7^e *Conférence Internationale de MOdélisation et SIMulation - MOSIM'08 - du* 31 *mars au* 2 *avril* 2008 *- Paris - France*
«Modélisation, Optimisation et Simulation des Systèmes : Communication, Coopération et Coordination»

WEARINESS AND LOYALTY LOSS IN RECURRENT SERVICE MODELS

Olivier GALLAY, Max-Olivier HONGLER

Ecole Polytechnique Fédérale de Lausanne (EPFL)
STI-IPR-LPM
Station 17
CH-1015 Lausanne
olivier.gallay@epfl.ch, max.hongler@epfl.ch

ABSTRACT : *For recurrent service providers (fast-food, entertainment, medical care,...), retaining loyal customers is obviously a key issue. The customers' loyalty essentially depends on their service satisfaction defined via an ad-hoc utility function. Among several criteria, the utility function strongly depends on the past perceived waiting time. Moreover, the patience that customers consent to allow in waiting does often decrease as a function of the successive utilizations of the service (i.e. weariness). We propose here an idealized queueing model in which the customers' loyalty is determined only by the individual experience gained during the successive visits to a service (i.e. the waiting time and the number of services yet received). For regimes where the law of large numbers holds, a deterministic approach enables to analytically discuss the resulting multi-agent dynamics governing the customers' flows. One is able, in particular, to fully calculate, analytically, the characteristics of the emerging complex patterns (i.e. here structured temporal oscillations) which are observed to be strongly structurally stable.*

KEYWORDS : *Recurrent Services, Leisure and Hospitality, Loyalty Loss, History-Based Routing, Queueing Networks.*

1. INTRODUCTION

Whatever the type of service, waiting in queues before being attended reduces the utility perceived by the customers. As time is valued for both the server and the customers, the complex relations between the waiting times and the consumers' satisfaction is a central topic in marketing, (Bielen and Demoulin, 2007) and the references therein. Accordingly, when for a given task, competing facilities are available, a server's ability to reduce the actual or perceived waiting time of incoming customers increases its attractiveness and may drastically modifies the market sharing proportions, an example with two servers is formalized in (Gallay and Hongler, 2007). Lowering the service duration, to enhance customers' satisfaction, does require investments which have to be counterbalanced by an extra inflow of new customers. A synthetic and quantitative formulation of such heuristic observations requires a set of mathematical tools available from a "fusion" between game and queueing theoretical approaches, a general point of view adopted by R. Hassin and M. Haviv in (Hassin and Haviv, 2003). One of the fundamental lessons taught by game theory is that the distinction between games played once only with those played repeatedly is mandatory, as it leads to drastically different optimal strategies (Aumann, 1987). Similarly, for competition between queueing nodes, the fact that customers pay a single or several successive visits to the servers does strongly influence the resulting traffic flows. In (Hassin and Haviv, 2003), customers always pay a single visit to the servers and the resulting stationary equilibria (i.e. the *Nash* equilibria) are thoroughly discussed. In the present paper, we will focus on the new dynamical features emerging from repeated visits to a server. Our modeling approach has been stimulated by several recent (yet mostly experimental) studies of systems where recurrent service requirements occur. Among them, actual situations ranging from medical care (Bielen and Demoulin, 2007) to leisure and hospitality facilities such as fast-food restaurants (Law et al., 2004), ski resorts (Pullman and Thompson, 2002) and attraction parks (Kawamura et al., 2004a), (Kawamura et al., 2004b), (Kataoka et al., 2005) are some well-known illustrations. Often in recurrent service systems, the cost of retaining an existing customer is comparatively less than the cost of acquiring a new one and hence the customers' loyalty is a central issue in optimizing gains.

Along the lines paved by these applications, let us consider, as an illustration, an attraction park in which, among other attractions, a roller-coaster is offered. This roller-coaster entertains people at a limited flow rate which, due to the high demand, is responsible for the formation of a queueing process. We here assume that the park entrance fare offers to visitors the possibility to attend any attraction repeatedly and without limitation. Due to the exciting sensations generated by the roller-coaster, customers agree to line-up and are fairly patient when attending the coaster for the first time. Repeated runs however does wear their patience. The trade-off between the excitement delivered by a roller-coaster trip and the waiting burden incurred before boarding, can be quantified by a (usually individual) utility function. When the utility is negative, the customers are deterred and leave the roller-coaster for another spot. The previous roller-coaster example belongs to the highly profitable leisure and hospitality (L&H) sector, which includes the entertainment and recreation, the tourism and accommodation, and the food services. Far from being exceptional, note that the previous customers' behavior is in fact quite common in the L&H sector. Ski traffic management offers another world-wide illustration. Generically, people will change slope either when they have suffered a large waiting-time at the ski lift and/or when they are bored of having done the same ski run several times. Despite to the apparent simplicity of the above roller-coaster illustration, the resulting queueing dynamics is highly complex. It does indeed simultaneously depend on a routing feedback loop, i.e. an intrinsic **non-linearity** due to the customers lining for a new trip and on **history-based** (HB) routing decisions. The decision to come back (i.e. to remain **loyal**) or to quit is taken according to a patience threshold, itself depending on the number of previous runs already achieved. The presence of HB mechanisms (i.e. mechanisms in which memory enters) gives a non-Markovian character to the dynamics. Beside that, the individual routing decisions confer to this queueing mechanism the basic feature characterizing multi-agent systems (Bonabeau, 2002). Indeed, the server is visited by a population of **autonomous decision-making agents individually assessing their situation and making decision on the basis of a HB rule**. At first sight, little hope is left regarding the possibility to characterize analytically the traffic flows resulting from this complex dynamics. Keeping the central features, namely **non-linearity and HB decision making**, we are nevertheless able, for a somehow simplified class of models, to describe analytically the resulting dynamics. Basically, two simplifications of the original situation are introduced. On one hand, we limit to agents having a common utility function. On the other hand, we decouple the role played by the waiting-time and the number of repeated visits to the server by introducing two separate utility thresholds. When exceeded, these thresholds trigger the loss of the agents' loyalty. This simplified multi-agent dynamics generates an **emergence of generically stable time-dependent periodic queue contents**.

Our paper is organized as follows. In section 2, we start by introducing the formal model and we present experimental results obtained by simulation. Section 3 is then devoted to the analytical approach which produces a complete understanding of the experimental results.

2. SIMPLIFIED MODEL OF A RECURRENT SERVICE

We model the customers' behavior faced to a recurrent service with a single queueing network (QN) composed, as illustrated in figure 1, of a single server and a feedback queue. An incoming flow of customers, described by a renewal process with mean inter-arrival time $\frac{1}{\lambda}$ and probability distribution $A(x)$ with density $dA(x)$, is served by a processing unit which service times are i.i.d. random variables with mean $\frac{1}{\mu}$, probability distribution $B(x)$ and density $dB(x)$. Accordingly, the parameters λ and μ are respectively the incoming and service rates of the renewal processes. We assume that the distributions $A(x)$ and $B(x)$ have finite moments. Here, we suppose the traffic intensity $\rho = \frac{\lambda}{\mu} < 1 \Leftrightarrow \lambda < \mu$, which ensures the stability of the queueing system when there is no feedback loop. Assume also that the waiting room capacity is unlimited and that the service discipline is first-in-first-out (FIFO). After being served at the decision node n, each customer has to choose among two possibilities, namely:

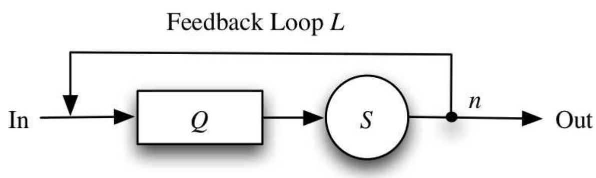

Figure 1: A single stage queueing system with feedback loop.

i) either to follow the feedback loop and line up again for being served once more *ii)* or to quit the system definitively.

Several contributions (Takàcs, 1963), (D'Avignon and Disney, 1976) and (Peköz, 2002) consider the situation arising when the decision between the choices *i)* and *ii)* is taken randomly. When this is the case, by imposing a stationary flow balance (i.e. incoming equals outgoing flow), we drive the system into a self-consistant *stationary* regime. As we will now see, such purely stationary flows strongly differ from the queue dynamics that can be observed when "intelligent" agents, able to take HB routing decisions, circulate in the network.

Our simplified model of recurrent services assumes that the customers' loyalty is based on their individual experience with the service provider. More particularly, the decision of a customer, at n, either to come back for another service or to leave the system depends on

1) the last sojourn time W it has spent in the system in order to be served (i.e. W is the sum of the queueing and the processing times)

and on

2) the number of services N_{it} it has already received (we say that the customer is at its i^{th} iteration).

We suppose furthermore that the influence played by these two measures is decoupled. The customers, who share a common utility function (i.e. we consider homogenous customers), consider two separate thresholds. They possess first a common patience parameter P to which they will compare their last experienced sojourn time W. Secondly, they will check that they have received less services than a common weariness parameter $N_{\max}$. It leads to the introduction of two independent rules $\mathcal{R}_1$ and $\mathcal{R}_2$, that the customers will apply when they decide their routing at decision node n. The first one is controlled by the sojourn time and is given by:

$$\mathcal{R}_1 = \begin{cases} \text{follow alternative } i) & \text{if } \; W \leq P, \\ \text{follow alternative } ii) & \text{if } \; W > P. \end{cases}$$

The second rule is driven by the number of already received services, it is defined as:

$$\mathcal{R}_2 = \begin{cases} \text{follow alternative } i) & \text{if } \; N_{\text{it}} < N_{\max}, \\ \text{follow alternative } ii) & \text{if } \; N_{\text{it}} \geq N_{\max}. \end{cases}$$

Combining these two independent rules, the customers will hence choose their routing at n following:

$$\begin{aligned} \mathcal{R} &= \mathcal{R}_1 \cap \mathcal{R}_2 \\ &= \begin{cases} \text{follow alternative } i) & \text{if } \; W \leq P \text{ and } \\ & \qquad N_{\text{it}} < N_{\max}, \\ \text{follow alternative } ii) & \text{otherwise.} \end{cases} \end{aligned}$$

When alternative *i)* is chosen, we speak of **loyal customers**, as they are pleased with the server and then return to it for another service. The rule $\mathcal{R}$ states that, providing its sojourn time remains below its patience parameter and providing it has already received less than a limit number of services, a customer comes back for another service.

The dynamics involving $\mathcal{R}_1$ alone is fully discussed in (Filliger and Hongler, 2005). In this case and when P is large enough, quasi-deterministic cyclo-stationary regimes emerge, i.e. *stable temporal oscillations* of the queue level $Q(t)$ are observed and this independently of the detailed nature of the probability laws $A(x)$ and $B(x)$.

Note that our simplified model assumes that P_i, the customers' common patience parameter when they receive their i^{th} service, has the following form:

$$P_i = \begin{cases} P & \text{if } \; i \leq N_{\max}, \\ 0 & \text{if } \; i > N_{\max}. \end{cases}$$

The next step to get closer to real-life customers behavior will be to consider more general forms for P_i. Typically, this patience parameter could be monotonically decreasing in i, denoting that the customers' loyalty suffers from a progressive weariness over time. Likewise, $N_{\max}$ is here common to all customers. A natural generalization would be to consider that each customer possesses its own behavior when faced to weariness.

2.1. Experimental Observations

Figures 2 and 3 show the typical dynamics of the

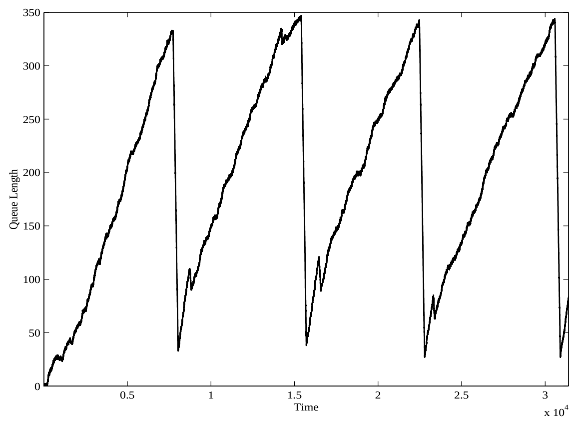

Figure 2: Queue length dynamics when inter-arrival times are uniformly distributed in $[1, 17]$ with mean $\frac{1}{\lambda} = 9$ and CV $= 0.51$, service times are uniformly distributed in $[0.1, 1.7]$ with mean $\frac{1}{\lambda} = 0.9$ and CV $= 0.51$ ($\rho = 0.1$), $P = 300$ and $N_{\max} = 12$.

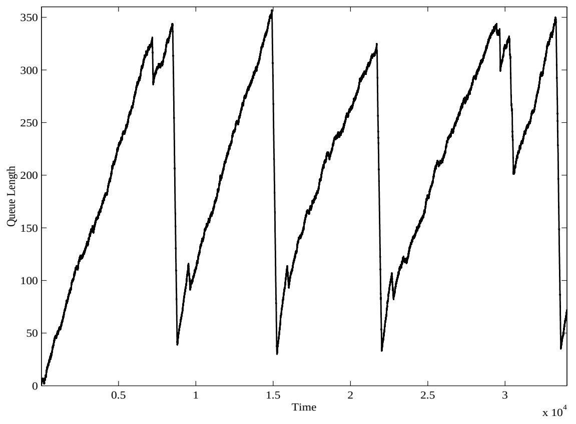

Figure 3: Queue length dynamics when inter-arrival times are Erlang(3) with mean $\frac{1}{\lambda} = 9$ and CV $= 0.57$, service times are Erlang(3) with mean $\frac{1}{\mu} = 0.9$ CV $= 0.57$ ($\rho = 0.1$), $P = 300$ and $N_{\max} = 12$.

queue length when customers follow the HB rule $\mathcal{R}$ to choose their routing at n.

Independently of the inter-arrival and service times distributions and when P is large enough (see (Filliger and Hongler, 2005) for a detailed discussion on the parameter P), we observe the emergence of quasi-deterministic cyclo-stationary regimes, i.e. stable temporal oscillations of the queue content. Indeed, despite to the presence of strong fluctuations, this robust and quasi-deterministic behavior is a direct consequence of the law of large numbers (LLN). The smoothing effect due to the underlying LLN is manifestly observable in figures 2 and 3, a further discussion on this aspect can be found in (Filliger and Hongler, 2005). According to the above considerations, for sufficiently large P, the dynamics can be approximatively discussed via a deterministic approach. The dynamics resulting from deterministic inter-arrival and service times is illustrated in figure 4. It is furthermore remarkable that the oscillations exhibit an extra peak during their increasing phase. This peak is entirely due to rule $\mathcal{R}_2$.

A restricted range of the values of the control parameters (i.e. λ, μ, P and $N_{\max}$) produces the fully complex dynamics visible in figures 2, 3 and 4. Indeed, when $N_{\max}$ is large, the customers remain in the system for a long time before being wearied. Hence, the queue length increases (new customers arrivals) and eventually reaches a level implying sojourn times larger than P. All the customers hence leave the system due to routing rule $\mathcal{R}_1$ (i.e. waiting-time) and none following $\mathcal{R}_2$ (i.e. maximum number of received services). The resulting dynamics is illustrated in figures 5 and 6 (random and de-

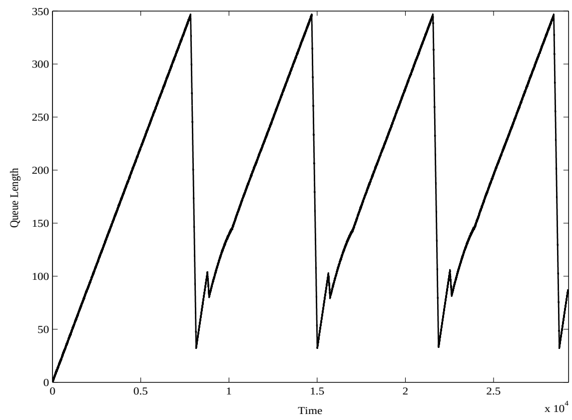

Figure 4: Queue length dynamics for deterministic inter-arrival times $\frac{1}{\lambda} = 9$, deterministic service times $\frac{1}{\mu} = 0.9$ ($\rho = 0.1$), $P = 300$ and $N_{\max} = 12$.

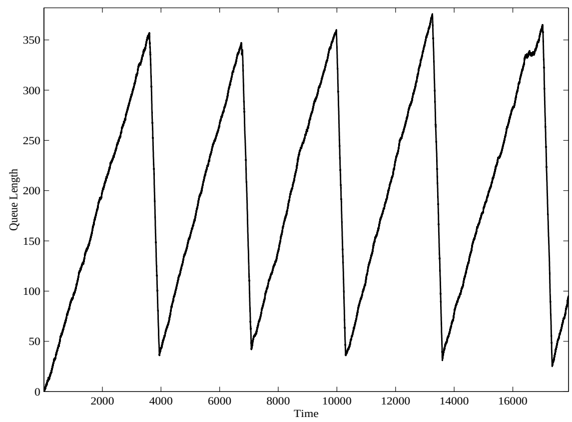

Figure 5: Queue length dynamics when inter-arrival times are Erlang(3) with $\frac{1}{\lambda} = 9$, service times are Erlang(3) with $\frac{1}{\mu} = 0.9$ ($\rho = 0.1$), $P = 300$ and $N_{\max} = 25$.

terministic cases). As we shall see later, the regime where no extra peak appears, emerges when:

$$\rho\,(1+\rho)^{N_{\max}-1} \geq 1. \tag{1}$$

Whenever condition (1) holds, the model is similar to the one where only rule $\mathcal{R}_1$ is implemented (Filliger and Hongler, 2005).

We have focused here on homogenous agents behavior. However, note that the emergence of a global behavior does also arise for agents with individual patience parameter, see (Filliger and Hongler, 2005).

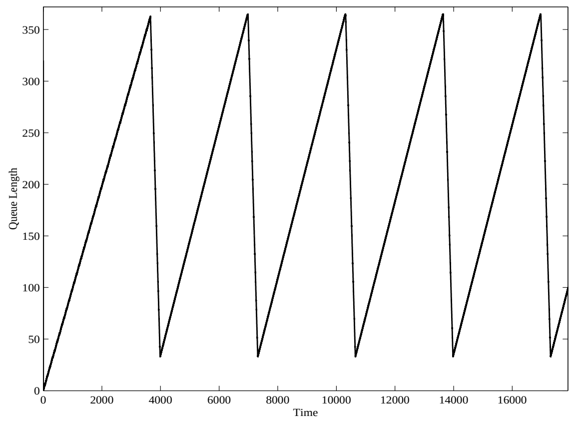

Figure 6: Queue length dynamics for deterministic inter-arrival times $\frac{1}{\lambda} = 9$, deterministic service times $\frac{1}{\mu} = 0.9$ ($\rho = 0.1$), $P = 300$ and $N_{\max} = 25$.

3. ANALYTICAL DISCUSSION

As illustrated in section 2.1, experimental results show that the dynamics exhibit stable temporal oscillations of the queue content even in presence of strong fluctuations. Hence from now on, we focus on deterministic dynamics. We decompose the oscillatory dynamics in five distinct phases. Figure 7 gives an illustration of these five phases on a generic oscillation of the queue length.

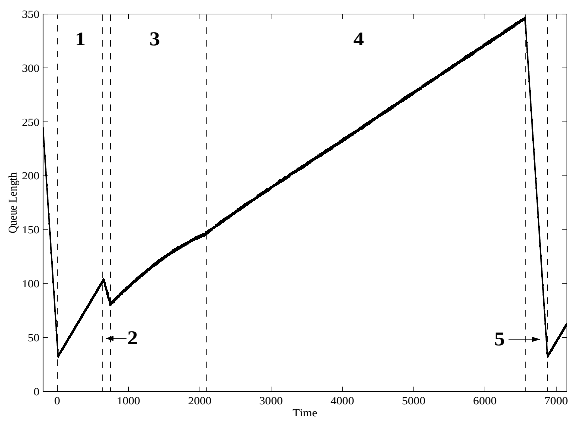

Figure 7: A generic oscillation of the queue content dynamics for deterministic inter-arrival times $\frac{1}{\lambda} = 9$, deterministic service times $\frac{1}{\mu} = 0.9$ ($\rho = \frac{\lambda}{\mu} = 0.1$), $P = 300$ and $N_{\max} = 12$.

- **First Phase: Pure Feeding**

As we will see in the fifth phase below, the queue is initially populated with an offset of $P\lambda$ fresh customers, who haven't received any service yet. During the first phase, which begins without loss of generality at time $t = 0$, the queue length remains small enough so that the customers wait less than their patience parameter P (and hence rule $\mathcal{R}_1$ is satisfied). Furthermore, since there is initially only fresh customers, no customer leaves the system due to the maximum number $N_{\max}$ of iterations. The queue length hence increases at rate λ (arrival rate of new customers in the system). This first phase ends when the original $P\lambda$ customers have received $N_{\max}$ services and thus will start to leave the system. To compute the resulting time T_1, we focus on $Q(k)$, the number of customers in the queue when all the $P\lambda$ original customers have completed k iterations, $k = 0, ..., N_{\max}$. We first have that:

$$Q(1) = P\lambda \frac{1}{\mu} \lambda + P\lambda = P\lambda\,(1 + \rho)\,.$$

Indeed, the necessary time to serve the $P\lambda$ original customers is equal to $T(0) = P\lambda\frac{1}{\mu}$. During that time, $T(0)\lambda$ fresh customers join the queue, which is still populated by the $P\lambda$ original customers, who have now received one service. Following this iterative reasoning, we find that:

$$Q(k) = P\lambda\,(1 + \rho)^k\,, \quad k = 0, ..., N_{\max}.$$

Accordingly, we find that:

$$T_1 = \frac{1}{\mu} \sum_{k=0}^{N_{\max}-1} Q(k) = P\left[(1 + \rho)^{N_{\max}} - 1\right].$$

□

- **Second Phase: Offset Purging**

During the second phase, beginning at T_1, a proportion of the customers receiving service leaves the system because they have been provided the maximum number of services (i.e. wearied customers). At the end of the first phase, $Q(N_{\max})$ customers populate the queue. Among them, $P\lambda$ have already done $N_{\max}$ iterations (the original offset of customers). Accordingly, they will leave the system after the next service. We define thus the effective offset purging rate

$$\mu_{\text{eff}}(N_{\max}) = \mu \frac{P\lambda}{Q(N_{\max})} = \frac{\mu}{(1 + \rho)^{N_{\max}}}$$

as the rate at which the original customers leave the system due to the maximum number of iterations. In the second phase, the slope of the queue content is $s = \lambda - \mu_{\text{eff}}(N_{\max})$. Note that, as soon as $s \geq 0$, the extra peak disappears. This phase ends when the $Q(N_{\max})$ initial customers have been served. Its duration is given by:

$$T_2 = \frac{Q(N_{\max})}{\mu}.$$

□

- **Third Phase: First Residual Offset Purging**

Again, in the third phase, weared customers leave the system. In this phase, the presence of the first $P\lambda$ original customers still influences the dynamics. At the beginning of phase 3, there are

$$Q(N_{\max}+1) = \overbrace{Q(N_{\max})\frac{1}{\mu}\lambda}^{\text{freshly arrived customers}} + \underbrace{Q(N_{\max}) - \underbrace{P\lambda}_{\text{leaving customers}}}_{\text{loyal customers}}$$

populating the queue. After the passage of these $Q(N_{\max}+1)$ initial customers, there is

$$Q(N_{\max}+2) = \overbrace{Q(N_{\max}+1)\frac{1}{\mu}\lambda}^{\text{freshly arrived customers}} + \underbrace{Q(N_{\max}+1) - \underbrace{P\lambda\frac{1}{\mu}\lambda}_{\text{leaving customers}}}_{\text{loyal customers}}$$

customers in the queue. Iteratively, we find that:

$$\begin{aligned} Q(N_{\max}+k) &= P\lambda\left[(1+\rho)^{N_{\max}+k}\right. \\ &\quad \left. -(1+\rho)^{k-2}(1+k\rho)\right], \\ &\quad 1 \le k \le N_{\max}+2, \end{aligned}$$

where $Q(N_{\max}+k)$ is the queue content after all the $Q(N_{\max}+k-1)$ customers previously in the queue have received service, $1 \le k \le N_{\max}+2$. The resulting effective purging rate (here piecewise linear) reads as:

$$\begin{aligned} \mu_{\text{eff}}(N_{\max}+k) &= \mu\frac{\rho(1+\rho)}{(1+\rho)^{N_{\max}+2}-(1+k\rho)}, \\ &\quad 1 \le k \le N_{\max}+1, \end{aligned}$$

with $\lambda - \mu_{\text{eff}}(N_{\max}+k)$ is the rate at which the queue raises between levels $Q(N_{\max}+k)$ and $Q(N_{\max}+k+1)$. The length of each iteration, during which the queue content increases from $Q(N_{\max}+k)$ to $Q(N_{\max}+k+1)$, is given by:

$$T(N_{\max}+k) = \frac{Q(N_{\max}+k)}{\mu}, \quad 1 \le k \le N_{\max}+1.$$

Hence, the total duration of the third phase is equals to:

$$\begin{aligned} T_3 &= \sum_{k=1}^{N_{\max}+1} T(N_{\max}+k) \\ &= P\left\{(1+\rho)^{2N_{\max}+2}\right. \\ &\quad \left. -\left[1+\rho(N_{\max}+2)\right](1+\rho)^{N_{\max}}\right\}. \end{aligned}$$

□

- **Fourth Phase: Following Residual Offset Purging, Constant Growth**

The influence played by the $P\lambda$ original customers is further reduced and becomes, in a first order approximation, negligible. This leads to a quasi-constant queue content increase rate $\lambda - \mu_{\text{CG}}$. Note that, if required, higher order approximations are analytically available. Accordingly, the rate of customers having reached the maximum number of iterations and thus leaving the system is quasi-constant over time. The purging rate $\mu_{\text{eff}}(2N_{\max}+2)$ (the exact value of the rate at the beginning of this phase) yields a good approximation for μ_{CG}. It is given by:

$$\mu_{\text{eff}}(2N_{\max}+2) = \frac{\mu\rho\left[(1+\rho)^{N_{\max}+1}-1\right]}{(1+\rho)^{2N_{\max}+2}-(1+\rho)^{N_{\max}}\left[1+(N_{\max}+2)\rho\right]}.$$

Table 1 gives the accuracy of this approximation for several values of the control parameters.

External Parameters	$\mu_{\text{eff}}(2N_{\max}+2)$	μ_{CG}	Error
$\frac{1}{\lambda}=9$, $\frac{1}{\mu}=0.9$, $P=300$, $N_{\max}=12$	0.0632	0.0656	4%
$\frac{1}{\lambda}=9$, $\frac{1}{\mu}=0.9$, $P=300$, $N_{\max}=17$	0.0312	0.0297	5%
$\frac{1}{\lambda}=9$, $\frac{1}{\mu}=0.8$, $P=300$, $N_{\max}=12$	0.0788	0.0836	6%
$\frac{1}{\lambda}=9$, $\frac{1}{\mu}=1.1$, $P=300$, $N_{\max}=12$	0.0419	0.0425	1.5%

Table 1: Comparison between the approximation and the experimental value of the quasi-constant purging rate of the fourth phase.

The queue length approximatively raises at rate $\lambda - \mu_{\text{eff}}(N_{\max}+2)$ until it reaches a P-dependent siphoning threshold. This phase ends when the queue length reaches the level, see (Filliger and Hongler, 2005),

$$Q(2N_{\max}+3) = \overbrace{P\mu}^{\text{critical level}} + \underbrace{P\left(\lambda - \mu_{\text{eff}}(2N_{\max}+2)\right)}_{\text{delay mechanism}}.$$

Its duration is hence given by:

$$T_4 = \frac{Q(2N_{\max}+3) - Q(2N_{\max}+2)}{\lambda - \mu_{\text{eff}}(2N_{\max}+2)}.$$

□

- **Fifth Phase: Siphon Mechanism**

  At time

  $$\tau = T_1 + T_2 + T_3 + T_4 - P\left(\lambda - \mu_{\text{eff}}(2N_{\max}+2)\right),$$

  the queue length is large enough ($= P\mu$) to get sojourn times W larger than P. As a consequence, at time $\tau + P$, **a siphon purging** with rate $\mu - \lambda$ is triggered. When the queue content exceeds a critical level, it autonomously releases its emptying. This behavior is fully analogous to the hydrodynamic self-siphoning device discussed in (Pikovsky et al., 2001), When the siphon purging happens before there are wearied customers leaving the system, which happens whenever condition (1) holds, regimes where only phases 1 and 5 are visible emerge (see figures 5 and 6). Note that, in phase 5, all customers leave the system due to rule $\mathcal{R}_1$. At the end of the fifth phase, there remain $P\lambda$ customers in the queue (Filliger and Hongler, 2005). All these $P\lambda$ customers are new incomers, who have never been served yet. The duration of this last phase is given by:

  $$T_5 = \frac{Q(2N_{\max}+3) - P\lambda}{\lambda - \mu} = \frac{P\left[\mu - \mu_{\text{eff}}(2N_{\max}+2)\right]}{\lambda - \mu}.$$

  □

Resuming the situation and grouping the above informations, it is possible to compute the period

$$\Pi = T_1 + T_2 + T_3 + T_4 + T_5$$

and the amplitude

$$\Delta = Q(2N_{\max}+3) - P\lambda = P\left[\mu - \mu_{\text{eff}}(2N_{\max}+2)\right]$$

of the stable temporal oscillations that the considered dynamics exhibit.

Table 2 gives a comparison, for the numerical values used in figure 7, between the analytical results given by the formulas derived in this section and the simulation results.

Analytical Results	Matching With Simulation
$T_1 = 641.46$	below 1% error
$Q(N_{\max}) = 104.61$	below 1% error
$\lambda - \mu_{\text{eff}}(N_{\max}) = -0.2429$	below 1% error
$T_2 = 94.15$	below 1% error
$Q(N_{\max}+1) = 81.73$	below 1% error
$\lambda - \mu_{\text{eff}}(N_{\max}+1) = 0.0658$	below 1% error
$\lambda - \mu_{\text{eff}}(2N_{\max}+1) = 0.0295$	below 1% error
$Q(2N_{\max}+2) = 146.18$	below 1% error
$T_3 = 1315.81$	below 1% error
$\lambda - \mu_{\text{eff}}(2N_{\max}+2) = 0.0479$	below 5% error
$Q(2N_{\max}+3) = 347.70$	below 1% error
$T_4 = 4207.09$	below 3% error
$T_5 = 314.37$	below 1% error
$\Pi = 6572.88$	below 3% error
$\Delta = 314.37$	below 1% error

Table 2: Comparison between analytical and simulation results for deterministic inter-arrival times $\frac{1}{\lambda} = 9$, deterministic service times $\frac{1}{\mu} = 0.9$ ($\rho = 0.1$), $P = 300$ and $N_{\max} = 12$..

4. PERSPECTIVES AND CONCLUSION

The present model is, by many aspects, oversimplified. In particular, to assume that all agents share a common patience threshold is obviously a pale reflect of reality. Everybody has its own perception of waiting time, which will directly affect the associated utility function. To further approach real situations and therefore to confer a more direct practical impact to our present modeling framework, it will now be required to actually characterize the underlying utility functions, a task which will obviously strongly depend on the particular situations to be investigated. Nevertheless, our model has so far the merit to allow for an analytical approach to a noticeably complex dynamics. It shows, once more, the structural emergence of macroscopic temporal patterns resulting from elementary, though non-linear, individual interactions. Similarly to ants which act according to the concentration of pheromones, here our agents decide in view of their waiting times and this feature confers to our dynamics its *stigmergic* self-organizing character.

To close, let us emphasize that the very strong structural stability (i.e. *the high insensitivity to external noise sources*) of the oscillations reported in figures 2, 3 and 4 does definitely increase the modeling power offered by this class of multi-agent non-linear dynamics. Models that enjoy strong structural stability evolution are the cornerstones of a *synergetic* approach which, with a limited number of salient relevant features, are able to encompass under a common modeling framework a wide range of transdisciplinary situations.

ACKNOWLEDGEMENTS

This work is partially supported by the FNS (Fonds National Suisse pour la Recherche).

REFERENCES

Aumann R. J., 1987. Game Theory. In *The New Palgrave, A Dictionary of Economics*, J. Eatwell, M. Milgate and P. Newman (Eds.), Macmillan, London and Basingstoke, 2, pp. 460-482.

D'Avignon G. R. and R. L. Disney, 1976. Single-Server Queues with State-Dependent Feedback. *INFOR Journal*, 14(1), pp. 71-85.

Begon M., 2007. Routage basé sur l'historique dans des files d'attente avec réinjection. *Semester Project*, EPFL-STI-LPM.

Bielen F. and N. Demoulin, 2007. Waiting time influence on the satisfaction-loyalty relationship in services. *Managing Service Quality*, 17(2), pp. 174-193.

Bonabeau E., 2002. Agent-Based Modeling Methods and Techniques for Simulating Human Systems. *Proceedings of The National Academy of Sciences USA*, 99(supplement 3), pp. 7280-7287.

Filliger R. and M.-O. Hongler, 2005. Syphon Dynamics - A Soluble Model of Multi-Agents Cooperative Behavior. *Europhysics Letters*, 70, pp. 285-291.

Gallay O. and M.-O. Hongler, 2007. Market Sharing Dynamics Between Two Service Providers. *European Journal of Operational Research*, in press.

Gallay O. and M.-O. Hongler, 2007. Cooperative Dynamics of Loyal Customers in Queueing Networks. Submitted to *Journal of Systems Science and Systems Engineering*.

Hassin R. and M. Haviv, 2003. To Queue or Not to Queue. *Kluwer*.

Kataoka T., H. Hidenori, K. Kurumatani and A. Ohuchi, 2005. Distributed Visitors Coordination System in Theme Park Problem. *MMAS 2004*, LNAI 3446, pp. 335-348.

Kawamura H., T. Kataoka, K. Kurumatani and A. Ohuchi, 2004. Investigation of Global Performance Affected by Congestion Avoiding Behavior in Theme Park Problem. *IEEJ Transactions EIS*, 124(10), pp. 1922-1929.

Kawamura H., K. Kurumatani and A. Ohuchi, 2004. Modeling of Theme Park Problem with Multiagent for Mass User Support. *Lecture Notes in Computer Science*, 3012, pp. 48-69.

Law A. K. Y., Y. V. Hui and X. Zhao, 2004. Modeling repurchase frequency and customer satisfaction for fast food outlets. *International Journal of Quality and Reliability Management*, 21(5), pp. 543-563.

Peköz E. A. and N. Joglekar, 2002. Poisson Traffic Flow in a General Feedback Queue. *Journal of Applied Probability*, 39(3), pp. 630-636.

Pikovsky A., M. Rosenblum and J. Kurths, 2001. Synchronization. A universal concept in nonlinear sciences. *Cambridge University Press*, pp. 41-42.

Pullman M. E. and G. M. Thompson, 2002. Evaluating capacity - and demand - management decisions at a Ski Resort. *Cornell Hotel and Restaurant Administration Quaterly*, 43(6), pp. 25-36.

Takàcs L., 1963. A Single-Server Queue with Feedback. *The Bell System Technical Journal*, 42, pp. 505-519.