2(t)$ とする。確率過程 $\{\hat{L}(t)\}$ は直列型待ち行列 $\mathbf{T}$ に滞在する合計の客数過程であるため、以下の等式が成立する。

$$\hat{L}(t) = \hat{L}_1(t) + \hat{L}_2(t), \quad t \geq 0. \quad (3.12)$$

$\hat{\tau}_m^{(1)}, m \in \mathbb{N}$, を確率過程 $\{(\hat{L}_1(t), J(t))\}$ が状態 $(0, i_0)$ を訪問する時刻とする。ただし、 $0 < \hat{\tau}_1^{(1)} < \hat{\tau}_2^{(1)} < \dots$ 。確率過程 $\{(\hat{L}_1(t), J(t))\}$ は $\{\hat{\tau}_m^{(1)}; m \in \mathbb{N}\}$ を再生点とする再生型確率過程である。

ここで、確率過程 $\{(\hat{L}_1(t), J(t))\}$ は $\mathbf{BMAP}/\mathbf{M}/\infty$ 待ち行列の系内客数過程と背後マルコフ連鎖の結合確率過程である。さらに、[19, Theorem 1] より、 $\mathbf{LBSM}$ 条件は $\mathbf{BMAP}/\mathbf{M}/\infty$ 待ち行列の安定条件である。よって、補題 2.1 より、 $\mathbf{LBSM}$ 条件のもとで以下の式が成立する。

$$\mathbb{E}[\hat{\tau}_2^{(1)} - \hat{\tau}_1^{(1)}] < \infty. \quad (3.13)$$

4. 安定性の十分条件

本節では、再生理論の諸定理を用いて、 $\mathbf{LBSM}$ 条件のもとで裾の軽いサービス分布を持つ $\mathbf{BMAP}/\mathbf{G}/\infty$ 待ち行列が安定であることを示す。以下の定理は、本論文の主結果である。

定理 4.1 以下の不等式を満たす $C \in (0, \infty)$ が存在すると仮定する。

$$\sum_{n=1}^{\infty} \log(n+e) \mathbf{D}(n) e \leq C e. \quad (4.1)$$

このとき、 $\{L(t); t \geq 0\}$ は唯一つの定常分布を持つ。

(4.1) 式は、集団の大きさの対数の期待値が有界であることを意味している。すなわち、定理 4.1 は裾の軽いサービス分布を持つ $\mathbf{BMAP}/\mathbf{G}/\infty$ 待ち行列の安定性の十分条件が $\mathbf{LBSM}$ 条件であることを述べている。

定理 4.1 の証明。 補題 2.1 と補題 3.1 より、確率過程 $\{(\hat{L}(t), J(t))\}$ が唯一つの定常分布を持つならば、 $\{(L(t), J(t))\}$ もまた唯一つの定常分布を持つ。よって、 $\{(\hat{L}(t), J(t))\}$ の平均再生間隔 $\mathbb{E}[\hat{\tau}_2 - \hat{\tau}_1]$ が有限な値を持つこと示せば、定理 4.1 の証明は完了する。再生報酬定理 [18, Chapter 2, Theorem 2] より、以下の等式が成立する。

$$\frac{1}{\mathbb{E}[\hat{\tau}_2 - \hat{\tau}_1]} \cdot \mathbb{E} \left[\int_{\hat{\tau}_1}^{\hat{\tau}_2} \mathbb{1}(\hat{L}(t) = 0, J(t) = i_0) dt \right] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbb{1}(\hat{L}(t) = 0, J(t) = i_0) dt. \quad (4.2)$$

ただし、 $\mathbb{E}[\hat{\tau}_2 - \hat{\tau}_1] = \infty$ の場合には、 $1/\mathbb{E}[\hat{\tau}_2 - \hat{\tau}_1] = 0$ とする。(4.2) 式を利用して、 $\mathbb{E}[\hat{\tau}_2 - \hat{\tau}_1]$ が有限な値を取ることを示す。

確率過程 $\{(\hat{L}(t), J(t))\}$ が時間 $[\hat{\tau}_1, \hat{\tau}_2)$ に状態 $(0, i_0)$ に留まる時間は、率 $-D_{i_0, i_0}(0)$ の指数分布に従うため、以下の式が成立する。

$$\begin{aligned} \mathbb{E} \left[\int_{\hat{\tau}_1}^{\hat{\tau}_2} \mathbb{1}(\hat{L}(t) = 0, J(t) = i_0) dt \right] &= \mathbb{E} \left[\int_0^{\infty} \mathbb{1}(\hat{\tau}_1 \leq t < \hat{\tau}_2, \hat{L}(t) = 0, J(t) = i_0) dt \right] \\ &= \int_0^{\infty} \mathbb{P}(\hat{\tau}_1 \leq t < \hat{\tau}_2, \hat{L}(t) = 0, J(t) = i_0) dt \\ &= \int_0^{\infty} \exp\{D_{i_0, i_0}(0)t\} dt. \end{aligned} \quad (4.3)$$

(4.2) 式に (4.3) 式を代入すると、以下の式を得る。

$$\frac{1}{\mathbb{E}[\hat{\tau}_2 - \hat{\tau}_1]} \times \left\{ \int_0^{\infty} \exp\{D_{i_0, i_0}(0)t\} dt \right\} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbb{1}(\hat{L}(t) = 0, J(t) = i_0) dt. \quad (4.4)$$

(4.4) 式右辺の極限を計算するために、以下の積分を変形していく。

$$R(T) := \int_0^T \mathbb{1}(\widehat{L}(t) = 0, J(t) = i_0) dt, \quad T \geq 0.$$

(3.12) 式より、 $R(T)$ は以下のように変形される。

$$R(T) = \int_0^T \mathbb{1}(\widehat{L}_1(t) = 0, \widehat{L}_2(t) = 0, J(t) = i_0) dt, \quad T \geq 0. \quad (4.5)$$

直列型待ち行列 $\mathbf{T}$ の二つ目の待ち行列でのサービス時間は α の固定長であるため、時間 α が経過する間に一度も到着が無いとき、二つ目の待ち行列は空となる。さらに、二つ目の待ち行列は一つ目の待ち行列の退去過程を入力としているため、任意の $u \in [t - \alpha, t)$ で確率過程 $\{(\widehat{L}_1(u), J(u))\}$ が状態 $(0, i_0)$ ならば $L_2(t) = 0$ となる。よって、(4.5) 式より、 $R(T)$ について以下の不等式が成立する。

$$\begin{aligned} R(T) &\geq \int_0^T \mathbb{1}(\widehat{L}_1(u) = 0, J(u) = i_0, \forall u \in [t - \alpha, t)) dt \\ &\geq \int_0^{\widehat{\tau}_{M(T)}^{(1)}} \mathbb{1}(\widehat{L}_1(u) = 0, J(u) = i_0, \forall u \in [t - \alpha, t)) dt. \end{aligned} \quad (4.6)$$

ただし、 $M(T), T \geq 0$, は以下の式で定義される確率変数で、 $\widehat{\tau}_0^{(1)} = 0$ とする。

$$M(T) := \sup\{n \in \mathbb{Z}_+; \widehat{\tau}_n^{(1)} \leq T\}.$$

ここで、確率変数 $V_n, n \in \mathbb{N}$, を以下のように定義する。

$$V_n = \int_{\widehat{\tau}_{n-1}^{(1)}}^{\widehat{\tau}_n^{(1)}} \mathbb{1}(L_1(u) = 0, J(u) = i_0, \forall u \in [t - \alpha, t)) dt, \quad n \in \mathbb{N}. \quad (4.7)$$

(4.7), (4.6) 式より、 $R(T)$ と V_n に関する以下の式が導かれる。

$$R(T) \geq \sum_{n=1}^{M(T)} V_n. \quad (4.8)$$

(4.4), (4.5), (4.8) 式より、以下の不等式を得る。

$$\frac{1}{\mathbb{E}[\widehat{\tau}_2 - \widehat{\tau}_1]} \times \left\{ \int_0^\infty \exp\{D_{i_0, i_0} t\} dt \right\} \geq \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{n=1}^{M(T)} V_n. \quad (4.9)$$

(4.9) 式右辺の極限に再生報酬定理を適用するために、 $V_n, n \in \mathbb{N}$ が $\{(\widehat{L}_1(t), J(t)); t \in [\widehat{\tau}_{n-1}^{(1)}, \widehat{\tau}_n^{(1)})\}$ の値のみで定まる確率変数であることを示す。 $\widehat{\tau}_{n-1}^{(1)}$ は状態 $(0, i_0)$ を訪問する時刻であるため、 $t - \alpha < \widehat{\tau}_{n-1}^{(1)}$ ならば、確率過程 $\{(\widehat{L}_1(t), J(t))\}$ はある時刻 $u \in [t - \alpha, \widehat{\tau}_{n-1}^{(1)})$ で $(0, i_0)$ 以外の状態をとる。よって、以下の式が成立する。

$$\mathbb{1}(\widehat{\tau}_{n-1}^{(1)} > t - \alpha, L_1(u) = 0, J(u) = i_0, \forall u \in [t - \alpha, t)) = 0.$$

上式を用いると、 V_n は以下のように変形できる。

$$\begin{aligned} V_n &= \int_{\widehat{\tau}_{n-1}^{(1)}}^{\widehat{\tau}_n^{(1)}} \mathbb{1}(\widehat{\tau}_{n-1}^{(1)} \leq t - \alpha, L_1(u) = 0, J(u) = i_0, \forall u \in [t - \alpha, t)) dt \\ &\quad + \int_{\widehat{\tau}_{n-1}^{(1)}}^{\widehat{\tau}_n^{(1)}} \mathbb{1}(\widehat{\tau}_{n-1}^{(1)} > t - \alpha, L_1(u) = 0, J(u) = i_0, \forall u \in [t - \alpha, t)) dt \\ &= \int_{\widehat{\tau}_{n-1}^{(1)}}^{\widehat{\tau}_n^{(1)}} \mathbb{1}(\widehat{\tau}_{n-1}^{(1)} \leq t - \alpha, L_1(u) = 0, J(u) = i_0, \forall u \in [t - \alpha, t)) dt, \quad n \in \mathbb{N}. \end{aligned} \quad (4.10)$$

(4.10) 式より, V_n は $\{(\hat{L}_1(t), J(t)); t \in [\hat{\tau}_{n-1}^{(1)}, \hat{\tau}_n^{(1)}]\}$ の値のみで定まる確率変数である. よって, 遅れのある再生点 $\{\hat{\tau}_n^{(1)}\}_{n \in \mathbb{Z}_+}$ と時刻 $\hat{\tau}_n^{(1)}$, $n \in \mathbb{N}$, に得る報酬 V_n に対して, 再生報酬定理を適用すると, 以下の等式を得る.

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{n=1}^{M(T)} V_n = \frac{1}{\mathbb{E}[\hat{\tau}_2^{(1)} - \hat{\tau}_1^{(1)}]} \cdot \mathbb{E}[V_2].$$

ただし, $\mathbb{E}[\hat{\tau}_2^{(1)} - \hat{\tau}_1^{(1)}] = \infty$ の場合には, $1/\mathbb{E}[\hat{\tau}_2^{(1)} - \hat{\tau}_1^{(1)}] = 0$ とする. 上式を (4.9) 式に代入すると, 以下の不等式が得られる.

$$\frac{1}{\mathbb{E}[\hat{\tau}_2 - \hat{\tau}_1]} \times \left\{ \int_0^\infty \exp\{D_{i_0, i_0}(0)t\} dt \right\} \geq \frac{1}{\mathbb{E}[\hat{\tau}_2^{(1)} - \hat{\tau}_1^{(1)}]} \cdot \mathbb{E}[V_2]. \quad (4.11)$$

最後に, $\mathbb{E}[V_2]$ を求める. 確率過程 $\{(\hat{L}_1(t), J(t))\}$ は, ある時刻 t で状態 $(0, i_0)$ ならば, 時刻 t の直前の再生点から時刻 t までの間は状態 $(0, i_0)$ に留まり続ける. よって, (4.10) 式より, 以下の式が導かれる.

$$\begin{aligned} \mathbb{E}[V_2] &= \mathbb{E} \left[\int_{\hat{\tau}_1^{(1)}}^{\hat{\tau}_2^{(1)}} \mathbb{1} \left(\hat{\tau}_1^{(1)} \leq t - \alpha, \hat{L}_1(u) = 0, J(u) = i_0, \forall u \in [t - \alpha, t] \right) dt \right] \\ &= \mathbb{E} \left[\int_{\hat{\tau}_1^{(1)}}^{\hat{\tau}_2^{(1)}} \mathbb{1} \left(\hat{\tau}_1^{(1)} \leq t - \alpha, \hat{L}_1(u) = 0, J(u) = i_0, \forall u \in [\hat{\tau}_1^{(1)}, t] \right) dt \right]. \end{aligned}$$

上式において, 変数変換を行うと,

$$\begin{aligned} \mathbb{E}[V_2] &= \mathbb{E} \left[\int_0^{\hat{\tau}_2^{(1)} - \hat{\tau}_1^{(1)}} \mathbb{1} \left(0 \leq t - \alpha, \hat{L}_1(u + \hat{\tau}_1^{(1)}) = 0, J(u + \hat{\tau}_1^{(1)}) = i_0, \forall u \in [0, t] \right) dt \right] \\ &= \mathbb{E} \left[\int_\alpha^\infty \mathbb{1} \left(t < \hat{\tau}_2^{(1)} - \hat{\tau}_1^{(1)}, \hat{L}_1(u + \hat{\tau}_1^{(1)}) = 0, J(u + \hat{\tau}_1^{(1)}) = i_0, \forall u \in [0, t] \right) dt \right]. \end{aligned}$$

確率過程 $\{(\hat{L}_1(t), J(t))\}$ が時間 $[\hat{\tau}_1^{(1)}, \hat{\tau}_2^{(1)})$ に状態 $(0, i_0)$ に留まる時間は, 率 $-D_{i_0, i_0}(0)$ の指数分布に従うため, $\mathbb{E}[V_2]$ は以下の通りに計算される.

$$\begin{aligned} \mathbb{E}[V_2] &= \int_\alpha^\infty \mathbb{P} \left(t < \hat{\tau}_2^{(1)} - \hat{\tau}_1^{(1)}, \hat{L}_1(u + \hat{\tau}_1^{(1)}) = 0, J(u + \hat{\tau}_1^{(1)}) = i_0, \forall u \in [0, t] \right) dt \\ &= \int_\alpha^\infty \exp\{D_{i_0, i_0}(0)t\} dt. \end{aligned} \quad (4.12)$$

よって, (4.12) 式を (4.11) 式に代入すると, 以下の不等式を得る.

$$\frac{1}{\mathbb{E}[\hat{\tau}_2 - \hat{\tau}_1]} \times \left\{ \int_0^\infty \exp\{D_{i_0, i_0}(0)t\} dt \right\} \geq \frac{1}{\mathbb{E}[\hat{\tau}_2^{(1)} - \hat{\tau}_1^{(1)}]} \times \left\{ \int_\alpha^\infty \exp\{D_{i_0, i_0}(0)t\} dt \right\}. \quad (4.13)$$

(4.13) 式より, $\mathbb{E}[\hat{\tau}_2^{(1)} - \hat{\tau}_1^{(1)}]$ が有限な値を持つとき, $\mathbb{E}[\hat{\tau}_2 - \hat{\tau}_1]$ もまた有限な値を持つことが分かる. よって, (3.13) 式より, LBSM 条件のもとで $\mathbb{E}[\hat{\tau}_2 - \hat{\tau}_1]$ は有限の値をもつ. そして, 補題 2.1 と補題 3.1 より, $\{(L(t), J(t))\}$ は唯一つの定常分布を持つ. $\square$

5. 結論

本論文では, 裾の軽いサービス分布を持つ BMAP/G/ ∞ 待ち行列の安定性について議論を行った. そして, 集団の大きさの対数期待値が有界であるとき, この待ち行列は安定になることを示した. 証明の際には, 客数過程が再生型確率過程の構造を持つことに着目して, 再生理論の諸定理を用いた. 今後の課題としては, 裾の軽いサービス分布を持つ BMAP/G/ ∞ 待ち行列の安定性の必要条件を構築することが挙げられる.

参考文献

- [1] S. Asmussen and G. Koole. Marked point processes as limits of Markovian arrival streams. *Journal of Applied Probability*, 30(2):365–372, 1993.
- [2] O. Berman and E. Kim. Stochastic models for inventory management at service facilities. *Stochastic Models*, 15(4):695–718, 1999.
- [3] U. Chatterjee and S.P. Mukherjee. On the non-homogeneous service system $M^X/G/\infty$. *European Journal of Operational Research*, 38(2):202–207, 1989.
- [4] T. D. Cong. On the $M^X/G/\infty$ queue with heterogeneous customers in a batch. *Journal of Applied Probability*, 31(1):280–286, 1994.
- [5] S. Foss, D. Korshunov, and S. Zachary. *An Introduction to Heavy-Tailed and Subexponential Distributions*. Springer, New York, 2011.
- [6] D. F. Holman, M. L. Chaudhry, and B. R. K. Kashyap. On the service system $M^X/G/\infty$. *European Journal of Operational Research*, 13(2):142–145, 1983.
- [7] J. Keilson and A. Seidmann. $M/G/\infty$ with batch arrivals. *Operations Research Letters*, 7(5):219–222, 1988.
- [8] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. SIAM, Philadelphia, PA, 1999.
- [9] D. M. Lucantoni. New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, 7(1):1–46, 1991.
- [10] D. M. Lucantoni, K. S. Meier-Hellstern, and M. F. Neuts. A single-server queue with server vacations and a class of non-renewal arrival processes. *Advances in Applied Probability*, 22(3):676–705, 1990.
- [11] W. A. Massey. The analysis of queues with time-varying rates for telecommunication models. *Telecommunication Systems*, 21(2-4):173–204, 2002.
- [12] H. Masuyama and T. Takine. Analysis of an infinite-server queue with batch Markovian arrival streams. *Queueing Systems*, 42(3):269–296, 2002.
- [13] V. Ramaswami and M. F. Neuts. Some explicit formulas and computational methods for infinite-server queues with phase-type arrivals. *Journal of Applied Probability*, 17(2):498–514, 1980.
- [14] D. N. Shanbhag. On infinite server queues with batch arrivals. *Journal of Applied Probability*, 3(1):274–279, 1966.
- [15] L. Takács. Queues with infinitely many servers. *RAIRO Recherche Opdrationnelle*, 14(2):109–113, 1980.
- [16] H. Thorisson. Construction of a stationary regenerative process. *Stochastic processes and their applications*, 42(2):237–253, 1992.
- [17] T. Van Woensel and N. Vandaele. Modeling traffic flows with queueing models: a review. *Asia-Pacific Journal of Operational Research*, 24(4):435–461, 2007.
- [18] R. W. Wolff. *Stochastic Modeling and the Theory of Queues*. Prentice Hall, Englewood Cliffs, NJ, 1989.
- [19] M. Yajima, T. Phung-Duc, and H. Masuyama. The stability condition of BMAP/M/ ∞ queues. In *Proceedings of the 11th International Conference on Queueing Theory and Network Applications (QTNA2016)*, Article number 5, New York, 2016. ACM.
- [20] M. Yajima, T. Phung-Duc, and H. Masuyama. The stability condition of BMAP/PH/ ∞ queues. In *Proceedings of the 12th International Conference on Queueing Theory and Network Applications (QTNA2017)*, 82–85, 2017.

BGP ネットワークの障害状況のマッピング

伊藤翔馬 小林香

富山県立大学 工学部 電子・情報工学科

概要：

本論文では、BGP ネットワークで発生する障害をマッピングにより可視化することで、人間の目で障害発生の有無の判断をすることができる手法を提案する。

AS(Autonomous System) と AS の間での経路選択には BGP(Border Gateway Protocol) が用いられ、この BGP によって形成されるネットワークを BGP ネットワークと呼ぶ。この BGP ネットワーク内では経路選択によるネットワーク障害や、悪意のある BGP ハイジャックなどが発生している。これら BGP に関するネットワーク障害や BGP ハイジャックは、異常検知が難しく、検知の手法も確立されていない。そのため、本研究では、ネットワークの状態を地図上にマッピングすることで、ネットワーク障害の発生している地域や規模を人間の目で判断し、その障害の詳細解析につなげることができるかについて検討する。

1. はじめに

ネットワークを介する通信は端末間のルータが経路情報を保持していることが前提であり、全世界を1つのネットワークとみなすインターネットにおいては、BGP(Border Gateway Protocol)[1] がその経路情報を正しく維持する役目を担っている。統一された運用ポリシーに従うネットワークを AS(Autonomous System) と称し、各々の AS には識別するために番号が与えられている。この AS と AS の間で経路制御を行なっているプロトコルが BGP である。「インターネットのカタチ」[2] には「BGP は、各 AS が『最適な経路』を選択しつつ、それを隣の AS に伝えていく、伝言ゲーム的な制御手法なのです。」と表現されているように、BGP は通信経路を確立するために、AS 間で必要な情報のやり取りと通信に最適な経路の選択を行なっている。

宛先となる IP ネットワークをプレフィックスと呼び、BGP ルータは、そのルータからプレフィックスに至るまでに経由する全ての AS 番号を ASpath リストとして保持している。このプレフィックスまでの経路が障害などによって失われた場合は BGP は withdraw としてメッセージを隣り合うルータに通知する。また、障害箇所を迂回した経路によってプレフィックスまでの通信経路が確立されると update として新しい経路を隣り合うルータに通知する。

BGP ルータからの誤った経路情報の発信や、BGP ルータ間を繋いでいるケーブルの断絶などの原因で通信経路が確立できず通信障害が発生する。現在全世界で稼働しているネットワークのうち、宛先となるプレフィックスの数は既に 70 万を超え、また、AS の数も図 1 に示すように 6 万を超える。そのため、BGP ネットワーク構造は複雑化しており、それに伴って BGP ネットワーク内で発生する通信障害は解析が難しい。

ネットワークの解析ツールには BGPmon[3] や BGPlay[4] などが実際によく活用されているが、これは解析の対象となる AS や、障害が発生していることが明らかな AS がわかっている場合でなければ解析作業を行えないほか、AS 番号を用いた関係図を用いた表現を行なっているため、地域や障害規模などの理解が難しい。

そこで、まず BGP ネットワークに発生する障害をいかに解析しやすくするかについて、地図上へのマッピングを用いた障害の可視化手法について提案する。次に、障害の状況を人間が簡単に理解することができるように、AS と AS 間でのつながりや変化の様子をマッピングすることで、ネットワークの状態やその変化を表現する手法について提案する。

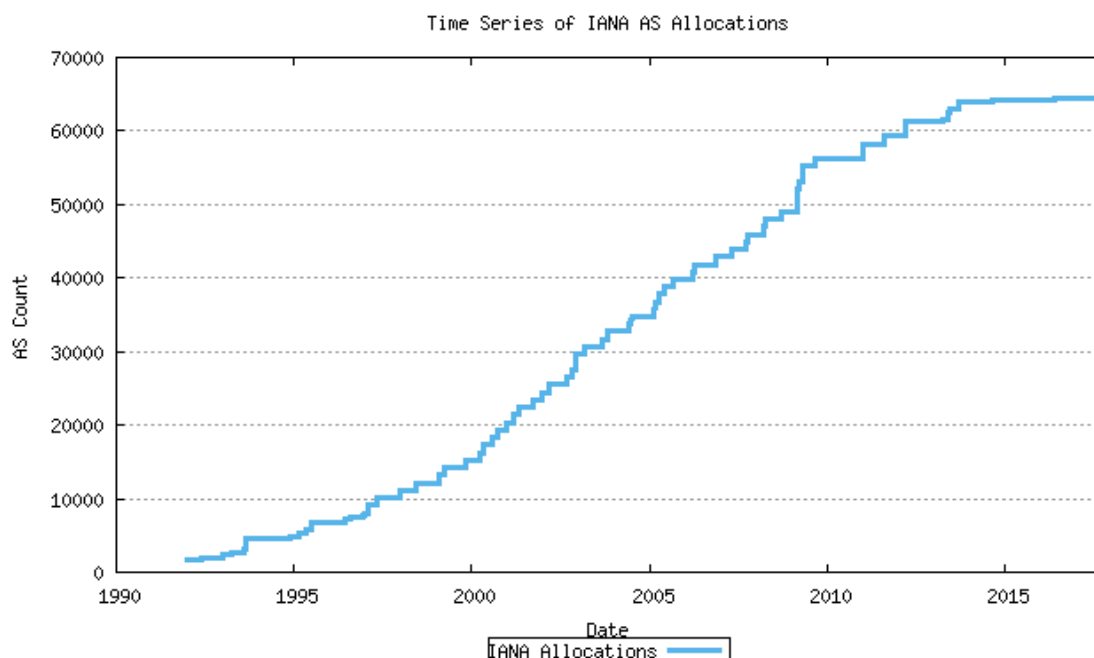


図 1: IANA による RIR への AS 番号払い出し数の時系列変化 (出典: <http://www.potaroo.net/tools/asn16/>)

2. 障害の可視化

一般的なネットワーク分析ソフトウェアにおいて、ノードの接続状況を図示する場合、バネ均衡などの描画アルゴリズムを用いて各ノードが最適な位置に配置される。BGP の経路においては、ノード数が 6 万弱、エッジ数が 70 万弱と非常に多いため、図 2 のような安定配置に収束するまでに長時間を必要とし、注目したいノードかエッジを選択して再描画させるとまた同じ程度の時間を要することから実用的ではない。またノードやエッジの位置は描画ごとに初期化されるため、異なる時刻における経路表を描画するごとに注目するノードやエッジの位置が異なることとなり、時間に従った変化を調べることもに適していない。

解析を目的とした描画のためには、まずネットワークの全体像が短時間で描画できること、次に何らかの変化が生じている箇所が明らかにできること、最後に変化の詳細が時間を追って解析可能な形で可視化できること、が条件となる。

2.1. 経路データの前処理

描画する経路表は BGPView[5] を用いて、30 分ごとに取得したものを用いる。この経路表は、プレフィックスとそのプレフィックスに至るまでに経由する AS のリストである ASpath リストで構成され、2017 年 11 月末現在でプレフィックス数は 70 万弱存在する。

BGP ルータが経路消失を知らせる withdraw を含むメッセージを受信すると、BGP ルータは経路表からそれらのプレフィックスとその ASpath を削除し、経路更新を知らせる announce を含む BGP メッセージを受信すると、すでにある経路と比較してより適した経路を経路表として更新する。

可視化に用いる数値は、図 3 に示すように時刻ごとの AS 番号を含んだプレフィックス数を計測し、隣り合う時間での差分を用いる。

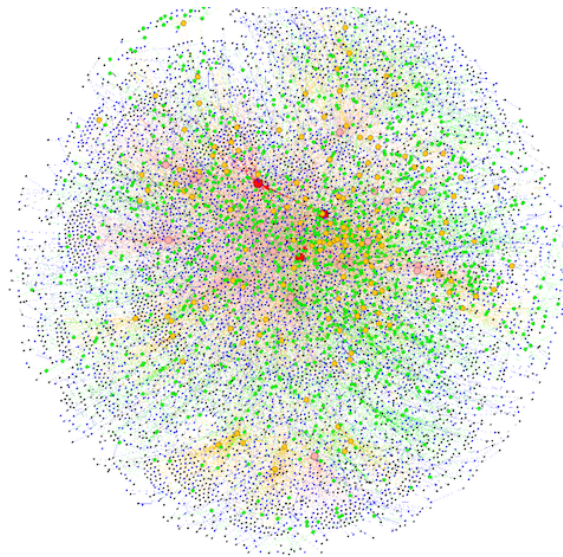


図 2: バネ均衡モデルを用いた BGP 経路情報の可視化例. AS を丸, AS 間のリンクを線で描画している.

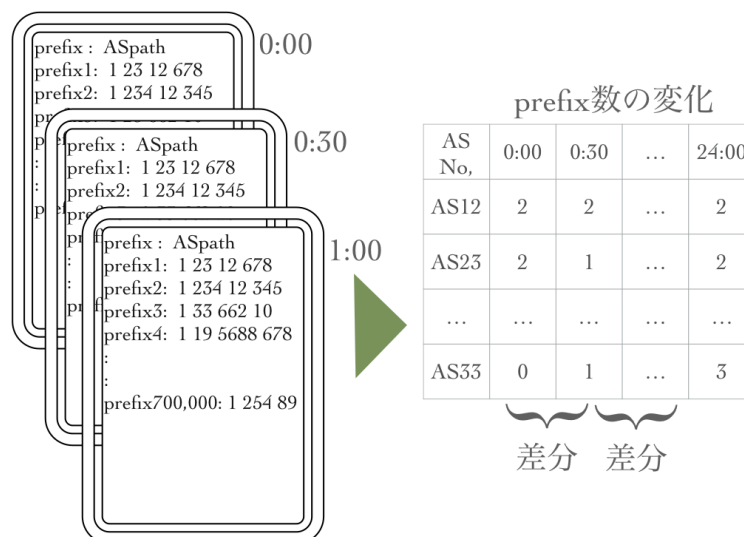


図 3: 経路データの前処理

2.2. AS の地理情報の関連付け

隣り合う時刻で、AS 番号とその AS 番号を ASpath リストにもつプレフィックス数の変化量の形式で経路表を整理すると、AS を通る経路の消失と更新状況が理解しやすくなる。しかし、経路消失・更新が発生した AS の位置を知りたい場合は、その地理情報を参照して地図上に表現する必要がある。

ASpath リストに含まれているのは AS の番号であるから、AS 番号よりその AS の地理情報を得るために、AS 番号からその AS を割り当てられている管理組織 (ISP:Internet Service Provider) 名とその所在地、次に所在地から緯度・経度などの地理情報を取得する。

まず AS 番号からその ISP の情報を得るために、AS 番号の割り当てを担当する 5 つの RIR(Regional Internet Registry) の whois データベースに AS 番号の問い合わせを行う。データが登録されている RIR であれば、登録されている ISP 名や所在地の情報が返され、そうでなければ登録なしの返答が返される。今回は使用する経路表の提供元が RIR の中でアジア地域を管理する APNIC に所属することから、APNIC, ARIN, RIPE, LACNIC, AFRINIC の順に問い合わせを実行する。

次に、所在地から地理情報への変換作業を行う。今回は Esri 社が提供する API の内、findAddressCandidate[6] のサービスを使用した。このサービスに対して Geocoding の対象となる住所を含んだ URL を生成してリクエストを行うことで、JSON 形式でのレスポンスが得られる。AS の管理を行なっている ISP と、その ISP の所在地を保存したデータファイルからリクエスト URL を生成し、リクエストを送信する。レスポンスとして返される JSON ファイルから、緯度・経度の値を地理情報として取得する。

2.3. 地図への可視化

背景に白地図を入れておくことが AS の地理情報を直感的に理解する上で必要なため、今回の研究では、CraftMap[7] サービスを使用して地図画像の作成を行った。その際、地図の設定はミラー図法基調を選択し、中央位置は経度 180 度、表示北緯・南緯はどちらも 90 度とする。また、緯度・経度を扱いやすくするため、緯度補正、球型補正はどちらも 0 とした。

マッピングを行う値の対象は、30 分ごとの時系列データから、30 分間で数値に変化があった場合の AS のみとし、表示する情報を減らす。マッピングの基準となる時刻を指定すると、その時刻から 30 分後の時刻を変化後の時刻として、変化後の値と基準となる時刻の値との差分を算出し、差分の量を AS の所在地に描画する円の直径に反映させる。

2.4. 具体例を用いた ASpath リストのマッピング

具体的に、2017 年 8 月 25 日 12 時から 12 時 30 分におけるネットワークの変化を可視化する。実際にこの時間帯には、OCN が管理する AS 番号 4713 の AS へ誤った BGP 経路が流入したことにより、日本全域で 12 時 20 分ごろから大規模な通信障害が発生していたことが報道等で知られている。

30 分毎のプレフィックス数の差分の値は、数 1000 にまで及び、そのままの数値を直径とするピクセル数で円を描画すると大きすぎるが、対数を用いると図 4 に示す通り円が小さくなりすぎて変化があることすら判断できない。また、変化量の 1/10 のピクセル数で円を描画した図 5 では、変化があった位置は明らかとなるものの、円の大小の違いがわかりにくく、視覚的な印象が薄れてしまっている。さまざまな試行錯誤ののちプレフィックス数の変化量を半分にしたピクセル数の直径の円を描くこととした。

図 6 には円形が大きく現れており、何らかの大規模なネットワークの変化が発生したことが一目で理解できる。日本及び台湾に大きな径の円が描画され、これらの AS で何らかの障害が発生したのではないかと推測される。赤色の円径が大きくなっていることから、基準時間である 12 時から 12 時 30 分にかけて BGP ルータの経路テーブルに日本や台湾に存在する AS 番号の出現回数が多くなっていることを表しており、これらの AS を経由する経路が増えたことを表している。このことから、さまざまな宛先プレフィックスまでの経路の中に日本

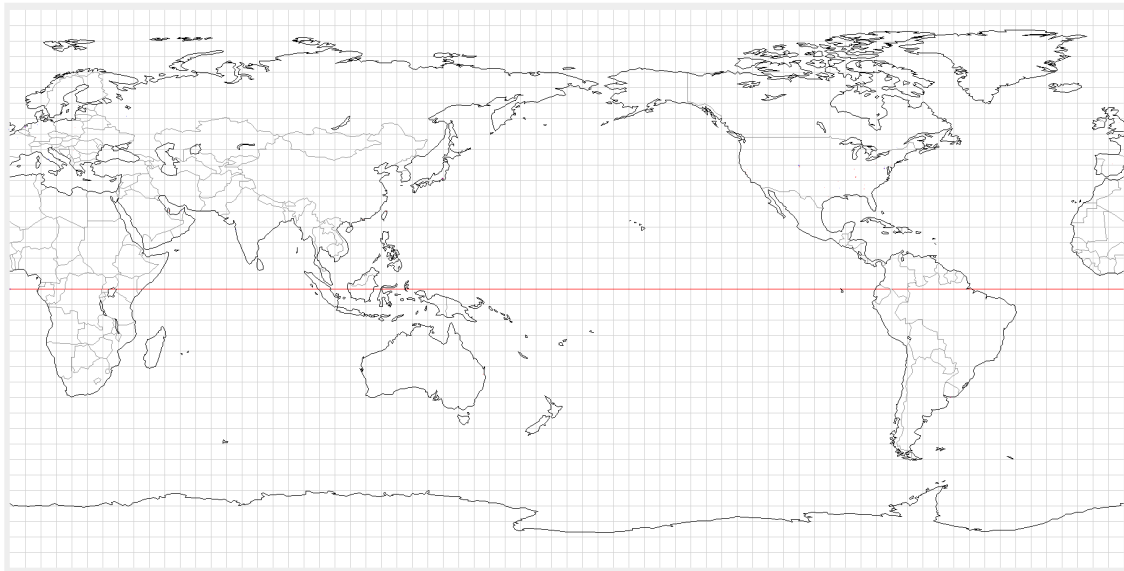


図 4: 2017 年 8 月 25 日 12 時 00 分から 12 時 30 分における経路テーブル内の AS 番号の出現回数の差の値の対数をとった場合の変化

や台湾に存在する AS を含んだ経路が大量に発生したことがわかる。その結果回線が圧迫され、通信速度の低下や通信の切断などが発生したのではないかと推測される。

BGP 経路障害時のマッピングを図 6 に示したが、対象事例として大きな障害がみられなかった 2017 年 8 月 25 日の 23 時 30 分から 24 時の情報のマッピングを図 7 に示す。図 6 と図 7 を比較すると、地図での可視化が有効であることが明らかである。

マッピングを行うとネットワークの状態が人間の目でも理解することができるようになり、BGP メッセージや BGP 経路テーブルのような文字列よりもより効率的に BGP ネットワークの状態の理解ができると考えられる。また、マッピングを時系列に従って複数並べることで、他の時間帯と比較して明らかに大きな円表示されれば、障害が発生している状態であると判断できる。

変化が起きている AS を調べるためのマッピングは AS を“点”として捉えているだけであり、次に知りたいのは、障害が発生している ASpath の情報であるから、AS-AS 間のリンク情報をマッピングすることが必要となる。そこで、BGP ルータから 30 分ごとに取得される経路テーブルより、宛先となるプレフィックスとその宛先までに経由する AS 番号 (経路) を取得し、直線を用いたマッピングを行う。

AS の所在地及び背景の地図画像は、2.2 節で取得した位置情報及び、緯度・経度へ変換したデータと同一のものをを用いる。

リンクの描画をする際に、BGP ルータから取得されるフルルートを描画すると、その中に含まれる経路の量の多さから、その関係を表す線が多くなってしまい、背景の白地図が埋まってしまうため変化を捉えることができない。そのため、フルルートを取得している時間間隔と同じ 30 分間で経路変化があった ASpath リストのみをフルルートから抽出し、さらに、図 6 で障害が発生している状態と判断した地域の AS に絞り込んで、それらの AS を含んだ ASpath リストだけを地図上へマッピングする。

AS 間は直線でリンクを表し、色分けすることで 30 分間の ASpath リストの変化を 1 つの地図上にマッピングする。マッピングの基準となる時間の経路を青色、基準となる時間の 30 分後の経路を赤色で描画する。赤色で描画する経路は、基準となる時間の 30 分後のフルルートから、30 分間に経路変化があったもの、かつ 2.2 節で障害が発生している状態と判断された地域にある AS 番号を含んでいる ASpath リストのみを抽出してマッピングする。また、青色で描画するリンクは、赤色で描画する際に抽出した ASpath リストの宛先プレフィックス

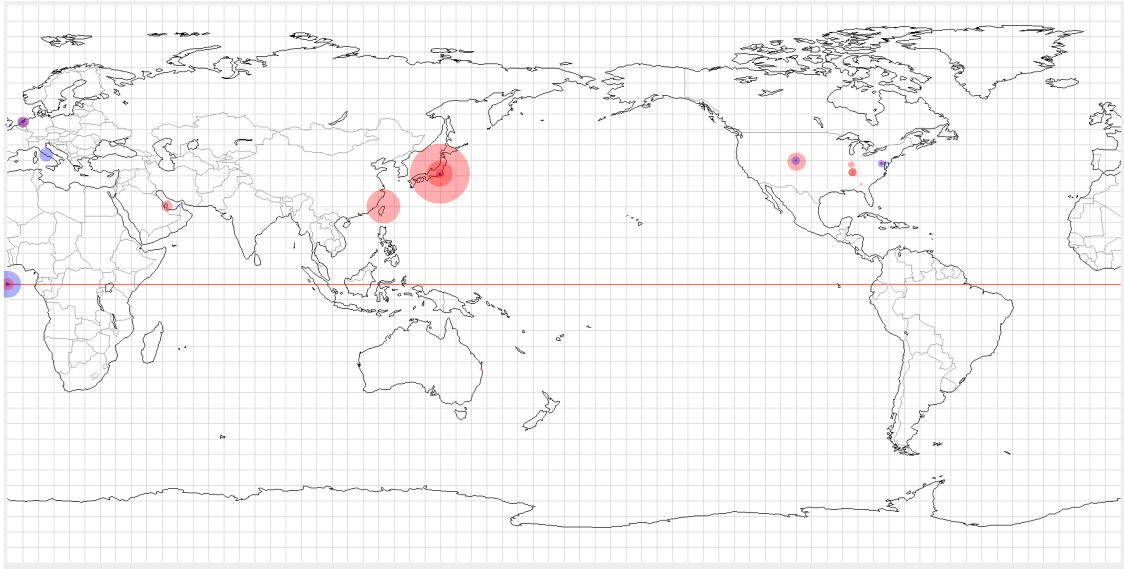


図 5: 2017 年 8 月 25 日 12 時 00 分から 12 時 30 分における経路テーブル内の AS 番号の出現回数の差の値を 10 分の 1 倍にした値を用いた場合の変化

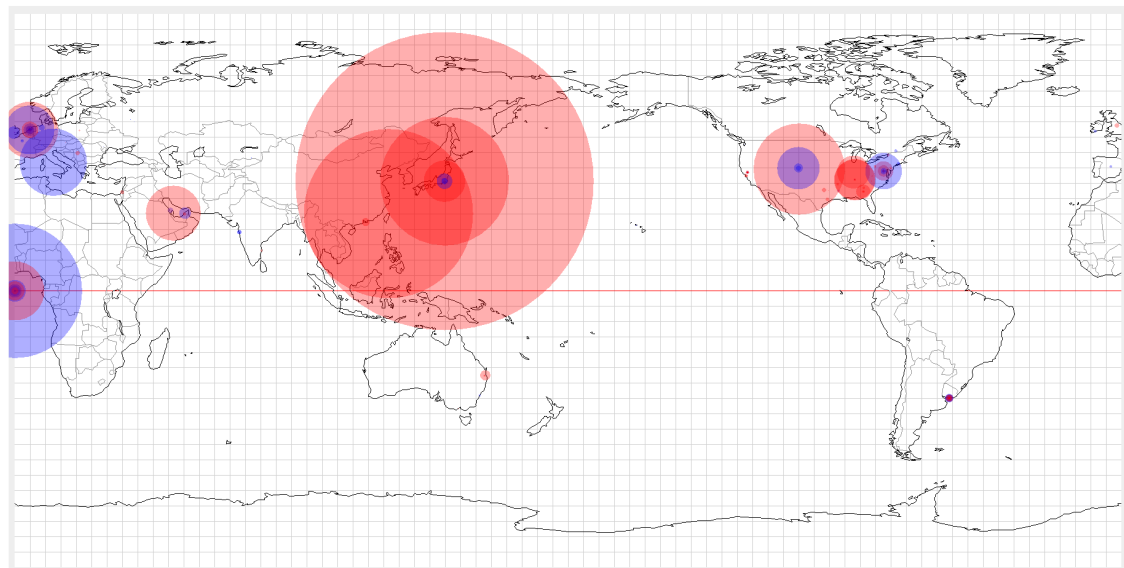


図 6: 2017 年 8 月 25 日 12 時 00 分から 12 時 30 分における経路テーブル内の AS 番号の出現回数の差の変化

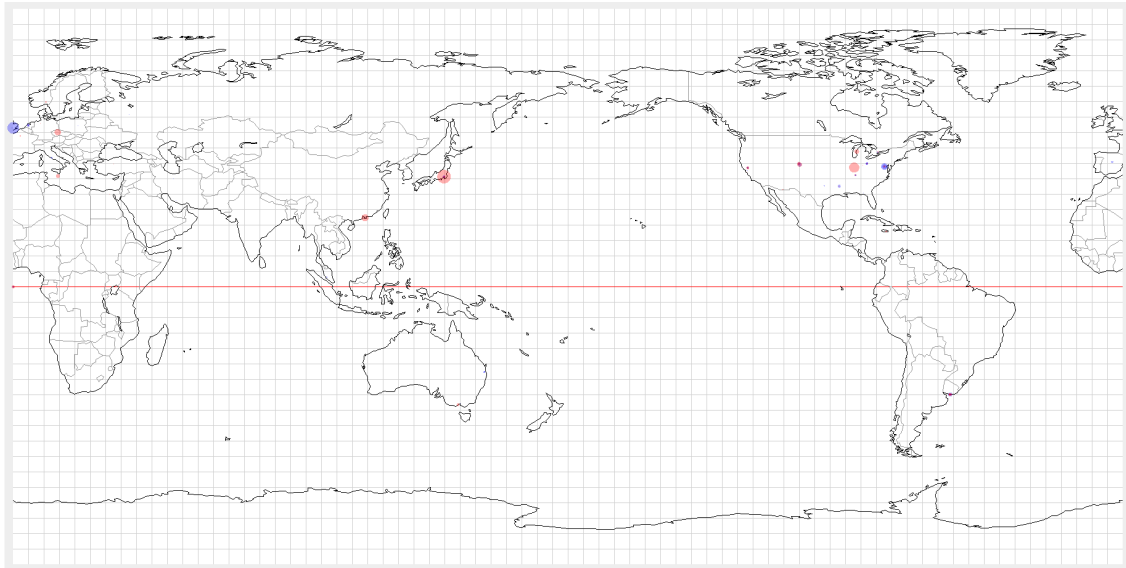


図 7: 2017 年 8 月 25 日 23 時 30 分から 24 時における経路テーブル内の AS 番号の出現回数の差の変化

と同じプレフィックスの ASpath リストを，基準となる時間のフルルートから取り出してマッピングするものとする。

また，経路が重なることを考慮して描画する線には透過処理を行い，経路が重なる場合は線の色が濃く現れ，基準となる時間の ASpath と基準から 30 分後の ASpath が重なる場合は青と赤の混ざった色で表現される。

図 8 に示す画像は，2017 年 8 月 25 日 12 時から 12 時 30 分にかけての ASpath リストの変化のマッピングである．ノードに含まれる AS は図 6 に現れる円のうち，円の中心が日本にある AS を抽出しており，これらの AS 番号を含んだ経路のみを 12 時 30 分のフルルートから抽出して赤色の線で描画している．また，12 時 30 分の経路テーブルから抽出された ASpath リストの宛先であるプレフィックスと同じプレフィックスの経路を 12 時のフルルートから抽出して青色の線で描画している．図 8 に見られる日本とアメリカを繋ぐ濃い赤い線は，この間の経路が多く通信に使用されていることを表しており，濃い青色の線は消失した経路を表している．図 6 に現れている通り，日本に存在する AS 番号を ASpath にもつプレフィックス数が増えており，12 時 30 分の時点で多くの通信が日本の AS を迂回路として経由するようなネットワークの状態になっていたことが図 8 に現れている。

このマッピングを行なった結果，ASpath リストの変化をリンクで可視化することはできたと考えられる．濃い青のリンクが確認できる場所は複数の経路が消失したことを表しており，濃い赤色のリンクは，迂回路が現れたことを表す．また，宛先プレフィックスや AS 番号をさらに絞り込んでマッピングを行うことで，より詳細な解析につなげることができると考えられる．しかし，AS 同士のリンクを地図上にマッピングしただけでは，経路が集中している場所や経由する場所は明らかとなるが，どの AS を経由しているかを一目で理解するのは難しい．また，世界地図でのマッピングを行うと，多くの AS は首都圏に集中することから，近い位置にある AS 間の関係を表すことは難しく，それらの AS はより狭域な地図上で描画する必要があることが明らかとなった。

3. 今後の課題

今回マッピングに用いたデータは，日本に存在するとある AS からのデータ供給を受けて利用しているものであり，その AS から得られるフルルートは，その AS を始まりとしてプレフィックスまでの ASpath が記録されている．そのためマッピングを行うことができるネッ

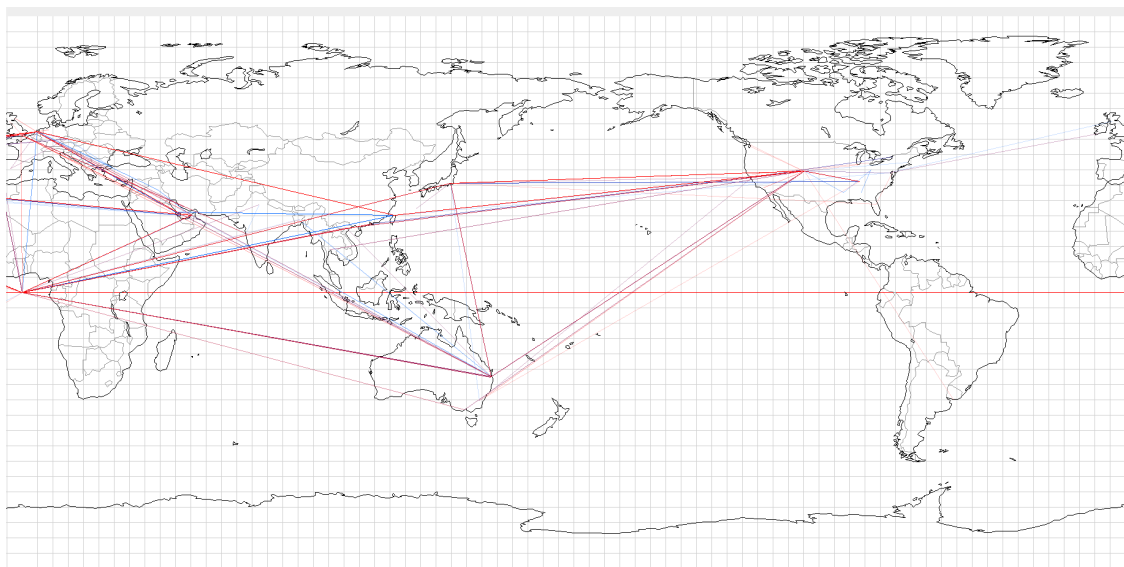


図 8: 2017 年 8 月 25 日 12 時 00 分と 12 時 30 分の経路の差

トワークはそのとある AS が中心となっており，世界全体をマッピングするには日本中心のネットワークに偏っている．また，その AS からではたどり着くことのできないプレフィックスが存在しており，今回使用したデータだけでは全てのネットワークを網羅できていない．この問題を解消するためには，データ取得点を増やす必要がある．データの取得点を増やしてフルルートを取得しマッピングを行うことで，今回使用したデータのみでは網羅しきれなかった障害を可視化することができるようになる．

さらに，AS の管理組織や所在地は，RIR の whois データベースへの問い合わせからデータを取得したが，RIR へ登録されている AS の所在地が実際の BGP ルータの設置されている所在地とは限らず，地図へのマッピングには本来の AS の位置との誤差を含んでいる．実際の AS の所在地は各 AS を管理している組織の機密であるため公開されることは少なく，所在地の誤差を修正することは難しい．しかし，より詳細な所在地データを取得することができた場合，世界規模でのマッピングののちに，より狭域な地域のマッピングも可能となる．

今回使用した AS の位置情報は，2017 年 12 月 1 日現在での情報であり，稼働している AS は時間の経過とともに，ISP から RIR へ AS 番号の管理を返還されている場合がある．この場合，一度取得した AS の管理組織や所在地などのデータを更新する必要がある．

4. 参考

参考文献

- [1] RFC4271, <https://tools.ietf.org/html/rfc4271>
- [2] あきみち・空閑洋平 2001 「インターネットのカタチ もろさが織りなす粘り強い世界」オーム社出版
- [3] BGPmon, <https://bgpmon.net>
- [4] BGPlay, <http://bgplay.routeviews.org>
- [5] BGPView, <http://www.bgpview.org/>
- [6] Esri 社 API, findAddressCandidate, <https://developers.arcgis.com/rest/geocode/api-reference/geocoding-find-address-candidates.html>
- [7] CraftMap, <http://www.craftmap.box-i.net>

行列式点過程がつくる無線干渉場の巨視的極限

油山 拓生 三好 直人
東京工業大学大学院 情報理工学研究科

概要： d 次元 Euclid 空間上に、行列式点過程に従う、点同士の間には斥力が働いて配置されるランダムな点配置があり、その配置の各点（基地局）から信号が発信されているモデルを考える。本稿では、空間上の任意の点における干渉信号（全無線通信基地局からの受信信号の総和）について、ズームアウトにより得られる巨視的な極限の導出を行う。ここでズームアウトとは、基地局の強度（密度）と信号の強さに関わるスケーリングパラメータ $\rho \in (0, 1)$ を減少させ、強度をより大きく、信号の強さをより弱くさせる操作を指す。結果として、巨視的には、点同士の間に関係のない Poisson 点過程を基に考えた場合と同様の振舞いを示すことがわかった。

1. はじめに

無線通信ネットワークの性能を評価する上で、干渉信号の強さは重要な問題である。2020 年には通信トラフィックが現在の約 1000 倍に増大することが予想されており、次世代の 5G ネットワークでは、現在の無線通信ネットワークより遥かに基地局の高密度化が進んだ超高密度ネットワーク (UDN: Ultra-dence networks) が必要とされている [1]。UDN では基地局が大量にあるため、干渉信号の強さはより興味深い対象となる。無線通信ネットワークは、基地局の配置という不規則性と、無線電波の回折・散乱等による変動（フェイディング、シャドウイング）という不規則性をはらんでいるため、しばしば、（マーク付き）空間点過程を用いてモデル化・解析されている [2, 3, 4]。本稿では、Poisson 点過程と行列式点過程を用いた 2 つのモデルにおいて、点（基地局）の強度（密度）を無限大に、各信号の強さを 0 に近づけた場合の、干渉信号の強さ（無線干渉場）に関するスケーリング極限をそれぞれ導出する。ただし本稿では、干渉信号の強さは全ての基地局からの受信信号の強さの総和であるとする。

類似のモデルに、Random balls model と呼ばれるものがある。これは、 $\mathbb{R}^d \times \mathbb{R}_+$ 上の点過程 Φ が生成する、中心 $x \in \mathbb{R}^d$ 、半径 $r \in \mathbb{R}_+$ の閉球 $B(x, r) = \{y \in \mathbb{R}^d \mid \|y - x\| \leq r\}$ の族 $\mathcal{B}$ のことを指す。このモデルは、1 次元の場合、半球 $[x, x+r]$ を接続セッション（時刻 x から r だけ通話）と見なすことで通信ネットワークにおけるトラフィックを表現することができ、2 次元の場合は、点 x を基地局、閉球 $B(x, r)$ を基地局の被覆領域と見なすことで、これも無線通信ネットワークを表現できる。更に 3 次元の場合は、多孔質媒体の表現に用いられる。このモデルと本稿で述べるモデルの違いは、各点のマークとして領域を持つか信号の強さ（スカラー）を持つかという点である。

Random balls model において、スケーリングにより得られるランダム場は、点配置が Poisson 点過程に従う場合 [5, 6, 7]、行列式点過程に従う場合 [8] 共に極限が導出されており、閉球の体積が分散を持つか否かと、持たない場合は更に強度と半径の極限の飛ばし方によって挙動が変化するということが示されている。また、本稿で考えるモデルにおいては、点配置が Poisson 点過程に従い、更に信号の伝搬損失関数（距離による信号の減衰を表す関数） $\ell: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ が $\ell(r) = r^{-\beta}$ という形の場合についての導出がなされている [9]。この形の伝搬損失関数は、無線通信ネットワークの性能評価をする上で、計算が容易になるためしばしば用いられるが、基地局に近づけば近づくほど信号の強さが青天井で増幅されることを意味してしまうという短所がある。本稿では伝搬損失関数は $\int_{\mathbb{R}_+} \ell(r) r^{d-1} dr < \infty$ を満たすような、基地局付近での過度な増幅がないものとして解析を行う。

本稿の構成は次の通りである。第 2 節で無線干渉場の定義と説明を行い、第 3 節で行列式点過程の定義と証明で用いるいくつかの性質を紹介する。そして第 4 節でスケーリングの定式化と無線干渉場のスケーリング極限に関する結果を紹介し、第 5 節でその証明を行う。

2. 無線干渉場

$\Phi = \{(X_i, H_i)\}_{i \in \mathbb{N}}$ を $\mathbb{R}^d \times \mathbb{R}_+$ 上の定常点過程とする。すなわち、 X_i は $\mathbb{R}^d$ 値、 H_i は $\mathbb{R}_+$ 値をとる確率変数で、それぞれ基地局の位置と、その基地局が発信する信号の強さに変動（フェイディング

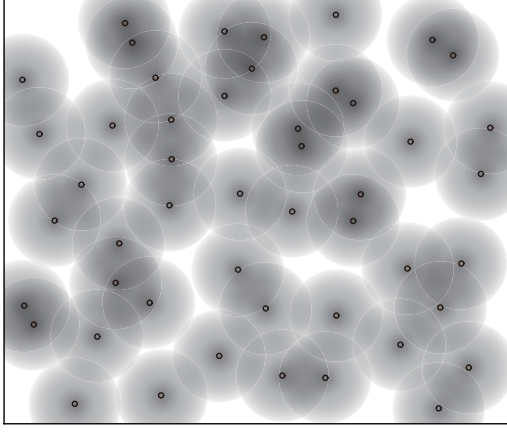


図 1: 無線干渉場

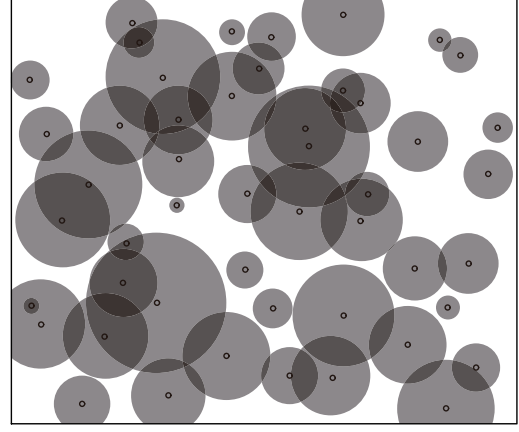


図 2: Random balls model

等) がかったものに対応している. H_i は, i.i.d. かつ Φ と独立であるとし, 分布 F (密度 f) に従うと仮定する. 各基地局から発信された電波は距離に応じてだんだん弱くなり, その減衰を伝搬損失関数 $\ell: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ で表す. ℓ は単調減少であり, べき乗則 $\ell(r) \sim cr^{-\alpha}$ ($c > 0$, $\alpha > 2$) に従うことが経験的に知られている. 各基地局から発信される電波の周波数帯域が同じであるとする, 通信相手以外の全ての基地局が干渉源となるため, $y \in \mathbb{R}^d$ における干渉信号の強さ (無線干渉場) を次の形で定義する.

$$I(\ell, y) := \sum_{(X_i, H_i) \in \Phi} H_i \ell(\|X_i - y\|) = \int_{\mathbb{R}^d \times \mathbb{R}_+} h \ell(\|x - y\|) \Phi(dx, dh) \quad (1)$$

3. 行列式点過程

本節では行列式点過程の定義と性質を与える. 詳細は [10, 8] 参照.

Φ を $\mathbb{R}^d$ 上の単純な点過程であるとし, $\rho_n: (\mathbb{R}^d)^n \rightarrow \mathbb{R}_+$, $n \in \mathbb{N}$ を Φ の Lebesgue 測度上の結合強度とする. すなわち, 任意の連続かつ対称でコンパクトな台を持つ関数 $f: (\mathbb{R}^d)^n \rightarrow \mathbb{R}$ に対して次が成り立つとする.

$$\mathbb{E} \left[\sum_{X_1, \dots, X_n \in \Phi}^{\neq} f(X_1, \dots, X_n) \right] = \int_{(\mathbb{R}^d)^n} f(x_1, \dots, x_n) \rho_n(x_1, \dots, x_n) dx_1 \cdots dx_n$$

定義 3.1 (行列式点過程). 結合強度 ρ_n が核 $K: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{C}$ を用いて,

$$\rho_n(x_1, \dots, x_n) = \det(K(x_i, x_j))_{1 \leq i, j \leq n}$$

と表される点過程 Φ を, 核 K を持つ行列式点過程と呼ぶ

核 K を持つ行列式点過程において, $\rho_2(x, x) = 0$ より, K が連続であると仮定すると, $x_1 \simeq x_2$ のとき $\rho_2(x_1, x_2) \simeq 0$ となるため, 行列式点過程は点同士の反発がある配置をサンプルパスに持つことが観察される.

本稿で扱う行列式点過程は核がある特別な性質を満たすものとする. その性質を述べるためにいくつかの準備をする. まず, 内積 $\langle \cdot, \cdot \rangle$ を備えた Hilbert 空間 H 上のコンパクト作用素 T に対して, トレース Tr を次で定める. ただし, $(e_n)_{n \in \mathbb{N}}$ は H の正規直交基底である.

$$\text{Tr}(T) := \sum_{n=1}^{\infty} \langle T e_n, e_n \rangle$$

更に T が次を満たすとき, T をトレースクラス作用素と呼ぶ. ここで $|T| = \sqrt{T^* T}$ で, T^* は T の Hermite 共役である.

$$\|T\| := \text{Tr}(|T|) < \infty$$

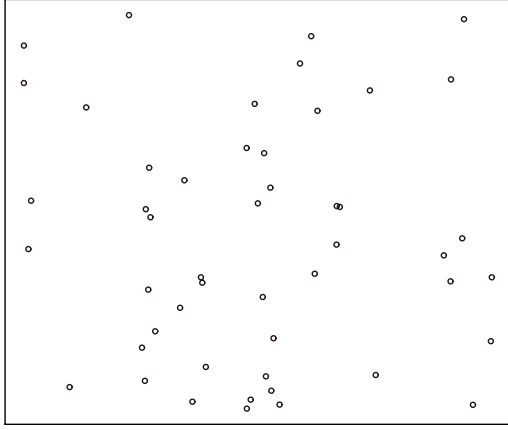


図 3: Poisson 点過程のサンプルパス

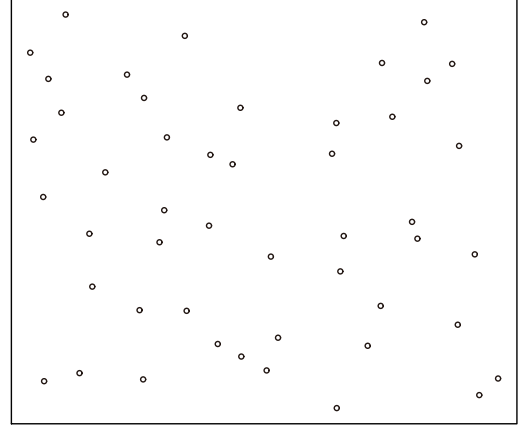


図 4: 行列式点過程のひとつである Ginibre 点過程のサンプルパス

$\|T\|_1 < 1$ となるトレースクラス作用素 T に対して, $I + T$ の Fredholm 行列式を次で定める.

$$\text{Det}(I + T) := \exp \left(\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \text{Tr}(T^n) \right) \quad (2)$$

$$= \sum_{n=1}^{\infty} \frac{1}{n!} \int_{(\mathbb{R}^d)^n} \det((T(x_i, x_j))_{1 \leq i, j \leq n}) dx_1 \cdots dx_n \quad (3)$$

$\forall x \in \mathbb{R}^d$ と $\forall f \in L^2(\mathbb{R}^d)$ に対して, 次の積分変換写像 $\mathbb{K}$ を導入する.

$$\mathbb{K}f(x) := \int_{\mathbb{R}^d} K(x, y) f(y) dy \quad (4)$$

本稿で扱う行列式点過程は次の仮定を満たすとする.

仮定 3.1. $\mathbb{K} : L^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R}^d)$ は有界 Hermite 積分作用素であり, スペクトルが $[0, 1]$ に含まれる. 更に $\mathbb{K}$ は局所トレースクラス作用素である. すなわち任意のコンパクト集合 $\Lambda \subset \mathbb{R}^d$ に対して, 制限 $\mathbb{K}_\Lambda := \int_\Lambda K(x, y) f(y) dy$ がトレースクラス作用素である.

K が仮定 3.1 を満たす写像 $\mathbb{K}$ の核であるとき $x \mapsto K(x, x)$ は非負であり, 更に本稿で重要になる Laplace 汎関数に関する次の定理が成り立つ.

定理 3.1 ([10] の Th.1.2 の特殊ケース). $\mathbb{K}$ が仮定 3.1 を満たす $L^2(\mathbb{R}^d)$ 上の積分作用素であるとき, 行列式点過程 ϕ が一意に存在し, 任意のコンパクトな台を持つ $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ に対して次が成り立つ.

$$\mathbb{E} \left[\exp \left(- \int f(x) \phi(dx) \right) \right] = \text{Det} \left(I - \mathbb{K}[1 - e^{-f}] \right) \quad (5)$$

ここで, $\mathbb{K}[1 - e^{-f}]$ は核 $K[1 - e^{-f}](x, y) = \sqrt{1 - \exp(-f(x))} K(x, y) \sqrt{1 - \exp(-f(y))}$ を持つ作用素である.

この定理における, f がコンパクトな台を持つという条件は緩和することが可能である.

定理 3.2 ([11] の Lam.2). $\mathbb{K}$ が仮定 3.1 を満たす $L^2(\mathbb{R}^d)$ 上の積分作用素であるとき, 行列式点過程 ϕ が一意に存在し, 次の条件を満たす $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ に対して (5) 式が成り立つ.

$$\lim_{|x| \rightarrow \infty} f(x) = 0, \quad \lim_{r \rightarrow \infty} \int_{\mathbb{R}^d \setminus B(0, r)} K(x, x) f(x) dx = 0, \quad \int_{\mathbb{R}^d} K(x, x) (1 - e^{-f(x)}) dx < \infty$$

系として次が得られ, 行列式点過程の強度関数が $K(x, x)$ で与えられることがわかる.

系 3.1. 行列式点過程 ϕ の核 K が仮定 3.1 を満たす作用素を持つとき, 任意のコンパクト集合 $\Lambda \in \mathbb{R}^d$ と任意の非負関数 f に対して,

$$\mathbb{E} \left[\int_{\Lambda} f(x) \phi(dx) \right] = \int_{\Lambda} f(x) K(x, x) dx = \text{Tr}(K_{\Lambda}[f])$$

以下の命題はマーク付き行列式点過程もまた行列式点過程であることを示している.

命題 3.1 (マーク付き行列式点過程. [8] の Prop.A.6). $\phi = \{X_i\}_{i \in \mathbb{N}}$ を核 K を持つ $\mathbb{R}^d$ 上の行列式点過程とし, $\{H_i\}_{i \in \mathbb{N}}$ を i.i.d. で ϕ と独立で密度 f を持つ $\mathbb{R}_+$ 値確率変数の族とする. このとき, $\Phi = \{(X_i, H_i)\}_{i \in \mathbb{N}}$ も $\mathbb{R}^d \times \mathbb{R}_+$ 上の行列式点過程で, 核 $\hat{K}((x, h), (x', h')) = \sqrt{f(h)} K(x, x') \sqrt{f(h')}$ を持つ.

命題 3.2 ([8] の Prop.A.7). $\mathbb{K}$ が仮定 3.1 を満たすとき, $\hat{\mathbb{K}}$ も仮定 3.1 を満たす.

最後に定理 3.1 に現れる $K[1 - e^f]$ についての命題を挙げる.

命題 3.3 ([8] の Prop.A.9). K を仮定 3.1 を満たす作用素の核とし, $g: \mathbb{R}^d \rightarrow \mathbb{R}_+$ をコンパクトな台を持つ有界関数とすると,

$$K[g](x, y) := \sqrt{g(x)} K(x, y) \sqrt{g(y)}$$

は Hilbert-Schmidt 作用素の核である.

4. スケーリングと結果

4.1. Poisson 点過程の場合

$\Phi = \{(X_i, H_i)\}_{i \in \mathbb{N}}$ を $\mathbb{R}^d \times \mathbb{R}_+$ 上の強度測度 $\lambda dx f(h) dh$ を持つマーク付き Poisson 点過程とする. またスケーリングパラメータ $\rho \in (0, 1]$ を導入し, Φ_{ρ} を $\mathbb{R}^d \times \mathbb{R}_+$ 上の強度測度 $\lambda_{\rho} \lambda dx \frac{f(h/\rho)}{\rho} dh$ を持つマーク付き Poisson 点過程とする. ただし $\lim_{\rho \rightarrow 0} \lambda_{\rho} = \infty$, $\lambda_1 = 1$ とする. Φ_{ρ} についての無線干渉場 $I_{\rho}(\ell, y)$ は次で与えられる.

$$I_{\rho}(\ell, y) := \int_{\mathbb{R}^d \times \mathbb{R}_+} h \ell(\|x - y\|) \Phi_{\rho}(dx, dh) \quad (6)$$

また, $\mathbb{E}[H] < \infty$, $\int_{\mathbb{R}_+} \ell(r) r^{d-1} dr < \infty$ のとき, $\mathbb{E}[I_{\rho}(\ell, y)]$ が存在し次の形で表される. [9] では ℓ がこの条件を満たさない形 ($\ell(r) = r^{-\beta}$, $\beta > d$) であるため, $\mathbb{E}[I_{\rho}(\ell, y)]$ は存在しない.

$$\mathbb{E}[I_{\rho}(\ell, y)] = \int_{\mathbb{R}^d \times \mathbb{R}_+} h \ell(\|x - y\|) \lambda_{\rho} \lambda dx \frac{f(h/\rho)}{\rho} dh \quad (7)$$

$I_{\rho}(\ell, y)$ の偏差のスケーリング極限について次が得られた. なお証明は 5.1 節で行う. [9] では $\mathbb{E}[I_{\rho}(\ell, y)]$ が存在しないため, 偏差ではなく $I_{\rho}(\ell, y)$ そのもののスケーリング極限を導出している.

定理 4.1. $\mathbb{E}[H^2] < \infty$ のとき, $\int_{\mathbb{R}_+} \ell(r)^2 r^{d-1} dr < \infty$ を満たす ℓ に対して, 有限次元分布の意味で,

$$\frac{I_{\rho}(\ell, y) - \mathbb{E}[I_{\rho}(\ell, y)]}{\rho \lambda_{\rho}^{1/2}} \rightarrow W(\ell) \quad \text{as } \rho \rightarrow 0 \quad (8)$$

が成り立つ. ここで, $W(\ell) = \int_{\mathbb{R}^d} \ell(\|x\|) M_2(dx)$ である. M_2 は $\lambda \mathbb{E}[H^2] dx$ を control measure に持つ Gauss ランダム測度で, $A \in \mathcal{B}(\mathbb{R}^d)$ に対して, $M_2(A) \sim \mathcal{N}(0, \lambda \mathbb{E}[H^2] |A|)$ を満たす.

定理 4.2. $f(h) \sim C_{\beta} h^{-\beta-1}$ as $h \rightarrow \infty$ ($\beta \in (1, 2)$) のとき, $\int_{\mathbb{R}_+} \ell(r)^{\beta} r^{d-1} dr < \infty$ を満たす ℓ に対して, 有限次元分布の意味で,

$$\frac{I_{\rho}(\ell, y) - \mathbb{E}[I_{\rho}(\ell, y)]}{\rho \lambda_{\rho}^{1/\beta}} \rightarrow Z(\ell) \quad \text{as } \rho \rightarrow 0 \quad (9)$$

が成り立つ. ここで,

$$Z(\ell) = \int_{\mathbb{R}^d} \ell(\|x\|) M_{\beta}(dx)$$

である. M_β は $\lambda C_\beta(\int_0^\infty (1 - \cos r)r^{-1-\beta}dr)dx$ を *control measure* に持つ β -安定ランダム測度で, $A \in \mathcal{B}(\mathbb{R}^d)$ に対して,

$$M_\beta(A) \sim S_\beta \left(\left(\lambda C_\beta \int_0^\infty \frac{1 - \cos r}{r^{1+\beta}} dr |A| \right)^{\frac{1}{\beta}}, 1, 0 \right)$$

を満たす. ただし $X \sim S_\beta(\sigma, \gamma, \delta)$ であるとき, $\mathbb{E}[\exp(i\theta X)] = \exp(i\delta\theta - \sigma|\theta|^\beta(1+i\gamma\text{sgn}(\theta)\tan(\pi\beta/2)))$ が成り立つ [13].

4.2. 行列式点過程の場合

本稿では, 核 K は 2 点間の距離のみに依存するとする. すなわち $|K(x, y)| = |K(x - y)|$ と書けるとし, $\lambda = K(0)$ する. 更に K は仮定 3.1 を満たすものとする.

$\Phi = \{(X_i, H_i)\}_{i \in \mathbb{N}}$ を $\mathbb{R}^d \times \mathbb{R}_+$ 上の核 $\hat{K}((x, h), (x', h')) = \sqrt{f(h)}K(x, x')\sqrt{f(h')}$ を持つマーク付き行列式点過程とする. またスケールパラメータ $\rho \in (0, 1)$ を導入し, Φ_ρ を $\mathbb{R}^d \times \mathbb{R}_+$ 上の核 $\hat{K}_\rho((x, h), (x', h')) = \sqrt{\frac{f(h/\rho)}{\rho}}K_\rho(x, x')\sqrt{\frac{f(h'/\rho)}{\rho}}$ を持つマーク付き行列式点過程とする. ただし $\lim_{\rho \rightarrow 0} \lambda_\rho = \infty$, $K_\rho(0) = \lambda_\rho K(0)$ とする. このとき, Φ_ρ についての無線干渉場 $I_\rho(\ell, y)$ はやはり (6) 式で与えられる. また, $\mathbb{E}[H] < \infty$, $\int_{\mathbb{R}_+} \ell(r)r^{d-1}dr < \infty$ のとき, $\mathbb{E}[I_\rho(\ell, y)]$ が存在し次の形で表される.

$$\mathbb{E}[I_\rho(\ell, y)] = \int_{\mathbb{R}^d \times \mathbb{R}_+} h\ell(\|x - y\|)K_\rho(0)dx \frac{f(h/\rho)}{\rho}dh \quad (10)$$

$I_\rho(\ell, y)$ の偏差のスケール極限について, Poisson 点過程の場合と同様の次の結果が得られた. 証明は 5.2 節で行う.

定理 4.3. $\mathbb{E}[H^2] < \infty$ かつ $\exists \epsilon > 0$ について $\int_{\mathbb{R}^d} |K_\rho(x)|^2 dx = O(\lambda_\rho^{1-\epsilon})$ as $\rho \rightarrow 0$ であるとき, $\int_{\mathbb{R}_+} \ell(r)^2 r^{d-1} dr < \infty$ かつ $\int_{\mathbb{R}_+} (1 - e^{\ell(r)})r^{d-1} dr < \infty$ を満たす ℓ に対して, 有限次元分布の意味で,

$$\frac{I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)]}{\rho\lambda_\rho^{1/2}} \rightarrow W(\ell) \quad \text{as } \rho \rightarrow 0 \quad (11)$$

が成り立つ. ここで, $W(\ell) = \int_{\mathbb{R}^d} \ell(\|x\|)M_2(dx)$ である. M_2 は定理 4.1 のものと同じである.

定理 4.4. $f(h) \sim C_\beta h^{-\beta-1}$ as $h \rightarrow \infty$ ($\beta \in (1, 2)$) かつ $\int_{\mathbb{R}^d} |K_\rho(x)|^2 dx = O(\lambda_\rho)$ as $\rho \rightarrow 0$ であるとき, $\int_{\mathbb{R}_+} \ell(r)^2 r^{d-1} dr < \infty$ かつ $\int_{\mathbb{R}_+} (1 - e^{\ell(r)})r^{d-1} dr < \infty$ を満たす ℓ に対して, 有限次元分布の意味で,

$$\frac{I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)]}{\rho\lambda_\rho^{1/\beta}} \rightarrow Z(\ell) \quad \text{as } \rho \rightarrow 0 \quad (12)$$

が成り立つ. ここで, $Z(\ell) = \int_{\mathbb{R}^d} \ell(\|x\|)M_\beta(dx)$ で, M_β は定理 4.2 のものと同じである.

4.3. Random balls model との比較

Random balls model では次の量に関するスケール極限が調べられている [5, 6, 7, 8]. ここで, μ は有限な総変動を持つ符号付き測度で, v_d は d 次元単位球の体積である.

$$M(\mu) := \int_{\mathbb{R}^d \times \mathbb{R}_+} \mu \left(\frac{v}{v_d} B(0, 1) + x \right) \Phi(dx, dv) \quad (13)$$

結果として, $\mathbb{E}[H^2] < \infty$ の場合は Random balls model と同じ結果が, $f(h) \sim C_\beta h^{-\beta-1}$ as $h \rightarrow \infty$ ($\beta \in (1, 2)$) の場合は Random balls model での small-balls scaling ($\lambda_\rho \rho^\beta \rightarrow 0$ の場合) と同じ結果が得られた. 後者で, Random balls model とは異なり $\lambda_\rho \rho^\beta$ の挙動による分岐が起こらなかった原因として, 無線干渉場では, (信号の強さ) $\times$ (伝搬損失関数) という形の総和をとっているのに対し, Random balls model では, 球を μ で測ったものの総和をとっているために, 球の体積が大きい場合に μ が効くことで $M(\mu)$ への寄与が抑えられることがありと推測される.

5. 証明

本節では、定理 4.1-4.4 の証明を行う。

5.1. Poisson 点過程の場合

定理 4.1, 4.2 の証明は、 $C_\rho^{-1}(I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)])$ ($C_\rho = \rho\lambda_\rho^{1/2}$ または $\rho\lambda_\rho^{1/\beta}$) の特性汎関数を用いて行うため、まずはそのための準備を行う。

強度測度 η を持つ Poisson 点過程 Φ と、決定的な関数 g について、積分 $\int g(d\Phi - d\eta)$ を考える。この積分は $\int (|g| \wedge g^2) d\eta < \infty$ のとき存在することが知られており、その特性汎関数は $\Psi(u) = e^{iu} - 1 - iu$ を用いて、

$$\mathbb{E} \left[\exp \left(i \int g(d\Phi - d\eta) \right) \right] = \exp \left(\int (\Psi \circ g) d\eta \right) \quad (14)$$

と表される [12]。 Ψ については、 $\forall u \in \mathbb{R}$ に対して、

$$|\Psi(u)| \leq (2|u| \wedge |u|^2/2), \quad |\Psi(u) + u^2/2| \leq (u^2 \wedge |u|^3/6) \quad (15)$$

という不等式が成り立つ。また、 $f(h) \sim C_\beta h^{-\beta-1}$, ($h \rightarrow \infty$) ($\beta > 0$) なる $\mathbb{R}_+$ 上の確率密度関数 f に対して、次の補題が成り立つ。

補題 5.1 ([6] の Lem.2). $\mathbb{R}_+$ 上の連続関数 g が、 $0 < p < \beta < q$ に対して、

$$\limsup_{h \rightarrow \infty} h^{-p} |g(h)| < \infty, \quad \limsup_{h \rightarrow 0} h^{-q} |g(h)| < \infty$$

を満たすとき次が成り立つ。

$$\int_{\mathbb{R}_+} g(h) \frac{f(h/\rho)}{\rho} dh \sim C_\beta \rho^{-\beta} \int_{\mathbb{R}_+} g(h) h^{-\beta-1} dh \quad \text{as } \rho \rightarrow 0 \quad (16)$$

5.1.1. 定理 4.1 の証明

(14) 式より、 $(\rho\lambda_\rho^{1/2})^{-1}(I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)])$ の特性汎関数は、

$$\begin{aligned} \mathbb{E} \left[\exp \left(i \frac{I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)]}{\rho\lambda_\rho^{1/2}} \right) \right] &= \exp \left(\int_{\mathbb{R}^d \times \mathbb{R}_+} \Psi(\rho^{-1}\lambda_\rho^{-1/2}h\ell(\|x-y\|)) \lambda_\rho \lambda dx \frac{f(h/\rho)}{\rho} dh \right) \\ &= \exp \left(\int_{\mathbb{R}^d \times \mathbb{R}_+} \Psi(\lambda_\rho^{-1/2}h\ell(\|x-y\|)) \lambda_\rho \lambda dx f(h) dh \right) \\ &\stackrel{\rho \rightarrow 0}{\sim} (a) \exp \left(-\frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}_+} h^2 \ell(\|x-y\|)^2 \lambda dx f(h) dh \right) \\ &= \exp \left(-\frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}_+} \ell(\|x-y\|)^2 \lambda \mathbb{E}[H^2] dx \right) \end{aligned} \quad (17)$$

と表せる。(a) では、(15) 式と優収束定理により、 $\rho \rightarrow 0$ のとき、 $\Psi(\lambda_\rho^{-1/2}h\ell(\|x-y\|))$ を $-\frac{1}{2}\lambda_\rho^{-1}h^2\ell(\|x-y\|)^2$ で近似できることを用いている。(17) 式が $W(\ell)$ の特性汎関数であることから、 $(\rho\lambda_\rho^{1/2})^{-1}(I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)])$ が $\rho \rightarrow 0$ において有限次元分布の意味で $W(\ell)$ に収束することが示された。

5.1.2. 定理 4.2 の証明

5.1.1 節と同様に、 $(\rho\lambda_\rho^{1/\beta})^{-1}(I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)])$ の特性汎関数を計算すると、

$$\begin{aligned} \mathbb{E} \left[\exp \left(i \frac{I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)]}{\rho\lambda_\rho^{1/\beta}} \right) \right] &= \exp \left(\int_{\mathbb{R}^d \times \mathbb{R}_+} \Psi(\rho^{-1}\lambda_\rho^{-1/\beta}h\ell(\|x-y\|)) \lambda_\rho \lambda dx \frac{f(h/\rho)}{\rho} dh \right) \\ &= \exp \left(\int_{\mathbb{R}^d \times \mathbb{R}_+} \Psi(h\ell(\|x-y\|)) \lambda_\rho \lambda dx \lambda_\rho^{1/\beta} f(\lambda_\rho^{1/\beta}h) dh \right) \\ &\stackrel{\rho \rightarrow 0}{\sim} (b) \exp \left(\int_{\mathbb{R}^d \times \mathbb{R}_+} \Psi(h\ell(\|x-y\|)) \lambda dx C_\beta h^{-\beta-1} dh \right) \end{aligned} \quad (18)$$

と表せる. (b) では, 補題 5.1 で $p = 1, q = 2$ とすることにより, $\rho \rightarrow 0$ のとき,

$\int_{\mathbb{R}^d \times \mathbb{R}_+} \Psi(h\ell(\|x - y\|)) \lambda_\rho \lambda dx \lambda_\rho^{1/\beta} f(\lambda_\rho^{1/\beta} h) dh$ が $\int_{\mathbb{R}^d \times \mathbb{R}_+} \Psi(h\ell(\|x - y\|)) \lambda dx C_\beta h^{-\beta-1} dh$ で近似できることを用いている. (18) 式は更に [13] の Exercise 3.24 により次のように変形できる.

$$\begin{aligned} & \exp \left(\int_{\mathbb{R}^d} |\ell(\|x - y\|)|^\beta \frac{\Gamma(2 - \beta)}{\beta(\beta - 1)} \cos \frac{\pi\beta}{2} \left(1 - i \operatorname{sgn}(\ell(\|x - y\|)) \tan \frac{\pi\beta}{2} \right) \lambda C_\beta dx \right) \\ &= \exp \left(\int_{\mathbb{R}^d} |\ell(\|x - y\|)|^\beta \left(1 - i \tan \frac{\pi\beta}{2} \right) \lambda C_\beta \left(\int_0^\infty \frac{1 - \cos r}{r^{1+\beta}} dr \right) dx \right) \end{aligned} \quad (19)$$

(19) 式が $Z(\ell)$ の特性汎関数である ([13] の Chapter.3) ことから, $(\rho \lambda_\rho^{1/\beta})^{-1} (I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)])$ が $\rho \rightarrow 0$ において有限次元分布の意味で $Z(\ell)$ に収束することが示された.

5.2. 行列式点過程の場合

定理 4.3, 4.4 の証明は, $C_\rho^{-1} (I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)])$ ($C_\rho = \rho \lambda_\rho^{1/2}$ または $\rho \lambda_\rho^{1/\beta}$) の Laplace 変換 $\mathcal{L}[C_\rho^{-1} (I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)])](\theta)$ を用いて行う. そこで定理 3.1 が鍵を握るのだが, これは被積分関数がコンパクトな台を持つ場合でないと適用できない. そのためまずは次の $I_\rho^M(\ell, y)$ を導入する.

$$I_\rho^M(\ell, y) := \int_{\mathbb{R}^d \times \mathbb{R}_+} h \mathbb{1}_{[0, M]}(h) \ell(\|x - y\|) \Phi_\rho(dx, dh) \quad (20)$$

$h \mathbb{1}_{[0, M]}(h)$ が $\mathbb{R}_+$ 上コンパクトな台を持ち, 仮定より ℓ も定理 3.2 の条件を満たすため, $I_\rho^M(\ell, y)$ については定理 3.1 を用いることができる.

$$\lim_{\rho \rightarrow 0} \mathcal{L}[C_\rho^{-1} (I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)])] = \lim_{\rho \rightarrow 0} \lim_{M \rightarrow \infty} \mathcal{L}[C_\rho^{-1} (I_\rho^M(\ell, y) - \mathbb{E}[I_\rho^M(\ell, y)])]$$

となるが, $\lim_{M \rightarrow \infty} \mathcal{L}[C_\rho^{-1} (I_\rho^M(\ell, y) - \mathbb{E}[I_\rho^M(\ell, y)])]$ が実は ρ によらないため, $C_\rho^{-1} (I_\rho(\ell, y) - \mathbb{E}[I_\rho(\ell, y)])$ の Laplace 変換の $\rho \rightarrow 0$ の極限が求まる.

5.2.1. 定理 4.3 の証明

$(\rho \lambda_\rho^{1/2})^{-1} (I_\rho^M(\ell, y) - \mathbb{E}[I_\rho^M(\ell, y)])$ の Laplace 変換は,

$$\mathbb{E} \left[\exp \left(-\theta \frac{I_\rho^M(\ell, y) - \mathbb{E}[I_\rho^M(\ell, y)]}{\rho \lambda_\rho^{1/2}} \right) \right] = \exp \left(\theta \mathbb{E}[(\rho \lambda_\rho^{1/2})^{-1} I_\rho^M(\ell, y)] \right) \mathbb{E} \left[\exp \left(-\theta (\rho \lambda_\rho^{1/2})^{-1} I_\rho^M(\ell, y) \right) \right]$$

と表される. 右辺の第 2 項について, $g_{\ell, y}^M(x, h) = h \mathbb{1}_{[0, M]}(h) \ell(\|x - y\|)$ とすると定理 3.1 から,

$$\begin{aligned} \mathbb{E} \left[\exp \left(-\theta (\rho \lambda_\rho^{1/2})^{-1} I_\rho^M(\ell, y) \right) \right] &= \operatorname{Det} \left(I - \hat{K}_\rho \left[1 - e^{-\theta (\rho \lambda_\rho^{1/2})^{-1} g_{\ell, y}^M} \right] \right) \\ &= \exp \left(-\sum_{n=1}^{\infty} \frac{1}{n} \operatorname{Tr} \left(\hat{K}_\rho \left[1 - e^{-\theta (\rho \lambda_\rho^{1/2})^{-1} g_{\ell, y}^M} \right]^n \right) \right) \end{aligned}$$

となり, 特に和の第 1 項について, 系 3.1 から,

$$\operatorname{Tr} \left(\hat{K}_\rho \left[1 - e^{-\theta (\rho \lambda_\rho^{1/2})^{-1} g_{\ell, y}^M} \right] \right) = \int_{\mathbb{R}^d \times \mathbb{R}_+} \left(1 - e^{-\theta (\rho \lambda_\rho^{1/2})^{-1} g_{\ell, y}^M(x, h)} \right) K_\rho(x, x) \frac{f(h/\rho)}{\rho} dx dh \quad (21)$$

が成り立ち, このため次が成り立つ.

$$\begin{aligned} & \exp \left(\theta \mathbb{E}[(\rho \lambda_\rho^{1/2})^{-1} I_\rho^M(\ell, y)] \right) \exp \left(-\operatorname{Tr} \left(\hat{K}_\rho \left[1 - e^{-\theta (\rho \lambda_\rho^{1/2})^{-1} g_{\ell, y}^M} \right] \right) \right) \\ &= \exp \left(\int_{\mathbb{R}^d \times \mathbb{R}_+} \psi(\theta \rho^{-1} \lambda_\rho^{-1/2} g_{\ell, y}^M(x, h)) \lambda_\rho K(0) dx \frac{f(h/\rho)}{\rho} dh \right) \end{aligned}$$

ここで, $\psi(u) = \Psi(iu) = e^{-u} - 1 + u$ である. これは Φ は Poisson 点過程の場合の $(\rho\lambda_\rho^{1/2})^{-1}(I_\rho^M(\ell, y) - \mathbb{E}[I_\rho^M(\ell, y)])$ の Laplace 変換と M の存在を除いて等しく, Poisson 点過程の場合と同様に変形することによって ρ によらない形になり, $M \rightarrow \infty$ とすると同様の結果が得られる.

従って, 残りの $\exp\left(-\sum_{n=2}^{\infty} \frac{1}{n} \text{Tr}\left(\hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M}\right]^n\right)\right)$ が 1 に収束することを示せば証明が完了する. 次の 2 つの補題が成り立つ.

補題 5.2. $\forall n \geq 2$ について,

$$\text{Tr}\left(\hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M}\right]^n\right) \leq \text{Tr}\left(\hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M}\right]^2\right)^{n/2}$$

Proof. [8] の Lem.2.8 と全く同様である. □

補題 5.3. $\exists \rho^* > 0$ と $\exists C_K \in (0, \infty)$ が存在し, $\forall \rho \in (0, \rho^*)$ について,

$$\text{Tr}\left(\hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M}\right]^2\right) \leq C_K \left(\int_{\mathbb{R}^d} \ell(\|x - y\|)^2 dx\right) (\mathbb{E}[H])^2 \theta^2 \lambda_\rho^{-\epsilon}$$

Proof. 紙面の都合上概略のみ記述する.

作用素 $\hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M}\right]^2$ は次の核を持つ.

$$\begin{aligned} & \hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M}\right]^2((x, h), (x', h')) \\ &= \int_{\mathbb{R}^d \times \mathbb{R}_+} \hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M}\right]((x, h), (x'', h'')) \hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M}\right]((x'', h''), (x', h')) dx'' dh'' \\ &= \sqrt{1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M(x, h)}} \sqrt{\frac{f(h/\rho)}{\rho}} \\ & \quad \times \left(\int_{\mathbb{R}^d \times \mathbb{R}_+} (1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M(x'', h'')}) \sqrt{\frac{f(h''/\rho)}{\rho}} K_\rho(x, x'') K_\rho(x'', x') dx'' dh'' \right) \\ & \quad \times \sqrt{1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M(x', h')}} \sqrt{\frac{f(h'/\rho)}{\rho}} \end{aligned}$$

この作用素のトレースは次で与えられる.

$$\begin{aligned} & \text{Tr}\left(\hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M}\right]^2\right) \\ &= \int_{\mathbb{R}^d \times \mathbb{R}_+} \hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M}\right]^2((x, h), (x, h)) dx dh \\ &= \int_{(\mathbb{R}^d \times \mathbb{R}_+)^2} (1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M(x, h)}) (1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M(x', h')}) \frac{f(h/\rho)}{\rho} \frac{f(h'/\rho)}{\rho} |K_\rho(x, x')|^2 dx dh dx' dh' \end{aligned}$$

ℓ の条件と, Cauchy-Schwarz の不等式と, $\exists \epsilon > 0$ について $\int_{\mathbb{R}^d} |K_\rho(x)|^2 dx = O(\lambda_\rho^{1-\epsilon})$ as $\rho \rightarrow 0$ から, 次が得られる.

$$\begin{aligned} & \int_{(\mathbb{R}^d \times \mathbb{R}_+)^2} (1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M(x, h)}) (1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M(x', h')}) |K_\rho(x, x')|^2 dx dx' \\ & \leq C_K \theta^2 \lambda_\rho^{-\epsilon} \rho^{-2} \mathbb{1}_{[0, M]}(h) \mathbb{1}_{[0, M]}(h') \left(\int_{\mathbb{R}^d} \ell(\|x - y\|)^2 dx \right) hh' \end{aligned}$$

$\text{Tr} \left(\hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M} \right]^2 \right)$ にこれを適用すると,

$$\text{Tr} \left(\hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M} \right]^2 \right) \leq C_K \left(\int_{\mathbb{R}^d} \ell(\|x-y\|)^2 dx \right) \left(\int_0^{M/\rho} h f(h) dh \right) \theta^2 \lambda_\rho^{-\epsilon}$$

となり, $\rho \rightarrow 0$ とすると証明が完了する. □

この2つの補題により,

$$\begin{aligned} \left| -\sum_{n=2}^{\infty} \frac{1}{n} \text{Tr} \left(\hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M} \right]^n \right) \right| &\leq \sum_{n=1}^{\infty} \frac{1}{n} \left(\sqrt{\text{Tr} \left(\hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M} \right]^2 \right)} \right)^n \\ &= -\ln \left(1 - \sqrt{\text{Tr} \left(\hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M} \right]^2 \right)} \right) \\ &\rightarrow 0 \quad \text{as } \rho \rightarrow 0 \end{aligned}$$

となり, 定理 4.3 の証明が完了する.

5.2.2. 定理 4.4 の証明

定理 4.3, 4.2 の証明とほとんど同様であるが, 補題 5.3 が次のものに置き換わる.

補題 5.4. $\exists \rho^* > 0$ と $\exists C_K \in (0, \infty)$ が存在し, $\forall \rho \in (0, \rho^*)$ について,

$$\text{Tr} \left(\hat{K}_\rho \left[1 - e^{-\theta(\rho\lambda_\rho^{1/2})^{-1}g_{\ell,y}^M} \right]^2 \right) \leq C_K \int_{\mathbb{R}^d} \ell(\|x-y\|)^2 dx (\mathbb{E}[H])^2 \theta^2 \lambda_\rho^{1-2/\beta}$$

参考文献

- [1] Osseiran, Afif, et al. "Scenarios for 5G mobile and wireless communications: the vision of the METIS project." IEEE Communications Magazine 52.5 (2014): 26-35.
- [2] Haenggi, Martin. Stochastic geometry for wireless networks. Cambridge University Press, 2012.
- [3] Miyoshi, Naoto, and Tomoyuki Shirai. "A cellular network model with Ginibre configured base stations." Advances in Applied Probability 46.3 (2014): 832-845.
- [4] Nakata, Itaru, and Naoto Miyoshi. "Spatial stochastic models for analysis of heterogeneous cellular networks with repulsively deployed base stations." Performance Evaluation 78 (2014): 7-17.
- [5] Biermé, Hermine, and Anne Estrade. "Poisson random balls: self-similarity and X-ray images." Advances in applied probability 38.4 (2006): 853-872.
- [6] Kaj, Ingemar, et al. "Scaling limits for random fields with long-range dependence." The Annals of Probability (2007): 528-550.
- [7] Biermé, Hermine, Anne Estrade, and Ingemar Kaj. "Self-similar random fields and rescaled random balls models." Journal of theoretical probability 23.4 (2010): 1110-1141.
- [8] Breton, Jean-Christophe, Adrien Clarenne, and Renan Gobard. "Macroscopic analysis of determinantal random balls." arXiv preprint arXiv:1504.04513 (2015).
- [9] Baccelli, François, and Anup Biswas. "On scaling limits of power law shot-noise fields." Stochastic Models 31.2 (2015): 187-207.
- [10] Shirai, Tomoyuki, and Yoichiro Takahashi. "Random point fields associated with certain Fredholm determinants I: fermion, Poisson and boson point processes." Journal of Functional Analysis 205.2 (2003): 414-463.

- [11] Li, Yingzhe, et al. "Statistical modeling and probabilistic analysis of cellular networks with determinantal point processes." *IEEE Transactions on Communications* 63.9 (2015): 3405-3422.
- [12] Kallenberg, Olav. *Foundations of modern probability*, 2nd ed. Springer Science & Business Media, 2002.
- [13] Samorodnitsky, Gennady, and Murad S. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*. Vol. 1. CRC press, 1994.

BGP データを使ったネットワーク障害のパターン推定

奥寺 諒

小林 香

富山県立大学 工学部 電子・情報工学科

概要：本研究では、インターネットに経路障害が起こった時、BGP メッセージからその規模を推定する手法を検討する。インターネットの経路制御を行う BGP において、経路表を作成するためにルータ間で交換される BGP メッセージの 1 種類である、経路更新情報を通知する UPDATE メッセージから、障害によって大きく変化すると考えられる 2 つのパラメータを分毎に抽出し、これらのパラメータについて分布の中心からの外れの程度を指標化し、複数の計測地点において処理した統計値で障害地点と障害規模を推定する。

1. はじめに

今日、我々が利用しているインターネットは通信サービス事業者等のネットワークが相互接続することによって構成された巨大なネットワークである。個々のネットワークは各々の統一したポリシーのもと運用されており、それらを総じて AS (Autonomous system) と呼ぶ。ピアリングしている AS 間は BGP (Border Gateway Protocol) [1] メッセージによって互いの状態や経路情報をやりとりし、それに基づいて最適経路の決定を行う。しかし、BGP では間違った情報や意図的に流された偽の情報も受け取ってしまうため、許容できないレベルの遅延が発生する経路を採用してしまう等の経路障害が発生することがある。

BGP に関する障害検知の一例として、BFD (Bidirectional Forwarding Detection) [2] が挙げられる。BFD は、ピア間で数十～数百 msec の間隔でパケット交換を行うことにより障害を検知するシステムである。しかし BFD では、ピアの生存以外の情報を得られないので、経路障害には対応できない。

経路障害が発生してしまった時には、適切な対処のために障害の発生地点と規模を把握しなければならないが、一つの AS で得られる情報からでは難しい。したがって本研究では経路障害は発生地点を中心に多 AS に影響を及ぼすことを仮定し、複数地点で受信される BGP メッセージからパラメータを抽出し、障害規模と発生地点の推定を行う。

2. 障害規模推定に用いるデータ

2.1. 本研究で扱う BGP メッセージの種類

BGP メッセージには、経路制御を行うために 5 種類のタイプが規定されている [1]。この中から、経路情報に変更が生じた時に BGP ルータ間で交換される経路更新情報を伝える UPDATE タイプのメッセージに着目する。

2.2. 本研究で取り扱う BGP データ

本研究室では BGPview[3]によりある ISP (通信事業者) での経路情報を取得しているが、後述する目的のために、多地点で取得された BGP の経路情報を必要とする。よって、米国オレゴン大学が公開している Route Views Project[4]と RIPE NCC が公開している RIS Raw Data[5]の

BGP データを使用する。オレゴン大学の他にもこのプロジェクトに協力する多くの組織（日本では WIDE project が参加）が、自組織の BGP ルータで受信している BGP データから、受信時刻、送受信する peer の情報、経路喪失（withdraw）した prefix 情報、経路更新（announce）した prefix とその経路リストを抽出して記録し、MRT 形式[6]に変換したものを公開している。図 1 に、MRT 形式の BGP メッセージの一例を示す。

```

TIME: 08/25/17 00:00:09
TYPE: BGP4MP/MESSAGE/Update
FROM: 27.111.228.159 AS24482
TO: 27.111.228.208 AS6447
ORIGIN: IGP
ASPATH: 24482 174 6762 28329 264386 264386 264386
NEXT_HOP: 27.111.228.159
WITHDRAW
170.245.126.0/23
186.12.248.0/22
186.182.234.0/24
ANNOUNCE
131.108.174.0/24
131.108.175.0/24

```

図 1. MRT 形式の BGP メッセージの一例

これは、2017/08/25 00:00:09 (UTC)におけるシンガポール (AS 24482) からオレゴン大学 (AS 6447) への BGP メッセージで、170.245.126.0/23 を含む 3 つの prefix が経路喪失し、131.108.174.0/24 を含む 2 つの prefix が 24482-174-6762-2839-264386-264386-264386 に経路変更したことを伝えている。

データ公開サイトにはかなりの長期間のデータが蓄積、公開されているが、本研究では 2016 年 1 月 1 日から 2017 年 10 月 31 日までの期間の BGP データを解析対象とする。

3. 障害規模推定手法

3.1. パラメータ推定

使用するデータが、BGP メッセージの部分を抽出した MRT 形式のものであるため、ここから解析可能な情報は、1 つのメッセージ中の経路リスト (AS path) に含まれる AS の数、AS path 長、withdraw した prefix の数、announce した prefix の数、情報の送信元 peer の AS 番号である。経路喪失があれば withdraw が、新しい経路の更新があれば announce が自分につながる peer から届く。また、メッセージの到着についての情報を含めるために、単位時間あたりの message 数、その中で withdraw prefix を含んだ message 数と announce prefix を含んだ message 数に分け、withdraw した prefix 数、announce した prefix 数などを解析対象とした。これら 8 つのパラメータ候補を表 1 にまとめる。

表 1. MRT 形式の UPDATE メッセージから抽出できるパラメータ

p1	総 message 数	p5	announce した prefix 数
p2	withdraw を含んだ message 数	p6	平均 announce path 長
p3	announce を含んだ message 数	p7	平均 withdraw prefix 長
p4	withdraw した prefix 数	p8	平均 announce prefix 長

多量のデータを取り扱う上で 8 つのパラメータをさらに絞り込むために、既知の BGP 経路障害発生時の各パラメータの変化を調べる。

2017 年 8 月 25 日 03:30~03:40 (UTC) に Google による内部経路が外部にリークしたことで OCN 及び KDDI が関係するネットワークに大規模な遅延が発生したので、Route Views Project の WIDE の同日付の BGP データで解析を行う。

パラメータを数値化する単位時間は 1 分とする。BGP 経路障害の発生した時間を含む 2017 年 8 月 25 日 03:00~04:00 の各パラメータの変化を図 2 に示す。

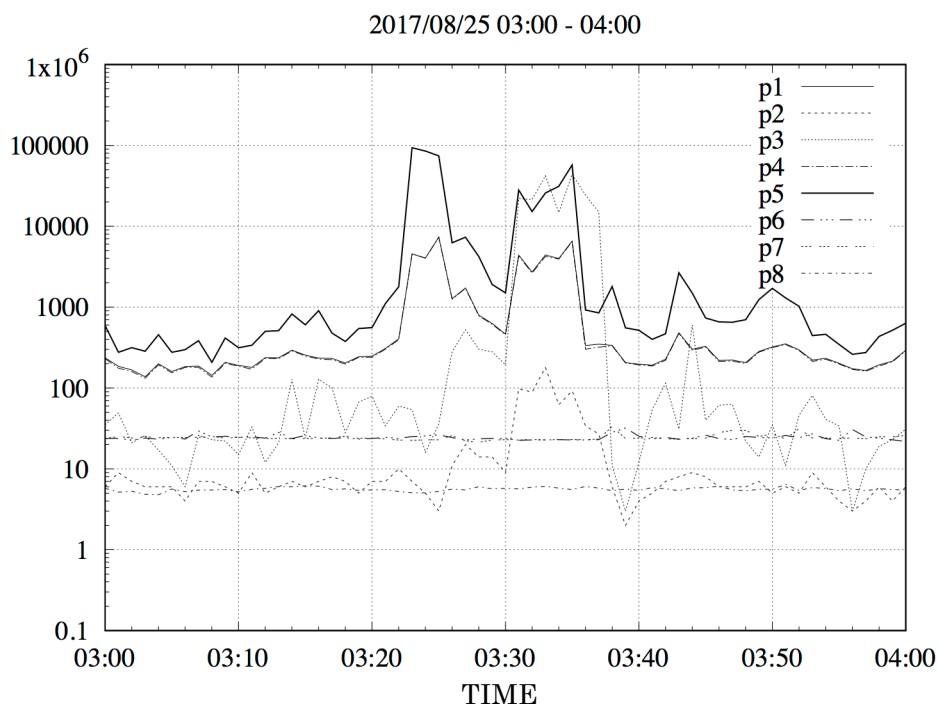


図 2. 2017 年 8 月 25 日の 3:00~4:00 (UTC) における WIDE の BGP データから算出した各パラメータ (p1~p8 は表 1 参照)

各パラメータの変化に注目すると、実際に障害が起きたとされる 3:30~3:40 の間で分毎の message 数(p1), withdraw を含んだ message 数(p2), announce を含んだ message 数(p3), withdraw した prefix 数(p4), announce した prefix 数(p5)が継続的に増えている。また、p6~p8 の 3 つのパラメータはあまり変化していないため、解析対象外とする。

WIDE の BGP ルータは 4 つの peer からメッセージを受信しており、メッセージを peer 毎に分けてパラメータを計算したところ、peer 毎に扱えるメッセージサイズの最大値が違うことが判明した。最大値を超えるメッセージが生起された時、最大サイズに収まるようにメッセージを分割して送るため、見かけ上のメッセージの数が peer によって変化してしまう。よって、総 message 数(p1), withdraw を含んだ message 数(p2), announce を含んだ message 数(p3)はパラメータには適さないと判断し、解析対象から外すことにした。

3.2. パラメータの分布

2016 年 1 月 1 日から 2017 年 10 月 31 日の WIDE の BGP データの分毎の withdraw した prefix 数と announce した prefix 数の相対度数分布と，さまざまな関数を試したなかで最もよく近似できた対数正規分布の確率密度関数を図 3, 4 に示す．

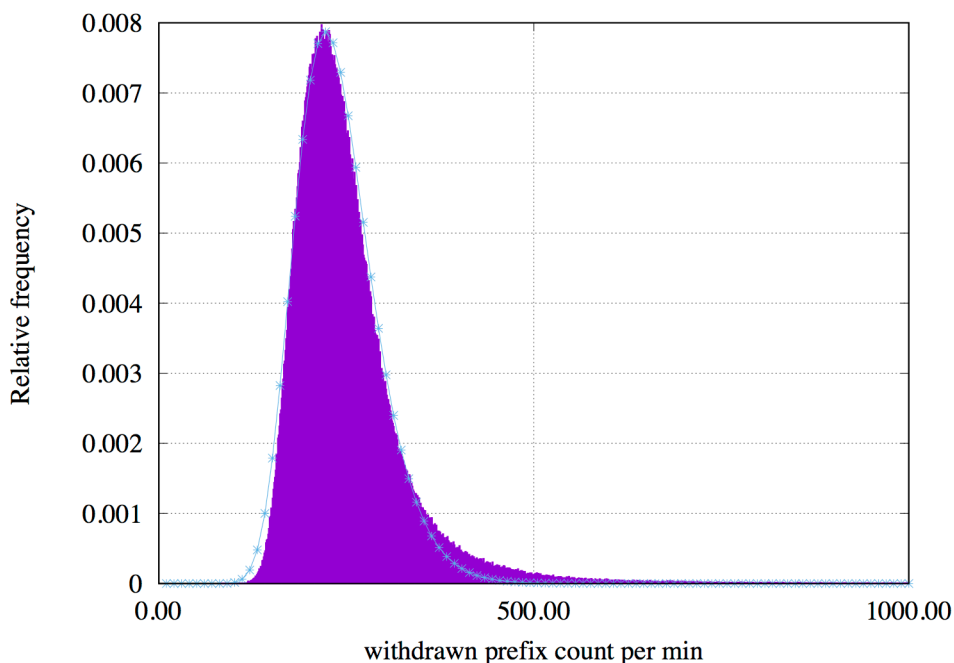


図 3. withdraw した prefix 数の相対頻度分布と対数正規分布でフィッティングさせた結果

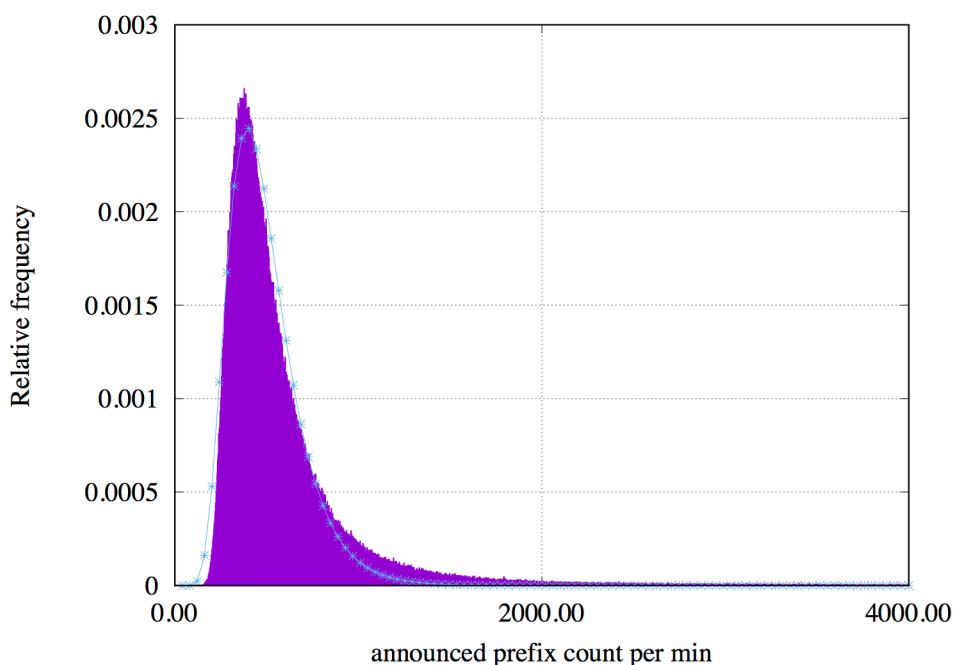


図 4. announce した prefix 数の相対頻度分布と対数正規分布でフィッティングさせた結果

ある確率変数 X が対数正規分布に従うと仮定すれば $\ln(X)$ は正規分布に従うため，withdraw し

た prefix 数と announce した prefix 数に対して自然対数をとればその頻度分布は正規分布になると考えた。自然対数をとった値の相対度数分布と正規分布の確率密度関数でフィッティングしたものを図 5, 6 に示す。

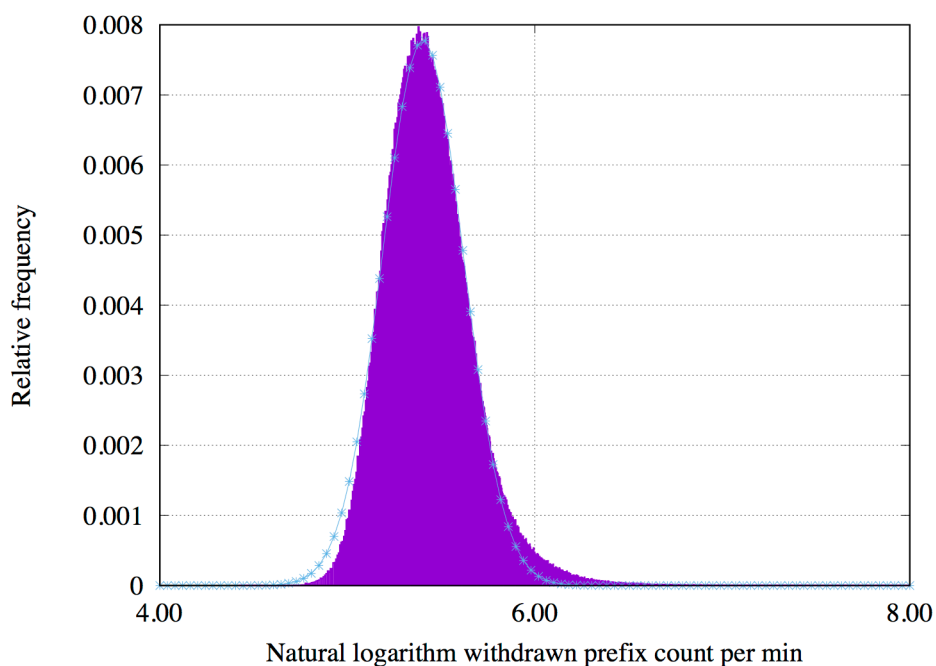


図 5. withdraw した prefix 数の自然対数の相対頻度分布と正規分布 $N(\mu_w, \sigma_w^2)$ でフィッティングさせた結果

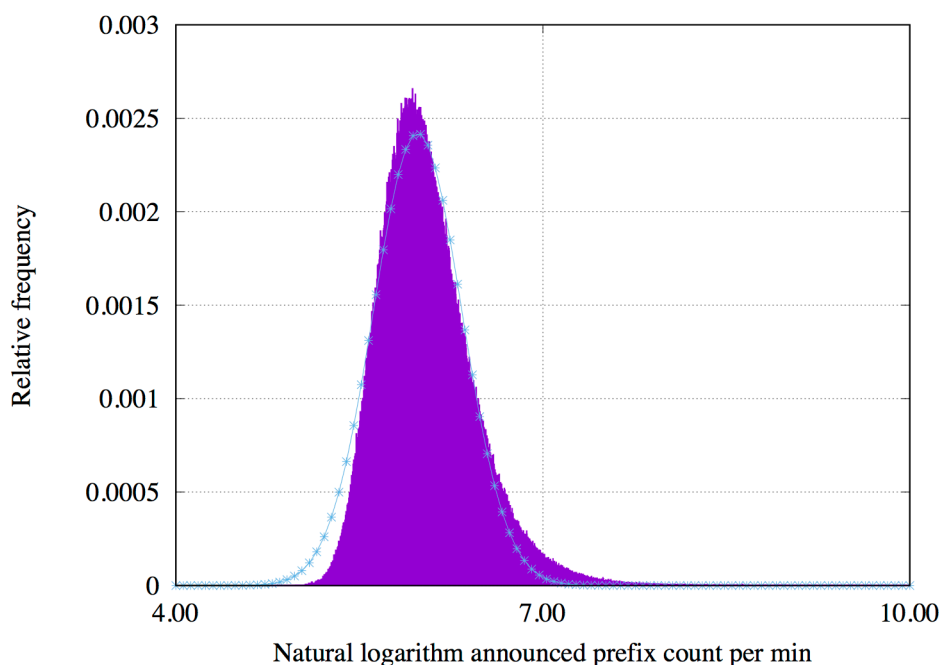


図 6. announce した prefix 数の自然対数の相対頻度分布と正規分布 $N(\mu_a, \sigma_a^2)$ でフィッティングさせた結果

フィッティング結果に注目するとそれぞれ正規分布 $N(\mu_w, \sigma_w^2)$, 正規分布 $N(\mu_a, \sigma_a^2)$ に十分従っているといえる。また、時々データに含まれる BGP ルータの動作停止に起因する異常値を回避するために 10 分の移動平均を取り x_w , x_a とする。よって、次のような計算式で指標値を計算する。

観測点 i における withdraw での指標値

$$IW_i = \frac{\ln(x_w) - \mu_w}{\sigma_w} \quad (1)$$

観測点 i における announce での指標値

$$IA_i = \frac{\ln(x_a) - \mu_a}{\sigma_a} \quad (2)$$

これらの指標値を複数地点で求め $\text{Max}(IW_i)$ 及び $\text{Max}(IA_i)$ と $\sum_i IW_i$ 及び $\sum_i IA_i$ を算出することによって、障害発生地点と障害規模を推定する。

4. 障害規模推定の結果

Route Views Project 及び RIS Raw Data から 2016 年 1 月 1 日～2017 年 10 月 31 日における WIDE(日本), isc(カリフォルニア州), nwax(オレゴン州), rrc13(ロシア)の BGP データを取得した。障害の時はまず withdraw した prefix 数に変化が現れるため、式(1)のみを使用し IW_i を算出した。

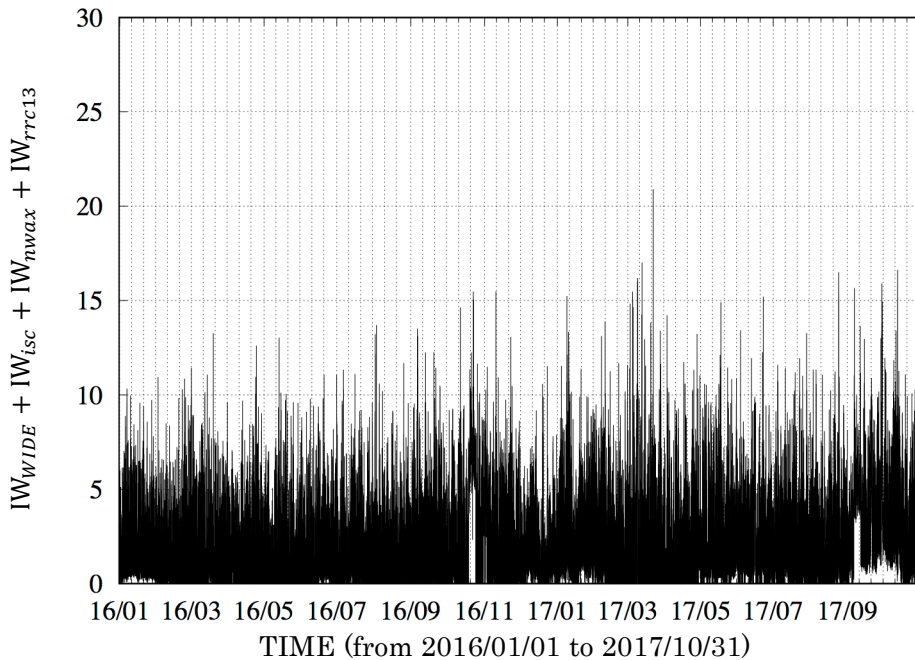


図 7. 4 観測地点での $\sum_i IW_i$

既知の障害である 2017 年 8 月 25 日の指標値について図 8 の上図に IW_{WIDE} , 下図に IW_{isc} を, 図 9 の上図に IW_{nwax} , 下図に IW_{rrc13} を示す. また, 特に高い指標値を示している 2017 年 3 月 22 日の指標値についての図 10 の上図に IW_{WIDE} , 下図に IW_{isc} を, 図 11 の上図に IW_{nwax} , 下図に IW_{rrc13} を示す.

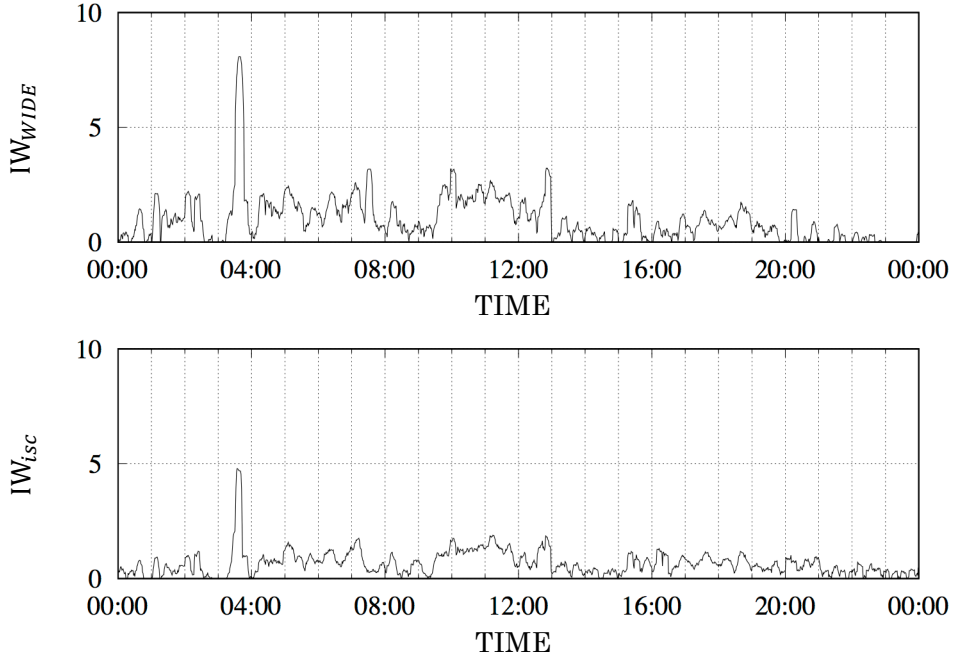


図 8. 2017 年 8 月 25 日 (UTC) における指標値 (上図: IW_{WIDE} , 下図 IW_{isc})

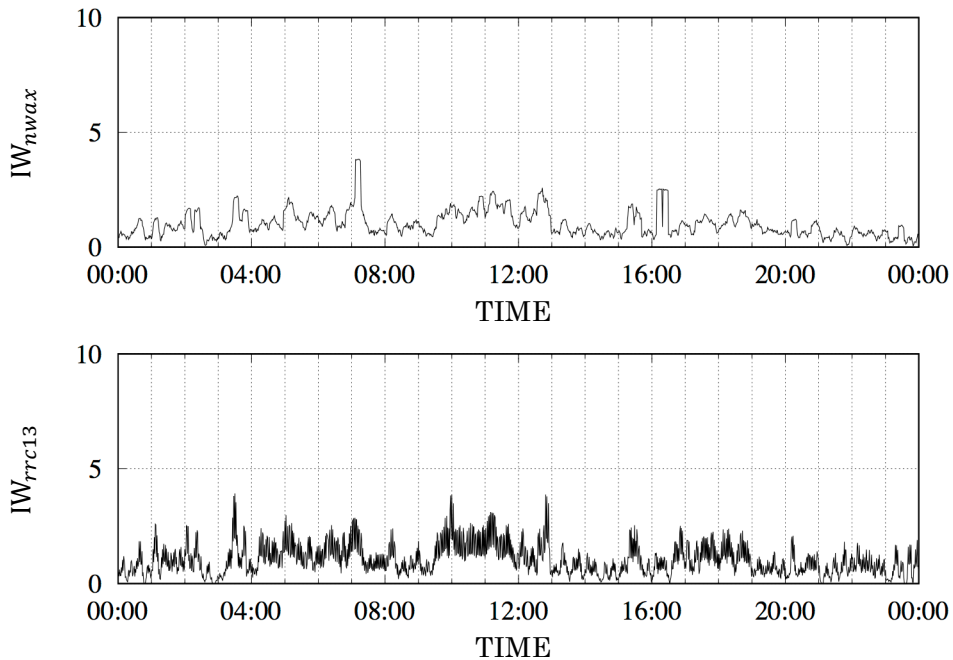


図 9. 2017 年 8 月 25 日 (UTC) における指標値 (上図: IW_{nwax} , 下図 IW_{rrc13})

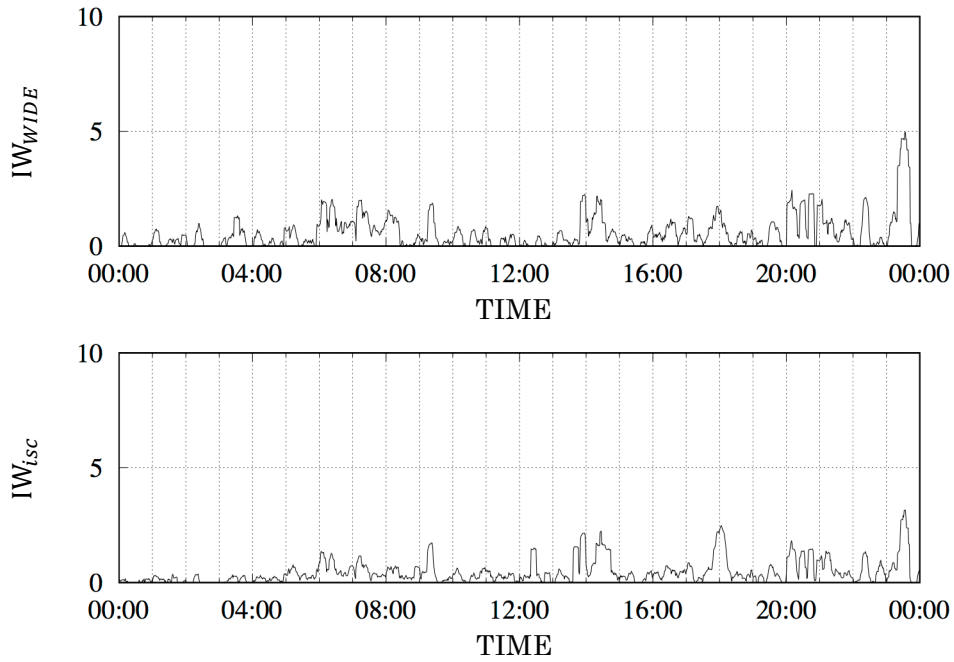


図 10. 2017 年 3 月 22 日 (UTC)における指標値 (上図: IW_{WIDE} , 下図: IW_{isc})

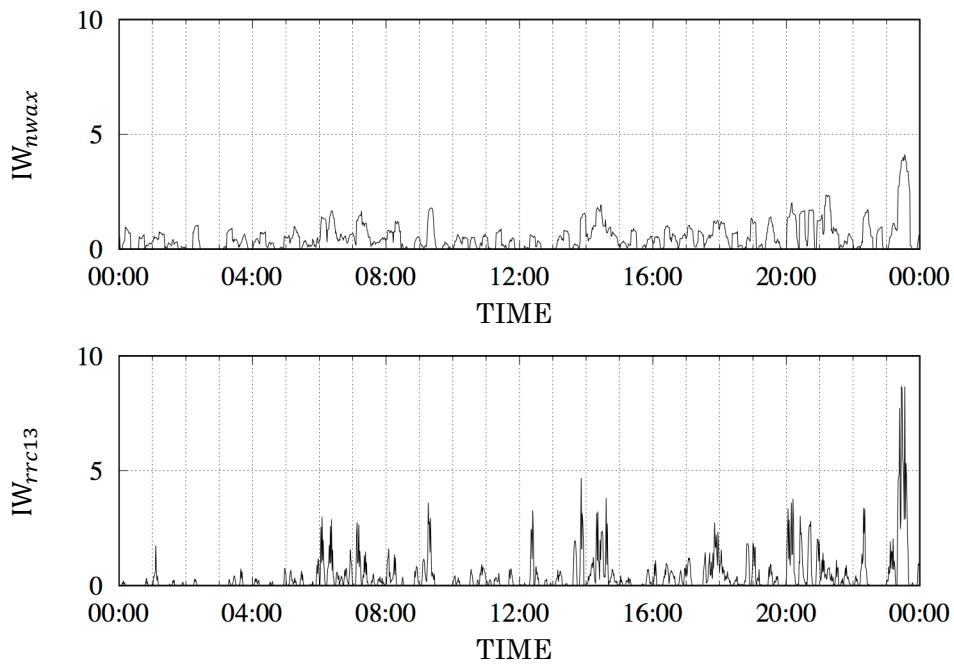


図 11. 2017 年 3 月 22 日 (UTC)における指標値 (上図: IW_{nwax} , 下図: IW_{rrc13})

図 7, 8 に注目すると, 3:32(UTC)をピークに IW_{WIDE} が高くなっており, 他地点もそれに追従して断続的に高くなっている. よって, Google 社の誤広告によって引き起こされた経路障害の様子が推定できている.

図 9, 10 に注目すると, IW_{rrc13} を中心とした 23:30(UTC)がピークの経路障害が断続的に起き

ていると推定できる.

4. 結論と今後の展望

複数地点の BGP データからパラメータを抽出し, 分布の中心からの外れの程度を指標化することで経路障害の発生地点とその規模をある程度推定することができた. また, 扱いやすい正規分布を用いて分布中での位置を数値化することで BGP データベースにデータが公開され次第解析することが可能であることを示した.

今回考案した手法は観測地点が少ないと詳細な推定ができないため, 可能な限り観測地域を増やすことが望ましい. また, 異常値の算出方法が簡易的なのでより精度の高い手法にする必要がある. 同時に経路障害が発生した時に混同してしまうことも課題である.

最後に本研究では扱えなかったが, 2017 年 8 月 25 日の Google 社の誤広告による経路障害の直前に, 平常時と全く違う内容の announce を含んだメッセージが大量に発生しており, それによって経路障害に繋がったのではないかと考える. よって announce を含んだメッセージ内容をパラメータ化することができれば, 経路障害の予測に繋がるのではないかと推測する.

参考文献

- [1] RFC4271 A Border Gateway Protocol 4 (BGP-4), <https://tools.ietf.org/html/rfc4271>
- [2] RFC5880 Bidirectional Forwarding Detection (BFD), <https://tools.ietf.org/html/rfc5880>
- [3] BGPView, <http://www.bgpview.org>
- [4] Route Views Project, <http://www.routeviews.org>
- [5] RIS Raw Data, <https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/ris-raw-data>
- [6] RFC6396 MRT routing information export format, <https://tools.ietf.org/html/rfc6396>

Ginibre cluster process and its application in wireless networks

Takuya Kobayashi

Department of Mathematical and Computing Sciences
Tokyo Institute of Technology

Abstract This paper proposes a Ginibre Thomas Cluster process (GTCP) defined similar to the Thomas cluster process. While the parent point process of Thomas process follows the Poisson point process (PPP), that of GTCP follows the Ginibre point process (GPP), which can express the repulsion between points of each other. In addition, we propose a device-to-device (D2D) network model in which the D2D transmitter users (D2D-Txs) are located according to a GTCP. In this model, the D2D users only communicate with users inside the same cluster. For this model, we derive the approximations of the coverage probability and the area spectral efficiency (ASE).

1. Introduction

There are several benefits in the direct device-to-device (D2D) communications between closely users different from cellular communications via a base station. For example, the spectral efficiency is improved due to smaller communication distance and the network traffic can be eliminated by not passing through a base station. In the D2D communications, the D2D users can cash popular files shared among nearby D2D users [1, 2]. If the popular file depends on the location of user, the set of popular files may not be the same across cluster consist of nearby users. In this paper, we consider a D2D network model in which D2D communication between users located across different cluster is limited.

The most popular spatial D2D network model is Poisson Dipole Process, in which the corresponding receivers are located at fixed distance from each transmitters [3, 4]. However, these model does not capture the clustering effect and the randomness of communication distance. On the other hand, the D2D network model in which the D2D users are located according to the Thomas cluster process was proposed [5]. The Thomas cluster process consists of a parent point process according to Poisson point process (PPP) and daughter point processes which normally scattered around each parent point. Furthermore, the performance of the D2D network may depends on the configuration of the cluster centers [6]. In fact, the performance of the cellular network model of Ginibre point process (GPP) tends to be different from the result of the PPP [7, 8, 9]. The GPP can express the repulsion between points of each other [10, 11, 12]. Now we introduce a Ginibre Thomas Cluster process (GTCP) defined similar to the Thomas cluster process. While the parent point process of Thomas process follows the PPP, that of GTCP follows the GPP. In addition, we propose a D2D network model in which the D2D transmitter users (D2D-Txs) are located according to a GTCP. In this model, the D2D users only communicate with users inside the same cluster. For this model, we derive the approximations of the coverage probability and the area spectral efficiency (ASE).

The rest of this paper is organized as follows. In Section 2, we introduce the GTCP and its properties. In Section 3, we propose the D2D network model in which the locations of D2D users follows the GTCP. The numerically computable representations of the coverage probability and the ASE is described in Section 4. The results of numerical experiments are described in Section 5, and conclusion is given in Section 6.

2. Ginibre Thomas cluster process

Now we introduce a Ginibre Thomas cluster process extended from Thomas cluster process. Let $\Phi_p = \{X_i\}_{i=0}^{\infty}$ denotes a GPP with intensity λ_p , which represents the locations of the parent points. Each parent point X_i has a mark $\Psi_i = \{Y_{ij}\}_{j \in [N_i]}$, where $[N_i] = \{1, 2, \dots, N_i\}$. The mark Ψ_i , $i \in [N_i]$, denotes a Poisson point process (PPP) with intensity measure $\lambda_{\overline{m}}(x) =$

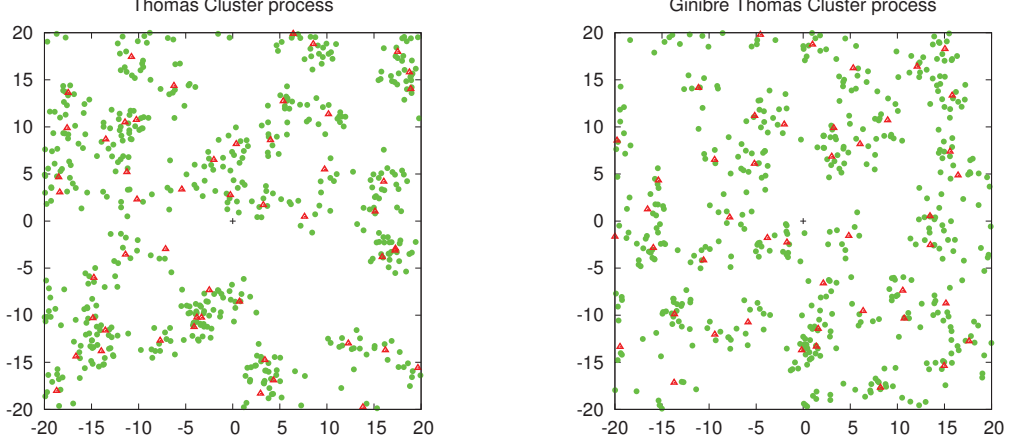


Figure 1: A sample path of Thomas cluster process and Ginibre Thomas cluster process

$\bar{m} \exp(-\|x\|^2/2\sigma^2)/2\pi\sigma^2, x \in \mathbb{R}^2$, for $\bar{m} > 0, \sigma^2 > 0$. The PPPs $\{\Psi_i\}_{i \in \mathbb{N}}$ are called daughter point processes. Then the Ginibre Thomas cluster process (GTCP) is given by

$$\Phi^{(GC)} = \bigcup_{X_i \in \Phi_p} (X_i + \Psi_i), \quad (1)$$

where $(X_i + \Psi_i) = \{X_i + Y_{ij} | Y_{ij} \in \Psi_i\}$. The number of daughter points N_i of a parent point located at X_i follows a Poisson distribution with mean $\bar{m}$ and the daughters are assumed to be i.i.d. according to a radially symmetric normal distribution with variance σ^2 .

Lemma 1 (Distribution of i.i.d. sequence $\{\|Y_j\|\}$). *Let $\Psi = \{Y_j\}_{j \in [N]}$ be a PPP with intensity measure $\lambda_{\bar{m}}(\cdot)$. Then the probability density function of the distance $\|Y_j\|$, $j \in [N]$, follows Rayleigh distribution; that is,*

$$f_V(\nu) = \frac{\nu}{\sigma^2} \exp\left(-\frac{\nu^2}{2\sigma^2}\right). \quad (2)$$

Lemma 2 (Distribution of i.i.d. sequence $\{\|X + Y_j\|\}$). *Let X be the location of the cluster center and $\Psi = \{Y_j\}_{j \in [N]}$ be a PPP with intensity measure $\lambda_{\bar{m}}(\cdot)$. Conditioned on the distance $\nu = \|X\|$, the probability density function of the distance $\|X + Y_j\|$, $j \in [N]$, is*

$$f_U(u|\nu) = \frac{u}{\sigma^2} \exp\left(-\frac{u^2 + \nu^2}{2\sigma^2}\right) I_0\left(\frac{u\nu}{\sigma^2}\right), \quad (3)$$

where $I_0(z) = \pi^{-1} \int_0^\pi e^{z \cos \psi} d\psi$ is the modified Bessel function of the first kind with order 0.

3. System model

We assume that the locations of all simultaneously active D2D-Txs follow the GTCP $\Phi^{(GC)} = \bigcup_{i \in \mathbb{N}} (X_i + \Psi_i)$. Without loss of generality, the typical user, which is randomly chosen from a randomly chosen cluster, termed representative cluster, is located at the origin due to the stationarity of the GPP and PPP.

3.1. Channel model

The path-loss function representing the attenuation of the signal with distance is given by $\ell(r) = r^{-2\beta}$, $r > 0$, where $\beta > 1$ is the path-loss exponent. Let $\{h_{ij}\}$ denote the propagation random effect of Rayleigh fading; that is, the $\{h_{ij}\}$ are mutually independent and exponentially distributed with unit mean and independent of the point process $\Phi^{(GC)}$. Let P_d denote the constant transmit power of each device. Then the received power at the typical device located at X_0 from D2D transmitter (D2D-Tx) located at Y_{00} is given by

$$P_r = P_d h_{00} \ell(\|X_0 + Y_{00}\|).$$

In this D2D network, the total interference at the typical D2D user (D2D-Rx of interest) can be divided into two parts; that is, intra-cluster interference and inter-cluster interference. The intra-cluster interference is caused by the interfering D2D-Txs belonging to the representative cluster, and the inter-cluster interference is caused by simultaneously active D2D-Txs outside the representative cluster. For the D2D-Tx of interest located at Y_{00} and the respective cluster center located at X_0 , the intra-cluster interference can be expressed as

$$I_c = \sum_{j \in [N_0]} P_d h_{0j} \ell(\|X_0 + Y_{0j}\|). \quad (4)$$

On the other hand, the inter-cluster interference from the active D2D-Txs outside the representative cluster can be expressed as

$$I_c^c = \sum_{i \in \mathbb{N}} \sum_{j \in [N_i]} P_d h_{ij} \ell(\|X_i + Y_{ij}\|). \quad (5)$$

The SIR received at the typical user is given by

$$\text{SIR} = \frac{P_d h_{00} \ell(R)}{I_c + I_c^c}, \quad (6)$$

where $R = \|X_0 + Y_{00}\|$ denotes the serving distance. Since the transmit power is same for each D2D-Tx and is canceled in the expression of SIR, we can set $P_d \equiv 1$ without loss of generality.

3.2. Distance distribution

In this subsection, we derive the distributions of various geometric distance in D2D network model. We assume that the locations of all simultaneously active D2D-Txs follow the GTCP $\Phi^{(GC)}$ defined above and the typical user located at the origin. From the property of the GTCP given in the lemma 1 and the lemma 2, we obtain the following lemmas.

Lemma 3 (Distance from representative cluster center to typical user). *The distance $\|X_0\|$ follows the Rayleigh distribution and the probability density function is given by $f_V(\nu_0)$ in lemma 1.*

Lemma 4 (Serving distance ditribution). *Conditioned on the distance $\nu_0 = \|X_0\|$ between the cluster center and the typical D2D user, the conditional probability density function of the serving distance $R = \|X_0 + Y_{00}\|$ is $f_R(r|\nu_0) = f_U(r|\nu_0)$ in lemma 2.*

The following lemma gives the distribution of distances between interfering D2D users and typical user. The probability density function conditioning the distance between the cluster center and the typical user are expressed in the same form as lemma 4 by using the lemma 2.

Lemma 5 (Interference distance distribution). *Conditioned on the distance $\nu = \|X\|$ between the cluster center and the typical D2D user, the distances $\{U_{ij} = \|X_i + Y_{ij}\|\}_{j \in [N_i]}$ are i.i.d. with probability density function $f_U(r|\nu)$ in lemma 2.*

Since the cluster centers are located according to GPP, the distances between the representative cluster center and the inter-cluster centers are mutually independent and the squared distance follows Gamma distribution.

Lemma 6 (Distances between representative cluster center and inter-cluster centers). *The set of distances between the representative cluster center and the inter-cluster centers $\{\|X_i - X_0\|\}_{i \in \mathbb{N}}$ has the same distribution as $\{\sqrt{Z_j/\pi\lambda_p}\}_{j=2}^\infty$, where Z_j , $j = 2, 3, \dots$, are mutually independent and each Z_j follows the j th Gamma distribution with parameter 1, denoted by $Z_j \sim \text{Gamma}(j, 1)$.*

Proof. The proof is derived from the Palm version of the kostlan's theorem (see Proposition 4). $\square$

4. Coverage probability and ASE

In this section, we derive the coverage probability of the typical user and the ASE of the whole D2D network by using the distance distributions. From the definition of SIR, the coverage probability is

$$\begin{aligned} P_c(\theta) &= \mathbb{P}\left(\frac{P_d h_{00} \ell(\|X_0 + Y_{00}\|)}{I_c^c + I_c} > \theta\right) \\ &= \mathbb{E}\left[\mathbb{E}\left[\exp(-\theta R^{2\beta}(I_c^c + I_c)|R)\right]\right] \\ &= \mathbb{E}\left[\mathcal{L}_{I_c^c}(\theta R^{2\beta})\mathcal{L}_{I_c}(\theta R^{2\beta}|R)\right]. \end{aligned} \quad (7)$$

Therefore, in order to obtain the coverage probability, it is necessary to derive the Laplace transforms of interference.

4.1. Laplace transform of interference

Lemma 7. *The conditional Laplace transform of the intra-cluster interference I_c conditioned on the distance $\nu_0 = \|X_0\|$ between the typical user and the representative cluster center is given by*

$$\mathcal{L}_{I_c}(s|\nu_0) = \exp\left(-(\bar{m}-1) \int_0^\infty \frac{f_U(u|\nu_0)}{1+s^{-1}u^{2\beta}} du\right). \quad (8)$$

Proof. From the definition of I_c , the Laplace transform of intra-cluster interference at the typical user is

$$\begin{aligned} \mathcal{L}_{I_c}(s|\nu_0) &= \mathbb{E}\left[\exp\left(-s \sum_{j \in [N_0]} h_{0j} \ell(\|X_0 + Y_{0j}\|)\right)\right] \\ &= \mathbb{E}\left[\prod_{j \in [N_0]} \mathbb{E}\left[\exp(-s h_{0j} \|X_0 + Y_{0j}\|^{-2\beta}) \middle| Y_{0j}\right]\right] \\ &\stackrel{(a)}{=} \mathbb{E}\left[\prod_{j \in [N_0]} (1 + s \|X_0 + Y_{0j}\|^{-2\beta})^{-1}\right] \\ &\stackrel{(b)}{=} \exp\left(-\int_{\mathbb{R}^2} \left(1 - \frac{1}{1 + s \|X_0 + y\|^{-2\beta}}\right) \lambda_{\bar{m}-1}(y) dy\right) \\ &\stackrel{(c)}{=} \exp\left(-(\bar{m}-1) \int_0^\infty \frac{f_U(u|\nu_0)}{1+s^{-1}u^{2\beta}} du\right), \end{aligned}$$

where (a) follows from $h_{0j} \sim \text{Exp}(1)$ and $\mathcal{L}_h(s) = (1+s)^{-1}$, (b) follows from the Laplace functional of Poisson point process Ψ_0 , and (c) follows from the change of variable $\|X_0 + y\| \rightarrow u$ and the distance distribution given by lemma 5. $\square$

Corollary 1 (Approximation). *The Laplace transform of intra-cluster interference at the typical user is*

$$\tilde{\mathcal{L}}_{I_c}(s) = \exp\left(-(\bar{m}-1) \int_0^\infty \frac{f_R(u)}{1+s^{-1}u^{-2\beta}} du\right), \quad (9)$$

where $f_R(u) = \frac{u}{2\sigma^2} \exp\left(-\frac{u^2}{4\sigma^2}\right)$.

Corollary 2 (Lower bound). *The lower bound on Laplace transform of intra-cluster interference at the typical user is*

$$\tilde{\mathcal{L}}_{I_c}(s) \geq \exp\left(-\frac{\bar{m}-1}{4\sigma^2} s^{\frac{1}{\beta}} C(2\beta)\right), \quad (10)$$

where $C(x) = \frac{2\pi/x}{\sin(2\pi/x)}$.

Proof. See Corollary 4 in [5]. $\square$

On the other hand, the Laplace transform of inter-cluster interference at the typical user is given by

$$\begin{aligned}
\mathcal{L}_{I_c^c}(s|X_0) &= \mathbb{E} \left[\exp \left(-s \sum_{X_i \in \Phi_p \setminus \{X_0\}} \sum_{j \in [N_i]} h_{ij} \ell(\|X_i + Y_{ij}\|) \right) \middle| X_0 \right] \\
&= \mathbb{E} \left[\prod_{X_i \in \Phi_p \setminus \{X_0\}} \prod_{j \in [N_i]} \mathbb{E} \left[\exp \left(-s h_{ij} \|X_i + Y_{ij}\|^{-2\beta} \right) \middle| \Phi^{(GC)} \right] \middle| X_0 \right] \\
&= \mathbb{E} \left[\prod_{X_i \in \Phi_p \setminus \{X_0\}} \prod_{j \in [N_i]} (1 + s \|X_i + Y_{ij}\|^{-2\beta})^{-1} \middle| X_0 \right] \\
&= \mathbb{E} \left[\prod_{X_i \in \Phi_p \setminus \{X_0\}} \exp \left(- \int_{\mathbb{R}^2} \left(1 - \frac{1}{1 + s \|X_i + y\|^{-2\beta}} \right) \lambda_{\bar{m}}(y) dy \right) \middle| X_0 \right] \\
&= \mathbb{E} \left[\prod_{X_i \in \Phi_p \setminus \{X_0\}} \exp \left(- \bar{m} \int_0^\infty \frac{f_U(u \|X_i\|)}{1 + s^{-1} u^{2\beta}} du \right) \middle| X_0 \right]. \tag{11}
\end{aligned}$$

Since $\{\|X_i\|\}_{i \in \mathbb{N}}$ depends on the location X_0 , it is difficult to think of this expectation. However, if $\|X_0\|$ and λ_p are small enough, $\|X_i\|$ can be approximated as $\|X_i - X_0\|$. The following proposition is given by applying lemma 6 to (11).

Proposition 1 (Approximation). *The Laplace transform of inter-cluster interference at the typical user is*

$$\tilde{\mathcal{L}}_{I_c^c}(s) = \prod_{j=2}^{\infty} \int_0^\infty \frac{z^{j-1} e^{-z}}{(j-1)!} \exp \left(- \bar{m} \int_0^\infty \frac{f_U(u \sqrt{z/\pi\lambda_p})}{1 + s^{-1} u^{2\beta}} du \right) dz. \tag{12}$$

Corollary 3 (Lower bound). *The lower bound on Laplace transform of intra-cluster interference at the typical user is*

$$\tilde{\mathcal{L}}_{I_c^c}(s) \geq \exp \left(- \bar{m} \lambda_p \pi s^{\frac{1}{\beta}} C(2\beta) + \bar{m} \int_0^\infty \frac{2\pi \mu u e^{-\pi \mu u^2}}{1 + s^{-1} u^{2\beta}} du \right), \tag{13}$$

where $\mu = \lambda_p / (1 + 2\pi \lambda_p \sigma^2)$.

Proof. This result is obtained by applying Jensen's inequality to (11). $\square$

4.2. Coverage probability

From (7), the coverage probability is given by the following proposition.

Proposition 2 (Coverage probability). *Let $\mathcal{L}_{I_c^c}(s)$ be the Laplace transform of intra-interference and $\mathcal{L}_{I_c}(s|\nu_0)$ be the conditional Laplace transform of inter-cluster interference conditioned on the distance $\nu_0 = \|X_0\|$ between the cluster center and the typical user. Then the coverage probability is given by*

$$P_c(\theta) = \int_0^\infty \int_0^\infty \mathcal{L}_{I_c^c}(\theta r^{2\beta}) \mathcal{L}_{I_c}(\theta r^{2\beta}|\nu_0) f_R(r|\nu_0) f_V(\nu_0) dr d\nu_0, \tag{14}$$

We can apply the approximation and lower bound of Laplace transform in the previous subsection to the coverage probability.

Theorem 1 (Approximation of coverage probability). *The approximation of the coverage probability in Proposition 2 is given by*

$$\tilde{P}_c(\theta) = \int_0^\infty \tilde{\mathcal{L}}_{I_c^c}(\theta r^{2\beta}) \tilde{\mathcal{L}}_{I_c}(\theta r^{2\beta}) f_R(r) dr. \tag{15}$$

Theorem 2 (Lower bound). *The lower bound of the approximation value of the coverage probability in Theorem 1 is given by*

$$\tilde{P}_c(\theta) \geq \int_0^\infty f_R(r) \exp \left(-\frac{\bar{m}-1}{4\sigma^2} \theta^{\frac{1}{\beta}} r^2 C(2\beta) - \bar{m} \lambda_p \theta^{\frac{1}{\beta}} r^2 C(2\beta) + \bar{m} \int_0^\infty \frac{2\pi\mu e^{-\pi\mu u^2}}{1 + \theta^{-1}(\frac{u}{r})^{2\beta}} du \right) dr. \quad (16)$$

4.3. Area spectral efficiency

The ASE denotes the average number of bits transmitted per unit time per unit bandwidth per unit area. We can use Shannon's capacity formula to define $ASE = \lambda \log_2(1 + \theta) P_c$, where λ is the density of the active transmitters and P_c is the coverage probability of the typical user. For our model, the ASE is given by

$$ASE = \bar{m} \lambda_p \log_2(1 + \theta) P_c(\theta), \quad (17)$$

where $\bar{m} \lambda_p$ is the average intensity of the active D2D-Txs and $P_c(\theta)$ is the coverage probability of the typical user given by Proposition 2. The approximation and lower bound are also defined as well as the coverage probabilities.

5. Numerical experiments

Here, we show the results of numerical experiments on computing the coverage probabilities. For all numerical experiments, we set the SIR threshold θ as 0 dB. Figure 2 shows the comparison of the analytical and simulation results of the coverage probability. From figure 2, we observe that the approximation for coverage probability given by Theorem 1 is very tight except $\bar{m} = 1$. The lower bound given by Theorem 2 is slightly loose.

In Fig. 3 and Fig. 5, we present the coverage probability as a function of the average number of simultaneously active D2D-Tx per cluster $\bar{m}$. We compare the coverage probabilities in three cases; only inter-cluster interference, only intra-cluster interference, and total interference. From these figures, we can see that the coverage probability is strongly dictated by the intra-cluster interference and the intra-cluster interference is independent of the scattering variance. In the case of intra-cluster interference, it can be seen that the smaller σ provides higher coverage probability. We can also see that the smaller the cluster center density provides higher coverage probability.

In Fig. 4 and Fig. 6, we show the ASE, defined by applying (15) to (17), as a function of the average number of simultaneously active D2D-Tx per cluster $\bar{m}$. While the coverage probability monotonically decreases with respect to $\bar{m}$, the ASE has an optimal number of D2D transmitters.

6. Conclusion

In this paper, we proposed the GTCP and applying the GTCP to a D2D network model. The GTCP can simultaneously express the clustered points and the repulsion between hotspots. For this model, we derive the coverage probability, its approximation, and its lower bound. The numerical experiment shows that the approximation is sufficiently practical. For the ASE, there exists an optimal number of D2D transmitters as well as the result of Thomas cluster process [5].

References

- [1] N. Golrezaei, P. Mansourifard, A. Molisch, and A. dIMAKIS, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3665–3676, 2014.
- [2] Y. Zhang, E. Pan, L. Song, W. Saad, Z. Dawy, and Z. Han, "Social network aware device-to-device communication in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 177–190, 2015.
- [3] H. Feng, H. Wang, X. Xu, and C. Xing, "A tractable model for device-to-device communication underlying multi-cell cellular networks," *IEEE Intl. Conf. Commun.*, pp. 587–591, 2014.

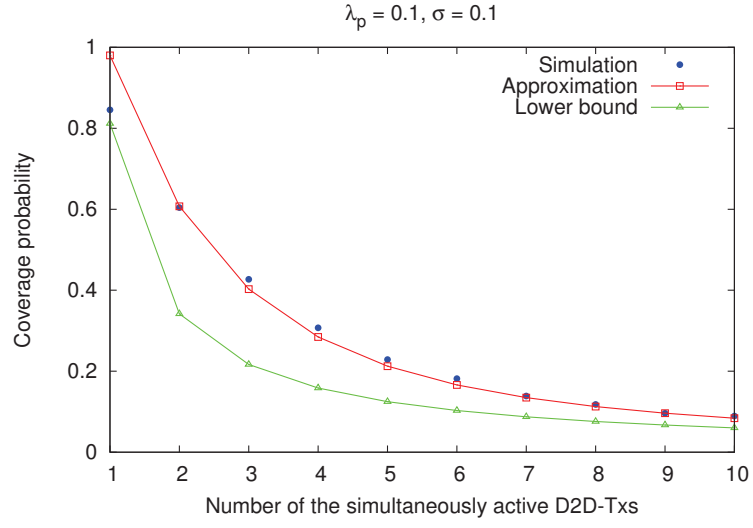


Figure 2: Coverage probability as a function of number of simultaneously active D2D-Txs, where $\lambda_p = 0.5$, $\sigma = 0.1$.

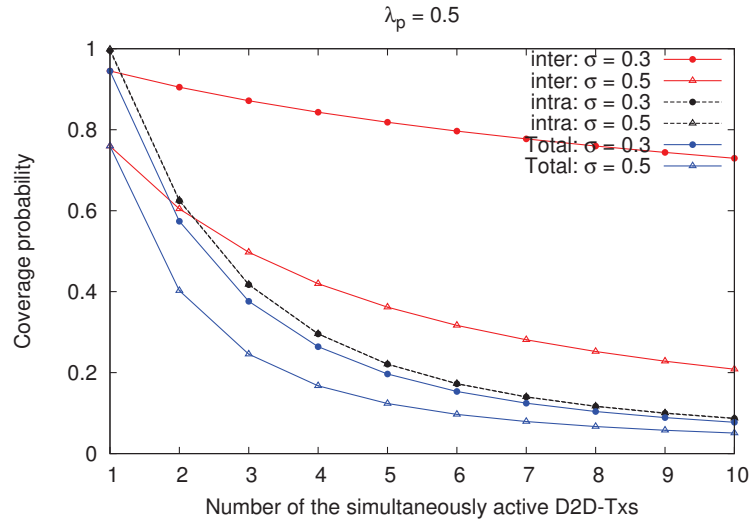


Figure 3: Coverage probability as a function of number of simultaneously active D2D-Txs, where $\lambda_p = 0.5$ and $\sigma = 0.3, 0.5$.

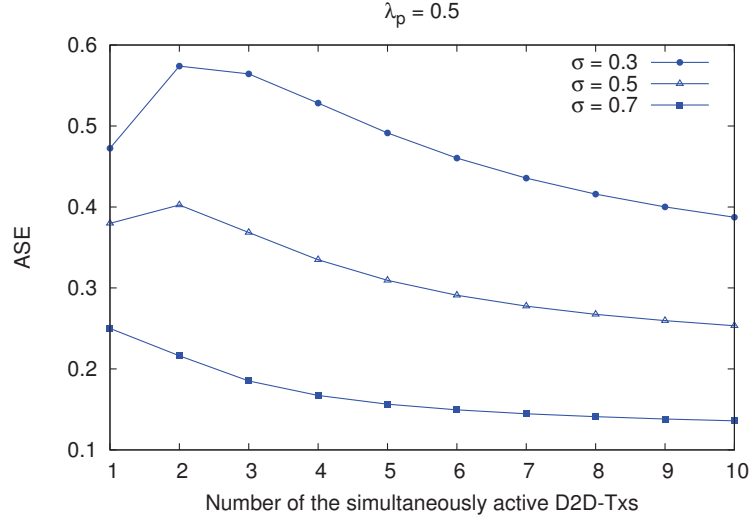


Figure 4: ASE as a function of number of simultaneously active D2D-Txs, where $\lambda_p = 0.5$ and $\sigma = 0.3, 0.5, 0.7$.

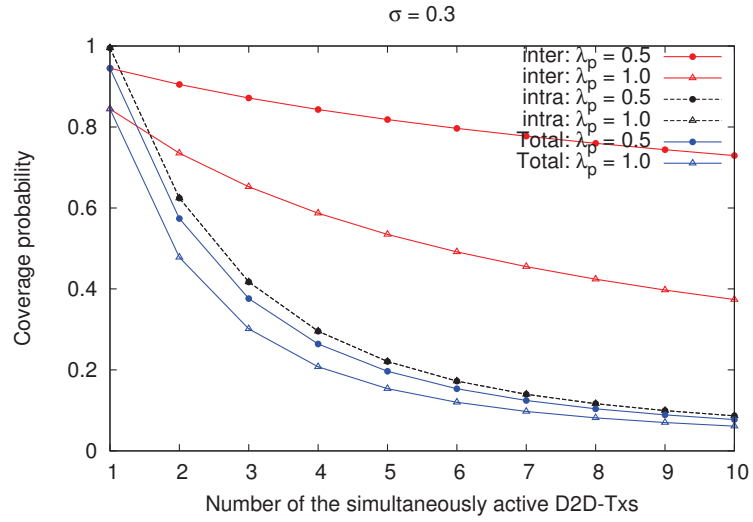


Figure 5: Coverage probability as a function of number of simultaneously active D2D-Txs, where $\sigma = 0.3$ and $\lambda_p = 0.5, 1.0$.

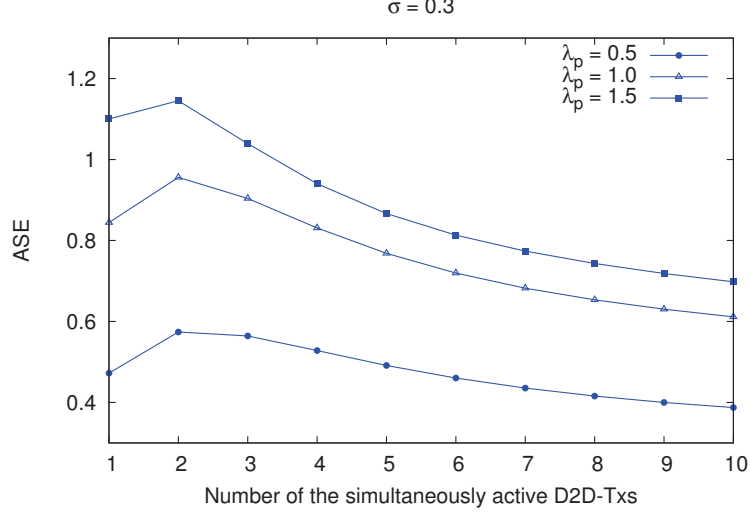


Figure 6: ASE as a function of number of simultaneously active D2D-Txs, where $\sigma = 0.3$ and $\lambda_p = 0.5, 1.0, 1.5$.

- [4] G. George, R. K. Mungara, and A. Lozano, “An analytical framework for device-to-device communication in cellular networks,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6297–6310, 2015.
- [5] M. Afshang, H. S. Dhillon, and P. H. J. Chong, “Modeling and performance analysis of clustered device-to-device networks,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4957–4972, 2016.
- [6] J. Møller and R. P. Waagepetersen, “Statistical inference and simulation for spatial point process,” Chapman & Hall/CRC, 2004.
- [7] N. Miyoshi and T. Shirai, “A cellular network model with Ginibre configured base stations,” *Adv. Appl. Probab.*, vol. 46, pp. 832–845, 2014.
- [8] N. Miyoshi and T. Shirai, “Downlink coverage probability in a cellular network with Ginibre deployed base stations and Nakagami- m fading channels,” *WiOpt 2015*, pp. 483–489, 2015.
- [9] T. Kobayashi, and N. Miyoshi, “Uplink cellular network models with Ginibre deployed base stations,” *26th International Teletraffic Congress (ITC26)*, pp. 1–7, 2015.
- [10] T. Shirai and Y. Takahashi, “Random point fields associated with certain Fredholm determinants I: Fermion, Poisson and Boson processes,” *J. Funct. Analysis*, vol. 205, pp. 414–463, 2003.
- [11] J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág, *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*. American Math. Soc., 2009.
- [12] A. Goldman, “The Palm measure and the Voronoi tessellation for the Ginibre process,” *Annals Appl. Probab.*, vol. 20, pp. 90–128, 2010.
- [13] E. Kostlan, “On the spectra of Gaussian matrices,” *Linear Algebra and its Appl.*, vol. 162, pp. 385–388, 1992.

A. Ginibre point processes

In this section, we define the GPP and describe their tractable properties. The GPP is a DPP on the complex plane $\mathbb{C}$, which can express repulsion between the points of each other. Let Φ denote a simple point process on $\mathbb{C}$ and $\rho_n: \mathbb{C}^n \rightarrow \mathbb{R}_+$, $n \in \mathbb{N}$, denote its joint intensities with respect to some locally finite measure ν on $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$. That is, for any continuous symmetric function f on

$\mathbb{C}^n$ with compact support, we have

$$\mathbb{E} \left[\sum_{\substack{X_1, \dots, X_n \in \Phi \\ \text{distinct}}} f(X_1, X_2, \dots, X_n) \right] = \int_{\mathbb{C}^n} f(z_1, \dots, z_n) \rho_n(z_1, \dots, z_n) \nu(dz_1) \cdots \nu(dz_n).$$

The point process Φ is said to be a determinantal point process with kernel $K: \mathbb{C}^2 \rightarrow \mathbb{C}$ with respect to the reference measure ν if the ρ_n , $n \in \mathbb{N}$, satisfy

$$\rho_n(z_1, \dots, z_n) = \det(K(z_i, z_j))_{1 \leq i, j \leq n},$$

for $z_1, \dots, z_n \in \mathbb{C}$, where $\det$ denotes the determinant. Furthermore, Φ is said to be the GPP if the kernel K is given as following definition.

Definition 1 (Ginibre point process). *Φ is said to be the GPP if the kernel K is given by $K(z, w) = e^{z\bar{w}}$, $z, w \in \mathbb{C}$, with respect to the Gaussian measure $\nu(dz) = \pi^{-1} e^{-|z|^2} m(dz)$, where $\bar{w}$ denotes the complex conjugate of $w \in \mathbb{C}$ and m denotes the Lebesgue measure on $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$.*

This choice of pair of K and ν is not unique. Indeed, the determinantal point process associated with the kernel $\tilde{K}(z, w) = \pi^{-1} e^{-(|z|^2 + |w|^2)/2} e^{z\bar{w}}$ with respect to $\tilde{\nu}(dz) = m(dz)$ coincides with the GPP. From this expression, it is easy to see that $\tilde{\rho}_n(z_1, \dots, z_n) = \det(\tilde{K}(z_i, z_j))_{1 \leq i, j \leq n}$ is motion-invariant, or equivalently, the GPP is motion-invariant (stationary and isotropic). From its definition, we see that $\mathbb{E}[\Phi(C)] = m(C)/\pi$ for $C \in \mathcal{B}(\mathbb{C})$; i.e. the (first order) intensity is equal to π^{-1} with respect to the Lebesgue measure. In order to adjust the intensity, we consider a scaled GPP Φ_c with a scaling parameter $c > 0$. Φ_c is given by the kernel $K_c(z, w) = c e^{cz\bar{w}}$ with respect to the reference measure $\nu_c(dz) = \pi^{-1} e^{-c|z|^2} m(dz)$, or equivalently, $\tilde{K}_c(z, w) = (c/\pi) e^{-c(|z|^2 + |w|^2)/2} e^{cz\bar{w}}$ with respect to the Lebesgue measure. Since $\mathbb{E}[\Phi_c(C)] = (c/\pi) m(C)$, Φ_c has the intensity c/π .

One of the useful properties of the GPP comes from its radial symmetry, and it is described as follows.

Proposition 3 (Kostlan [13]). *Let X_i , $i \in \mathbb{N}$, denote the points of GPP Φ . Then, the set $\{\|X_i\|\}_{i \in \mathbb{N}}$ has the same distribution as $\{\sqrt{Z_i}\}_{i \in \mathbb{N}}$, where the Z_i , $i \in \mathbb{N}$, are mutually independent and each Z_i follows the i th Erlang distribution with a unit rate parameter, denoted by $Z_i \sim \text{Gam}(i, 1)$, $i \in \mathbb{N}$.*

Proposition 3 can be extended to scaled GPP Φ_c ; i.e., for the points X_i , $i \in \mathbb{N}$, of Φ_c , the set $\{\|X_i\|\}_{i \in \mathbb{N}}$ has the same distribution as $\{\sqrt{Z_i}\}_{i \in \mathbb{N}}$, where Z_i , $i \in \mathbb{N}$, are mutually independent and each Z_i follows the i th Erlang distribution with rate parameter c , denoted by $Z_i \sim \text{Gam}(i, c)$, $i \in \mathbb{N}$.

Moreover, we can extend Proposition 3 to the Palm version. From [10, 12], the kernel $\tilde{K}^o$ of the reduced Palm version $\Phi^o = (\Phi \mid \Phi(\{o\}) = 1) \setminus \{o\}$ of GPP is given by $\tilde{K}^o(z, w) = \pi^{-1} e^{-(|z|^2 + |w|^2)/2} (e^{z\bar{w}} - 1)$, with respect to the Lebesgue measure. The following proposition is obtained by applying this kernel to Theorem 4.7.1 in [11].

Proposition 4 (Extension of Kostlan's result to the Palm version). *Let X_i , $i \in \mathbb{N}$, denote the points of the reduced Palm version Φ^o of GPP. Then, the set $\{\|X_i\|\}_{i \in \mathbb{N}}$ has the same distribution as $\{\sqrt{Z_i}\}_{i=2}^\infty$, where the Z_i , $i \in \mathbb{N}$, are mutually independent and each $Z_i \sim \text{Gam}(i, 1)$, $i \in \mathbb{N}$.*

2次元反射型ランダムウォークにおける定常分布の積率母関数の収束領域:解析的証明

小林 正弘^a 宮沢 政清^b 小沢 利久^c

^a 東海大学 情報数理学科 ^b 東京理科大学 情報科学科 ^c 駒沢大学 市場戦略学科

概要： 本発表では反射型ランダムウォークの定常分布について論じる。我々は、反射型ランダムウォークの定常分布の裾の漸近特性を求めるために、積率母関数の収束領域を求めてきた。その収束領域を求めるときに、定常不等式と呼ばれる不等式を用いている。本発表では、定常不等式を使わず解析関数とその接続を用いて、新たな証明方法により収束領域を導出する。

1. 2次元反射型ランダムウォーク

まず、2次元反射型ランダムウォークの定義を紹介する。 $(\Omega, \mathcal{F}, \mathbb{P})$ を完備確率空間とし、以下の記号を定義する。

$$\mathbb{H} = \{0, \pm 1\}, \quad \mathbb{H}_+ = \{0, 1\}, \quad \mathbb{I} = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}, \quad \mathbb{I}_+ = \mathbb{I} \setminus \{\emptyset\}.$$

状態空間を $\mathcal{S} = \mathbb{Z}_+^2$ とする。状態空間 $\mathcal{S}$ に対する以下の有限分割を定義する。

$$\mathcal{S}^\emptyset = \{(0, 0)\}, \quad \mathcal{S}^{\{1\}} = \{(i, 0) \in \mathcal{S}; i \geq 1\}, \quad \mathcal{S}^{\{2\}} = \{(0, i) \in \mathcal{S}; i \geq 1\}, \quad \mathcal{S}^{\{1, 2\}} = \{(i, j) \in \mathcal{S}; i, j \geq 1\}.$$

$A \in \mathbb{I}$ に対して、 $\mathbf{X}^A = (X_1^A, X_2^A)$ を以下の $\mathbb{X}^A$ を値域にもつ2次元の確率変数ベクトルとする。

$$\mathbb{X}^\emptyset = \mathbb{H}_+ \times \mathbb{H}_+, \quad \mathbb{X}^{\{1\}} = \mathbb{H} \times \mathbb{H}_+, \quad \mathbb{X}^{\{2\}} = \mathbb{H}_+ \times \mathbb{H}, \quad \mathbb{X}^{\{1, 2\}} = \mathbb{H} \times \mathbb{H}.$$

2次元反射型ランダムウォークの定義は以下のように与えられる。

定義 1.1 状態空間 $\mathcal{S}$ 上マルコフ連鎖 $\{\mathbf{Y}_\ell; \ell \in \mathbb{Z}_+\} = \{(Y_{1\ell}, Y_{2\ell}); \ell \in \mathbb{Z}_+\}$ が

$$\mathbb{P}(\mathbf{Y}_{\ell+1} = \mathbf{y} + \mathbf{x} | \mathbf{Y}_\ell = \mathbf{y}) = \mathbb{P}(\mathbf{X}^A = \mathbf{x}), \quad A \in \mathbb{I}, \mathbf{y} \in \mathcal{S}^A, \mathbf{x} \in \mathbb{X}^A,$$

を満たすとき、 $\{\mathbf{Y}_\ell\}$ を2次元反射型ランダムウォークと呼ぶ。

1.1. 仮定

以下、本発表の仮定を紹介する。2次元反射型ランダムウォークの定常分布を考えるため、以下を仮定する。

仮定 1.1 2次元反射型ランダムウォーク $\{\mathbf{Y}_\ell\}$ が既約で正再帰的である。

$\{\mathbf{Y}_\ell\}$ の安定条件は、[2, 4] など知られている。本発表では $\{\mathbf{Y}_\ell\}$ の周期性については仮定をしない。仮定 1.1 の下で定常分布が唯一存在する。定常分布に従う確率変数ベクトルを $\mathbf{Y} = (Y_1, Y_2)$ とする。

次に $\mathbf{X}^A$ の仮定を考える。 $A \in \mathbb{I}$ に対して $\mathbf{X}^A$ の確率関数を $\{p_{\mathbf{x}}^A; \mathbf{x} = (x_1, x_2) \in \mathbb{X}^A\}$ とする。

$$p_{\mathbf{x}}^A = \mathbb{P}(\mathbf{X}^A = \mathbf{x}), \quad \mathbf{x} = (x_1, x_2) \in \mathbb{X}^A.$$

$\gamma^A: \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ を $\mathbf{X}^A$ の母関数とする。

$$\gamma^A(r_1, r_2) = \sum_{\mathbf{x} \in \mathbb{X}^A} p_{\mathbf{x}}^A r_1^{x_1} r_2^{x_2}, \quad A \in \mathbb{I}, (r_1, r_2) \in \mathbb{R}_+^2.$$

このとき、 $r_1 r_2 (\gamma^{\{1, 2\}}(r_1, r_2) - 1)$ は r_1, r_2 に関する多項式である。この多項式が可約もしくは少なくともどちらかの次数が1であるとき、特異な (*singular*) ランダムウォークと呼ぶ ([1] を参照)。特異なランダムウォークとそうでないランダムウォークは性質が異なる。よって、以下を仮定する。

仮定 1.2 反射型ランダムウォーク $\{\mathbf{Y}_\ell\}$ が特異でない (*non-singular*)。

仮定 1.2 の必要十分条件はパラメータ表現可能であることが知られている (詳細は [1] を参照)。

最後に、定常分布の裾が軽い条件を加える。 $\varphi: \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ を $\mathbf{Y}$ の母関数とする、すなわち、

$$\varphi(r_1, r_2) = E(r_1^{Y_1} r_2^{Y_2}), \quad (r_1, r_2) \in \mathbb{R}_+^2.$$

$\varphi(e^{\epsilon_1}, e^{\epsilon_2}) < \infty$ となる $(\epsilon_1, \epsilon_2) > (0, 0)$ が存在するとき、定常分布の裾が軽いと呼ぶ。以下の $\{\mathbf{Y}_\ell\}$ の平均ドリフト条件が定常分布の裾が軽い条件である。

仮定 1.3 $E(X_1^{\{1, 2\}}) \neq 0$ もしくは $E(X_2^{\{1, 2\}}) \neq 0$ が満たされる。

2. 主結果

以下、主結果を紹介する。 $\mathcal{D}$ を

$$\mathcal{D} = \text{the interior of } \{(\theta_1, \theta_2) \in \mathbb{R}^2; \varphi(e^{\theta_1}, e^{\theta_2}) < \infty\},$$

とする。 $\mathcal{D}$ は収束領域と呼ばれる。本発表では、解析接続を使って $\mathcal{D}$ を求めることを主目的とする。以下の領域を定義する。

$$\Gamma^A = \{(\theta_1, \theta_2) \in \mathbb{R}^2; \gamma^A(e^{\theta_1}, e^{\theta_2}) \leq 1\}, \quad A \in \mathbb{I}_+.$$

ただし、 $\mathbb{R}$ は拡大実数全体の集合を表す。明らかに、各 $A \in \mathbb{I}_+$ に対して、 Γ^A は凸集合となる。ここで、 $\Gamma^{\{1,2\}}$ の有界性は仮定しない。 $k = 1, 2$ に対して、 $\Gamma^{\{k\}} \cap \Gamma^{\{1,2\}}$ 上で θ_k が最大になるような点を $\theta^{(k,c)}$ とする。

$$\theta^{(k,c)} = (\theta_1^{(k,c)}, \theta_2^{(k,c)}) = \arg \max_{(\theta_1, \theta_2) \in \mathbb{R}^2} \{\theta_k; (\theta_1, \theta_2) \in \Gamma^{\{k\}} \cap \Gamma^{\{1,2\}}\}.$$

ここで、 $\Gamma^{\{1,2\}}$ が有界である場合は $\theta^{(k,c)} \in \mathbb{R}^2$ であるが、有界でないときは $\theta_k^{(k,c)} = \infty$ となる場合がある。 $\Gamma^{\{1,2\}}$ の凸性から、可能な場合分けは以下である。

$$\theta_1^{(2,c)} < \theta_1^{(1,c)}, \quad \theta_2^{(1,c)} < \theta_2^{(2,c)}, \quad (\text{C1}),$$

$$\theta_1^{(2,c)} \leq \theta_1^{(1,c)}, \quad \theta_2^{(2,c)} \leq \theta_2^{(1,c)}, \quad (\text{C2}),$$

$$\theta_1^{(1,c)} \leq \theta_1^{(2,c)}, \quad \theta_2^{(1,c)} \leq \theta_2^{(2,c)}, \quad (\text{C3}).$$

さらに、以下の点を定義する。

$$\bar{\xi}_1(\theta) = \sup\{\theta_1 \in \mathbb{R}; (\theta_1, \theta) \in \Gamma^{\{1,2\}}\}, \quad \bar{\xi}_2(\theta) = \sup\{\theta_2 \in \mathbb{R}; (\theta, \theta_2) \in \Gamma^{\{1,2\}}\}.$$

$\Gamma^{\{1,2\}}$ が有界でないときは $\bar{\xi}_k(\theta) = \infty$ となる場合もある。場合分けによって $(\tau_1, \tau_2) \in \mathbb{R}^2$ を以下の点とする。

$$(\tau_1, \tau_2) = \begin{cases} (\theta_1^{(1,c)}, \theta_2^{(2,c)}), & (\text{C1}), \\ (\bar{\xi}_1(\theta_2^{(2,c)}), \theta_2^{(2,c)}), & (\text{C2}), \\ (\theta_1^{(1,c)}, \bar{\xi}_2(\theta_1^{(1,c)})), & (\text{C3}). \end{cases}$$

さらに、以下の集合を定義する。

$$\mathcal{D}_0 = \{(\theta_1, \theta_2) \in \mathbb{R}^2; \theta_1 < \tau_1, \theta_2 < \tau_2, (\theta_1, \theta_2) \in \Gamma^{\{1,2\}}\},$$

$$\hat{\mathcal{D}}_0 = \{(\theta_1, \theta_2) \in \mathbb{R}^2; \exists(\theta'_1, \theta'_2) \in \mathcal{D}_0 \text{ s.t. } (\theta_1, \theta_2) < (\theta'_1, \theta'_2)\}.$$

このとき、以下の主結果を得る。

定理 2.1 仮定 1.1–1.3 の下、 $\mathcal{D} = \hat{\mathcal{D}}_0$ が成立する。

この定理は [3, 4] の収束領域と同じである。[3, 4] では、定常不等式を用いている。しかし、この定常不等式は 2 次元反射型ランダムウォークを背後過程によりマルコフ変調させたモデルへの拡張が困難であることがわかった。そこで本発表では、定常不等式を使わずに、 $\mathbb{C}^2$ 上の解析関数とその接続を用いて定理 2.1 を証明する。現在、マルコフ変調の場合にこの方法を用いて収束領域を求める研究を進めている。

参考文献

- [1] G. Fayolle, R. Iasnogorodski, and V. Malyshev. *Random Walks in the Quarter-Plane: Algebraic Methods, Boundary Value Problems and Applications*. Springer, New York, 1999.
- [2] G. Fayolle, V. Malyshev, and M. Menshikov. *Topics in the constructive theory of countable Markov chains*. Cambridge University Press, Cambridge, UK, 1995.
- [3] M. Kobayashi and M. Miyazawa. Revisiting the tail asymptotics of the double QBD process: Refinement and complete solutions for the coordinate and diagonal directions. *Chapter 8 in Matrix-Analytic Methods in Stochastic Models*, pages 145–185, 2013.
- [4] M. Kobayashi and M. Miyazawa. Tail asymptotics of the stationary distribution of a two dimensional reflecting random walk with unbounded upward jumps. *Advances in Applied probability*, 46:365–399, 2014.

連続時間非斉時マルコフ連鎖の特別なクラスに対する一様化法

井上 文彰
大阪大学工学研究科
電気電子情報工学専攻

概要：本稿では、連続時間非斉時マルコフ連鎖における過渡状態確率の数値計算法を考察する。遷移速度行列が非斉時であるとき、一般に過渡状態確率は行列指数形式のような単純な形で表現できず、通常の一様化法の適用は困難となる。先行研究において、非斉時マルコフ連鎖に適用可能な一様化法の拡張が考案されているが、これは時間軸を十分小さな幅で分割する、離散化の近似を含んでいる。本稿では、遷移速度行列がある特別な構造を有する場合、この一様化法は時間軸の離散化を要さない形式に書き換えられることを示す。

1. はじめに

本稿では、遷移速度行列が時刻によって変化する、すなわち非斉時である連続時間有限状態マルコフ連鎖 $\{X(t); t \geq 0\}$ を考える。 $X(t)$ は有限集合 $\mathcal{M} = \{1, 2, \dots, M\}$ ($M = 1, 2, \dots$) 上に値を取るものとする。 $Q(t)$ ($t \geq 0$) を時刻 t における遷移速度行列とし、その対角成分について次式を仮定する。

$$\hat{\theta} := \sup_{t \geq 0, i \in \mathcal{M}} |[Q(t)]_{i,i}| < \infty$$

このとき、過渡状態確率ベクトル $\pi(t) := (\Pr(X(t) = 1), \Pr(X(t) = 2), \dots, \Pr(X(t) = M))$ は

$$\frac{d\pi(t)}{dt} = \pi(t)Q(t), \quad t \geq 0$$

を満たす。斉時マルコフ連鎖では、この微分方程式の解 $\pi(t)$ ($t \geq 0$) は行列指数関数を用いて単純な形で与えられ、一様化法を用いた効率的な数値計算が可能である。一方、非斉時マルコフ連鎖では、

$$\pi(t) = \pi(0) \sum_{k=0}^{\infty} \int_{u_1=0}^t \int_{u_2=u_1}^t \cdots \int_{u_k=u_{k-1}}^t Q(u_1)Q(u_2) \cdots Q(u_k) du_k du_{k-1} \cdots du_1$$

という無限級数により $\pi(t)$ ($t \geq 0$) が与えられる。 $Q(t)$ の対角成分は負であるため、上式の積分を直接計算することは数値的に不安定である。そこで、[1] では、上式が次の形に書き換えられることを利用した、一様化法の非斉時マルコフ連鎖への拡張に相当する数値計算法が考案されている。

$$\pi(t) = \pi(0) \sum_{n=0}^{\infty} \frac{\exp[-\theta t](\theta t)^n}{n!} \cdot P_n(t) \quad (1)$$

ただし、 $\theta \geq \hat{\theta}$ であり、 $P_n(t)$ は $P(t) := I + \theta^{-1}Q(t)$ (I は単位行列) を用いて $P_0(t) = I$ および

$$P_n(t) = \int_{u_n=0}^t \int_{u_{n-1}=0}^{u_n} \cdots \int_{u_1=0}^{u_2} P(u_1)P(u_2) \cdots P(u_n) \cdot \frac{n!}{t^n} \cdot du_1 du_2 \cdots du_n, \quad n = 1, 2, \dots \quad (2)$$

で定義される。 (2) の積分を数値的に評価するため、[1] では十分小さな幅 $\Delta > 0$ を用いて、時間 $[0, t]$ を t/Δ 個の区間に分割する、時間軸の離散化による近似が提案されている。

本稿では、 $Q(t)$ ($t \geq 0$) が次式で表される特別な構造を有する場合、 (2) の右辺における積分が非負数の無限和で厳密に与えられ、時間軸の離散化による近似を要さないことを示す。

$$Q(t) = \sum_{m=0}^{\infty} \frac{\exp[-\gamma t](\gamma t)^m}{m!} \cdot Q^{(m)} \quad (3)$$

ただし $\gamma > 0$ であり、 $Q^{(m)}$ ($m = 0, 1, \dots$) は遷移速度行列を表す。

2. 主結果

$\theta = \sup_{m \in \{0,1,\dots\}} \| [Q^{(m)}]_{i,i} \|$ と置き, $P^{(m)} := I + \theta^{-1} Q^{(m)}$ ($m = 0, 1, \dots$) とすると, (3) より

$$P(t) = \sum_{m=0}^{\infty} \frac{\exp[-\gamma t](\gamma t)^m}{m!} \cdot P^{(m)} \quad (4)$$

が成立することがわかる. 定義より, $P^{(m)}$ ($m = 0, 1, \dots$) は確率行列であることが確かめられる.

$B_n(t)$ ($t \geq 0, n = 0, 1, \dots$) を, $B_0(t) = I$ ならびに次式により定義する.

$$B_n(t) = \frac{t^n}{n!} \cdot P_n(t) = \int_{u_n=0}^t \int_{u_{n-1}=0}^{u_n} \cdots \int_{u_1=0}^{u_2} P(u_1)P(u_2) \cdots P(u_n) du_1 du_2 \cdots du_n, \quad n = 1, 2, \dots$$

このとき, $B_n(t)$ は次の漸化式で与えられる.

$$B_0(t) = I, \quad B_n(t) = \int_0^t B_{n-1}(u)P(u)du, \quad n = 1, 2, \dots \quad (5)$$

$n = 1, 2, \dots, m = 0, 1, \dots$, および $k = 0, 1, \dots, m$ に対し, $\alpha_{m,n}(k)$ と $b_n(m, k)$ を次式で定義する.

$$\alpha_{m,n}(k) = \binom{n+k}{n} / \binom{m+n+1}{n+1}, \quad b_n(m, k) = \binom{m}{k} \left(\frac{n}{n+1} \right)^k \left(\frac{1}{n+1} \right)^{m-k}$$

定義より, $\sum_{k=0}^m \alpha_{m,n}(k) = \sum_{k=0}^m b_n(m, k) = 1$ が成立することに注意せよ.

(4) と (5) を用いた帰納法により, 次の補題 1 が示される (証明の詳細は省略する).

補題 1. $B_n(t)$ ($t \geq 0, n = 1, 2, \dots$) は次式で与えられる.

$$B_n(t) = \frac{t^n}{n!} \sum_{m=0}^{\infty} \frac{\exp[-n\gamma t](n\gamma t)^m}{m!} \cdot B_n^{(m)} \quad (6)$$

ただし, $B_n^{(m)}$ は次式で定められる確率行列を表す.

$$B_1^{(m)} = \frac{1}{m+1} \sum_{k=0}^m P^{(k)}, \quad B_n^{(m)} = \sum_{k=0}^m \alpha_{m,n-1}(k) \sum_{k_1=0}^k b_{n-1}(k, k_1) B_{n-1}^{(k_1)} P^{(k-k_1)}, \quad n = 2, 3, \dots$$

補題 1 ならびに $B_n(t)$ の定義より, 次式が成立する.

$$P_n(t) = \sum_{m=0}^{\infty} \frac{\exp[-n\gamma t](n\gamma t)^m}{m!} \cdot B_n^{(m)}, \quad t \geq 0, n = 1, 2, \dots$$

したがって, (1) より, 次の結果が導かれる.

定理 1. 過渡状態確率ベクトル $\pi(t)$ ($t \geq 0$) は次式で与えられる.

$$\pi(t) = \pi(0) \sum_{n=0}^{\infty} \frac{\exp[-\theta t](\theta t)^n}{n!} \sum_{m=0}^{\infty} \frac{\exp[-n\gamma t](n\gamma t)^m}{m!} \cdot B_n^{(m)} \quad (7)$$

3. まとめ

本稿では, 非斉時連続時間非斉時マルコフ連鎖に対する一様化法を考察した. 遷移速度行列 $Q(t)$ が (3) に示す特別な構造を有する場合, 先行研究で与えられている過渡状態確率の公式 (1) ならびに (2) における積分が (7) に示す無限和に置き換えられることを示した. 本稿の結果から, このクラスの非斉時マルコフ連鎖については, (7) を直接計算する数値計算法の構築が可能であると期待される.

謝辞

本研究の一部は JSPS 科研費 JP16H06914 の支援を受けている.

参考文献

- [1] A. P. A. van Moorsel and K. Wolter, Numerical solution of non-homogeneous Markov processes through uniformization. Proc. 12th European Simulation Multiconference, 710–717, 1998.

強相関近似による複雑ネットワーク上の情報拡散過程の解析

南川 雅人* 中島 圭佑† 塩田 茂雄*

* 千葉大学 大学院融合理工学府 都市環境システムコース

† 千葉大学 工学部 都市環境システム学科

概要：実社会で人々の関心事が発生すると、SNSに関連する書き込みが多数投稿され、ピークを迎えたのち沈静化する様子が観察される。本研究では、この現象は実社会での情報拡散過程が SNS 上に現れた結果であるとして、感染症の数理モデルとして使われる SIR モデルを用いて、SNS への投稿数の時間推移を解析的に近似評価する手法を提案する。提案手法は、隣接ノード間では情報保持状態にある事象の相関性が高いとする仮定を用いており（強相関近似）、この仮定の妥当性をシミュレーションにより検証する。

1. まえがき

実社会で人々の関心事が発生すると、関連する書き込みが SNS に投稿され、単位時間あたりの投稿数がピークを迎えた後、書き込みは次第に減少して、やがて沈静化する。この現象は、実社会で情報が拡散して、やがて広く知られるところとなり、それ以上の情報伝達（SNS への投稿）が不要になるまでの過程、つまり実社会における情報拡散過程が SNS 上に表出した現象と解釈できる。情報の拡散は感染症の流行と類似していることから、本研究では、SNS への投稿件数の時間変化を、感染症の数理モデルとして使われている SIR モデル [1] により表現し、モデルに基づいて投稿数の時間推移を解析的に近似評価する手法を提案する。

2. 情報拡散モデル

2.1. モデルの説明

N 個のノードからなる有向ネットワークにおいて、1 つないしは複数のノードを起点として、情報が有向リンクを経由してネットワーク内の各ノードに拡散していく現象について考察する。ノードは以下のいずれかの状態を取る。

状態 0: 情報を知らない (SIR モデルにおける Susceptible 状態)

状態 1: 情報を知っており、将来拡散予定 (SIR モデルにおける Infected 状態: 感染力あり)

状態 2: 情報は知っているが、拡散しない (SIR モデルにおける Infected 状態: 感染力なし)

状態 3: 情報を知り、拡散済み (SIR モデルにおける Recovered 状態)

時刻 0 では情報発信元のノードは状態 1 にあり、それ以外の全てのノードは状態 0 にある。ノード i は（隣接ノードから）情報を受信することで、状態 0 から確率 q_i で状態 1 に、確率 $1 - q_i$ で状態 2 に遷移する。状態 1 に遷移後は、平均 $1/\lambda_i$ の指数分布に従う時間経過後に、全ての隣接ノードに同時に情報転送を行い、状態 3 に遷移する。状態 3 に遷移後は、そのまま状態 3 に留まる。状態 0 から状態 2 に遷移した場合は、状態 2 に留まるが、別の隣接ノードから情報を受信すると確率 q_i で状態 1 に遷移する。なお $q_i = 1$ とした場合の結果は、文献 [2] で述べている。

隣接行列を $A = \{a_{ij}\}$ とし、ノード i からノード j への情報転送によりノード j が（状態 0 から）状態 1 に遷移する場合に 1、遷移しない場合に 0 を取る確率変数を Y_{ij} とする ($a_{ij} = 0$ のときは $Y_{ij} = 0$ とする)。時刻 t においてノード i がとる状態の番号を $Z_i(t)$ で表すこととすると、時刻 t でのネットワークの状態は、確率変数 $\mathbf{Y} \stackrel{\text{def}}{=} \{Y_{ij}\}_{i,j=1,\dots,N}$ の値および各ノードの状態の組 $(Z_1(t), Z_2(t), \dots, Z_N(t))$ で記述され、 $(Z_1(t), Z_2(t), \dots, Z_N(t))$ の遷移は連続時間マルコフ連鎖に従う。時刻 t でノード i が状態 1 か状態 3 にいる場合に 1、その他の場合に 0 を取る確率変数を $X_i(t)$ で表すと、時刻 t までに情報拡散を行ったノード数（SNS への投稿件数に相当） $L(t)$ の期待値は $\{E[X_i(t)]\}_{i=1,\dots,N}$ から次式で求まる。つまり、 $\{E[X_i(t)]\}_{i=1,\dots,N}$ から、投稿件数の時間変化が評価できる。

$$E[L(t)] = \sum_i \int_0^t \lambda_i E[X_i(s)] e^{-\lambda_i(t-s)} ds. \quad (1)$$

2.2. 解析手法

$(Z_1(t), Z_2(t), \dots, Z_N(t))$ の取り得る状態数は 4^N であるので, $(Z_1(t), Z_2(t), \dots, Z_N(t))$ の遷移を厳密に追うことは, 確率変数 $\mathbf{Y}$ の実現値が既知であっても困難である. 本節では, $p_i(t|\mathbf{Y}) \stackrel{\text{def}}{=} E[X_i(t)|\mathbf{Y}]$ の時間変化に関する微分方程式を近似的に導出し, 導出した微分方程式を数値的に解いて, ネットワークの状態遷移の過渡特性を解析的に評価することを試みる. 以下が得られる.

$$\frac{dp_i(t|\mathbf{Y})}{dt} = \sum_j a_{ji} Y_{ji} \lambda_j (p_j(t) - E[X_j(t)X_i(t)|\mathbf{Y}]). \quad (2)$$

ここで, 右辺の $E[X_j(t)X_i(t)|\mathbf{Y}]$ (以下, 相関項) について, 以下の上下限を与えることができる.

$$p_j(t|\mathbf{Y})p_i(t|\mathbf{Y}) \leq E[X_j(t)X_i(t)|\mathbf{Y}] \leq \min\{p_j(t|\mathbf{Y}), p_i(t|\mathbf{Y})\}.$$

本稿では, 相関項をその上限式で代用する, すなわち (2) において $E[X_j(t)X_i(t)|\mathbf{Y}] = \min\{p_j(t|\mathbf{Y}), p_i(t|\mathbf{Y})\}$ を仮定し, 数値的に $p_i(t|\mathbf{Y})$ を求める近似を検討する. これを強相関近似と呼ぶ [2]. なお, 多くの既存研究では, 相関項をその下限式で代用する, すなわち $E[X_j(t)X_i(t)|\mathbf{Y}] = p_j(t|\mathbf{Y})p_i(t|\mathbf{Y})$ を仮定している. これを独立近似と呼ぶ. 投稿件数の評価のためには, $E[X_i(t)] = E[p_i(t|\mathbf{Y})]$ を求める必要があるが, 全ての $\mathbf{Y}$ の実現値について $p_i(t|\mathbf{Y})$ を求めて平均を取ることは困難であるので, ランダムに 1 つ選んだ $\mathbf{Y}$ の実現値について $p_i(t|\mathbf{Y})$ を計算し, その結果を $E[X_i(t)]$ の近似値として用いる.

3. 数値例

2.2 節で述べた近似解析法の精度を数値的に検証するため, Twitter のトポロジーデータ [3] (ノード数 81306, リンク数 1768149) を用いて, 二つの異なるノードからそれぞれ情報を拡散させたときの書き込み数の時間変化を, シミュレーションおよび近似解析法により評価し, 結果を比較した. パラメタ値として, $\lambda_i = 1$, $q_i = E[Y_{ji}|a_{ji} = 1] = 0.3$ を仮定した.

結果を図 1 および図 2 に示す. シミュレーションは, 1000 回の結果の平均をプロットしたものである. 近似解析法では, ランダムに選んだ 5 通りの $\mathbf{Y}$ の実現値に対して結果を求めており, 図には各結果をそれぞれ異なる (5 本の) 実線で示した. $\mathbf{Y}$ の実現値を変えても近似解析法による結果にはほとんど差異は見られず, ランダムに 1 つ $\mathbf{Y}$ の実現値を選んで計算することは合理的であると考えられる. 図には, 独立近似を採用した場合の結果も示した. 強相関近似は独立近似よりもシミュレーションに近い結果を与えていること (特に発信ノードの出次数が大きい場合) が確認できる.

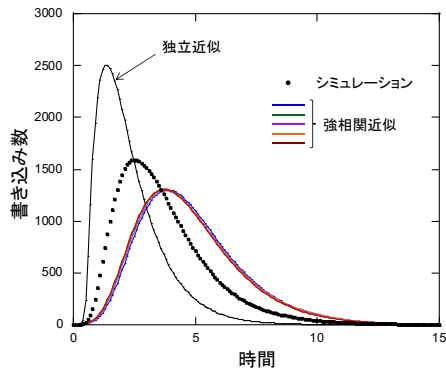


図 1: 書き込み数 (発信ノード出次数 = 59)

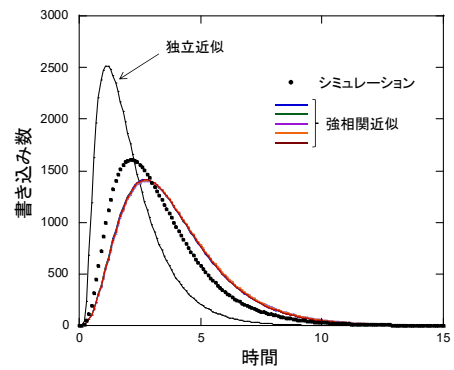


図 2: 書き込み数 (発信ノード出次数 = 1111)

参考文献

- [1] R. Pastor-Satorras, C. Castellano, P.V. Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Reviews of Modern Physics*, vol.87, pp.926–979, 2015.
- [2] 中島圭佑, 南川雅人, 塩田茂雄, “SNS における投稿件数推移分析のための情報拡散モデル,” 電子情報通信学会コミュニケーションクオリティ研究会, CQ2017-84, pp.79–84, 2017.
- [3] “Stanford large network dataset collection.” <http://snap.stanford.edu/data/>.