

Crowdsourcing and Stochastic Modeling

Srinivas R. Chakravarthy¹, Ph.D & Serife Ozkar²

¹ Department of Industrial and Manufacturing Engineering, Kettering University, Flint, Michigan, 48504, USA.

² Serife Ozkar, Department of Statistics, Dokuz Eylul University, Faculty of Science, Tinaztepe Campus, 35160, Buca Izmir, Turkey

Correspondence: Srinivas R. Chakravarthy, Ph.D, Professor and Department Head, Industrial and Manufacturing Engineering Department, Kettering University, Flint, MI 48504, USA

Received: May 2, 2016

Accepted: May 17, 2016

Online Published: May 20, 2016

doi:10.5430/bmr.v5n2p19

URL: <http://dx.doi.org/10.5430/bmr.v5n2p19>

This paper was invited for publication.

Abstract

Crowdsourcing has been used in different domains gaining significant exposure in many fields such as healthcare, computer science, environmental sciences, business and marketing. The meaning and interpretation of crowdsourcing is varied and despite its popularity among companies in many sectors it remains little understood. However, the modeling aspects of crowdsourcing in the context of stochastic and algorithmic methods have not been considered until recently. Thus, in this expository paper we intend to provide needed connection between these two areas.

Keywords: Crowdsourcing, Matrix-analytic method, Phase type services, Algorithmic probability

1. Introduction

Crowdsourcing is a concept that has been tested and used by many industries. According to Howe (2006), "Simply defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers."

Crowdsourcing has been used in different domains gaining significant exposure in many fields such as healthcare, computer science, environmental sciences, business, finance, and marketing and we refer the reader to a recent survey paper by Hosseini et al. (2015). A recent review of crowdsourcing literature in the context of manufacturing industry can be seen in Evans et al. (2016). Also, the meaning and interpretation of crowdsourcing is varied and despite its popularity among companies in many sectors it remains little understood (Djelassi and Decoopman, 2013). We refer the reader to Quora (2015) for a number of examples of crowdsourcing in many sectors.

With innovation, which is key to survival of any business, *crowdsourcing* plays a vital role (AyferBozat and Erenel, 2016) in business letting new ideas flow in from outer sources, thanks to open innovation.

Jayanti (2012) suggested dividing crowdsourcing into three types: content crowdsourcing, technical crowdsourcing and competition crowdsourcing. Content crowdsourcing occurs when organizations, both for profit and nonprofit, produce contents by directly connecting to users. These users are paid very little to nothing, and rarely have the opportunity for larger payments. The examples quoted for context crowdsourcing include Wikipedia, Facebook, Twitter and YouTube. Technical crowdsourcing occurs when organizations exploit external sources such as Websites. For example, Amazon Mechanical Turk (AMT or MTurk), Microworkers, Innocentive, oDesk and eLance are quoted in Jayanti (2012) as relevant ones. These and other similar websites are also referred as crowdsourcing platforms. Competition crowdsourcing occurs when companies challenge highly talented people to complete the company's strategic and crucial tasks, and reward the participants accordingly. Doritos has used the competition crowdsourcing method to develop content for a Superbowl ad (Jayanti, 2012). Another classification of crowdsourcing applications is given in four categories: voting system, information sharing system, game and creative system (Yuen, King, & Leung, 2011).

However, only recently stochastic models, more specifically queueing models, useful in crowdsourcing in the context of service sectors have been studied. The purpose of this expository article is two-fold. First and foremost is to bring the concept *crowdsourcing* to a wider audience, especially, researchers in stochastic modeling. Secondly, to show how stochastic modeling can be a natural companion to business, especially service sector ones, in the modeling of some aspects of *crowdsourcing*.

The rest of the paper is organized as follows. In Section 2, we will give a brief literature review of *crowdsourcing* in various fields and how stochastic modeling, especially queueing model, is applied to a specific type of *crowdsourcing*. Section 3 contains a very brief introduction to key concepts in stochastic modeling needed to integrate queueing models into the specific *crowdsourcing* under consideration and in Section 4 we will review the queueing models studied in this context of *crowdsourcing*. Some concluding remarks and future work are given in Section 5.

2. Literature Review

One of the most successful domains in which crowdsourcing has been used is in the financial services industry. Crowdsourcing in finance occurs in the form of crowdfunding. Crowdfunding is defined as raising monetary contributions from a large number of people and is considered as one of the new sources of firm financing along with shadow banking industry and trade credit (Riasi, 2015). Various studies have focused on the innovative forms of firm financing which are based on the concept of crowdsourcing (e.g., Belleflamme, et al., 2014; Riasi, 2015; Mollick, 2014). Hasteer, Bansal, & Murthy (2015) analyzed the process of software development at a Crowdsourced platform. They modelled the development process by using Markov Decision Process and presented an illustration of their proposed model. The temporal behavior of crowd performance to achieve the best performance in terms of accuracy and completion time of large batches of tasks was studied in Bhattacharya et al. (2015) with the help of time-series models. The authors, using simulation testbed, constructed an automated system for optimally scheduling a large batch of tasks on multiple crowdsourcing platforms.

The literature on crowdsourcing dealing with various definitions and classification is huge. On the other hand, the literature on the analysis of the crowdsourcing with the help of mathematical models is very small. Having quantitative models for crowdsourcing platforms help businesses to understand the underlying processes and predict behavior when the system parameters change. Survival analysis is popular model for predicting task completion time, especially in non-real-time scenarios (Faridani, Hartmann, & Ipeirotis, 2011). They modelled the arrival process of workers using a non-homogeneous Poisson process and derived the optimal pricing strategies for the employers. The model is based on a crawled dataset from MTurk. Wang et al. (2011) analyzed the completion time of crowdsourcing campaigns by using again a survival analysis based on a crawled dataset from MTurk. They were able to show the impact of time-independent factors, e.g., the payment or the type of the task, on the completion time. Hirth, Hossfeld and Tran-Gia (2012) analyzed a crowdsourcing platform by using majority decision approach and control group approach. Then, they developed the cost model for the platform. The material flow of crowdsourcing processes in manufacturing systems was studied by using stochastic petri nets (Wu, Rosen, & Schaefer, 2014).

However, only recently stochastic models, more specifically queueing models, useful in crowdsourcing in the context of service sectors have been studied. We want to stress that the paper by Bernstein, Karger, Miller, & Brandt (2012) deals with optimizing a real-time crowdsourcing and has nothing to do with the way the crowdsourcing is defined and modeled in the papers outlined in this expository article. Bernstein, Karger, Miller, & Brandt (2012) study a “retainer model” in which the system will try to recruit “servers” from the pool of “workers” who have been pre-determined and paid ahead of time to act as “on-call servers” as opposed to using the customers who visit the system to serve fellow customers. The model was validated in experiment with users on MTurk. Schwartz, Borchert, Hirth and Tran-Gia (2015) considered the crowdsourcing platform as a $M^{(X)}/M/c - \infty$ queue model and used a dataset directly provided from the commercial crowdsourcing platform Microworkers. They extended the study of Bernstein et al. (2012) by also considering the time until the processing of any task in the campaign begins.

The models described in this expository article are motivated by the leading retailer, Walmart which considered a plan that would get its in-store customers to deliver packages to its online customers (Forbes, 2013). To our knowledge these are the first set of queueing models (Chakravarthy & Dudin, 2015; Chakravarthy & Ozkar, 2015; Chakravarthy & Ozkar, 2016a; Chakravarthy & Ozkar, 2016b) incorporating the *crowdsourcing* concept in the sense of using the customers as possible servers after these customers are served by one of the system servers. For example, one type of customers visits the store to procure items while the other type of customers orders over some medium such as Internet and phone and expects them to be delivered. The store management may opt to use the customers visiting them as couriers to “serve” the other type of customers. Not all in-store customers may be willing and in

some cases not possible to act as servers for the store. Hence a probability is introduced for in-store customers to opt for servicing the other type.

3. Brief Introduction to Stochastic Modeling, Phase Type Distribution and Markovian Arrival Process

3.1 What is Stochastic Modeling?

Modeling plays an important role in practice. Models are representations of objects, processes or anything that we wish to describe or whose patterns of behavior we wish to analyze. They enable us to understand the behavior of complex systems. Modeling is an iterative process involving several stages. They can be summarized as follows:

- **Formulation Stage:** study the problem; propose a model.
- **Solution Stage:** choose appropriate tools to bring out essential characteristics of the model.
- **Validation Stage:** verification of the model and propose modification, if any.

There are basically two types of modeling: deterministic modeling and stochastic modeling. The deterministic models include topics such as mathematical programming (linear and nonlinear programming), scheduling and sequencing. The stochastic modeling (also referred to as probabilistic modeling) covers areas such as Markov processes, queueing theory, reliability theory, warranty analysis, inventory, and probabilistic scheduling. Note that statistical modeling that deals with the study of uncertainty and variability is subsumed in stochastic modeling in our presentation here.

3.2 Motivation to Study Stochastic Modeling

To motivate the need for the study of stochastic modeling in practice, let us look at some examples.

Example 3.2.1: Orders for a certain product are to be met by the manufacturer. There are costs for making, storing the products and for not meeting the demand. There is revenue for selling the product. From the manufacturer's point of view, how should the production be planned that will maximize the company's profit? Usually the orders are known only in some probabilistic way. That is, the number of orders will follow a certain stochastic process. Also the production time of a finished product will usually be a random variable.

Example 3.2.2: In practice the waiting room (area where arriving jobs or customers are placed before offering services) is usually finite. Arrival and service times of the jobs are usually random. In order to attract customers and serve efficiently the management has to control the service rates. That is, increase the service rate when the buffer exceeds a certain level, say, N , and then bring back to normal when the buffer gets back to a level, say, L . There are various costs such as cost of increasing the service rate, cost of lost jobs, and cost of waiting in the buffer. From the management point of view, how should the values of L and N be chosen (by fixing all other values) so that the total cost will be minimized?

Example 3.2.3: In both communications and industrial (manufacturing) engineering systems we see jobs that require primary and secondary services. For example, in manufacturing area jobs may have some additional processing such as special labeling, inspection procedure, and special handling. In communications a job may leave behind some secondary work of administrative in nature. These secondary jobs may be processed later (by storing them separately in an auxiliary buffer). To study these types of models in terms of the number of jobs waiting for primary and secondary services, the rate of jobs completed per unit of time, the fraction of idle time of the server, we need stochastic modeling.

Example 3.2.4: Complex high technology products, equipment, and systems have become a critical part of the infrastructure of today's businesses and homes. In recent years, companies selling high-tech manufactured products are realizing that a customer's product purchase decision is not only influenced by the product's value, but also by the service support available after the sale of the product. A major part of after-sales support includes installation, maintenance, and repair of equipment at customers' site, which is referred to as field service. Equipment failures reduce productivity. Reducing machine downtime by providing prompt and successful field service is very important for customer satisfaction. One of the factors that impacts downtime is the extent of worker cross-training employed. Firms have a choice of employing specialists who is dedicated to serve one type of customer, or generalists (or cross-trained workers) who can serve more than one type of customers. Cross-trained workers represent flexible capacity, since they can be readily assigned to serve any type of incoming customers. Although cross-training increases server flexibility and improves responsiveness, it also increases service costs and may reduce service efficiency. On the other hand, using specialized servers increases the downtime because of unbalanced server loads due to lack of work sharing. The question is: Given various costs and a fixed number of servers, how many dedicated servers of each type and how many cross-trained workers needed so as to minimize total expected cost per unit time?

Naturally, this type of problems can generally be solved only algorithmically due to the nature of the objective function that is not explicitly known in terms of the parameters of the model.

Example 3.2.5: One of the major aspects in any product development process is the design stage of a product. In this stage, different types of engineers, such as design release engineers, test engineers, and quality engineers are involved. We partition these stages into various nodes so that there is a possibility of feedback from the next node to the current node. This feedback is very important as it carries significant information such as modification of certain parameters of the design, and identification of improper specifications. The management's primary concerns are (a) the feasibility of making such a product and the process of making the products; (b) the proper allocation of resources at each node including release of resource from one node to another; and (c) timely release of the design to the manufacturing plant. Items (a) and (b) are very closely linked in the sense that there will be a lot of feedback from one to another. Modeling this very practical problem involves sophisticated queueing networks and the analysis of such networks for use in practice calls for a very careful algorithmic implementation.

3.3 Modeling to Algorithmic Probability

In practice, the qualitative aspects of stochastic models are studied mostly through computation. Computing is both an art and science. It has become an integral and important part of the analysis of any stochastic model. Days are gone when one has the need to have explicit analytical solutions. Algorithmic or computational probability plays a major role in the analysis and implementation of stochastic models in practice. Computational aspects are not luxury anymore! Modern day development of algorithmic probability is mainly due to Neuts (1981).

There are several such practical examples that fall under the topic of stochastic modeling. It is important to note that stochastic models (in general the mathematical models) play an important role in bridging the gap between theory and practice. Models often need to be modified to better approximate the practical situations.

Wherever we go and whatever we do, we invariably have to wait for our job (whatever it is) to be done. When we look at the total time spent in executing the job, it includes the actual time for executing the job plus the waiting time (the time spent before our jobs gets into service). The study of such waiting line problems both from the customer's and service facility's points of view, in order to best utilize the resources, falls under the general topic of queueing models.

Queueing theory plays an important role in many areas, notably telecommunications, production and manufacturing engineering, computer science, and other service offering facilities. Developed originally in the context of telephone traffic congestion by Danish Mathematician A.K. Erlang in 1909. He noticed that the telephone systems were generally characterized by either Poisson input and exponential holding times and multiple channels; or by Poisson input and constant holding times and a single channel. Erlang was responsible for the concept of stationary equilibrium. Molina, in 1927, published "Applications of the Theory of Probability to Telephone Trunking Problems". In early 1930s, F. Pollaczek did some fundamental work on Poisson input, arbitrary output, and single/multiple channel problems. At the same time, Kolmogorov, Khintchine (from Russia), Crommelin (France) and Palm (Sweden) did some additional work. Other additional contributions include works by D.R. Lindely, L. Takacs, D.G. Kendall, S. Karlin, J.L. McGregor, J.D.C. Little, M.F. Neuts, N.U.Prabhu, U.N. Bhat, T.L. Saaty, and R. Syski. Modern queueing theory including computational aspects include pioneering work by M.F. Neuts. We refer the reader to the latest queueing theory book by Bhat (2015).

During the 1970's Neuts with a genuine concern for the usefulness of stochastic models in practice developed algorithmic probability and matrix analytic methods. In the first stage of that development he introduced phase type distributions. Phase type distributions are natural generalizations of exponential distributions and have found tremendous applications among others in queueing, reliability, inventory, and warranty models. Furthermore, these distributions require simple matrix formalism not only in the analysis but also in algorithmic implementation. Phase type distributions can easily be incorporated into undergraduate curriculum as students are exposed to matrix theory and basic probability in almost all engineering curriculum. In the second stage, Neuts developed matrix geometric methods in queueing and popularized the algorithmic methods as a way to study stochastic models. Then he introduced Markovian arrival process, a versatile point process that not only generalized phase type renewal process but also provides a way to model correlated arrivals. Details of phase type distributions and matrix analytic methods in stochastic models can be seen in Neuts (1981). Details on Markovian arrival processes can be seen in Neuts (1989), Lucantoni (1991), Chakravarthy (2001), Chakravarthy (2010), Artalejo et. al.,(2010), and Bhat (2015). However, for the sake of completeness we will briefly summarize these in the following.

3.4 Phase type distributions

Poisson processes and exponential distributions have very nice mathematical properties that make queueing models with these as arrival processes or service time distributions very attractive and tractable. See for example, the book by Bhat (2015). However, these assumptions are highly restrictive in applications. Thus, seeing this limitation Neuts (1981) first developed the theory of phase type (PH) distributions and related point processes. In stochastic modeling, PH-distributions lend themselves naturally to algorithmic implementation. They have nice closure properties and a related matrix formalism that make them attractive for use in practice. Specifically, in queueing, reliability and warranty analysis, phase type distributions play a major role. Phase type distributions are defined in discrete time as well as in continuous time. For our purposes here we will very briefly describe PH-distributions in continuous time.

Suppose that $\{X_t\}$ is continuous-time Markov chain (CTMC) on $\{1, 2, \dots, m, m + 1\}$ with m transient states $1, 2, \dots, m$ and an absorbing state $m + 1$. For details on CTMC we refer the reader to, say, Ross (2014). The generator of the CTMC is of the form

$$\tilde{Q} = \begin{bmatrix} T & \mathbf{T}^0 \\ \mathbf{0} & 0 \end{bmatrix}, \tag{1}$$

where T is an $m \times m$ matrix, \mathbf{T}^0 is a column vector of order m such that $T\mathbf{e} + \mathbf{T}^0 = 0$. Assume that the matrix $T\mathbf{e} + \mathbf{T}^0\boldsymbol{\alpha}$ is irreducible. Let $(\alpha_1, \dots, \alpha_m, \alpha_{m+1}) = (\boldsymbol{\alpha}, \alpha_{m+1})$ be a probability vector giving the initial probabilities. That is, $P(X_0 = i) = \alpha_i, 1 \leq i \leq m + 1$. Suppose we start the CTMC in one of the m transient states. We are interested in finding the time until absorption into the absorbing state. Let Y denote the time until absorption into state $m + 1$. Then Y is a continuous random variable on $[0, \infty)$ with probability density given by $f(t) = \boldsymbol{\alpha}e^{Tt}\mathbf{T}^0, t \geq 0$. In this case, we say that Y follows a PH-distribution with representation $(\boldsymbol{\alpha}, T)$ of order m and denote this by $Y \sim PH(\boldsymbol{\alpha}, T)$ of order m . The mean and variance of Y are given, respectively, by $\mu_Y = \boldsymbol{\alpha}(-T)^{-1}\mathbf{e}$ and $\sigma_Y^2 = 2\boldsymbol{\alpha}T^{-2}\mathbf{e} - \mu_Y^2$. It should be pointed out the class of continuous PH-distributions is dense in the class of all distributions with support on the nonnegative real line.

Example 3.4.1: The classical exponential distribution with rate λ is a very simple PH-distribution with $m = 1, \boldsymbol{\alpha} = 1$, and $T = (-\lambda)$.

Example 3.4.2: The Erlang distribution of order m with parameter λ is a PH-distribution with representation:

$$\boldsymbol{\alpha} = (1, 0, \dots, 0) \text{ and } T = \begin{bmatrix} -\lambda & \lambda & 0 & \dots & 0 \\ 0 & -\lambda & \lambda & \dots & 0 \\ 0 & 0 & -\lambda & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda \end{bmatrix}.$$

Example 3.4.3: The generalized Erlang of order m with parameters $\lambda_1, \dots, \lambda_m$ is a PH-distribution with representation:

$$\boldsymbol{\alpha} = (1, 0, \dots, 0) \text{ and } T = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & \dots & 0 \\ 0 & -\lambda_2 & \lambda_2 & \dots & 0 \\ 0 & 0 & -\lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_m \end{bmatrix}.$$

Example 3.4.4: A hyperexponential, which is a mixture of m exponentials with parameters $\lambda_1, \dots, \lambda_m$ and mixing probabilities, respectively, given by p_1, \dots, p_m is a PH-distribution with representation:

$$\boldsymbol{\alpha} = (p_1, p_2, \dots, p_m) \text{ and } T = \begin{bmatrix} -\lambda_1 & 0 & 0 & \dots & 0 \\ 0 & -\lambda_2 & 0 & \dots & 0 \\ 0 & 0 & -\lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_m \end{bmatrix}.$$

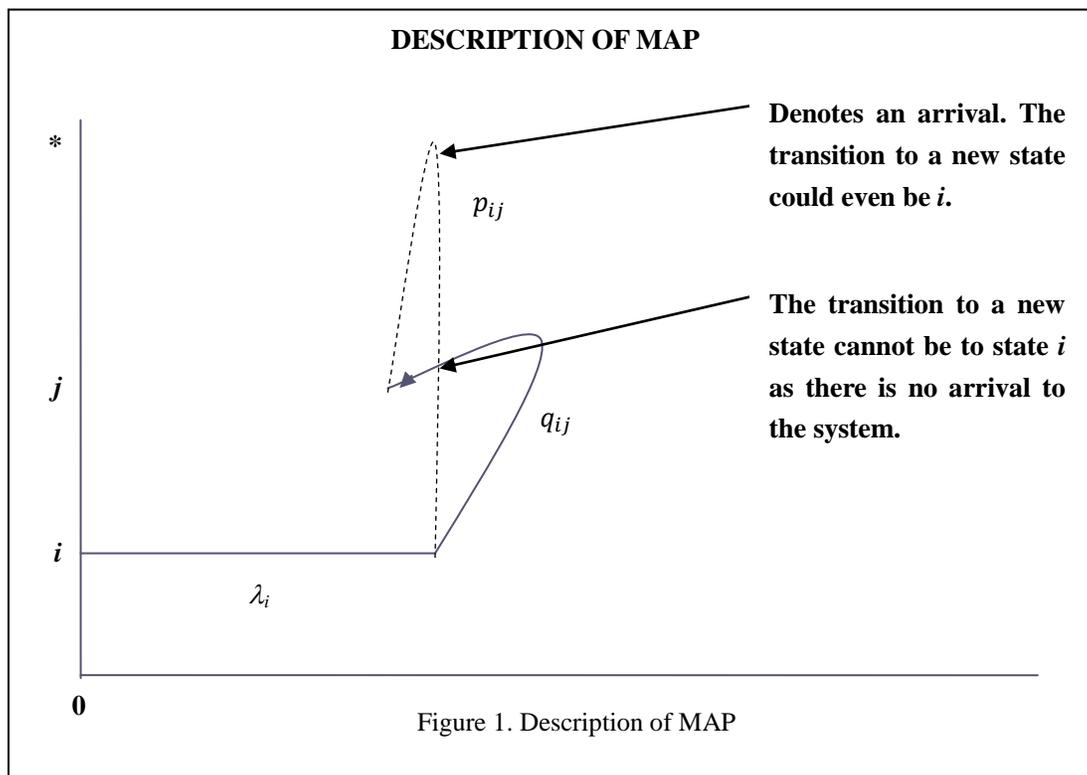
3.5 Markovian Arrival Process

If one looks at Example 3.2.1 dealing with the orders to be met by the manufacturer in practice, these orders arrive from different sources and even if each of these source generate demands for orders in renewal fashion, the pooled arrivals of demands to the order site may not necessarily form a renewal process. Also, in a typical assembly line problem, the output from station 1 will form the input to station 2. If we consider non-Poisson arrivals, say, to the first node, the output process will not necessarily be a renewal process. Also, in production line problems where arrivals from different sources form input to an assembly system, the arrival process may not necessarily be a

renewal process. So, how do we model these point processes? The answer is to use Markovian Arrival process or simply referred to as *MAP*. Neuts (1979) first introduced this process in the context of bulk arrivals as a *Versatile Markovian Point Process (VMPP)* with very heavy notation. Later on Lucantoni, et al., (1990) simplified the original notation and introduced Neuts' process (it is worth mentioning that Ramaswami (1980) coined *VMPP* as *N*-process and studied *N/G/1* queueing model) as *Batch Markovian Arrival Process (BMAP)*. These notation are now standard in describing a *BMAP* as well as *MAP*, which is a special case of *BMAP* by setting the group size of the arrivals to be one.

Consider an irreducible CTMC with m transient states. At the end of a sojourn in state i , that is exponentially distributed with parameter λ_i , there are two possibilities. The first possibility corresponds to an "event" (or an arrival) and the CTMC can visit any of the m transient states including the state from which this event occurred with probability p_{ij} . The second possibility corresponds to no arrival and the CTMC can visit any of the $(m - 1)$ transient states (all m except i) with probability q_{ij} . Thus, the CTMC can go from state i to state i only through an arrival.

Define matrices $D_0 = (d_{ij}^{(0)})$ and $D_1 = (d_{ij}^{(1)})$ such that $d_{ii}^{(0)} = -\lambda_i, 1 \leq i \leq m, d_{ij}^{(0)} = \lambda_i q_{ij}, j \neq i, 1 \leq i, j \leq m$, and $d_{ij}^{(1)} = \lambda_i p_{ij}, 1 \leq i, j \leq m$, with $\sum_j p_{ij} + \sum_{j \neq i} q_{ij} = 1$, for $1 \leq i \leq m$. By assuming that D_0 to be a nonsingular matrix, the successive times between "events" (or arrivals) will be finite with probability one and that the process will not terminate. A pictorial description of this process is given in Figure 1 below.



Thus, a *MAP* is described by the parameter matrices (D_0, D_1) of order m such that the transitions corresponding to no arrivals are governed by D_0 and the transitions corresponding to arrivals are governed by D_1 . The underlying CTMC has the generator given by $Q = D_0 + D_1$.

The point process described by the *MAP* is a special class of semi-Markov processes with transition probability matrix given by

$$\int_0^t e^{D_0x} D_1 dx = [I - e^{D_0t}](-D_0)^{-1} D_1.$$

Let π be the steady state probability vector of the generator Q governing the underlying CTMC satisfying (e denotes a column vector of 1's of appropriate dimension). That is, $\pi Q = 0, \pi e = 1$.

Let α be the initial probability vector of the underlying CTMC with generator Q . We can choose α in a variety of

ways to model different scenarios. For example, if want the time origin to coincide with an arbitrary arrival point, we can take $\alpha = c\pi D_1$, where c is the normalizing constant to make the vector α a probability vector. Other scenarios include the time origin to be the end of an interval during which there are at least k arrivals, the point at which the system is in specific state such as the busy period ends or busy period begins. The most interesting case is the one where we get the stationary version of the MAP by $\alpha = \pi$. The fundamental rate (the rate of arrivals per unit of time), λ , is given by $\lambda = \pi D_1 e$. Often in model comparisons it is convenient to select the time scale of the MAP so that λ has a certain value. This is accomplished by multiplying the parameter matrices D_0 and D_1 by the appropriate common constant.

Example 3.5.1: The classical Poisson process is a very special case of MAP Suppose that the inter-arrival times are exponentially distributed with parameter η . Then in this case, we have $m = 1, D_0 = -\eta, D_1 = \eta$.

Example 3.5.2: The PH-renewal process with the inter-renewal (arrival) times having a representation (α, T) of order m , is a particular case of MAP with $D_0 = T, D_1 = T^0 \alpha$.

Example 3.5.3: Markov Modulated Poisson Process (*MMPP*), which is very widely used in many fields, notably, in computer and communication process and Web traffic, is a doubly stochastic Poisson process in which the instantaneous arrival rate depends on an auxiliary Markov process. Suppose that the Markov process has the generator Q of order m . When that Markov process is in state i , arrivals occur at rate λ_i . Then, the *MMPP* with (single) arrivals is a *MAP* with $D_0 = Q - \Delta(\lambda), D_1 = \Delta(\lambda)$, where $\Delta(\lambda)$, is a diagonal matrix whose i^{th} diagonal entry is given by λ_i . Thus, in this *MAP* arrival transitions do not change the current state of the underlying *CTMC*. Note that a two-state *MMPP* in which only one diagonal of $\Delta(\lambda)$ is positive is the *interrupted Poisson process* (*IPP*).

Example 3.5.4: Markov Switched Poisson Process, a point process consisting of geometric runs of intervals that are exponentially distributed with rates depending on the state of the underlying discrete time Markov chain with m transient states and with transition probability matrix P , is a *MAP* with $D_0 = -\Delta(\lambda)$ and $D_k = \Delta(\lambda)P$, where $\Delta(\lambda)$ is a diagonal matrix whose i^{th} diagonal entry is given by λ_i . In this special case of *MAP*, only arrivals can change the current state of the underlying *CTMC*.

For more details on *MAP* and *BMAP* we refer the reader to Artalejo, et al. (2010), Bhat (2015), Chakravarthy (2001, 2010), Lucantoni, et al. (1990), Lucantoni (1991), Neuts (1979, 1989), and Ramaswami (1980). We also strongly recommend readers to refer to the book by Neuts (1995) that deals with a wealth of information and problems related to PH-distribution, *MAP*, and *BMAP*.

4. Queueing Models and Crowdsourcing

Only recently, queueing models with *crowdsourcing* useful in the context of service sectors have been studied. These studies involve queueing models of various types. In the next few subsections we will briefly describe these queueing models and some key observations made using those models.

4.1 Crowdsourcing and M/M/c Queueing Model

The first to study crowdsourcing with the help of a queueing model including from the management perspective was by Chakravarthy and Dudin (2015). The authors use crowdsourcing in the context of service sectors getting possible help from one group of customers who first receive service from them and then opt to execute similar services to another group of customers. For example, one type of customers visit the store to procure items while the other type of customers orders over some medium such as Internet and phone and expects them to be delivered. The store management can use the customers visiting them as couriers to "serve" the other type of customers (see Forbes (2013)). Not all in-store customers may be willing and in some cases not possible to act as servers for the store. Hence a probability is introduced for in-store customers to opt for servicing the other type.

In Chakravarthy and Dudin (2015) the authors consider a c -server queueing system in which two types, say, Type 1 and Type 2, of customers arrive. Type 1 customers arrive according to a Poisson process with rate λ_1 while Type 2 customers, who are essentially virtual ones ordering through phone and Internet, arrive according to a Poisson process with rate λ_2 . Type 1 customers are to be served by one of c servers and the service times are assumed to be exponential with parameter μ_1 . Type 2 customers may be served by a Type 1 customer having already been served and also available to act as a server or by one of c servers. Type 2 customer getting serviced by a Type 1 customer depends on the following conditions. First, that Type 1 customer should have just finished getting a service and opts to service a Type 2 customer. Secondly, at the time of opting to serve there is at least one Type 2 customer waiting to get a service. That is this Type 2 should not have started getting a service from one of c servers. We assume that a served Type 1 customer will be available for serving a Type 2 customer under the conditions mentioned above with

probability p , $0 \leq p \leq 1$. With probability $q = 1 - p$, the served Type 1 customer will leave the system. Upon completion of a service a free server will offer service to a Type 1 customer on a *FCFS* basis; however, if there are no Type 1 customers waiting, the server will serve a Type 2 customer if there is one present in the queue. If a Type 1 customer decides to serve a Type 2 customer, for our analysis purposes that Type 2 customer will be removed from the system immediately. This is due to the fact that the system no longer needs to track that Type 2 customer. Type 2 customers when served (again on a *FCFS* basis) by one c servers require exponential processing times with parameter μ_2 . They assume that services are offered in the order of the arrivals of the customers and Type 1 customers have non-preemptive priority over Type 2 customers.

Under the assumption that Type 1 customers have a limited waiting space, say, L , $1 \leq L < \infty$ and Type 2 customers have unlimited waiting area, the authors employed matrix-analytic methods to study the model in steady-state. In the following the term “ratio” of a measure refers to the ratio obtained by looking at that measure when $p > 0$ when compared to the same measure when $p = 0$. For example, the ratio of the mean waiting time, $MWTS_1$, in the system of a Type 1 customer, is obtained by looking at the $MWTS_1$ when $p > 0$ and $MWTS_1$ when $p = 0$.

Through various numerical examples, they pointed out the significant advantage, even for small p , in considering crowdsourcing by offering more traffic load (through increasing the rate of Type 2 customers without violating the stability condition) to the system resulting in more customers served (which in turn increasing the revenues when cost/profits are incorporated). The rate of increase in the offered load to the system is much higher for small values of c . This indicates that smaller businesses will probably benefit more since they normally will have less resources.

4.2 Crowdsourcing and MAP/PH/1 Queueing Model

Having motivated with the need for having more queueing models in the context of *crowdsourcing* Chakravarthy and Ozkar (2015) studied *MAP/PH/1*-type queueing models. The model assumptions are as follows.

- Arrivals occur according to a *MAP* for Type 1 arrivals (in-store customers).
- Arrivals occur according to a Poisson process for Type 2 arrivals (Online customers).
- The service scheme for both Type 1 and Type 2 customers is the same as that of the model described in Section 4.1. Note that here there is only one server in the system.
- The service times are assumed to be of phase type.

Employing matrix-analytic methods, the authors study the model in steady-state and some key highlights reported in Chakravarthy and Ozkar (2015) are as follows.

- The server busy probabilities (serving Type 1 and Type 2 customers) vary based on the type of arrival and the service processes as is to be expected; however, in all cases, they appear to be insignificant to p .
- The probability that a Type 2 customer leaves with a Type 1 customer (i.e., Type 2 is served by a Type 1 customer) is more sensitive to the service time distribution when L is small.
- Just because the busy probabilities are insignificant to p does not imply that p does not play a role in the model. In fact, they demonstrated that the fraction of Type 2 customers served by Type 1 customers is very significant. This would not have been discovered by looking only at the busy probabilities.
- Even when the system has a very low saturation level (i.e., when the traffic intensity is small) there appears to be a marked improvement in having some fraction of Type 1 customers acting as servers for Type 2 customers.
- For very small values of p , ($p < 0.2$), there appears to be no significant drop in the ratio for the mean waiting time, $MWTS_2$, in the system for Type 2 customers as L is increased; however, when $p \geq 0.2$, there is a significant rate of decrease as L increases. The rate of decrease is much higher when the service times have a higher variability like hyperexponential one.
- The role of correlation, especially positive one, present in the inter-arrival times in the system performance measures is highlighted.

4.3 Crowdsourcing and MAP/PH/1 Queueing Model with Pure and Working Vacations

Having studied *crowdsourcing* from single and multiple-server points of view by having the server (s) available to the system at all times, Chakravarthy and Ozkar (2016b) introduced vacation and working vacation in the context of *MAP/PH/1* with *crowdsourcing*. A number of papers have been studied in the literature on classical queueing models with vacations and working vacations [see e.g., Baba (2005), Servi and Finn (2002), Sreenivasan, et. al., (2013), Tian

and Zhang (2006), Tian and Li (2009), Wu and Takagi (2006), and Zang and Hou (2011).]

In Chakravarthy and Ozkar (2016b) the authors employ the N policy and working vacations similar to Sreenivasan, et.al., (2013), but in the context of *crowdsourcing*. The model assumptions are as follows.

- Arrivals occur according to a *MAP* for Type 1 arrivals (in-store customers).
- Arrivals occur according to a Poisson process for Type 2 arrivals (Online customers).
- The service scheme for both Type 1 and Type 2 customers is the same as that of the model described in Section 4.1. There is one server in the system.
- The service times are of phase type.
- When the system becomes empty at the time of a completion of a service, the server will go on a vacation. The duration of a vacation is assumed to be exponentially distributed. A vacation is interrupted only when a Type 1 customer arrives during that time. However, the server offers services to those (Type 1) customers arriving during a vacation at a lower rate. Note that only Type 1 customers can be served during a vacation mode. The service times of those customers (served at a lower rate) are also of phase type but with rate of services decreased by a (positive) factor that is less than 1. The server continues to serve at this rate until either the vacation expires or the number of Type 1 customers in the system hits a pre-determined threshold, say, N , $1 \leq N \leq L+1$. At this instant, the server instantaneously switches over to the normal rate and continues to serve at this rate until the system becomes empty. Note that once the service resumes at a normal rate (either through vacation getting over or when the number of Type 1 customers in the system hits N) the server will serve both types of customers. At the end of a vacation if there is no customer waiting for service, the server takes another vacation.

Employing matrix-analytic methods, the authors study the model in steady-state and some key highlights reported in Chakravarthy and Ozkar (2016b) are as follows.

- When the average vacation duration is large, the role played by p is apparent for the *pure* vacation model as compared to the model with no vacation and the model with working vacation.
- For Erlang type arrivals (which have less variability in the inter-arrival times that are independent of each other) and for Erlang type services, the role of the threshold N is not that significant.
- For (positively) correlated inter-arrival times, the role played by the threshold N is significant.
- Again, the authors demonstrate the role of correlation, especially, the positive ones.
- Various ratios for system performance measures such as the mean waiting time in the system of a Type 2 customer, probability of server is busy serving a Type i , $i = 1, 2$, customer, and the fraction of Type 1 customers serving Type 2 customers, are used to highlight the significance of *crowdsourcing*.

4.4 Crowdsourcing and MAP/PH/c Queueing Model

In Chakravarthy and Ozkar (2016a) the authors study the most general queueing model so far in the context of *crowdsourcing*. The model assumptions are as follows.

- Arrivals occur according to a *MAP* for Type 1 arrivals (in-store customers).
- Arrivals occur according to a Poisson process for Type 2 arrivals (Online customers).
- The service scheme for both Type 1 and Type 2 customers is the same as that of the model described in Section 4.1.
- There are c servers in the system.

Due to large state space requirement to study the model analytically, the authors employed simulation approach for studying this multi-server model. After validating the simulated model with the help of the queueing models listed in Sections 4.1 and 4.2, the authors report some interesting observations. Some specific ones are listed below.

In the sequel, *FATH*, stands for the fraction of Type 2 customers whose waiting time in the system exceeds r , $r \geq 2$, times the average service time by one of the system servers. That is, if Y denotes the waiting time in the system of a Type 2 customer and if μ denotes the rate of service, then $P(Y > r/\mu), r \geq 2$. Since there is no analytical expression available for the measure, *FATH*, dealing with a specific tail probability of the waiting time in the system of a Type 2 customer, we used the simulated result instead. It should be pointed out that one can compute algorithmically the tail probability for classical single-server model, *MAP/PH/1* using the matrix-analytic methods

(Neuts, 1981) but for a multi-server case it is highly complicated and hence the authors resort to simulation only for this particular measure.

- The ratio for $MWTS_1$ decreases as p increases but not as significantly as compared to $MWTS_2$. This is somewhat surprising since one would expect the ratio to be close to 1 since Type 1 service is not affected by the value of p due to non-preemptive nature of services. However, as p increases, Type 1 customers have a higher probability of serving Type 2 customers resulting in a relatively fewer Type 2 customers to be served by one of the system servers and hence a reduction in the mean waiting time in the system.
- The ratio for $FATH$ is decreasing at a significantly at a higher rate as p increases (for all scenarios) which is again very important from both management as well as customers' points of view and can also be used by the management to provide some kind of a guarantee on the service times of Type 2 customers.

5. Concluding Remarks

While there is a vast literature on Crowdsourcing in many areas such as healthcare, computer science, environmental sciences, business, finance, and marketing, only recently stochastic models, more specifically queueing models, useful in crowdsourcing in the context of service sectors have been studied. In this expository article dealing with *crowdsourcing* and stochastic modeling, we presented various queueing models that will be of help to businesses, especially in service sectors. There are several interesting models that are available for future study. For example, one can incorporate production aspects to the demands (of services) which will require having various inventories to offer services.

Acknowledgements

Serife Ozkar would like to acknowledge the support of her host institution in Turkey to visit Kettering University, and also the hospitality provided to her by the Industrial and Manufacturing Engineering Department and Kettering University during her stay in Flint, Michigan, USA.

References

- Artalejo, J. R., Gomez-Correl, A., & He, Q. M. (2010). Markovian arrivals in stochastic modelling: a survey and some new results. *SORT*, 34(2), 101-144.
- AyferBozat, Z., & Erenel, F. (2016). Open innovation in human resources and the use of crowdsourcing. *Modern Management Science & Engineering*, 4, 43-52.
- Baba, Y. (2005). Analysis of a GI/M/1 queue with multiple working vacations. *Operations Research Letters*, 33, 201-209. <http://dx.doi.org/10.1016/j.orl.2004.05.006>
- Belleflamme, P., Lambert, T., & Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. *Journal of Business Venturing*, 29(5), 585-609. <http://dx.doi.org/10.1016/j.jbusvent.2013.07.003>
- Bernstein, M. S., Karger, D. R., Miller, R. C., & Brandt, J. (2012). Analytic methods for optimizing realtime crowdsourcing, Proceeding of Collective Intelligence Conference held at MIT, April 18-20, 2012.
- Bhat, U. N. (2015). *An introduction to queueing theory* (2nd ed.). New York, NY: Springer. <http://dx.doi.org/10.1007/978-0-8176-8421-1>
- Bhattacharya, S., Celis, L. E., Chander, D., Dasgupta, K., Karanam, S., & Rajan, V. (2015). Crowds of crowds: Performance based modeling and optimization over multiple crowdsourcing platforms. *Human Computation*. 2(1), 105-131. <http://dx.doi.org/10.15346/hc.v2i1.6>
- Chakravarthy, S. R. (2001). The Batch Markovian Arrival Process: A Review and Future Work. *Advances in Probability Theory and Stochastic Processes*, Eds., A. Krishnamoorthy, et al., 21-49, Notable Publications, Inc., New Jersey.
- Chakravarthy, S. R. (2010). Markovian arrival processes. *Wiley Encyclopedia of Operations Research and Management Science*. Published Online: 15 JUN 2010.
- Chakravarthy, S. R., & Dudin, A. N. (2015). A queueing model for crowdsourcing. Submitted for Publication.
- Chakravarthy, S. R., & Ozkar, S. (2015). Analysis of a stochastic model for crowdsourcing using MAP arrivals and phase type services. Submitted for Publication.
- Chakravarthy, S. R., & Ozkar, S. (2016a). Simulation of a queueing model useful in crowdsourcing. Submitted for Publication.

- Chakravarthy, S. R., & Ozkar, S. (2016b). MAP/PH/1 queueing model with working vacations and crowdsourcing. Submitted for Publication.
- Djelassi, S., & Decoopman, I. (2013). Customers' participation in product development through crowdsourcing: Issues and implications. *Industrial Marketing Management*, 42, 683–692. <http://dx.doi.org/10.1016/j.indmarman.2013.05.006>
- Evans, R. D., Gao, J. X., Mahdikhah, S. Messaadia, M., & Baudry, D. (2016). A review of crowdsourcing literature related to the manufacturing industry. *Journal of Advanced Management Science*, 4(3), 224-231. <http://dx.doi.org/10.12720/joams.4.3>
- Faridani, S., Hartmann, B., & Ipeiros, P. G. (2011). What's the right price? Pricing tasks for finishing on time. Human Computation: Papers from the Association for the Advancement of Artificial Intelligence (AAAI)
- Forbes (2013). "Forbes". Accessed March 26, 2016. <http://www.forbes.com/sites/erikamorph/2013/03/28/about-walmarts-idea-to-crowdsource-its-same-day-delivery-service/#5b25486f1ff1>.
- Hasteer, N., Bansal, A., & Murthy, B.K. (2015). Crowdsourced software development process: Investigation and modeling through Markov decision theory. *International Journal of Software Engineering and Its Applications*, 9(9), 41-54. <http://dx.doi.org/10.14257/ijseia.2015.9.9.05>
- Hosseini, M., Shahri, A., Phalp, K., Taylor, J. & Ali, R. (2015). Crowdsourcing: A taxonomy and systematic mapping study. *Computer Science Review*, 17, 43–69. <http://dx.doi.org/10.1016/j.cosrev.2015.05.001>
- Howe, J. (2006). "Crowdsourcing: A definition". Accessed April 26, 2016. <http://www.crowdsourcing.com/cs/>
- Jayanti, E. (2012). Open sourced organizational learning: Implications and challenges of crowdsourcing for human resource development (HRD) practitioners. *Human Resource Development International*, 15(3), 375-384. <http://dx.doi.org/10.1080/13678868.2012.669235>
- Lucantoni, D. M., Meier-Hellstern, K. S. & Neuts, M. F. (1990). A single-server queue with server vacations and a class of nonrenewal arrival processes. *Advances in Applied Probability*, 22, 676–705. <http://dx.doi.org/10.2307/1427464>
- Lucantoni, D. (1991). New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, 7, 1-46. <http://dx.doi.org/10.1080/15326349108807174>
- Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of business venturing*, 29(1), 1-16. <http://dx.doi.org/10.1016/j.jbusvent.2013.06.005>
- Neuts, M. F. (1979). A versatile Markovian point process. *Journal of Applied Probability*, 16, 764-779. <http://dx.doi.org/10.2307/3213143>
- Neuts, M. F. (1981). *Matrix geometric solutions in stochastic models -An algorithmic approach*. Dover Publications, 1995 (originally published by Johns Hopkins University Press, 1981).
- Neuts, M. F. (1989). *Structured Stochastic Matrices of M/G/1 type and their applications*. New York, NY: Marcel Dekker.
- Neuts, M. F. (1995). *Algorithmic probability: A collection of problems*. New York, NY: Chapman and Hall.
- Quora (2015). <http://www.quora.com/What-are-the-best-examples-of-crowdsourcing> (accessed on April 28, 2016).
- Ramaswami, V. (1980). The N/G/1 queue and its detailed analysis. *Advances in Applied Probability*, 12, 222-261. <http://dx.doi.org/10.2307/1426503>
- Riasi, A. (2015). Competitive advantages of shadow banking industry: An analysis using porter diamond model. *Business Management and Strategy*, 6(2), 15-27. <http://dx.doi.org/10.5296/bms.v6i2.8334>
- Ross, S. M. (2014). *Introduction to probability models* (11th ed.). California, USA: Academic Press.
- Schwartz, C., Borchert, K., Hirth, M., & Tran-Gia, P. (2015). Modeling crowdsourcing platforms to enable workforce dimensioning. International Telecommunication Networks and Applications Conference (ITNAC), 30-37. <http://dx.doi.org/10.1109/ATNAC.2015.7366785>
- Servi, L. & Finn, S. (2002). M/M/1 queue with working vacations (M/M/1/WV). *Performance Evaluation*, 50, 41-52. [http://dx.doi.org/10.1016/S0166-5316\(02\)00057-3](http://dx.doi.org/10.1016/S0166-5316(02)00057-3)
- Sreenivasan, C., Chakravarthy, S. R., & Krishnamoorthy, A. (2013). MAP/PH/1 queue with working vacations,

- vacation interruptions and N Policy. *Applied Mathematical Modelling*, 37, 3879-3893. <http://dx.doi.org/10.1016/j.apm.2012.07.054>
- Tian, N. S & Zhang, Z. G. (2006). *Vacation queueing models: Theory and applications*. New York, NY: Springer Publishers.
- Tian, N. S. & Li, J. H. (2009). Matrix analytic method and working vacation queues - A survey. *International Journal of Information and Management Sciences*, 20, 603-633.
- Yuen, M. C, King, I., & Leung, K. S. (2011). A Survey of Crowdsourcing Systems. IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing. <http://dx.doi.org/10.1109/passat/socialcom.2011.203>
- Wang, J., Faridani, S., & Ipeiritos, P. G. (2011). Estimating the completion time of crowdsourced tasks using survival analysis models. Workshop on Crowdsourcing for Search and Data Mining, Hong Kong, China, Feb. 9.
- Wu, D., Rosen, D. W., & Schaefer, D. (2014). Modelling and Analyzing the Material Flow of Crowdsourcing Processes in Cloud-Based Manufacturing Systems Using Stochastic Petri Nets. The ASME 2014 International Manufacturing Science and Engineering Conference, June 9-13, Ann Arbor, Michigan, USA.
- Wu, D., & Takagi, H. (2006). M/G/1 queue with multiple working vacations. *Performance Evaluation*, 63, 654-681. <http://dx.doi.org/10.1016/j.peva.2005.05.005>
- Zang, M., & Z. Hou. (2011). Performance analysis of MAP/G/1 queue with working vacations and vacation interruption. *Applied Mathematical Modelling*, 35, 1551-1560. <http://dx.doi.org/10.1016/j.apm.2010.09.031>