

TDK PAPER

A generalization of Barabási priority model

Lajos Vágó

Supervisors:

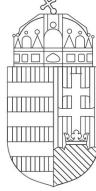
Dr. Károly Simon	Júlia Komjáthy
Professor of Mathematics	Research Fellow
Department of Stochastics,	
Institute of Mathematics	

**BME
2011**

A GENERALIZATION OF BARABÁSI PRIORITY MODEL

BY LAJOS VÁGÓ *

Technical University of Budapest.



Nemzeti
Fejlesztési Ügynökség

In this paper we generalize the A.-L. Barabási priority queuing model, in which tasks are executed according to their priorities. In his work, A. Vázquez obtained exact results for the Barabási model. First, we add some new results related to the expected waiting time using the heuristics of Vázquez. Further, we give a probabilistic interpretation of that result of him, which determines the waiting time distribution of active tasks when the waiting list has length two. On the other hand, we introduce a natural generalization of the Barabási model, and calculate the expected waiting time at stationarity. Furthermore, we calculate numerically the distribution of active tasks with fixed parameters.

1. Introduction. Nowadays, many human-driven phenomena has become momentous in different areas of life, for example in economics and in social sciences. Classical models of human dynamics are based on Poisson processes, so thus in these models the elapsed time between two consecutive events, the inter-event time is exponentially distributed. On the contrary, in many real cases the inter-event time has a heavy-tailed distribution, which distribution decreases in power-law. These heavy tailed distributions give relatively large weight to long inter-event times, and there are bursts of high activity between these inactive intervals. This is in sharp contrast with the exponential distribution, which gives low weight to the longer inter-event times.

Albert-László Barabási showed in his paper [1] that the bursty nature of human behavior can be obtained from a model based on individual decisions. Let us introduce his model: a human has tasks to do, what he/she can store in a buffer of length L . At time 0 there are L tasks in the buffer, and in every discrete time step $(1, 2, \dots)$ two things happen. A task is selected for execution according to a selection protocol, and at the same time an other task arrives to the place of the leaving task. Arriving tasks, including the tasks in the 0-th

*This research was supported by the grant KTIA-OTKA # CNK 77778, funded by the Hungarian National Development Agency (NFÜ) from a source provided by KTIA.

step, get an independent and identically distributed (i.i.d.) priority from some underlying distribution. The selection protocol is as follows. With probability p , where $0 \leq p \leq 1$, the task with the highest priority is selected for execution and leaves the system. Otherwise, a random task is selected. If $p = 1$, then we use the highest priority first selection protocol, and if $p = 0$, then the selection is completely random. Using computer simulations, Barabási found that in the $p \rightarrow 1$ limit the probability distribution of the waiting time τ of a task has a power law tail with exponent 1. This asymptotics already holds for $L = 2$. In the case $L = 2$ Vázquez [9] showed that this asymptotics means exactly that

$$\lim_{p \rightarrow 1} \mathbb{P}(\tau = k) = \begin{cases} 1 + O\left(\frac{1-p}{2} \ln(1-p)\right) & \text{if } k = 1, \\ O\left(\frac{1-p}{2}\right) \frac{1}{k-1} & \text{if } k > 1. \end{cases}$$

This work is organized as follows. In section 2 we review the results of Vázquez [9], then we add some new results related to the expected waiting time using the heuristics of Vázquez. Further, we give a probabilistic interpretation of that result of him, which determines the waiting time distribution of active tasks when the waiting list has length two. In section 3 we introduce a natural generalization of the Barabási model, and calculate the expected waiting time at stationarity. Furthermore, we calculate numerically the distribution of active tasks with some fixed parameters. Finally, in section 4 we sketch directions where we aim to continue our research.

2. Exact results.

2.1. The results of Vázquez. In this section, we review some results of A. Vázquez [9] obtained in the case the waiting list has length $L = 2$. In every step, the task newly added to the list will be called *new*, and the other one *old*. Furthermore, the density function (DF) of the priority of *new* is denoted by $r(x)$, and its cumulative distribution function (CDF) is $R(x)$, similarly *old*'s DF and CDF are $r_1(x, t)$ and $R_1(x, t)$ at step t . Then, the dynamics of the system is given by the following selection protocol. With probability p , where $0 \leq p \leq 1$, the task with the highest priority is selected for execution and leaves the system. Otherwise, a random task is selected. Thus, if we want to determine the CDF of the old task at the $t + 1$ -th step given its CDF at time t , there are two cases: either the old task remains in the system and remains old, or it is selected and the arriving new will become old in the next step. So, the following equation holds:

$$(2.1) \quad R_1(x, t + 1) = \int_0^x r_1(s, t) q(s) ds + \int_0^x r(s) q_1(s, t) ds,$$

where $q(s) = p(1 - R(s)) + \frac{1-p}{2}$ is the probability that the new task is selected given the old task has priority s , and $q_1(s, t) = p(1 - R_1(s, t)) + \frac{1-p}{2}$ is the probability that the old task is selected given the new task has priority s . We will later see that the system has a stationary distribution for the old task's priority if $p < 1$. In this case equation (2.1) turns

into

$$(2.2) \quad R_1(x) = \int_0^x r_1(s)q(s)ds + \int_0^x r(s)q_1(s)ds,$$

where *old*'s DF and CDF are $r_1(x)$ and $R_1(x)$, furthermore $q_1(s) = p(1 - R_1(s)) + \frac{1-p}{2}$. Using the definitions of $q(s)$ and $q_1(s)$, $R_1(x)$ can be exactly determined in the following way:

$$\begin{aligned} R_1(x) &= \int_0^x \frac{1+p}{2}(r_1(s) + r(s)) - p(r_1(s)R(s) + r(s)R_1(s))ds \\ &= \frac{1+p}{2}(R_1(x) + R(x)) - p \int_0^x (R_1(s)R(s))'ds \\ &= \frac{1+p}{2}(R_1(x) + R(x)) - pR_1(x)R(x). \end{aligned}$$

From here $R_1(x)$ can be easily obtained:

$$(2.3) \quad R_1(x) = \frac{(1+p)R(x)}{1-p+2pR(x)} = \frac{1+p}{2p} \left(1 - \frac{1}{1 + \frac{2p}{1-p}R(x)} \right).$$

This $R_1(x)$ is absolutely continuous and satisfies the stationary equation, so the system has a stationary distribution for the old task's priority if $p < 1$.

If we consider a task arriving into the system, it is our main interest to compute its waiting time till execution. The distribution of the waiting time satisfies

$$(2.4) \quad \mathbb{P}(\tau = k) = \begin{cases} \int_0^1 (1 - q_1(x))dR(x) & \text{if } k = 1, \\ \int_0^1 q_1(x)(1 - q(x))q(x)^{k-2}dR(x) & \text{if } k > 1. \end{cases}$$

If $k = 1$, then the task has to be executed immediately, so the probability of this is exactly $1 - q_1(x)$ if the task has priority x . Otherwise, if $k > 1$, then the event that a task with priority x stays exactly k steps on the waiting list is the intersection of k independent events. When the task is added to the list, it must not be executed, this happens with probability $q_1(x)$, then it becomes old and has to stay $k - 2$ more steps on the list, the probability of this event is $q(x)^{k-2}$, and finally, the task has to be executed, which event has probability $1 - q(x)$. It turns out that these integrals can be explicitly computed, hence we obtain

$$(2.5) \quad \mathbb{P}(\tau = k) = \begin{cases} 1 - \frac{1-p^2}{4p} \ln \frac{1+p}{1-p} & \text{if } k = 1, \\ \frac{1-p^2}{4p} \left(\left(\frac{1+p}{2} \right)^{k-1} - \left(\frac{1-p}{2} \right)^{k-1} \right) \frac{1}{k-1} & \text{if } k > 1. \end{cases}$$

As it was expected, the waiting time is independent of $R(x)$. This phenomenon can be explained by the fact that the waiting time of a task depends only on its relative priority respect to the others, so in the case of absolute continuous distributions, with a single change of variables we can transform every other DF back to the uniform case.

From equation (2.5) we can derive a few more consequences:

1. If $p = 0$, i.e. the protocol is completely random, which we also got back in the limit:

$$\lim_{p \rightarrow 0} \mathbb{P}(\tau = k) = \left(\frac{1}{2} \right)^k.$$

2. On the other hand, if $p \rightarrow 1$, then we select only according to priority. After some simple analytical calculations we obtain:

$$\lim_{p \rightarrow 1} \mathbb{P}(\tau = k) = \begin{cases} 1 + O\left(\frac{1-p}{2} \ln(1-p)\right) & \text{if } k = 1, \\ O\left(\frac{1-p}{2}\right) \frac{1}{k-1} & \text{if } k > 1. \end{cases}$$

Thus we see that the waiting time essentially follows a power law distribution, which was also expected by the numerical simulations of A.-L. Barabási [1].

3. When $0 \leq p < 1$ is fixed and $k \rightarrow \infty$

$$\mathbb{P}(\tau = k) \sim \frac{1-p^2}{4} \frac{1}{k} e^{(-\frac{k}{k_0})},$$

where

$$k_0 = \left(\ln \frac{2}{1+p} \right)^{-1},$$

and $a_n \sim b_n$ means that $\frac{a_n}{b_n} \rightarrow 1$ as $n \rightarrow \infty$.

4. Finally, the expectation of τ , when $0 \leq p < 1$:

$$\mathbb{E}(\tau) = \sum_{i=1}^{\infty} i \mathbb{P}(\tau = i) = 2.$$

2.2. Probabilistic interpretation. Consider the selection protocol defined in Section 2.1 with $p < 1$. Now, we give a probabilistic interpretation of this process, and we determine the CDF $R_1(x)$ of the priority of the old task in the stationary case in this manner. Finally, this reasoning gives back the result obtained by Vázquez [9].

We are looking back in time and count how many other tasks had an old task to compete with in order to stay in the list. First, if the selection protocol turns out to be the random choice and a new task is executed, it leaves the list immediately. Thus, the old task does not have to compete with its priority at all, so we do not have to count these cases. This event R_n happens in each step with probability $\frac{1-p}{2}$. Secondly, if random protocol is chosen, and the old task is selected (event R_o), then the arriving new task becomes old and the old task before it leaves the system. Since the priority distribution of the new task is just $R(x)$, considering the old task's priority distribution, the system restarts at each event R_o . The third event which can happen at each step is just priority selection, P . Thus the process can be coded by sequences of type $\{R_n, R_o, P\}^{\mathbb{N}}$.

Let X_t be the t -th element of this sequence. Then the renewal in every R_o event can be formalized as follows. Using the fact that the sequence $\{X_t\}_{t=1}^{\infty}$ is independent of the priorities, for every $x \in \mathbb{R}$

$$\mathbb{P}(O_{t+1} < x \mid X_t = R_o) = \mathbb{P}(N_t < x \mid X_t = R_o) = \mathbb{P}(N_t < x) = R(x),$$

where O_t and N_t is the priority of the *old* and the *new* task at time step t .

Since the priority distribution of the old task is renewed at every event R_o , in order to determine the priority distribution of the old task in the system for a big enough T , we have to count how many P -s we had after the last R_o event.

It is easy to see that the arrival time of the last R_o is just $T - GEO\left(\frac{1-p}{2}\right) + 1$. Similarly, if we only want to count the number of P -s after the last R_o , we have to re-normalize the probabilities to exclude cases of R_n , so it's distribution is

$$GEO\left(\frac{\frac{1-p}{2}}{p + \frac{1-p}{2}}\right) - 1 = GEO\left(\frac{1-p}{1+p}\right) - 1.$$

Thus, the old task in the system at time T had to compete with the other tasks at each event P and had to "win", i.e its priority had to be less than all of them, in order to stay in. So, its priority had to "defeat" $GEO\left(\frac{1-p}{1+p}\right) - 1$ many other tasks to stay in the list so far. Hence the distribution we were looking for is the minimum of $X \sim GEO\left(\frac{1-p}{1+p}\right)$ many independent random variables $(\{Y_i\}_{i=1}^X)$, of distribution $R(x)$. Let us compute it's CDF using the law of total probability:

$$\begin{aligned} F(x) &= \mathbb{P}(\min_{i=1\dots X} Y_i < x) = \mathbb{P}(\exists i \leq X : Y_i \leq x) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(\exists i \leq X : Y_i \leq x \mid X = k) \mathbb{P}(X = k) \\ &= \frac{1-p}{1+p} \sum_{k=1}^{\infty} \left(\frac{2p}{1+p}\right)^{k-1} (1 - (1 - R(x))^k) \\ &= 1 - (1 - R(x)) \frac{1-p}{1+p} \sum_{k=1}^{\infty} \left(\frac{2p(1 - R(x))}{1+p}\right)^{k-1} \\ &= \frac{1+p}{2} \left(1 - \frac{1}{1 + \frac{2p}{1-p} R(x)}\right), \end{aligned}$$

as in the paper [9], since the summation converges if $p < 1$.

2.3. Records and expected waiting time. We have seen that the system has a stationary distribution for the old task's priority if $p < 1$. We have also seen that the expected waiting time of a task in the stationary system is 2. Using the arguments introduced by Vázquez [9] let us denote by τ_i the waiting time of the task executed at the i -th step, and by $\tau'_{1,t}$ the resident time of the task that is still in the buffer at the t step. Since the total buffer time is $2t$ and in each time step exactly one task leaves the system, we have that

$$\sum_{i=1}^t \tau_i + \tau'_{1,t} = 2t.$$

Hence

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \tau_i = 2 - \lim_{t \rightarrow \infty} \frac{\tau'_{1,t}}{t}$$

if these limits exist. If $p < 1$, then we have shown that $\tau'_{1,t}$ has a stationary distribution given by formulas (2.3). So $\frac{\tau'_{1,t}}{t} \rightarrow 0$ as $t \rightarrow \infty$ holds almost surely (a.s.) and in L_1 .

As a consequence we get that the ergodic theorem also holds in this case, that is,

$$(2.6) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \tau_i = \mathbb{E}\tau = 2. \quad \text{a.s. and in } L_1.$$

Let us introduce the notation $\{M_t\}_{t=1}^\infty$ of the old tasks' priorities. Clearly, the sequence is an aperiodic Markov-chain on the state space $[0, 1]$. When $p < 1$, irreducibility of the chain is the consequence of the argument in Section 2.2: the system restarts at each event R_0 , which happens infinitely often a.s.. Thus the Markov-chain is ergodic, and this fact immediately implies that (2.6) holds.

This reasoning verifies rigorously the heuristics of Vázquez [9], and finds the expectation of τ in the stationary case. These arguments can be extended to the case when the buffer length L is greater than 2. In this case, there is no analytic formula for the distribution of waiting time yet, and there is no rigorous argument showing that its expectation is $\mathbb{E}_L(\tau) = L$.

THEOREM 2.1. [Ergodic theorem for $L \geq 2, p < 1$.] *The priority queueing system described in (2.1) is ergodic for any $L > 2$ and $p < 1$ and the following limit holds:*

$$(2.7) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \tau_i = \mathbb{E}_L(\tau) = L. \quad \text{a.s. and in } L_1.$$

PROOF. Set $p < 1$ and $L > 2$. As before, denote by τ_i the waiting time of the task executed at the i -th step, and by $\tau'_{l,t}$ ($l = 1, \dots, L-1$) the resident times of the tasks that are still in the buffer at the t -th step. Then the analogue of equation (2.3) is

$$(2.8) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \tau_i = L - \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{l=1}^{L-1} \tau'_{l,t}$$

if these limits exist. Similarly to the $L = 2$ case, the process renews whenever there are $L - 1$ successive R_0 events, where R_0 means that the oldest task is selected for execution because of the random selection. For every t the probability that there is a renewal section of the process starting at the step t is $\left(\frac{1-p}{L}\right)^{L-1}$. In addition, two renewal sections starting at time t_1 and t_2 are independent if $|t_1 - t_2| \geq L - 1$. So the system renews infinitely often a.s., implying that it has a stationary distribution for the $L - 1$ not newly added task's priority. Then, as before, it follows that

$$(2.9) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{l=1}^{L-1} \tau'_{l,t} = 0.$$

There remains the proof of ergodicity. Let us denote by $\underline{M}_t = (M_t^1, \dots, M_t^{L-1})$ the $L - 1$ dimensional vector of the oldest elements in the buffer at time t , such that M_t^1 is the oldest, M_t^i is the i -th oldest task in the system. Then clearly the sequence $\{\underline{M}_t\}_{t=1}^\infty$ is Markovian with state space $[0, 1]^{L-1}$. It is clearly aperiodic, and since the system renews, it is irreducible. Thus ergodicity and equation (2.8) and (2.9) immediately implies (2.7) as it was suggested by the heuristic arguments of Vázquez [9]. $\square$

2.3.1. Records. In this section we investigate the case $p = 1$, i.e. the case when the selection protocol has no randomness and always picks the task in the system with higher priority. We show that the system achieves no stationary distribution in this case.

It is easy to see that when a task arrives whose priority is less than all the priorities before, then the old task in the system will be chosen for execution and this task will remain in the buffer. This implies that $R_1(x, t)$ (the distribution of the old task's priority at time t) is the distribution of the minimum, i.e. the lower record value of the first $t + 2$ tasks' priority. That is, we need to consider the minimum of $t + 2$ i.i.d. random variables with distribution $R(x)$.

On the other hand, a stationary distribution $R_1^s(x)$ should satisfy equation (2.2), which has only a degenerate solution $R_1(x) = 0$ a.s. if $p = 1$. Thus, the stationary distribution $R_1^s(t)$ will never be achieved, since

$$\mathbb{P}(\forall t \ R_1(x, t) \neq 0) = 1.$$

Hence, the properties of the waiting time τ are related to the process of records.

In order to introduce some classical theorems of record theory, we will need some definitions.

DEFINITION 2.2. *Let $\{X_t\}_{t=1}^\infty$ be i.i.d. random variables with absolutely continuous CDF F . Then we define the followings:*

1. *The sequence $\{T_k\}_{k=1}^\infty$ of lower record times is defined by*

$$\begin{aligned} T_1 &= 1, \text{ and for } k > 1 \\ T_k &= \min\{t \mid X_t < X_{T_{k-1}}\}. \end{aligned}$$

2. *The sequence $\{\Delta_k\}_{k=2}^\infty$ of inter-record times is defined by*

$$\Delta_k = T_k - T_{k-1}.$$

3. *The sequence $\{I_t\}_{t=1}^\infty$ of record indicators is defined by*

$$\begin{aligned} I_1 &= 1, \text{ and for } t > 1 \\ I_t &= \mathbb{1}\{X_t < \min\{X_1, \dots, X_{t-1}\}\}. \end{aligned}$$

4. *The sequence $\{x_k\}_{k=1}^\infty$ of lower record values is defined by*

$$x_k = X_{T_k}.$$

If we do not care about the record values, without loss of generality one can assume that the distribution of the X_i -s is $U[0, 1]$. Thus one can easily compute the joined distribution of the I_t -s. Let $n, 1 = t_1 < t_2 < \dots < t_n$ be arbitrary positive integers. Then

$$\begin{aligned} &\mathbb{P}(I_{t_1} = 1, \dots, I_{t_n} = 1) \\ &= \int_{0 < x_1 < \dots < x_n < 1} \mathbb{P}(I_{t_1} = 1, \dots, I_{t_n} = 1 \mid X_{t_1} = x_1, \dots, X_{t_n} = x_n) dx_1 \dots dx_n \\ &= \int_{0 < x_1 < \dots < x_n < 1} x_1^{t_2-1} x_2^{t_3-t_2-1} \dots x_n^{t_n-t_{n-1}-1} dx_1 \dots dx_n \\ &= \frac{1}{t_1} \frac{1}{t_2} \dots \frac{1}{t_n}. \end{aligned}$$

Since the integers $n, 1 = t_1 < t_2 < \dots < t_n$ were arbitrary, and the I_t -s are discrete random variables, it is evident that these random variables are independent, and $I_t \sim \text{BER}(\frac{1}{t})$.

Now consider the ratio $\frac{\Delta_k}{T_k}$. The distribution of $\frac{\Delta_k}{T_k}$ conditioned on the value of T_{k-1} can be determined by the reasoning of Tata [8]:

$$\begin{aligned} \mathbb{P}\left(\frac{\Delta_k}{T_k} > x \mid T_{k-1} = t\right) &= \mathbb{P}\left(\frac{\Delta_k}{t + \Delta_k} > x \mid T_{k-1} = t\right) \\ &= \mathbb{P}\left(\Delta_k > \frac{x}{1-x}t \mid T_{k-1} = t\right) \\ &= \mathbb{P}(I_{t+1} = 0, \dots, I_{\lfloor t/(1-x) \rfloor} = 0) \\ &= \frac{t}{t+1} \frac{t+1}{t+2} \dots \frac{\lfloor t/(1-x) \rfloor - 1}{\lfloor t/(1-x) \rfloor} \\ &= \frac{t}{\lfloor t/(1-x) \rfloor}. \end{aligned}$$

Since $T_k \rightarrow \infty$ a.s. as $k \rightarrow \infty$, hence

$$\lim_{k \rightarrow \infty} \mathbb{P}\left(\frac{\Delta_k}{T_k} > x\right) = 1 - x,$$

so thus $\frac{\Delta_k}{T_k}$ and $\frac{T_{k-1}}{T_k}$ are asymptotically $U[0, 1]$.

Now we turn back to the Barabási model in the case $p = 1$ and $L = 2$. Also in this situation, equation (2.3) holds:

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \tau_i = 2 - \lim_{t \rightarrow \infty} \frac{\tau'_{1,t}}{t}$$

if these limits exist. But now, with the help of record theory one can prove that the limit on the right hand side does not exist. On the one hand if we choose $t = T_k + 1$ then we obtain:

$$\liminf_{t \rightarrow \infty} \frac{\tau'_{1,t}}{t} = 0 \text{ a.s.},$$

since the sequence of records is infinite a.s. (This is ensured by the continuity of the CDF F), so thus $\tau'_{1,t} = 0$ will hold for infinitely many t -s. On the other hand the *limsup* is taken if $t = T_k$ (we look the subsequence of records):

$$\begin{aligned} \limsup_{t \rightarrow \infty} \frac{\tau'_{1,t}}{t} &= \lim_{k \rightarrow \infty} \frac{\Delta_{k-1}}{T_k} = \lim_{k \rightarrow \infty} \frac{T_k - T_{k-1}}{T_k} \\ &= 1 - \lim_{k \rightarrow \infty} \frac{T_{k-1}}{T_k} = 1 - Z, \end{aligned}$$

where T_k and Δ_k are the k -th record and inter-record times, and Z is a $U[0, 1]$ random variable.

Thus, the heuristic of Vázquez does not work in this case. Note that the dynamics is not even ergodic, since the irreducibility does not hold, as the priority of the old task decreases monotonically.

Further, we can state precise asymptotic results about the waiting time of the records in the queue. It was proven by Holmes and Strawderman [5] that the Strong Law of Large

Numbers (SLLN), by Neuts [6] that the Central Limit Theorem (CLT) and by Strawderman and Holmes [7] that the Law of Iterated Logarithm (LIL) hold for the series of $\ln T_k$ and $\ln \Delta_k$ as well:

THEOREM 2.3. *Let z_k be either Δ_k or T_k . Then*

1. [SLLN][5]

$$\frac{\ln z_k}{k} \xrightarrow{a.s.} 1 \text{ as } k \rightarrow \infty.$$

2. [CLT][6]

$$\frac{\ln z_k - k}{\sqrt{k}} \Rightarrow N(0, 1) \text{ as } k \rightarrow \infty.$$

3. [LIL][7]

$$\begin{aligned} \limsup_{k \rightarrow \infty} \left\{ \frac{\ln z_k - k}{\sqrt{2k \ln \ln k}} \right\} &= 1 \text{ a.s.}, \\ \liminf_{k \rightarrow \infty} \left\{ \frac{\ln z_k - k}{\sqrt{2k \ln \ln k}} \right\} &= -1 \text{ a.s.} \end{aligned}$$

Thus we see that the k -th record spends roughly e^k time in the buffer.

3. A natural generalization of the model. In this section we consider an other variant of the Barabási model which were not analyzed in the literature before. In this new model, the dynamics of the system is given by the following selection protocol. With probability p , where $0 \leq p \leq 1$, a task proportional to the priorities is selected for execution and leaves the system. Otherwise, a random task is selected. So, if there are two tasks in the system with priorities x and y then if priority choice is chosen, then we pick

$$(3.1) \quad \begin{cases} x \text{ w.prob. } \frac{x}{x+y} \\ y \text{ w.prob. } \frac{y}{x+y} \end{cases}$$

The tasks in the buffer are called *new* and *old* as before. Let the priority of the new task be uniformly distributed on $[c, 1]$ ($0 \leq c < 1$), i.e. its DF and CDF is $r(x) = \frac{1}{1-c}$ and $R(x) = \frac{x-c}{1-c}$ if $x \in [c, 1]$. Note that in the original model of priority choice, the underlying distribution of the new tasks does not affect the dynamics, so without loss of generality we could assume that it is uniform on $[0, 1]$. Now this is not the case, i.e having a strictly positive lower bound on the new tasks priority has the effect that the probability of selecting the lower priority task in (3.1) is bounded from below.

Denoting *old*'s DF and CDF by $r_1(x, t)$ and $R_1(x, t)$ at step t the following equation holds:

$$R_1(x, t+1) = \int_c^x r_1(s, t) q(s) ds + \int_c^x r(s) q_1(s, t) ds,$$

where now

$$(3.2) \quad q(s) = \frac{p}{1-c} \int_c^1 \frac{y}{s+y} dy + \frac{1-p}{2} = \frac{1+p}{2} - \frac{ps}{1-c} \ln \frac{1+s}{c+s}$$

is the probability that the new task is selected given the old task has priority s , and

$$(3.3) \quad q_1(s, t) = p \int_c^1 \frac{y}{s+y} dR_1(y, t) + \frac{1-p}{2}$$

is the probability that the old task is selected given the new task has priority s .

Similarly to the model considered in section 2, it is easy to see that the system has a stationary distribution for the old task's priority if $p < 1$. This is because in every step the system renews with probability $\frac{1-p}{2}$ independently of each other. Let us denote the CDF of the old task's priority by $R_1(x)$ in this stationary case. Clearly, the stationary probability that the old task is selected given that the new task has priority s is

$$q_1(s) = p \int_c^1 \frac{y}{s+y} dR_1(y) + \frac{1-p}{2}.$$

Then the analogue of equation (2.4) is as follows:

$$(3.4) \quad \mathbb{P}(\tau = k) = \begin{cases} \int_c^1 1 - q_1(x) dR(x) & \text{if } k = 1 \\ \int_c^1 q_1(x)(1 - q(x))q(x)^{k-2} dR(x) & \text{if } k > 1, \end{cases}$$

with the newly defined $q(x)$ and $q_1(x)$. This equation can be explained as in the previous section.

We argue shortly that shifting the distribution of the new task's priority up by c has a real relevance in the literature. To see this, note that if $X, Y \sim U[c, 1]$, and $V, W \sim U[0, 1]$, then

$$\frac{X}{X+Y} \sim \frac{V(1-c) + c}{V(1-c) + c + W(1-c) + c} \sim \frac{V + \delta}{V + W + 2\delta},$$

where $\delta = \frac{c}{1-c} \geq 0$. This protocol is similar to the growth rule of the preferential attachment model. This model was introduced by Barabási and Albert [2], the first who investigated the model rigorously were Bollobás et al [4], and the mathematically rigorous construction was done by Bollobás and Riordan [3]. The preferential attachment model produces a random graph sequence $\{PA_t(1, \delta)\}_{t=1}^\infty$, where $PA_t(1, \delta)$ has t vertices $(v_1, v_2, \dots, v_t)$ and t edges. Then $PA_{t+1}(1, \delta)$ is generated from $PA_t(1, \delta)$ by adding a new vertex with a single edge to the graph. The procedure of connecting the new edge is very similar to our execution procedure. Namely, the new edge is connected to v_i with probability proportional to $D_i(t) + \delta$, where $D_i(t)$ is the degree of the vertex v_i in $PA_t(1, \delta)$ and $\delta \geq -1$ is constant. When $\delta > -1$ the preferential attachment model has a power-law degree sequence, so it shows similar features also in this sense.

3.1. Stationary equation. We would like to determine the distribution of the old task at stationarity.

First of all, it is easy to verify that a reasoning similar to the probabilistic interpretation in the previous section does not help. Looking back in time for the last renewal event (last R_o choice), similarly to the original version there are $GEO\left(\frac{1-p}{1+p}\right)$ many i.i.d. random variables which have to "compete" with each other. However, now these random variables

do not have the same chance to "win", i.e. not always the minimum stays in the buffer, but the probability of execution depends on the priorities and of the arrival time as well. This asymmetry makes it hard to determine the distribution of the old task's priority at stationarity, so only the analytic way remains to be carried through.

The stationary equation is

$$R_1(x) = \int_0^x r_1(s)q(s)ds + \int_0^x r(s)q_1(s)ds,$$

After substituting (3.3) and (3.4) and rearranging terms we obtain

$$\frac{1+p}{2p}R_1(x) = \frac{1}{1-c} \frac{1-p}{2p}x + \frac{1}{1-c} \int_{z=c}^x \int_{y=c}^1 \frac{y}{y+z}(r_1(z) + r_1(y))dydz.$$

Differentiation with respect to variable x yields the following equation for the stationary density function $r_1(x)$ of the old task's priority:

$$\frac{1+p}{2p}r_1(x) = \frac{1}{1-c} \frac{1-p}{2p} + \frac{1}{1-c} \int_{y=c}^1 \frac{y}{y+x}(r_1(x) + r_1(y))dy$$

We can simplify this equation in two steps:

$$r_1(x) \left(\frac{1+p}{2p} - \left(1 - \frac{x}{1-c} \ln \frac{1+x}{c+x} \right) \right) = \frac{1}{1-c} \frac{1-p}{2p} + \frac{1}{1-c} \int_c^1 \frac{y}{y+x} r_1(y) dy.$$

Easy to see that the coefficient of $r_1(x)$ on the left side cannot be zero, hence we arrive at:

$$(3.5) \quad \begin{aligned} r_1(x) = & \underbrace{\int_c^1 \underbrace{\frac{y}{y+x}}_{\alpha(x,y)=} \underbrace{\frac{1}{\frac{1+p}{2p} - \left(1 - \frac{x}{1-c} \ln \frac{1+x}{c+x} \right)}}_{g(x)=} \underbrace{\frac{1}{1-c}}_{K(x,y)=} r_1(y) dy}_{K(x,y)=} \\ & + \underbrace{\frac{1-p}{2p} \frac{1}{1-c} \frac{1}{\frac{1+p}{2p} - \left(1 - \frac{x}{1-c} \ln \frac{1+x}{c+x} \right)}}_{f(x)=}. \end{aligned}$$

For shorter formulas we introduce the following functions:

$$\begin{aligned} \alpha(x, y) &:= \frac{y}{y+x}, \\ g(x) &:= \frac{1}{\frac{1+p}{2p} - \left(1 - \frac{x}{1-c} \ln \frac{1+x}{c+x} \right)} \frac{1}{1-c}, \\ K(x, y) &:= \alpha(x, y)g(x) \quad \text{and} \\ f(x) &:= \frac{1-p}{2p} \frac{1}{1-c} \frac{1}{\frac{1+p}{2p} - \left(1 - \frac{x}{1-c} \ln \frac{1+x}{c+x} \right)}. \end{aligned}$$

Let us further introduce the integral operator A associated to kernel function K , given by

$$A\varphi(s) := \int_c^1 K(s, t)\varphi(t)dt.$$

With this notation, we see from (3.5) that the density function satisfies

$$(3.6) \quad r_1(x) = Ar_1(x) + f(x).$$

Hence, assuming that the operator A is invertible for certain choice of parameters (p, c) , we would like to determine the function $\varphi(s)$ which satisfies

$$(3.7) \quad \varphi(s) = (I - A)^{-1}f(s)$$

If $\sum_{n=0}^{\infty} A^n f(s)$ converges we could write

$$(3.8) \quad \varphi(s) = \sum_{n=0}^{\infty} A^n f(s).$$

It turns out that A^n is also an integral operator with some kernel function K_n . Let us compute this kernel $K_n(s, t)$:

$$\begin{aligned} A^2 h(s) &= \int_c^1 K(s, t) \int_c^1 K(t, u) h(u) du dt = \int_c^1 \underbrace{\int_c^1 K(s, t) K(t, u) dt}_{K_2(s, u) :=} h(u) du \\ &= \int_c^1 K_2(s, t) h(t) dt. \end{aligned}$$

Similarly

$$K_n(s, t) = \underbrace{\int_c^1 \dots \int_c^1}_{n-1} K(s, u_1) K(u_1, u_2) \dots K(u_{n-1}, t) du_1 \dots du_{n-1}.$$

Note that we can benefit from the fact that $f(s)$ and $g(s)$ (see equation (3.5)) are a constant factor of each other, so we get

$$\begin{aligned} A^n f(s) &= A^n g(s) \frac{1-p}{2p} = \frac{1-p}{2p} \int_c^1 K_n(s, t) g(t) dt \\ &= \frac{1-p}{2p} \int_c^1 \dots \int_c^1 \alpha(s, u_1) g(s) \dots \alpha(u_{n-1}, t) g(u_{n-1}) g(t) du_1 \dots du_{n-1} dt \\ &= \frac{1-p}{2p} g(s) \underbrace{\int_c^1 \dots \int_c^1 \underbrace{\alpha(s, u_1) g(u_1) \dots \alpha(u_{n-1}, u_n) g(u_n)}_{\tilde{K}(s, u_1) =} du_1 \dots du_n}_{H_n(s) =} \end{aligned}$$

Thus we see that it is useful for us to investigate the kernel function given by

$$\tilde{K}(s, t) := \alpha(s, t) g(t),$$

and the corresponding integral operator

$$\tilde{A}h(s) := \int_c^1 \tilde{K}(s, t) h(t) dt.$$

With this notations, the terms $A^n f(s)$ can be obtained as $\frac{1-p}{2p} g(s) H_n(s)$, where

$$\begin{aligned} H_n(s) &= \int_c^1 \dots \int_c^1 \alpha(s, u_1) g(u_1) \dots \alpha(u_{n-1}, u_n) g(u_n) du_1 \dots du_n \\ &= \int_c^1 \dots \int_c^1 \tilde{K}(s, u_1) \dots \tilde{K}(u_{n-1}, u_n) du_1 \dots du_n = \tilde{A}^n \mathbf{1}(s). \end{aligned}$$

So we see that (3.7) is equivalent to

$$(3.9) \quad \varphi(s) = \frac{1-p}{2p} g(s) (1 + H_1(s) + H_2(s) + \dots).$$

In the next section, we will determine $r_1(x)$ numerically, using the software *Mathematica*. We will compute the terms in (3.9) and give estimates on $H_n(s)$. From these we can derive estimates on parameters (p, c) where the series in (3.9) converges.

3.2. Numerical computations. We have seen that $H_n(s) = \tilde{A}^n \mathbf{1}(s)$. So in order to prove convergence of (3.9) it is enough to give an upper bound on the L^2 norm of $H_n(s)$. This can be achieved by estimating the L^2 norm of $\tilde{A}$ or the Hilbert-Schmidt norm of the kernel $\tilde{K}$. We want to give a good estimate on $\|\tilde{A}\|_2$.

For an integral operator T with kernel $\kappa(s, t)$ and a function $f(s)$ with $\|f(s)\|_2^2 = 1$ the L_2 norm of the operator is defined as:

$$(3.10) \quad \|T\|_2 := \sup_{f: \|f\|_2=1} \|Tf\|_2$$

If the operator T is an integral operator with a kernel κ then we have the estimate

$$\begin{aligned} \|(Tf)(s)\|_2^2 &= \int_c^1 \left(\int_c^1 \kappa(s, t) f(t) dt \right)^2 ds \leq \int_c^1 \int_c^1 \kappa^2(s, t) dt \int_c^1 f^2(t) dt ds \\ (3.11) \quad &= \int_c^1 \int_c^1 \kappa^2(s, t) dt ds = \|\kappa(s, t)\|_{HS}^2. \end{aligned}$$

Since $H_n(s) = \tilde{A}^n \mathbf{1}(s)$ we have from (3.10)

$$\|H_n(s)\|_2^2 \leq \|\tilde{A}\|_2^2 \cdot \|\mathbf{1}(s)\|_2^2.$$

We apply the estimates in (3.11) on the operator $\tilde{A}$ to arrive at

$$\|\tilde{A}\|_2^2 \leq \|\tilde{K}\|_{HS}^2.$$

If $\|\tilde{K}\|_{HS}^2 < 1$ then the sum in equation (3.9) converges, so it remains to find the parameter region (p, c) for which this holds.

For fixed $0 \leq c < 1$, $\|\tilde{K}\|_{HS}^2$ is a monotonously increasing function of p ($0 \leq p < 1$), since the norm is given by

$$\begin{aligned} \|\tilde{K}\|_{HS}^2 &= \int_c^1 \int_c^1 \left(\frac{t}{t+s} \frac{1}{\frac{1+p}{2p} - \left(1 - \frac{x}{1-c} \ln \frac{1+x}{c+x}\right)} \frac{1}{1-c} \right)^2 ds dt \\ &= \int_c^1 t^2 \left(\frac{1}{t+c} - \frac{1}{t+1} \right) \left(\frac{1}{\frac{1+p}{2p} - \left(1 - \frac{x}{1-c} \ln \frac{1+x}{c+x}\right)} \frac{1}{1-c} \right)^2 dt \end{aligned}$$

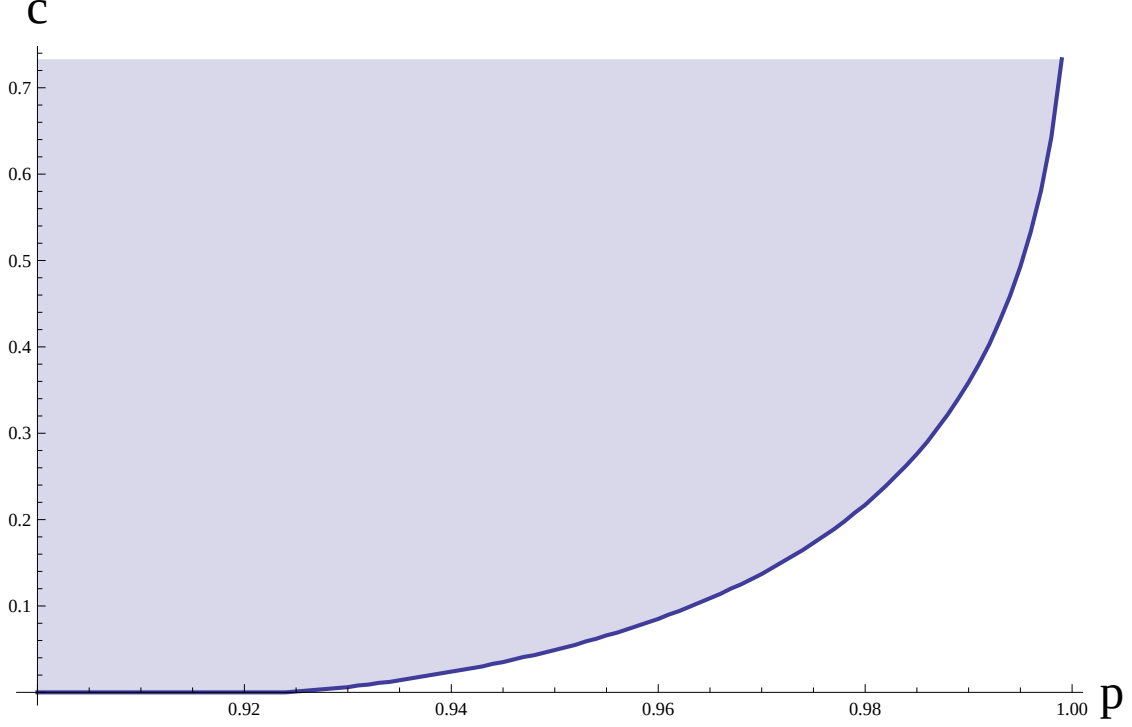


FIG 1. The pair of parameters (p, c) where $\|\tilde{K}(s, t)\|_{HS}^2 < 1$ and $p \geq 0.9$.

The expression increases in p since $\frac{1+p}{2p}$ in the denominator decreases. This implies that if we have found a pair (p, c) for which $\|\tilde{A}\|_{HS}^2 < 1$ then this inequality also holds for all pairs (q, c) where $0 < q \leq p$. The region of pair of parameters for which $\|\tilde{A}\|_{HS}^2 < 1$ can be determined numerically, and it is represented on Fig. 1. In addition, Fig. 2 and 3 illustrates the density $r_1(x)$ and the distribution of τ on a log-log plot for three different choice of parameters, respectively.

Now we give estimates on the distribution of τ when $k > 1$, which is given in equation (3.4). It is easy to see that $q'(x) \neq 0$, so we can write

$$\begin{aligned} \mathbb{P}(\tau = k) &= \int_c^1 q_1(x)(1 - q(x))q(x)^{k-2}dR(x) \\ &= \int_c^1 \underbrace{\frac{1}{1-c} \frac{q_1(x)(1 - q(x))}{-q'(x)}}_{K(x)} (-q'(x))q(x)^{k-2}dx. \end{aligned}$$

With the new notation $K(x) := -\frac{1}{1-c} \frac{q_1(x)(1-q(x))}{q'(x)}$, using the definition of $q(x)$ and $q_1(x)$ it is easy to verify that

$$\begin{aligned} \frac{1-p}{2} &\leq q_1(x) \leq 1, \\ 0 &< 1 - q(c) \leq 1 - q(x) \leq 1 - q(1) < 1, \\ 0 &< -\frac{1}{q'(1)} \leq -\frac{1}{q'(x)} \leq -\frac{1}{q'(c)} < 1. \end{aligned}$$

Thus

$$m \leq K(x) \leq M,$$

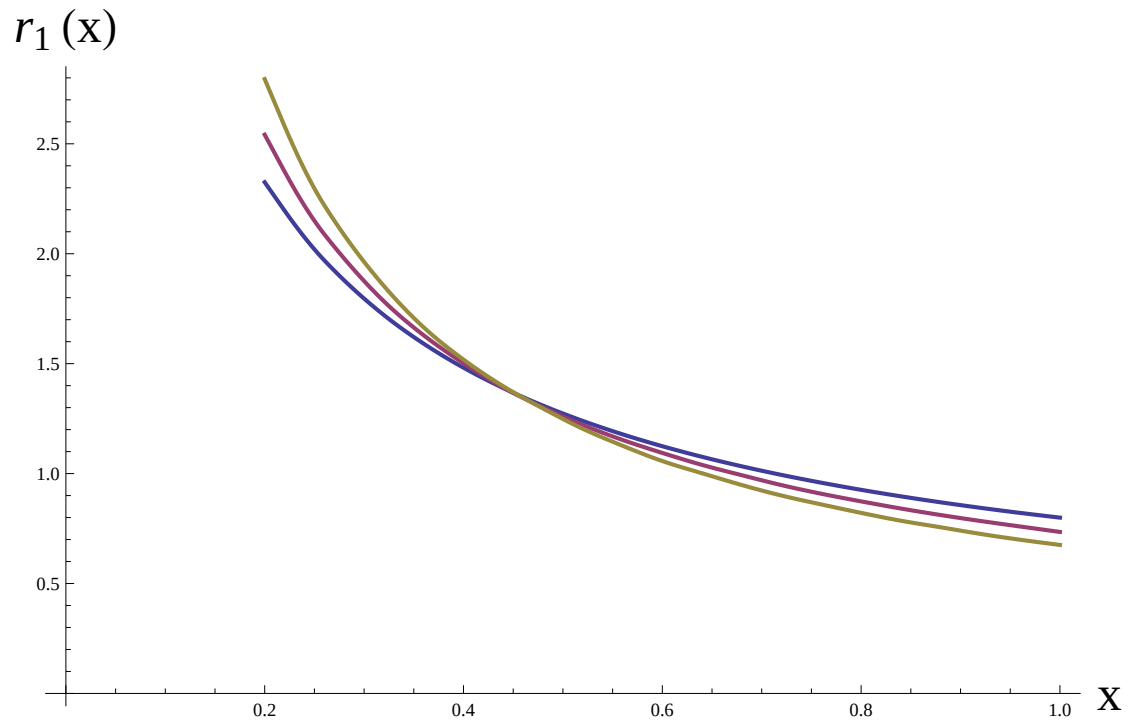


FIG 2. The DF of the old task's priority in the stationary case with $c = 0.2$ and different values of p . Blue: $p = 0.7$. Red: $p = 0.8$. Mustard: $p = 0.9$.

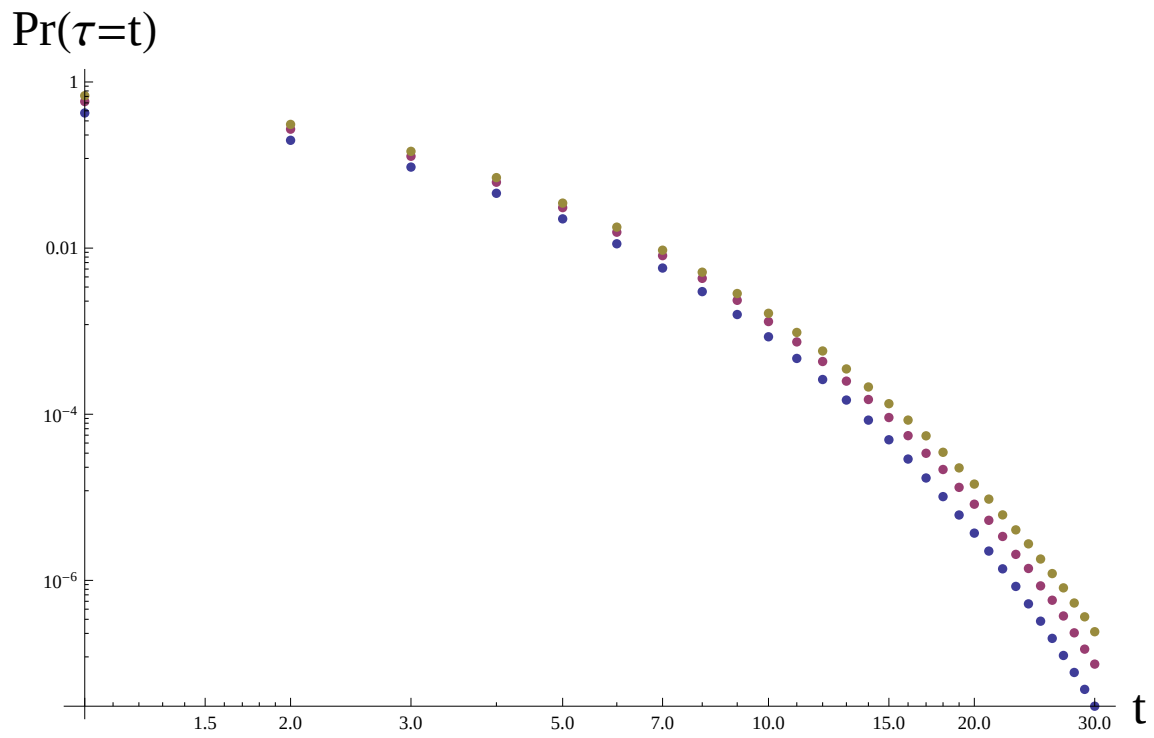


FIG 3. The distribution of τ (i.e. the waiting time of a new task in the stationary case) with $c = 0.2$ and different values of p . Blue: $p = 0.7$. Red: $p = 0.8$. Mustard: $p = 0.9$.

where

$$m = -\frac{1}{1-c} \frac{1-p}{2} \frac{1-q(c)}{q'(1)}, \text{ and } M = -\frac{1}{1-c} \frac{1-q(1)}{q'(c)}.$$

With these bounds, one can easily estimate $\mathbb{P}(\tau = k)$:

$$\mathbb{P}(\tau = k) \leq M \int_c^1 q'(x) q(x)^{k-2} dR(x) = M \frac{1}{k-1} (q^{k-1}(c) - q^{k-1}(1))$$

and similarly

$$\mathbb{P}(\tau = k) \geq m \frac{1}{k-1} (q^{k-1}(c) - q^{k-1}(1)).$$

Since $q(c) > q(1)$, thus

$$(3.12) \quad \mathbb{P}(\tau = k) \sim \text{const} \cdot \frac{1}{k} q^k(c) = \text{const} \cdot \frac{1}{k} e^{-k/k_0},$$

where

$$k_0 = -\frac{1}{\ln(q(c))}.$$

This heuristically means that the probability that the waiting time is of length k behaves approximately as $\frac{1}{k}$ until $k < k_0$, i.e. it obeys a power law-like behavior. Furthermore, if $c \neq 0$ and $(p, c) \rightarrow (1, 0)$, then $k_0 \rightarrow \infty$, i.e. the exponential cutoff in equation (3.12) shifts to infinity, and thus for (p, c) values close to $(1, 0)$ the distribution of τ will be close to a power-law distribution.

3.3. Expected waiting time. In this section we will determine the average waiting time for a task at stationarity. As we have shown in section 2.3 the expected waiting time should equal 2 for any selection protocol where the system is ergodic and has some stationary distribution. We will verify this analytically for our selection protocol. By equation (3.4) we have:

$$\begin{aligned} \mathbb{E}(\tau) &= \sum_{k=1}^{\infty} k \mathbb{P}(\tau = k) = \frac{1}{1-c} \int_c^1 1 - q_1(x) + \sum_{k=2}^{\infty} k q_1(x) (1 - q(x)) q^{k-2}(x) dx \\ &= \frac{1}{1-c} \int_c^1 1 - q_1(x) + \\ &\quad + q_1(x) (1 - q(x)) \left(\sum_{k=2}^{\infty} (k-1) q^{k-2}(x) + \sum_{k=2}^{\infty} q^{k-2}(x) \right) dx \\ &= \frac{1}{1-c} \int_c^1 1 - q_1(x) + q_1(x) (1 - q(x)) \frac{2 - q(x)}{(1 - q(x))^2} dx \\ &= \frac{1}{1-c} \int_c^1 \frac{(1 - q_1(x))(1 - q(x)) + q_1(x)(2 - q(x))}{(1 - q(x))} dx \\ &= \frac{1}{1-c} \int_c^1 \frac{q_1(x)}{1 - q(x)} dx + 1. \end{aligned}$$

Using the definitions (3.2), (3.3) and (3.5) one can obtain:

$$\begin{aligned}
\mathbb{E}(\tau) &= \frac{1}{1-c} \int_c^1 \frac{q_1(x)}{1-q(x)} dx + 1 = 1 + \int_c^1 \frac{g(x)}{p} q_1(x) dx \\
&= 1 + \int_c^1 \underbrace{\left(g(x) \frac{1-p}{2p} + \int_c^1 g(x) a(x, y) r_1(y) dy \right)}_{r_1(x)} dx \\
&= 1 + \int_c^1 r_1(x) dx = 2.
\end{aligned}$$

Note that this calculation was valid for all choices of $q(x) < 1$ and $q_1(x)$, so it stays valid for any protocol where $q(x) < 1 \forall x$ holds. This criterion is met for all $p < 1$, where also the stationarity holds. Furthermore, if $p = 1$ and $c > 0$, then clearly $\forall x \in [c, 1]$ $q(x) \leq 1 - \frac{c}{1+c} < 1$. And, since the probability that a task in the buffer is selected for execution can be bounded from below with $\frac{c}{1+c} > 0$, thus there exists a stationary distribution.

4. Work in progress. There are several ways to continue this research.

1. We gave some parameters (p, c) for which (3.9) converges. It is known that (3.9) could converge with parameters for which $\|\tilde{A}\|_2^2 > 1$ as well, thus the region of parameters could be expanded.
2. The case $L \geq 3$ could be a subject of further research. In this case we have to find the fixpoint of a non-linear operator acting on 2-dim vector field.
3. The models discussed in sections 2 and 3 could be generalized to the case where $L = 2$ and the priorities of the arriving task are not i.i.d. but come from a stationary process. Similar results for this case would be useful if $L \geq 3$ as the following heuristics shows it. If $L = 3$ then the oldest and the second-oldest task as a system behaves as a slowed down version of $L = 2$ where the new priorities are from a stationary process. Thus, in order to determine the behavior of $L \geq 3$ it is of crucial importance to understand the behavior of the system if the new priorities are not i.i.d..

5. History and context of the research. Five institutes (including the Math Institute at BME) received a consortial project with OTKA NKTH no. 77778 for research of the future internet. The leader of the project is professor Károly Simon. The research group aims to state new results on the field of internet related mathematical questions. I joined them a year ago when I was taking the course Random Graphs and Complex Networks. The research areas of the group are mainly concentrated around internet-related mathematical questions, most importantly it focuses on the theory of different network models and dynamics on networks. Our first aim is to describe and present network models that show similar properties to some real networks, e.g. they have hierarchical structure combined with scale-free degree distribution, or they consist of small, well-wired communities connected by weak edges. These networks try to model complex systems such as social networks. On the other hand, a very important other aim of ours is to model the behav-

ior of the spread of information in real networks. This is how modeling human behavior came to our interest, since it models the communication process on a single edge in social networks.

References.

- [1] BARABÁSI, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211.
- [2] BARABÁSI, A.-L. AND ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286**, 5439, 509–512.
- [3] BOLLOBÁS, B. AND RIORDAN, O. (2004). The diameter of a scale-free random graph. *Combinatorica* **24**, 1, 5–34.
- [4] BOLLOBÁS, B., RIORDAN, O., SPENCER, J., AND TUSNÁDY, G. (2001). The degree sequence of a scalefree random graph process. *Random Structures Algorithms* **18**, 3, 279–290.
- [5] HOLMES, P. T. AND STRAWDERMAN, W. E. (1969). A note on the waiting times between record observations. *Journal of Applied Probability* **6**, 3, 711–714.
- [6] NEUTS, M. F. (1967). Waiting time between record observations. *Journal of Applied Probability* **4**, 1, 206–208.
- [7] STRAWDERMAN, W. E. AND HOLMES, P. T. (1970). On the law of the iterated logarithm for inter-record times. *Journal of Applied Probability* **7**, 2, 432–439.
- [8] TATA, M. N. (1969). On outstanding values in a sequence of random variables. *Probability Theory and Related Fields* **12**, 9–20. 10.1007/BF00538520, <http://dx.doi.org/10.1007/BF00538520>.
- [9] VÁZQUEZ, A. (2005). Exact results for the barabási model of human dynamics. *Physical Review Letters* **95**, 248701.