

# 1 Conceptual model

Let's start with the *second* step of a typical discrete-event simulation: the **conceptual model** building.

First of all our system represents a network of nodes to which some application, running in set of external mobile devices, send their tasks, called **jobs**, because of performance or energy saving reasons.

Not all job sent to system are the same, in fact we make distinction between two types of task, denoted as **class 1 job** and **class two job**; differences between them are described in following sections.

To minimize mean response time experienced by mobile device users, the system is been built as a "*two-layer*" network in such a way that an arriving job is sent, if possible, firstly to the nearest node.

To be more precise, our system is made up of a set of nodes described below:

**Cloudlet** It represents an *edge cloud server* which, in addition to being located at "*one-hop*" distance from mobile devices users, and therefore nearest to them, is able to guarantee absence of interferences among tasks allocated to it as long as their number does not exceed a given threshold  $N$ . From our conceptual model point of view, cloudlet represents a **fixed-capacity multi-server service node with any queue**.

Its capacity, that is the maximum possible number of jobs in it, is fixed to  $N$

**Cloud** It represents a *remote cloud server* which, although it **suffers for greater network delay**, due of its high distance from mobile devices, has virtually **unlimited resources** so that he can to process any number of tasks allocated to it. Consequently we had modelled this subsystem as an **unlimited-capacity multi-server service node with any queue**.

Being made up of an unlimited number of servers, its capacity is also unlimited.

**Controller** It represent the entry point of our system because all task are sent to it from outside in order to decide about whether an arrived job should be sent to the cloudlet or the cloud according to a given access control algorithm.

It is located on cloudlet and has the capability to precess each arriving job **instantaneously**. So that it is modelled simply as **single-server service node with no queue**.

---

## Algorithm 1

---

```
1: function ACCESSCONTROLFUNCTION(arrivalJob)
2:   if ( $n_1 + n_2 = N$ ) then
3:     Send arrivalJob on the cloud.
4:   else
5:     Send arrivalJob on the cloudlet.
```

---

At this point, in order to properly define our conceptual model, we need to specify *state variables* of our system. Due to access control algorithms characteristics, described below, we have decided to represent the state of our system

as an ordered pair  $(n_1, n_2)$  where  $n_1$  represents the number of class one job while  $n_2$  the number of class two job both contained in cloudlet. Consequently we have decided to just ignore type and number of jobs contained in cloud node because not relevant in any access control algorithms and negligible for goal of our study.

A this point we can observe access control algorithms used by controller to make decisions which are fully described by their pseudo-code shown in Algorithm 1 and Algorithm 2.

Observe that to make its decision controller node can use only one access control algorithm at a time.

Note that Algorithm 2 use a further threshold variable to make its decision denoted by  $S$  which is less or equal than  $N$ . Remember that when a class 2 job is interrupted and sent on the cloud, a **setup time** has to be considered to restart the task on the cloud.

---

#### Algorithm 2

---

```

1: function ACCESSCONTROLFUNCTION(arrivalJob)
2:   if (arrivalJob.isClassOne) then
3:     if  $n_1 = N$  then
4:       Send arrivalJob on the cloud.
5:     else if  $n_1 + n_2 \leq S$  then
6:       Send arrivalJob on the cloudlet.
7:     else if  $n_2 \geq 0$  then
8:       Interrupt a class 2 job currently running on cloudlet.
9:       Send interrupted job to cloud.
10:      Send arrivalJob on the cloudlet.
11:    else
12:      Send arrivalJob on the cloudlet.
13:  else
14:    if ( $n_1 + n_2 \geq S$ ) then
15:      Send arrivalJob on the cloud.
16:    else
17:      Send arrivalJob on the cloudlet.

```

---

Table 1: Tempi esecuzione sperimentali (espessi in  $\mu s$ ) ottenuti da vari test

$\lambda_{controller}(1)$	Total mean arrival rate into <i>controller</i> of class 1 tasks
$\lambda_{controller}(2)$	Total mean arrival rate into <i>controller</i> of class 2 tasks
$\mu_{cloudlet}(1)$	Mean service rate at <i>cloudlet</i> of class 1 tasks
$\mu_{cloudlet}(2)$	Mean service rate at <i>cloudlet</i> of class 2 tasks
$\mu_{cloud}(1)$	Mean service rate at <i>cloud</i> of class 1 tasks
$\mu_{cloud}(2)$	Mean service rate at <i>cloud</i> of class 2 tasks

## 2 Specification Model

### 3 Analytical solution

In this last section we will develop an analytical solution to validate the results obtained previously through our simulations.

#### 3.1 System based on access control Algorithm 1

Let's start with presentation of analytical solution of system based on access control algorithm 1.

In order to compute all parameters and metrics of interest associated with above-mentioned system, due to cloudlet's limited resources according to which it can accept jobs until their number does not exceed a given threshold  $N$ , is crucial compute first the **fraction of jobs that are forwarded to cloudlet and to cloud**; to do it, we must compute at first probability according to which the sum of job of each class in cloudlet system is equal to that threshold.

To determine this probability, we had modelled cloudlet with a **continuous-time Markov chain (CTMC)**, of which you can see a graphical representation in Figure 1, where each chain's state, denoted with  $(n_1, n_2)$ , is represented by the number of class 1 job,  $n_1$ , and class 2 job,  $n_2$ , present in system at a certain moment.

##### 3.1.1 Balance equation computing

Obviously we can compute **limiting probabilities**  $\pi_{(n_1, n_2)}$ , namely the probability according to which the chain is in a certain state, say  $j$ , independently of the starting state, say  $i$ , via **balance** (or **stationary**) **equations**, in which we can equate the rate at which the system leaves state  $j$  with the rate at which the system enters state  $j$ <sup>1</sup>, **remembering that limiting probabilities sum to 1** (i.e.,  $\sum_{j=0}^{\infty} \pi_j = 1$ ). These balance equations, shown in table 2, are been resolved using a very simple MATLAB script called `MATLAB_ALG1_CTMC_ResolverScript.m` in which each equation, previously generated through a Java script<sup>2</sup>, is been resolved using a MATLAB function called `solve(eqn, var)`.<sup>3</sup>

---

<sup>1</sup>Cfr. Mor Harchol-Balter - *Performance Modeling and Design of Computer Systems* - Carnegie Mellon University, Pennsylvania, pag. 237

<sup>2</sup>Cfr. `CTMCResolverScriptGenerator.java` and `ResolverUsingRoutingAlgorithm1.java` files.

<sup>3</sup>Cfr. <https://www.mathworks.com/help/symbolic/solve.html>

Figure 1: Access control algorithm 1 based cloudlet's CTMC.

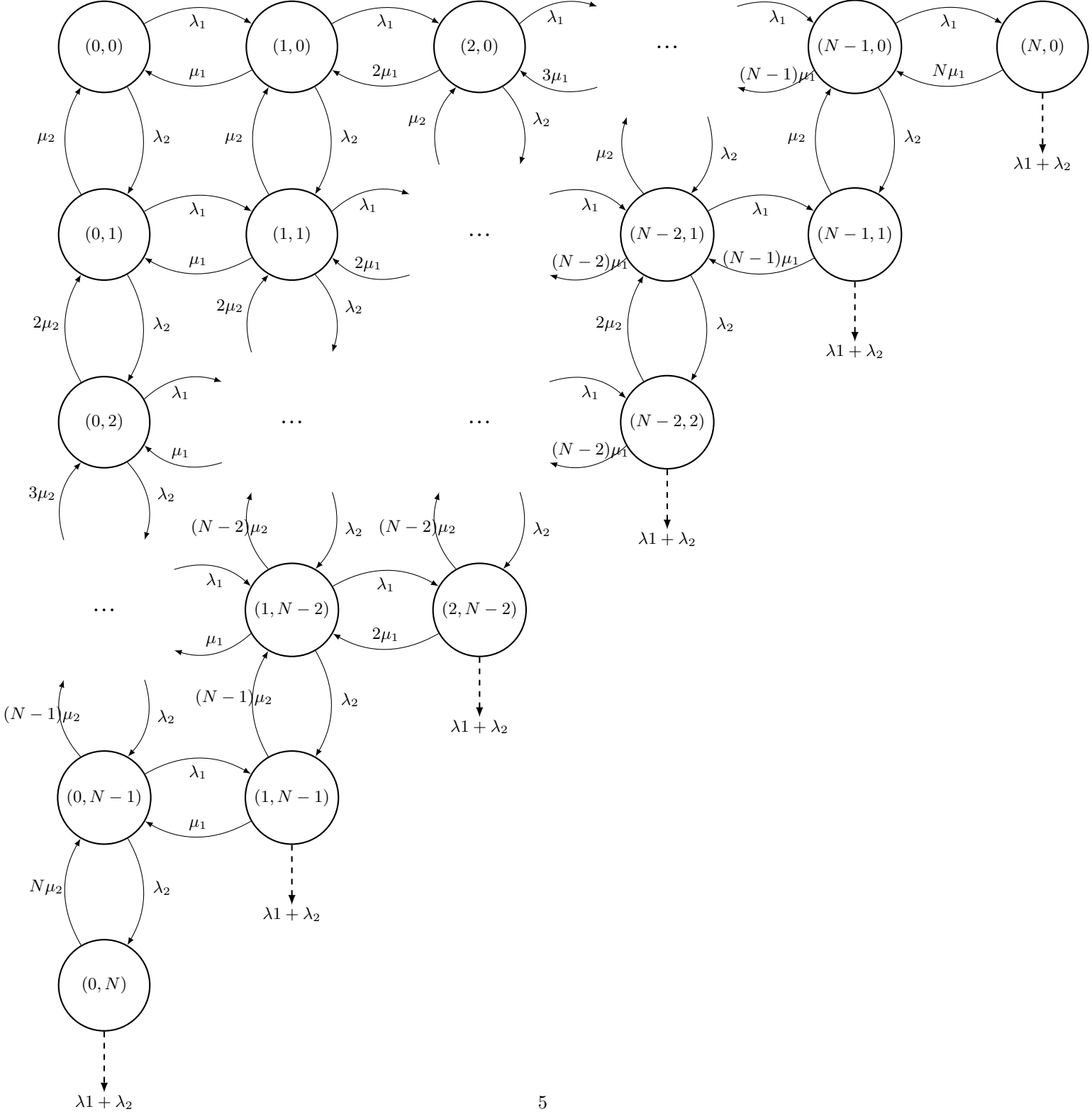


Table 2: Balance equations.

---

$(\lambda_1 + \lambda_2)\pi_{(0,0)}$	$=$	$\mu_1\pi_{(1,0)} + \mu_2\pi_{(0,1)}$	
$(\lambda_1 + \lambda_2 + n_1\mu_1)\pi_{(n_1,0)}$	$=$	$\lambda_1\pi_{(n_1-1,0)} + \mu_1(n_1+1)\pi_{(n_1+1,0)} + \mu_2\pi_{(n_1,1)}$	$\forall n_1 \in \mathbb{N} \cap [1, N-1]$
$(\lambda_1 + \lambda_2 + n_2\mu_2)\pi_{(0,n_2)}$	$=$	$\lambda_2\pi_{(0,n_2-1)} + \mu_1\pi_{(1,n_2)} + \mu_2(n_2+1)\pi_{(0,n_2+1)}$	$\forall n_2 \in \mathbb{N} \cap [1, N-1]$
$\mu_1 N \pi_{(N,0)}$	$=$	$\lambda_1 \pi_{(N-1,0)}$	
$\mu_2 N \pi_{(0,N)}$	$=$	$\lambda_2 \pi_{(0,N-1)}$	
$(n_1\mu_1 + n_2\mu_2)\pi_{(n_1,n_2)}$	$=$	$\lambda_1\pi_{(n_1-1,n_2)} + \lambda_2\pi_{(n_1,n_2-1)}$	$\forall n_1, n_2 \in \mathbb{N} \cap [1, N-1] \mid n_1 + n_2 = N$
$(\lambda_1 + \lambda_2 + n_1\mu_1 + n_2\mu_2)\pi_{(n_1,n_2)}$	$=$	$\lambda_1\pi_{(n_1-1,n_2)} + \lambda_2\pi_{(n_1,n_2-1)} + \mu_1(n_1+1)\pi_{(n_1+1,n_2)} + \mu_2(n_2+1)\pi_{(n_1,n_2+1)}$	$\forall n_1, n_2 \in \mathbb{N} \cap [1, N-1] \mid n_1 + n_2 < N$

---

### 3.1.2 Probabilities computing

Having found the stationary probabilities, we can now find  $\Pi_{\text{SendToCloud}}$ , that is the **probability that an arriving job on controller has to be forwarded to cloud**. Observe that the class to which an arrival job belongs to is not important because, according to access control Algorithm 1, **jobs of both classes have same probability to be sent to cloud**. To be more precise,  $\Pi_{\text{SendToCloud}}$  is the probability that an arrival job find that the number of jobs present in cloudlet has exceeded threshold  $N$ , which occurs when  $n_1 + n_2 = N$ . Formally:

$$\begin{aligned}
 \Pi_{\text{SendToCloud}} &= P\{\text{An arrival job on controller sees } N \text{ jobs in cloudlet}\} \\
 &= \text{Limiting probability that there are } N \text{ jobs in system} \\
 &= \sum_{\substack{n_1, n_2 \in \mathbb{N} \cap [0, N] \\ n_1 + n_2 = N}} \pi_{(n_1, n_2)}
 \end{aligned} \tag{1}$$

At this point we can easily compute  $\Pi_{\text{SendToCloudlet}}$ , which instead represents the **probability according to which an arriving job on controller has to be accepted on cloudlet** and it is same for both job classes too.

$$\begin{aligned}
 \Pi_{\text{SendToCloudlet}} &= P\{\text{An arrival job on controller sees less than } N \text{ jobs in cloudlet}\} \\
 &= 1 - P\{\text{An arrival job on controller sees } N \text{ jobs in cloudlet}\} \\
 &= 1 - \Pi_{\text{SendToCloud}}
 \end{aligned} \tag{2}$$

### 3.1.3 Average arrival rates

Is  $\lambda_i(c)$  the **total arrival rate into a system's component  $i$  of class  $c$  job**. Applying previous results, using an appropriate equation<sup>4</sup>, we can now compute per-class average arrival rates as follow:

$$\begin{aligned}\lambda_{\text{cloud}}(1) &= \lambda_1 \cdot \Pi_{\text{SendToCloud}} \\ \lambda_{\text{cloud}}(2) &= \lambda_2 \cdot \Pi_{\text{SendToCloud}} \\ \lambda_{\text{cloudlet}}(1) &= \lambda_1 \cdot \Pi_{\text{SendToCloudlet}} \\ \lambda_{\text{cloudlet}}(2) &= \lambda_2 \cdot \Pi_{\text{SendToCloudlet}}\end{aligned}\tag{3}$$

Then we get  $\lambda_i$ , that is **total arrival rate to system's component  $i$** , by summing the per-class rates as follows:

$$\begin{aligned}\lambda_{\text{cloud}} &= \lambda_{\text{cloud}}(1) + \lambda_{\text{cloud}}(2) \\ \lambda_{\text{cloudlet}} &= \lambda_{\text{cloudlet}}(1) + \lambda_{\text{cloudlet}}(2)\end{aligned}\tag{4}$$

### 3.1.4 Average population

Is  $E[N_i](c)$  the **average number of class  $c$  jobs into a component  $i$** .

We can use previously computed stationary probabilities to get average population for cloudlet; it's enough to sum each state's limiting probability multiplied by corresponding number of job as follows:

$$\begin{aligned}E[N_{\text{cloudlet}}](1) &= \sum_{(n_1, n_2) \in M} n_1 \cdot \pi(n_1, n_2) \\ E[N_{\text{cloudlet}}](2) &= \sum_{(n_1, n_2) \in M} n_2 \cdot \pi(n_1, n_2) \\ E[N_{\text{cloudlet}}] &= E[N_{\text{cloudlet}}](1) + E[N_{\text{cloudlet}}](2) \\ &= \sum_{(n_1, n_2) \in M} (n_1 + n_2) \cdot \pi(n_1, n_2)\end{aligned}\tag{5}$$

Since we have modelled cloud component as a  $M/M/\infty$  system, in which there is no job's waiting time due to presence of an infinite number of servers, we can simply to apply **Little's Law**<sup>5</sup> in order to get cloud's average population as shown in Equation 6.

$$\begin{aligned}E[N_{\text{cloud}}](1) &= \lambda_{\text{cloud}}(1) \cdot E[T_{\text{cloud}}](1) \\ &= \lambda_{\text{cloud}}(1) \cdot (E[T_{Q_{\text{cloud}}}](1) + E[S_{\text{cloud}}](1)) \\ &= \lambda_{\text{cloud}}(1) \cdot \left(\frac{1}{\mu_{\text{cloud}}(1)}\right) \\ &= \frac{\lambda_{\text{cloud}}(1)}{\mu_{\text{cloud}}(1)}\end{aligned}\tag{6}$$

<sup>4</sup> Cfr. Ivi, pag. 315, equation (18.1)

<sup>5</sup> Cfr. Ivi, pag. 95, theorem (6.1)

Similarly:

$$E[N_{\text{cloud}}](2) = \frac{\lambda_{\text{cloud}}(2)}{\mu_{\text{cloud}}(2)} \quad (7)$$

$$E[N_{\text{cloud}}] = E[N_{\text{cloud}}](1) + E[N_{\text{cloud}}](2) \quad (8)$$

Finally we can get global average job populations as follows:

$$\begin{aligned} E[N](1) &= E[N_{\text{cloudlet}}](1) + E[N_{\text{cloud}}](1) \\ E[N](2) &= E[N_{\text{cloudlet}}](2) + E[N_{\text{cloud}}](2) \\ E[N] &= E[N](1) + E[N](2) \end{aligned} \quad (9)$$

### 3.1.5 Average response time

Is  $E[T_i](c)$  the **mean response time experienced by a class  $c$  jobs into a component  $i$** .

In order to properly compute said metric for each system's component observe that:

- Knowing per-class average job population and per-class mean arrival rates, we can easily compute  $E[T_i](c)$  using Little's Law.
- Since our system haven't queues, because of there is no waiting time experienced by jobs, is true that  $E[T_i](c)$  is also equal to  $E[S_i](c)$ , that is **mean service time experienced by a class  $c$  jobs into a component  $i$** , which is equal to  $1/\mu_i(c)$ , where  $\mu_i(c)$  **means average service rate at which a class  $c$  jobs into a component  $i$  is served**.

Therefore we can get these metric as follows.

$$\begin{aligned} E[T_{\text{cloudlet}}](1) &= \frac{E[N_{\text{cloudlet}}](1)}{\lambda_{\text{cloudlet}}(1)} \\ &= E[S_{\text{cloudlet}}](1) \\ &= \frac{1}{\mu_{\text{cloudlet}}(1)} \end{aligned} \quad (10)$$

$$\begin{aligned} E[T_{\text{cloudlet}}](2) &= \frac{E[N_{\text{cloudlet}}](2)}{\lambda_{\text{cloudlet}}(2)} \\ &= E[S_{\text{cloudlet}}](2) \\ &= \frac{1}{\mu_{\text{cloudlet}}(2)} \end{aligned} \quad (11)$$



$$\begin{aligned}
E[T_{\text{cloud}}](1) &= \frac{E[N_{\text{cloud}}](1)}{\lambda_{\text{cloud}}(1)} \\
&= E[S_{\text{cloud}}](1) \\
&= \frac{1}{\mu_{\text{cloud}}(1)}
\end{aligned} \tag{12}$$

$$\begin{aligned}
E[T_{\text{cloud}}](2) &= \frac{E[N_{\text{cloud}}](2)}{\lambda_{\text{cloud}}(2)} \\
&= E[S_{\text{cloud}}](2) \\
&= \frac{1}{\mu_{\text{cloud}}(2)}
\end{aligned} \tag{13}$$

At this point we can get global per-class mean response times as follows:

$$\begin{aligned}
E[T](1) &= E[T_{\text{cloudlet}}](1) \cdot P\{\text{An arrival class 1 job is sent to cloudlet}\} \\
&\quad + E[T_{\text{cloud}}](1) \cdot P\{\text{An arrival class 1 job is sent to cloud}\} \\
&= E[T_{\text{cloudlet}}](1) \cdot \Pi_{\text{SendToCloudlet}} + E[T_{\text{cloud}}](1) \cdot \Pi_{\text{SendToCloud}}
\end{aligned} \tag{14}$$

Similarly:

$$\begin{aligned}
E[T](2) &= E[T_{\text{cloudlet}}](2) \cdot P\{\text{An arrival class 2 job is sent to cloudlet}\} \\
&\quad + E[T_{\text{cloud}}](2) \cdot P\{\text{An arrival class 2 job is sent to cloud}\} \\
&= E[T_{\text{cloudlet}}](2) \cdot \Pi_{\text{SendToCloudlet}} + E[T_{\text{cloud}}](2) \cdot \Pi_{\text{SendToCloud}}
\end{aligned} \tag{15}$$

Finally, using obtained results, we can get global mean response times as shown below:

$$E[T] = E[T](1) \cdot \frac{\lambda_1}{\lambda_1 + \lambda_2} + E[T](2) \cdot \frac{\lambda_2}{\lambda_1 + \lambda_2} \tag{16}$$

### 3.1.6 Throughput

To determine system's throughput let's prove if our system is stable. As we know, every queueing system in which its mean arrival rate is less than its mean service rate is known as a stable system; formally, are  $\lambda$  and  $\mu$  our system's mean arrival rate and mean service rate respectively, if  $\lambda \leq \mu$ , given system is stable. Cloudlet subsystem is clearly *not* stable due to its limited resources compared to its workload; we have already seen that exists a not null probability according to which an arriving job on controller sees  $N$  jobs in cloudlet which implying its forward to cloud subsystem. Since it haven't a queue, cloudlet stability

condition is achieved when  $\Pi_{\text{SendToCloud}}$  is null but, based on current system parameters, is not the case. Regarding cloud subsystem, no matter how high we make  $\lambda_{\text{cloud}}$  because it is made up of an infinite number of server for which completion rate is still bounded by the arrival rate. Accordingly to previous considerations we can conclude that **our system is stable** due to stability or its cloud subsystem therefore we can get system throughput as follows:

$$X = \lambda = \lambda_1 + \lambda_2 \quad (17)$$

To get per-subsystems throughput we proceed as shown below:

$$X_{\text{cloud}} = \lambda_{\text{cloud}} \quad (18)$$

$$X_{\text{cloudlet}} = X - X_{\text{cloud}} \quad (19)$$

### 3.1.7 Summary of analytical results

	Class 1 Jobs			Class 2 Jobs			Cloudlet	Cloud	Global
	Cloudlet	Cloud	System	Cloudlet	Cloud	System			
$\Pi_{\text{SendToCloudlet}}$									0.5861
$\Pi_{\text{SendToCloud}}$									0.4139
$X, \lambda$ (jobs/s)	2.3444	1.6556	4	3.663125	2.586875	6.25	6.007525	4.242475	10.25
$E[N]$ (jobs)	5.2040	6.6224	11.8264	13.5521	11.75852273	25.31062273	18.7561	18.38092273	37.13702273
$E[S], E[T]$ (s)	20/9	4	2.835407407	100/27	50/11	4.225763636	3460/1107	1954/451	3.683185595
$\rho$	0.2604888889			0.6783564815			0.9388453704		

### 3.2 System based on access control Algorithm 2

$$\begin{cases}
(\lambda_1 + \lambda_2)\pi_{(0,0)} = \mu_1\pi_{(1,0)} + \mu_2\pi_{(0,1)} \\
(\lambda_1 + \lambda_2 + i\mu_1)\pi_{(i,0)} = \lambda_1\pi_{(i-1,0)} + \mu_1(i+1)\pi_{(i+1,0)} + \mu_2\pi_{(i,1)} & \forall i \in \mathbb{N} \mid 1 \leq i \leq S-1 \\
(\lambda_1 + \lambda_2 + i\mu_2)\pi_{(0,i)} = \lambda_2\pi_{(0,i-1)} + \mu_1\pi_{(1,i)} + \mu_2(i+1)\pi_{(0,i+1)} & \forall i \in \mathbb{N} \mid 1 \leq i \leq S-1 \\
(S\mu_1 + \lambda_1)\pi_{(S,0)} = \lambda_1\pi_{(S-1,0)} + \lambda_1\pi_{(S-1,1)} + (S+1)\mu_1\pi_{(S+1,0)} \\
(\lambda_1 + S\mu_2)\pi_{(0,S)} = \lambda_2\pi_{(0,S-1)} \\
(i\mu_1 + j\mu_2 + \lambda_1)\pi_{(i,j)} = \lambda_1\pi_{(i-1,j)} + \lambda_2\pi_{(i,j-1)} & \forall i, j \in \mathbb{N} \mid \begin{matrix} 1 \leq i \leq S-1 \\ 1 \leq j \leq S-1 \end{matrix} \mid i+j = S \\
(\lambda_1 + \lambda_2 + i\mu_1 + j\mu_2)\pi_{(i,j)} = \lambda_1\pi_{(i-1,j)} + \lambda_2\pi_{(i,j-1)} + \mu_1(i+1)\pi_{(i+1,j)} + \mu_2(j+1)\pi_{(i,j+1)} & \forall i, j \in \mathbb{N} \mid \begin{matrix} 1 \leq i \leq S-1 \\ 1 \leq j \leq S-1 \end{matrix} \mid i+j < S \\
(i\mu_1 + \lambda_1)\pi_{(i,0)} = \mu_1(i+1)\pi_{(i+1,0)} + \lambda_1\pi_{(i-1,0)} & \forall i \in \mathbb{N} \mid S+1 \leq i \leq N-1 \\
N\mu_1\pi_{(N,0)} = \lambda_1\pi_{(N-1,0)} \\
\sum \pi_{(i,j)} = 1 & \forall i, j \in \mathbb{N}_0
\end{cases}$$

#### 3.2.1 Probabilities computing

although... To properly analyse this system, we need to compute some useful probabilities.

Similarly to the previous case we need to now following probabilities:

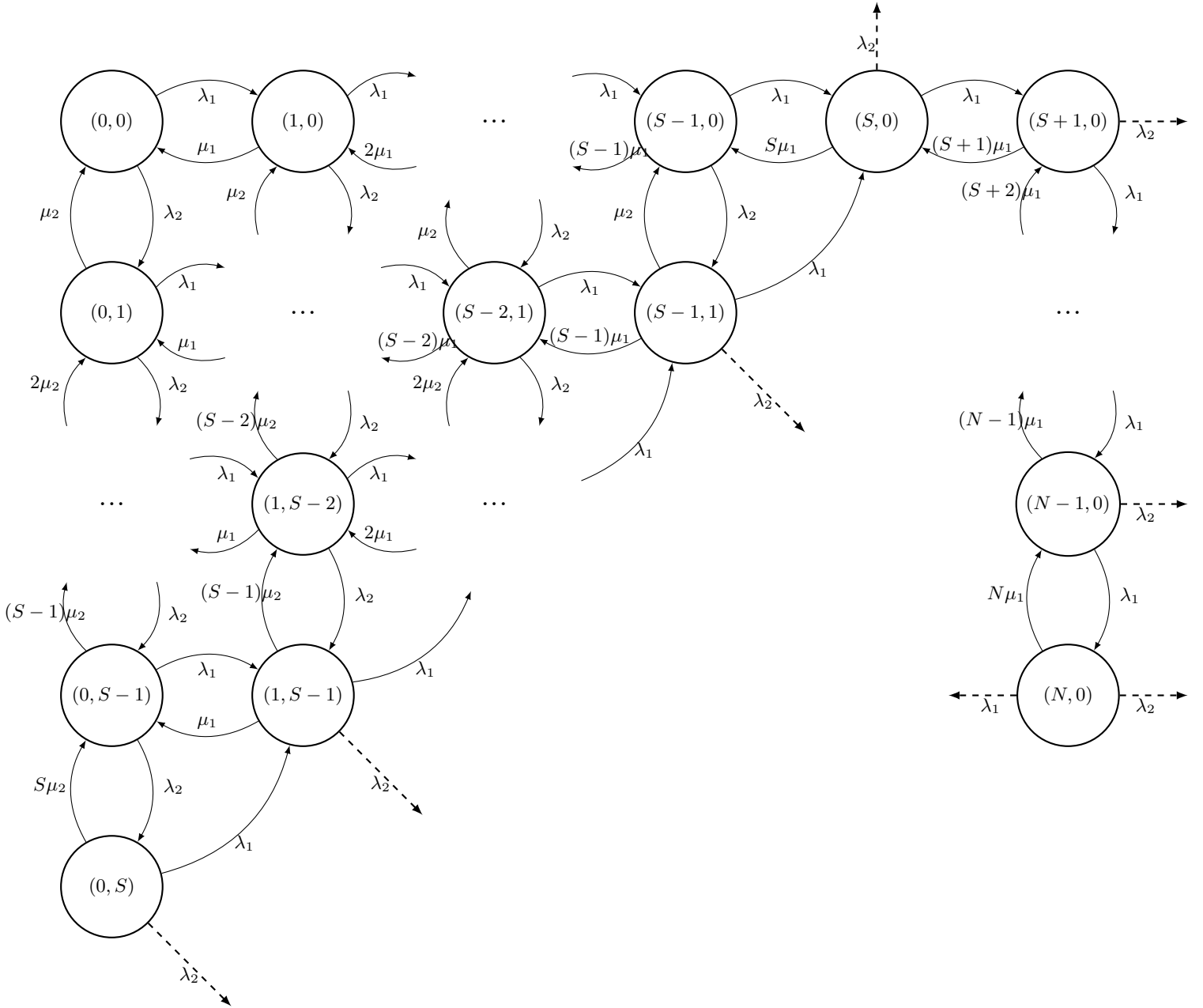
Table 3: Lista dei file del malware FASTCash

$\Pi_{\text{SendToCloudlet}}(k)$	Probability that an arriving job of class $k$ on controller has to be forwarded to cloudlet
$\Pi_{\text{SendToCloud}}(k)$	Probability that an arriving job of class $k$ on controller has to be forwarded to cloud
$\Pi_{\text{Class2JobInterruption}}(k)$	Probability that a job of class 2 running on cloudlet has to be interrupted and forwarded to cloud due of arriving class 1 job on cloudlet.

Differently from the previous case, owing to the use of a different access control algorithm, jobs of different classes have different probability according to which an arriving job on controller has to be send to cloud. From CTMC analysis, we can easily understand that:

$$\begin{aligned}
\Pi_{\text{SendToCloud}}(1) &= P\{\text{An arrival class 1 job on controller sees } N \text{ class 1 jobs in cloudlet}\} \\
&= P\{n_1 = N\} \\
&= \pi(N, 0)
\end{aligned} \tag{20}$$

Figure 2: Cloudlet system component modeled using a CTMC



$$\begin{aligned}
\Pi_{\text{SendToCloud}}(2) &= P\{\text{An arrival class 2 job on controller sees that number of jobs in cloudlet exceed or} \\
&= P\{n_1 + n_2 \geq N\} \\
&= \sum_{\substack{n_1, n_2=0 \\ n_1+n_2=S}}^S \pi(n_1, n_2) + \sum_{n_1=S+1}^N \pi(n_1, 0) \\
&= \sum_{\substack{0 \leq n_1 \leq N \\ 0 \leq n_2 \leq S \\ n_1+n_2 \geq S}} \pi(n_1, n_2)
\end{aligned} \tag{21}$$