Analysis of Television and Cinema Productions using Mean Field Models

P. Piazzolla [1]   M. Gribaudo [2]

Dipartimento di Informatica
Università di Torino
Torino, Italy

Abstract

Television and cinema productions have a critical influence in today's economy, but performance evaluation techniques are rarely used in these contexts. Both kind of productions share a similar key feature: the spectator. It is in fact the final user who determines the success of a production by deciding to spend his or her money for it. In this paper we use Mean Field techniques to model the behavior of the users, of the production and of the distribution. We use Generalized Stochastic Petri Nets as an upper level formalism to simplify the description of the proposed models. Finally we present comparisons with real data (where available) to prove the validity of the Mean Field approach.

Keywords: Mean field, Generalized Stochastic Petri Nets, Media Production, User Models.

1 Introduction

Television and cinema productions have a critical influence in today's economy, but performance evaluation techniques are rarely used in these contexts. Both kind of productions share a similar key feature: the spectator. It is him in fact, as the final user of the product, who determines the economic success of a production by deciding to gain access to it, in one form or another, usually spending his money or time in doing so. However, modelling the behavior of the users is not an easy task: the actual success of a TV or cinema production is determined by the behavior of a large number of users.

To accurately address the problem, each user must be modelled individually, since it is the choice of the majority of the users that determines the final behavior of the entire population. We imagine, through this paper, a sort of ideal spectator, well informed about the existence and the features of the product he or she is going

---

[1] Email: piazzolla.pietro@di.unito.it
[2] Email: marcog@di.unito.it

to buy. This is to exclude other sorts of influence to his or her behavior not explicitly part of our model.

The necessity of studying systems composed by several interacting objects is not new, and already emerged in several other applications like sensor networks, rich internet applications, social networks and molecular biology. However, the well known *state space explosion* problem has prevented most of the previously developed compositional techniques from being applied. For this reason new approximate analytical techniques have been introduced. For example *Fluid Process Algebra* [10] is an extension of the PEPA stochastic Process Algebra, that considers the number of components as continuous variables, and studies their evolution using ODE (Ordinary Differential Equations). *Mean Field Analysis* provides approximation of the counting process of objects described by both Discrete [7] and Continuous [6] Markov chains. *Markovian Agents* [9] instead considers a continuous population of entities (the agents) spread over a space that communicates by sending and receiving messages.

All the aforementioned techniques are based more or less on the same assumptions: a continuous (or *fluid*) approximation of counting process that computes the number of objects in each of the possible states. Although all these techniques are approximate, they provide very accurate solution as the population increases. A detailed study on the accuracy of such approximations as function of the population size can be found in [5] and [6].

In this paper we will use *Mean Field Analysis* to study the behavior of the users in two different contexts: cinema and television. We will use Generalized Stochastic Petri Nets (GSPN) [4] to specify the behavior of different agents involved in the models. For what concerns cinema, in Section 3 we will focus on the evolution of the number of spectators and seats available in theaters after the release date of a new movie. In Section 4 we will focus on the evolution of the audience of a TV Series, depending on the improvements effected by the production.

2   Preliminaries

In this work we will mainly use results from the mean field analysis [6] to consider a very large number of interacting components. In order to simplify the presentation of the agents involved in the models, we will describe their behavior using GSPN instead of Continuous Time Markov Chains (CTMC). A detailed description of a procedure that transforms a GSPN in a CTMC can be found in [4]. In our case however, the computation of the transition matrix of the CTMC associated to the GSPN must be done symbolically, in order to preserve the dependency on the state of the system.

2.1   Mean Field Analysis

Mean Field Analysis is a technique that approximates the counting process of sets of partially dependent similar objects. All the entities have the same behavior, which might depend either on their local state, or on the counting process of the

whole system. In particular, if there are N identical objects, each one described by m states, the evolution of the entire system can be approximated by its counting process $\mathbf{N}(\tau) = N_i(\tau)$. Here $N_i(\tau)$ counts the number of entities in state i, and is such that $\sum_{i=1}^{m} N_i(\tau) = N \; \forall \tau \geq 0$. If we call $X(\tau)$ the state of a given (tagged) object, following [6] we can define:

(1)
$$K_{ij}(\mathbf{N}(\tau)) =
\begin{cases}
\lim_{\Delta \to 0} \dfrac{Pr\{X(\tau + \Delta) = j | X(\tau) = i, N(\tau)\}}{\Delta} & \begin{array}{l} i \neq j, \\[4pt] N_i(\tau) > 0 \end{array} \\[20pt]
0 & \begin{array}{l} i \neq j, \\[4pt] N_i(\tau) = 0 \end{array} \\[16pt]
-\sum_{l \neq i} K_{il}(\mathbf{N}(\tau)) & i = j
\end{cases}$$

We can then compute the (approximate) transient evolution of the counting process by defining a matrix $\mathbf{K}(\mathbf{N}(\tau)) = [K_{ij}(\mathbf{N}(\tau))]$, and by solving the following ODE:

(2)
$$\frac{d\mathbf{N}(\tau)}{d\tau} = \mathbf{N}(\tau)\mathbf{K}(\mathbf{N}(\tau))$$

Note that $K_{ij}(\mathbf{N}(\tau))$ corresponds to the transition rate from state i to state j of one object in isolation when there is at least one entity in state i, and conditioned to the counts of the number of objects in each state. If the model is specified by a GSPN, this corresponds to the (symbolic) infinitesimal generator of the underlying CTMC. The approximation converges to the exact solution as $N \to \infty$.

Mean Field Analysis has also other interesting properties. Let us consider a system composed by C different types of entities, each one characterized by a different number of states $m^{(c)}$, $c = 1..C$, a counting vector $\mathbf{N}^{(c)}(\tau)$, and a different transition matrix $\mathbf{K}^{(c)}(\mathbf{N}(\tau))$. Each matrix $\mathbf{K}^{(c)}(\mathbf{N}(\tau))$ can depend on the whole state of the system $\mathbf{N}(\tau) = (\mathbf{N}^{(1)}(\tau), \ldots, \mathbf{N}^{(C)}(\tau))$. The evolution of the entire system can be computed by solving C independent ODEs, coupled only by the dependency on the entire vector $\mathbf{N}(\tau)$. Moreover, following [5], it is possible to add a vector of h additional *Measures* $\mathbf{M}(\tau)$, that are computed on the state space. Measures evolve following a function $\mathcal{M}$ that can depend both on the states of the objects, and on the previous value of the measures. In particular, the complete solution of the system can be computed by solving:

(3)
$$\begin{cases}
\dfrac{d\mathbf{N}^{(c)}(\tau)}{d\tau} = \mathbf{N}(\tau)\mathbf{K}(\mathbf{N}(\tau), \mathbf{M}(\tau)) & c = 1..C \\[10pt]
\dfrac{d\mathbf{M}(\tau)}{d\tau} = \mathcal{M}(\mathbf{N}(\tau), \mathbf{M}(\tau))
\end{cases}$$

2.2 Notations

In the specification of the models, we will use the graphical notation proposed in Figure 2. In particular, all the object classes considered in the model will be represented by big boxes. Each box will contain inside it a GSPN that describes

the behavior of the objects of that class. At the top-right corner of each box, we will write in a cell the multiplicity of the object: that is the number of objects of that kind that composes the system. In both the examples, Measures will be used to simulate a controller that guides the production or the distribution of a content. In particular, vector $\mathbf{M}(\tau)$ will be used to store the state distribution of a CTMC that describes the behavior of the controller. This behavior, will be visualized by another GSPN that is not included into a box.

When specifying the state dependent transition rate, we will use $r(T_i)$ to denote the rate of transition T_i. We will call $n(\mathcal{P})$ the number of objects that are characterized by marking $\mathcal{P}$, and in particular $n(P_i)$ will denote the number of agents that have place P_i marked. When expressing dependencies from the controller, we will call $\#(P_i)$, the number of tokens in place P_i (in this case P_i must be a place of the GSPN of the controller).

3   Cinema production

The first considered model aims at analyzing the behavior of the spectators of a newly released movie. Firstly we present some statistical data that characterizes this system, and then we describe the proposed model. Finally we compare the results and the data to discuss the effectiveness of the proposed model.

3.1   Spectators and Movies

We start investigating the income data of some of the most well known movie trilogies in their first 12 weeks at box offices across United States. The income data can be considered directly proportional to the number of spectators to the show, and thus can be a significant parameter of the considered system. The data we used are available through the net or specialized literature. In particular we used a data set from the Swivel web site [3]. Figure 1 shows the temporal evolution in earned USA dollars per week of 4 famous trilogies: *X-Men*, *Lord of The Rings*, *Mission Impossible* and *Shrek*. We believe that they can be a good representative of the main evolution tendencies seen among the movies. We observed 3 main ways in which the income may vary.

The first tendency, (which is also the most frequent one), is the one that characterizes the third episode of the top two film in Figure 1. We will refer to it through the paper as the A tendency. It displays a fast, nearly exponential, decay in income after the first days of film release. This behavior in our model is explained by a quick drop in the number of spectators wanting to see the movie because they are satisfied by the availability of cinemas screening it. As shown in Table 1, half of the total income is gained in the first week by films following this behavior.

The tendency somehow typical of the second episode of a trilogy is the one we will label as the B through the paper. In this case the number of people going to watch the film doesn't drop as quickly as in the first case, and remains almost constant for the first few weeks. Then, it drops down at a very fast rate (that can be even faster than in case A). The motivation for the initial constant rate
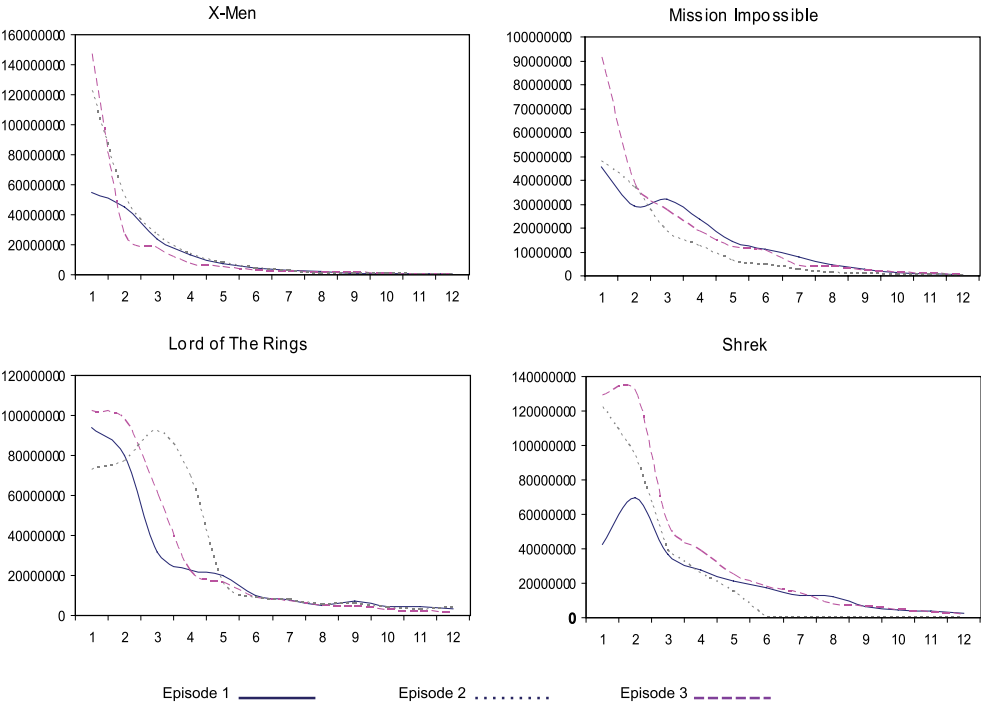
Fig. 1. Examples of typical spectator (y axis) distributions per week (x axis)

	Episode 1	Episode 2	Episode 3
Xmen	34.98	68.87	52.51
Mission Impossible	26.00	43.19	35.79
Lord of the Rings	32.29	30.73	19.72
Shrek	16.43	29.69	41.22

Table 1
Percentage of spectators in the first week

can be justified by the limited cinemas' capacity because all the available seats are saturated in the first weeks. However, as soon as the initial batch is over, the number of subsequent spectators drops down quickly. This is what we believe a typical "second episode effect": people who have enjoyed the first episode may have a great expectation toward the second one, wanting to see it as soon as it becomes available, thus saturating the cinemas' capacity. However, probably due to the lack of originality that most of the sequels seems to have, people tend to drop down quite quickly after the first few weeks of screening.

In the last tendency, that we called C, there is a reprise in the income after the usual initial decay. This can be justified by the positive feedback received from those who watched the film. Many people, who are not initially interested in the movie, change their mind and may decide to watch it thanks to the positive comments heard about it. We believe that this reprise is a form of *positive enforcement* due to the unexpected great matching of the movie features with the audience expectations.

This is for example the case of the first *Shrek* or the first *Impossible Mission* in the plots provided in Figure 1.

3.2 Model description

The considerations given in the previous section, can help us in defining the model. The model of the whole system is presented in Figure 2, and it is composed by three different entities: the spectators, the cinemas and the distribution. The last model is used as a control of the entire process and is implemented using the Measure vector $\mathbf{M}(\tau)$ as a measure of the controllers's probability of being in a specific state. The matrix with which measures are computed $\mathcal{M}(\mathbf{N}(\tau), \mathbf{M}(\tau))$ corresponds to the infinitesimal generator of the CTMC associated with the controller process.

We imagined a number N of potential spectators that can be either interested in the new film or not. Negative feedback received from those who already watched the film may affect the behavior of those interested. Those uninterested may become interested thanks to the positive feedback received from other satisfied spectators. For each week, only a limited number of people can really watch the movie, because of the limited capacity (in term of 'seats-per-day') of the cinemas. As the time passes, less people are interested in watching the film, for different reasons. Cinemas does tend to change their screenings, reducing the number of seats available for the considered movie. Each week the society in charge of the distribution of the film can decide to stop its screening in all cinemas.
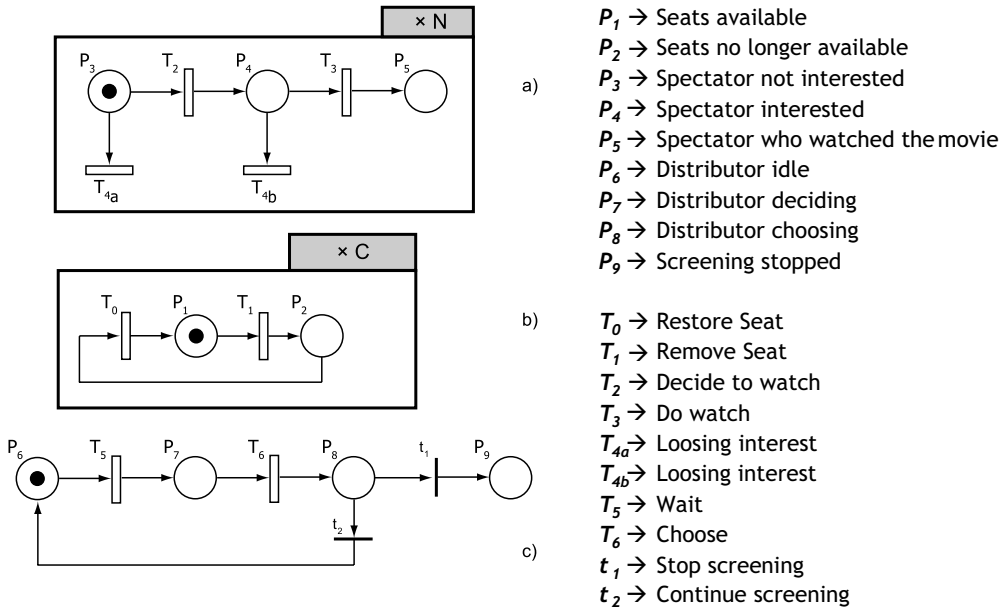

Fig. 2. A model of: (a) the audience, composed by N movie spectators; (b) the theaters, capable of serving C customers; and (c) the distributor of the movie.

Figure 2a shows the behavior of the potential N spectators of the movie: they can be interested (if place P_4 is marked), or not (place P_3) to the new movie. Those

interested are, as said before, those who intend to watch the movie. P_5 is the place that represents those who watched the movie. Transition T_3 models the action of watching the movie, and it rate $r(T_3) = K_2 \frac{n(P_1)}{n(P_4)}$ is expressed as a constant K_2 (which represents the number of people a 'seat' can serve per week) multiplied by the number of available seats $(n(P_1))$ and divided by the number of spectator that are competing to watch the movie $(n(P_4))$. Transition T_2, between P_3 and P_4, is based on the positive feedback received from those who already watched the movie. It's rate is $r(T_2) = \alpha_L n(P_5)^{\alpha_w} \mathbf{1}(n(P_3) > \theta_r)$, meaning that it is proportional to the α_w power of the number of spectators who have already watched the movie $n(P_5)$, by a constant α_L that represents how much a positive feedback does modify a person's decision. The exponent α_w has been introduced to allow non-linear effects on the customers feedbacks. The guard $\mathbf{1}(n(P_3) > \theta_r)$, is an indicator function that is equal to $1(\cdot)$ when the associated condition is verified. It is used to enable the reinforcement feedback only when the number of spectators that have seen the movie exceeds a threshold θ_r. The people interested (or not) to the film may also become definitely uninterested to it (by exiting the system) if they receive too much negative feedback. This is modelled by transitions T_{4a} and T_{4b}, whose rate is $r(T_{4a}) = r(T_{4b}) = \beta_D n(P_5)$. Constant β_D defines how much negative feedbacks modify a person's decision.

Figure 2b shows the behavior of cinemas screening the movie. Each cinema is capable of serving a given number of spectators each day it shows the movie. We used the concept of 'seats' as a measure of a cinema capacity, and we modelled the state of a 'seat'. The maximum number of possible 'seats' available in all cinema screening the film is C. When the distribution of the movie starts all tokens are in P_1. Tokens are moved then in P_2 by the firing of T_1 at rate:

$$(4) \qquad T_1 = \begin{cases} K_1 & \text{if } n(P_4) > \theta_N n(P_1) \\ K_2 \frac{n(P_1)}{n(P_4)} & \text{otherwise} \end{cases}$$

Place P_2 represents the seats no longer available to those interested in the movie: after some weeks from the film release, less people are interested to watch it and thus more cinemas stop showing it. The rate of transition $r(T_1)$ considers two possible cases: when the number of spectators wanting to see the movie is greater than a fraction θ_N of the number of seats, the seat reduction takes place at a (small) constant rate K_1. This is used to model the fact that cinemas do not reduce (or do it at a very slow rate) their capacity when there is a large number of people still wanting to watch the movie compared to the available seats. When however less people are interested in the movie, cinemas stops showing it at a rate proportional with constant K_2 to the ratio between cinemas' seats and people still liking to see the film. Positive reinforcement can induce cinemas to increase again the number of available seats. This is modelled by transition T_0, that fire with rate $r(T_0) = \alpha_C n(P_5)^{\alpha_w}$, that is with a rate proportional to the intensity of the feedback as defined by transition T_2.

The controller in this example (Figure 2c) models the behavior of the distributor of the movie. Its purpose is to decide if the film is to be stopped from screening

through cinemas. When place P_6 is marked, the distributor is idle for W_1. This is represented by transition T_5, whose firing rate is $r(T_5) = \frac{1}{W_1}$. In place P_7 the distributor decides after a W_2 time, moving the token to place P_8 (firing of T_6 transition, with rate $r(T_6) = \frac{1}{W_2}$). The decision is performed by two immediate transitions t_1 and t_2, whose weighting is

$$(5) \qquad w(t_1) = min\left(1, \frac{K_a n(P_1) + K_b n(P_5)}{n(P_3) + n(P_4)}\right)$$

and $w(t_2) = 1 - w(t_1)$. In Equation 5, K_a and K_b are two constants that define the importance given by the decision of stopping the screening respectively to the number of 'seats' still free and to the number of people who have already watched the movie. K_a represents the importance of keeping the movie in a cinema even if there is a large number of 'free' seats compared to number of people who already watched it. K_a takes into account how many people are going to see the film using the same seat and should be proportional to the number of expected screenings per week. If there are a lot of cinemas (i.e.: high number of seats compared to the $n(P_3)+n(P_4)$) showing a film no longer requested by many people, is not convenient for a cinema to screen it. This is taken into account by K_b, that represents the importance to keep a movie in cinemas considering how many people have already seen the movie, compared to those who haven't yet. Both constants are a positive real numbers and they are scaled by a factor that indicates the reactivity of the production. In our examples we used $0 \leq K_a, K_b \leq 1$.

Figure 3, compares the distribution of the N people (interested to the movie, not interested and who have seen it) to the distribution of the C available seats during 60 days of film screening. To make the comparison possible, Mean Field solutions are normalized by dividing them by the corresponding total population. We set N, the total number of potential spectators, to 80000 people, and C to 2000 free seats for the first week. Half of the N spectators started in P_3 and half in P_4, while all C seats started in P_1. Of course the Distributor started with place P_6 marked. W_1 is 6 days and W_2 is 1 day. We assumed only 1 screening per day in each cinema, and a rate of T_1 of about 2 weeks. The film being distributed is a very good one, so α_L is higher than β_D by a 100 factor, meaning that for every 100 persons liking the movie, just 1 disliked it. Figure 3 does not take into account the decision of the Distributor. The number of people interested in watching the movie, $n(P_4)$, raises rapidly in the first week. Because many of them do watch the movie, their number decrease in the following weeks. The audience raise almost constantly, ending around over 40% more the starting percentage of interested N. The number of cinema screening the movie, on the other hand, is fated to decrease until no more seats are available. Figure 4 shows the probability that the Distributor stops the screening of the movie through the cinemas according to several values of K_a and $K_b = 0.01$ as per Equation (5). At the beginning we have a fast increase of the probability to stop the movie because of the low number of people in P_5. After the first week the stabilization of the curve means that the distributor is satisfied by the number of people watching the film. Toward the end of the curve, the rapid raising of the probability represents the fact that the number of interested people
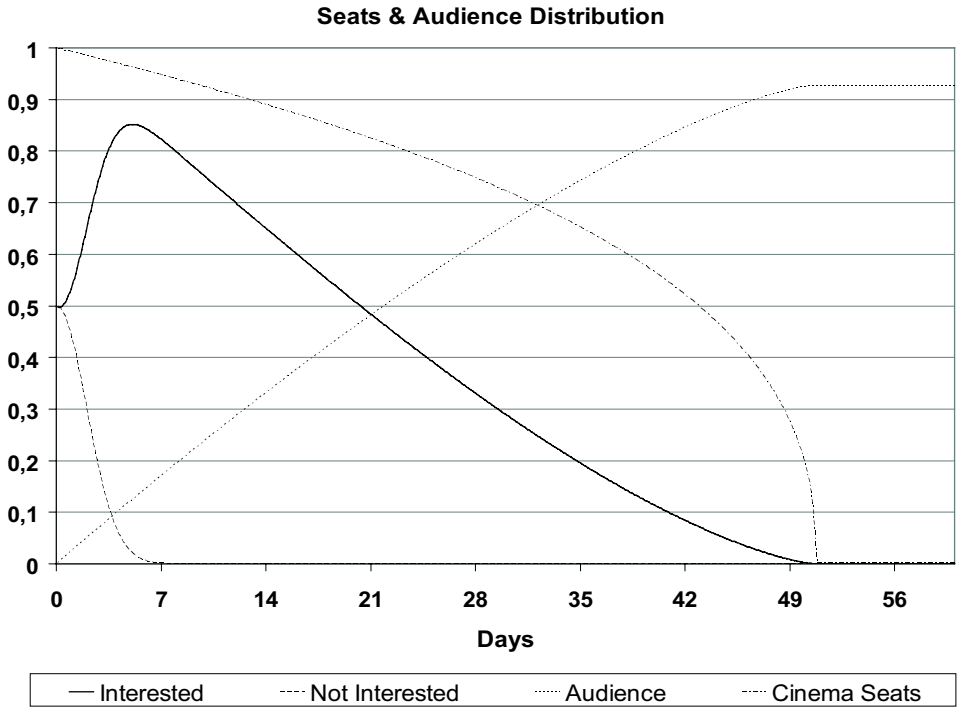
Fig. 3. Transient state distribution of the audience and of the cinema 'seats' models.

and the number of seats have radically dropped. Even if in Figure 3 we observe that 90% of people will potentially watch the movie, we have to take into account the probability that the distributor stops the screening before the deadline. To simplify the Mean Field model we have not included the feedback from the production in the spectator and cinemas components. Instead we will consider the production's decision during performance indexes computation. If we call T the random variable that represents the probability of ending the screening at a given time, we have that $T(\tau) = Pr\{T < \tau\} = \#(\tau, P_9)$. We can use this distribution to compute the distribution of the number of spectator S, noting that $\delta(s - n(\tau, P_5)) = Pr\{S = s|T = \tau\}$ (where $\delta(x)$ is the Dirac's delta function). We can then compute the mean audience $E[S]$ as:

$$(6) \qquad E[S] = \int E[n(\tau, P_5)]d\#(\tau, P_9)$$

Figure 5 shows $E[S]$ considering different values for K_a and K_b. It can be important to stop the screening when it is convenient to do it, it term of a costs - benefits ratio. In this case, the revenue is obtained by multiplying the mean audience by a factor representing the ticket cost c_s. The cost sustained by the cinemas instead is obtained by multiplying the mean number of seats needed to satisfy the demand by a constant representing an hypothetical sum expended for each seat c_t. The profit ξ of course, is obtained by subtracting costs from revenues, that is $\xi = c_s \cdot E[S] - c_t \cdot E[T]$. ξ as function of different K_a and K_b is shown in Figure 6 we show different profit
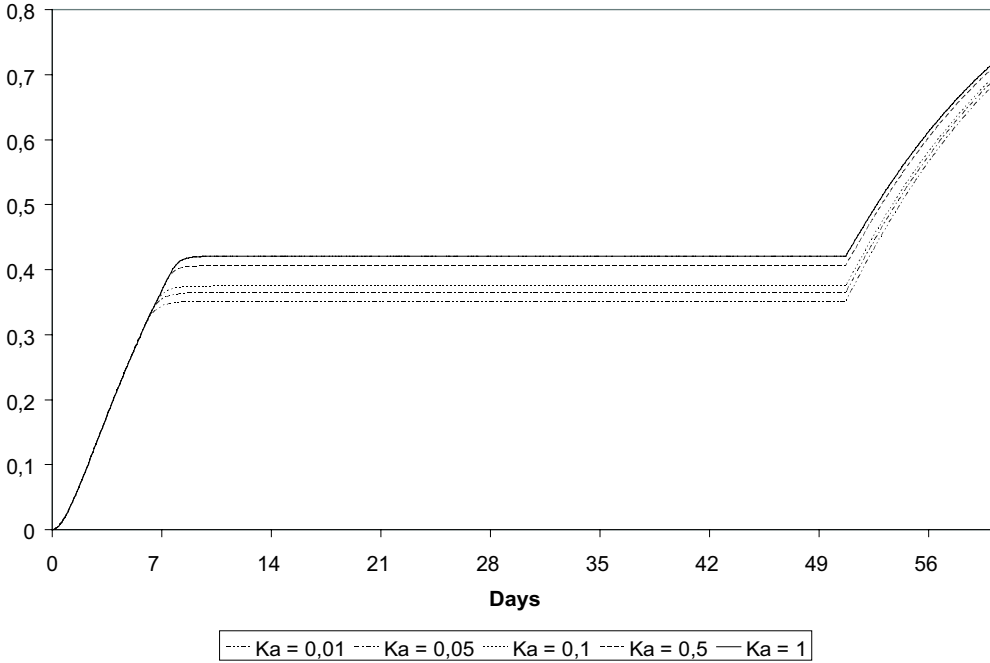
Distributor decision if Kb = 0,01

Fig. 4. Transient state distribution of the production.

levels obtained with different parameters influencing the Distributor's decisions.

3.3   Model Validation

It is very hard in practice to determine the values of the various parameters of the proposed model. However, we can compare the results that can be obtained playing with the model parameters in a 'reasonable' way with the data measured from the income of a real production. Figure 7 shows these results, comparing the three cases identified in Section 3.1 with the solution of the Mean Field model obtained choosing suitable parameters. In case A, we have set the reinforce parameters $\alpha_L = \alpha_C = 0$, and the threshold until which cinemas do not reduce the number of available seats $\theta_N = \infty$. For case B, we have lowered threshold $\theta_N = 5$, and put the reduction rate when the number of potential spectator is high at very low values $K_1 = \frac{1}{20}$ (1 seat every 20 weeks). This should model the fact that in case B cinema are saturated for the first weeks. Finally, in case C we have enabled the enforcement feedback, setting $\alpha_w = 1$, $\alpha_L = \frac{4}{N^{\alpha_w}}$ and $\alpha_C = \frac{\alpha_L}{2}$. Results have then been scaled to convert the number of spectator to the actual income. As we can see, results show some reasonable match. Some problems are visible in the beginning of the curves. We believe these are connected with the non-Markovian nature of most of the considered transitions that the exponential approximation (even if non-homogenous) is not able to capture.
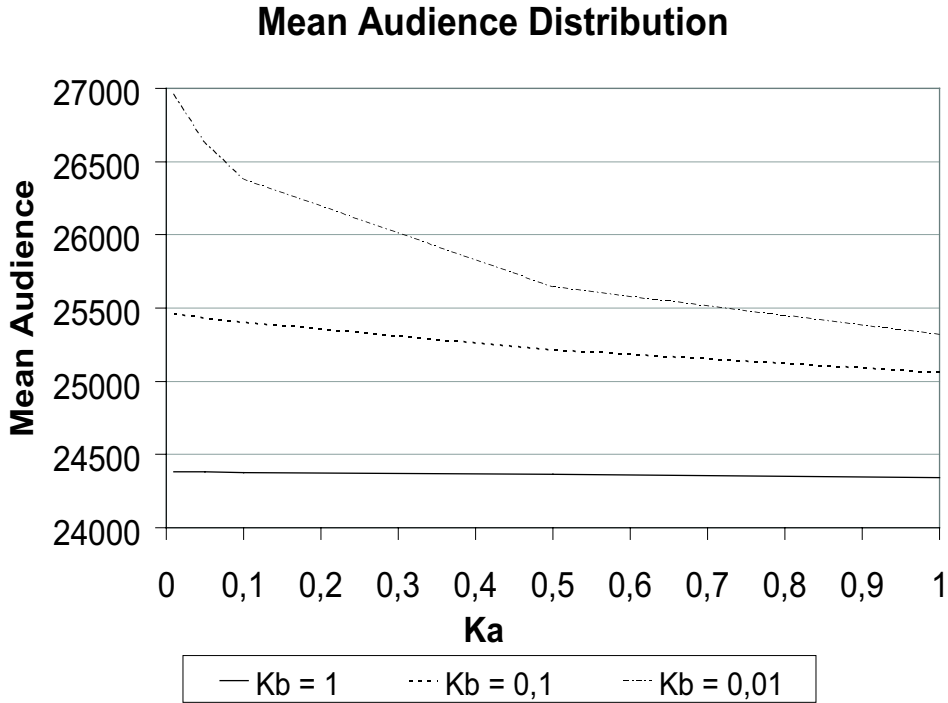
Mean Audience Distribution

Fig. 5. Distribution of the mean number of spectators as function of the production decision parameters.

4 Television production

Let us consider the analysis of the audience of a TV *series*. Every D time, a new episode is produced. In this example N spectators can decide to watch the next broadcasted episode of the series based on three factors: *quality, advertisement* and *other spectators feedback*. The production can decide after each episode, to change their investments, focusing either on improving the quality or on increasing the advertisement of the series. We measured the direction of the production with a variable α_{qL}. If $\alpha_{qL} = 0$ then the production will put all its resource on advertising. If instead $\alpha_{qL} = 1$, all the resources will be put on quality improvements. Intermediate values of α_{qL} represent an appropriate mixture of the two possible directions.

4.1 The spectator model

The behavior of one spectator is modelled by the GSPN in Figure 8a. The *Quality* of the episode, intended as a measure of how much the show contents matches with the needs/expectancies of a given spectator, is modelled by places P_1 and P_2. In particular, $n(P_2)$ represents the number of people satisfied by the Quality of the episode, while $n(P_1)$ represents those with a low matching between the show contents and their expectancies. The *Advertise Level* is the second factor: an abstract measure that quantifies effectiveness of the cost sustained by the Production to
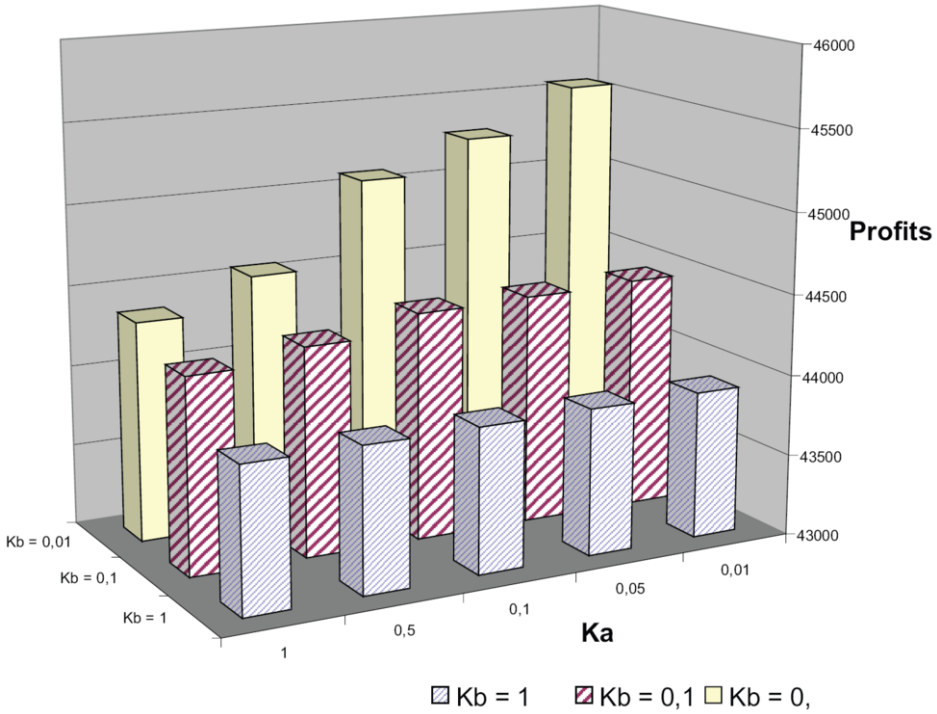
Fig. 6. Profits as function of the Distributor parameters.

convince the spectators to watch the next episode. This is modelled by places P_5 and P_6. The number of entities with the tokens in (P_6) is the number of people reached by the advertising of the episode, while tokens in (P_5) represents those not influenced by the advertising. Finally, the decision of the user to watch or not to watch the episode is modelled by places P_3 and P_4. In this way, the number of people watching the episode is $n(P_4)$, while $n(P_3)$ represents those who don't. They also represent the third factor that can influence the behavior of the spectators: the positive and the negative feedbacks. As time passes, people not watching the series might eventually exit the system. Their number is represented by $n(P_0)$.

The timed transition firing rate $r(T_1)$ is equal to $V_Q \alpha_{qL}$, where V_Q is a constant representing the speed of user decision based upon the Quality of the episodes. The firing rate of $r(T_2)$ is equal to $V_Q(1 - \alpha_{qL}Q_{min})$. The introduction of constant Q_{min} is to reflect that production will always ensure a minimum level of Quality per episode. In this way the minimum rate at which a spectator will be attracted by the Quality of the show will be $V_Q(1 - Q_{min})$. Transition T_3 and T_4, with firing rate $r(T_3) = V_{pp}\frac{n(P_3)}{N}$ and $r(T_4) = V_{pp}\frac{n(P_4)}{N}$, are both influenced by V_{pp}, a constant representing the *word-of-mouth* importance - that is the feedback from other spectators. The firing rates for transitions T_5 and T_6 are $r(T_5) = V_L(1 - \alpha_{qL})$ and $r(T_6) = V_L[\alpha_{qL}L_{min} + (1 - L_{min})]$. V_L represents the speed at which the Advertise Level influences user decisions. As for the Quality, production will always ensures a minimum Advertise level for each episode and we expressed this
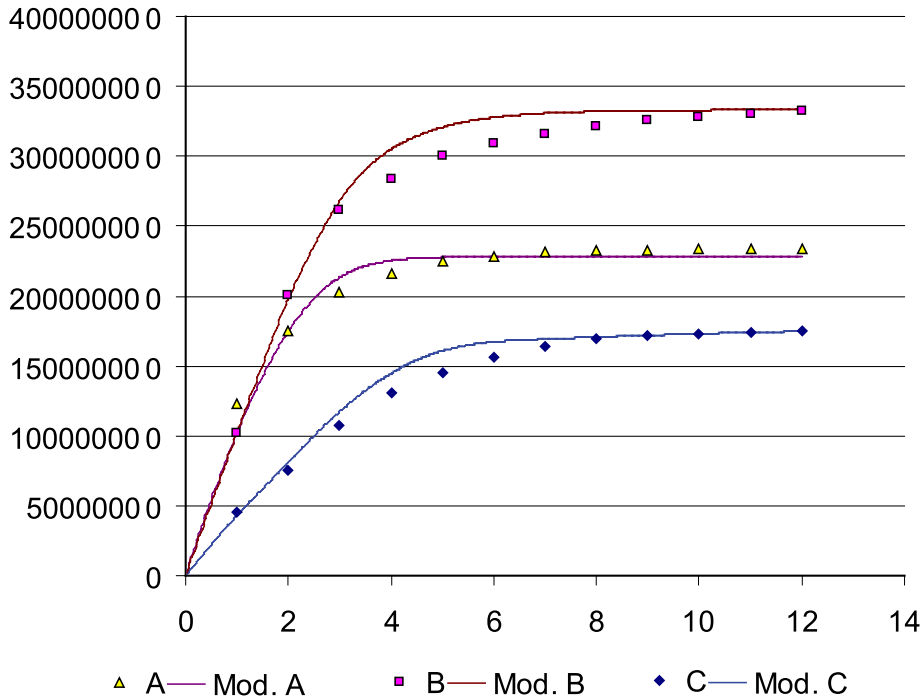
Fig. 7.   Validation of cinematographic Data: A = "X-Men: The Last Stand ", B = "Lord of the Rings: The Two Towers", C = "Mission Impossible I"

by the constant L_{min}. The firing rate of transition T_0 is equal to a constant V_{dis} representing the speed of the user disaffection to the show.

4.2   The production model

The goal of the Production is to maximize the number of spectators of each single episode constituting the series $n(P_4)$. The behavior of the Production is modelled by the DSPN (Deterministic Stochastic Petri Net) depicted in Figure 8b. The Production can allocate its resources to improve either the Quality or the Advertising level of the show. The marking of Place P_7 represents the number of resources allocated to the production's Quality, while the marking of Place P_8 the resources allocated to Advertising. They are initially set respectively to Q and L tokens. When the number of entities with P_1 marked falls below a given thresholds Q_{th}, the production at the beginning of the next episode (after a deterministic time D) decides to transfer some resources from Advertising to Quality (firing of transition t_4). Conversely, if the $n(P_5)$ falls below a threshold L_{th}, the production transfers resources from Quality to Advertising (firing of transition t_5). In case both levels fall below the threshold, a *policy* to resolve situation is established. This policy simply gives priority to one of the two levels. The total number of resources remains constant (that is $\#(P_7) + \#(P_8) = Q + L$), and t_6 represents the case in which no
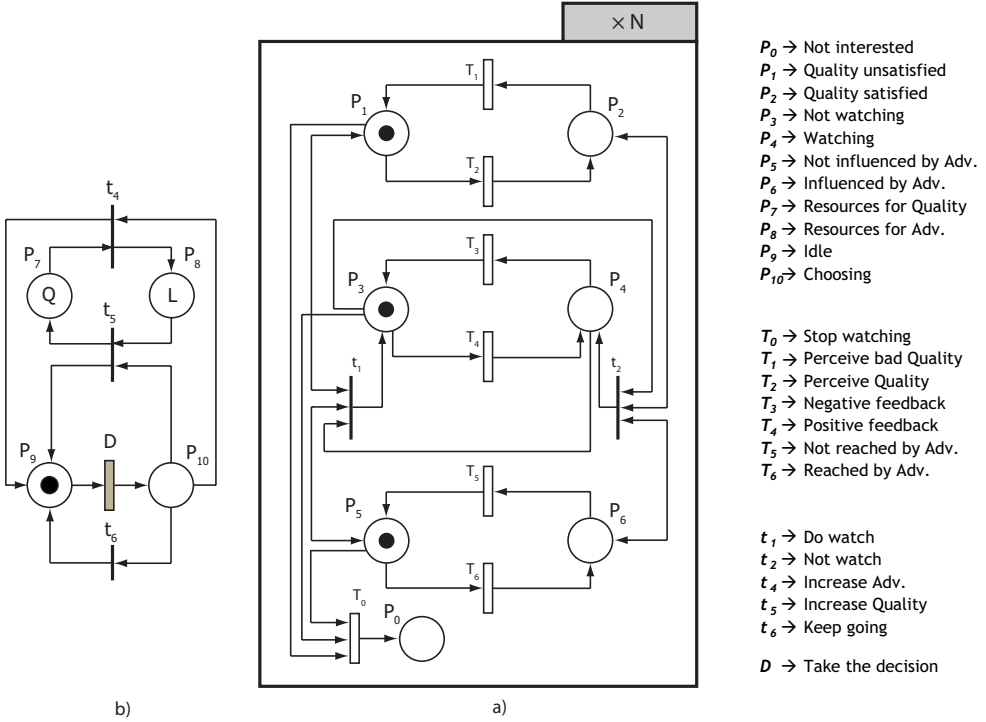
Fig. 8. A model of a television production and N TV-show spectators

Priority	Quality	Advertisement
$w(t_4)$	$\mathbf{1}(n(P_2) \leq Q_{th})$	$\mathbf{1}(n(P_6) > L_{th} \wedge$ $n(P_2) \leq Q_{th})$
$w(t_5)$	$\mathbf{1}(n(P_6) > L_{th} \wedge$ $n(P_2) > Q_{th})$	$\mathbf{1}(n(P_6) > L_{th} \wedge$ $n(P_2) > Q_{th})$
$w(t_5)$	$\mathbf{1}(n(P_6) \leq L_{th} \wedge$ $n(P_2) > Q_{th})$	$\mathbf{1}(n(P_6) \leq L_{th})$

Table 2
Weights of immediate transition of the model in Figure 8

change to resource allocation is needed. Table 2 shows how the weight of t_1, t_2 and t_3 changes according to the chosen policy. The level of resources α_{qL} (used to define the transition rate of the transitions modelling the spectators) is defined as:

$$(7) \qquad \alpha_{qL} = \frac{n(P_8)}{n(P_7) + n(P_8)}$$

4.3   Model analysis

The controller of the model proposed in Figure 8 has a deterministic clock of duration D. This could also be implemented as a Measure for the corresponding Mean Field Model. However, to simplify the presentation and avoid the introduction of Dirac's delta functional rate, we propose a slightly different approach. If we call $\boldsymbol{\pi}_u$ the vector that describes the state of the controller when $u \cdot D \leq \tau < (u+1) \cdot D$,

and with $\mathbf{C}(\mathbf{N}(\tau)) = |c_{ij}(\mathbf{N}(\tau))|$ a transition Matrix that encodes the evolution of $\boldsymbol{\pi}_u$ when the deterministic transition fires (that is $c_{ij}(\mathbf{N}(\tau))$ is the probability of jumping from state i to state j conditioned to the state of the Mean Field Part of the model $\mathbf{N}(\tau)$), we can compute the evolution of the complete process by solving:

$$(8) \qquad \begin{cases} \boldsymbol{\pi}_u = \boldsymbol{\pi}_{u-1}\mathbf{C}(\mathbf{N}(\tau)) & \tau = u \cdot D \\ \dfrac{d\mathbf{N}(\tau)}{d\tau} = \mathbf{N}(\tau)K(\mathbf{N}(\tau), \boldsymbol{\pi}_u) & \tau \neq u \cdot D \end{cases}$$

Note that $\boldsymbol{\pi}_u$ is an approximation to the vector of the state probabilities of the DSPN at time u.

In Figure 9 and 10 we presented the evolution of a system for the two considered policies evaluated with Mean Field analysis. The Production policy in Figure 9 is aimed toward maximizing the effects of Advertising, while in the second the Quality is maximized. In both cases we assumed a TV series composed by 25 episodes broadcasted every 7 days. The potential audience is made of 20000 people. The minimum Quality level parameter is $Q_{min} = 0.8$, while the minimum Advertise level parameter is $L_{min} = 0.92$. The thresholds Q_{th} and L_{th} are set, respectively, to 0.65 and 0.45. We set $Q + L = 10$ levels to represent the current state of the production decision. Parameter α_{qL} can be determined by the marking of place P_8 (that is L), ranging from 0 (Production allocating every possible resource in Quality improvement) to 10 (Production allocating every possible resource in Advertise improvement). We also assumed that about every 6 days people decide if their Quality expectancy can be satisfied by the next episode ($V_Q = 6days$), about every 3 days the advertising of the show can influence their decision ($V_L = 3days$) and about every day the word-of-mouth influence can persuade them ($V_{pp} = 1day$). After the 60% of the total number of episodes, a given spectator is presumed to became disaffected to the show.

We called $G = n(P_3)$ the number of people who actually watched the show. Please note that the curves labelled "Decision" in both figures are step functions since they represent the state of the television production that is a discreet time stochastic process. In particular this curve is proportional to the marking of place P_8 and it rises if the production aims to Advertise while it lowers if the focus is on Quality.

For the plot in Figure 10 the initial production decision level (marking of place P_8) is equal to 2, meaning a strong interest in promoting the Quality of the show. For the plot in Figure 9 the initial level (place P_8) is set to 8, because the interest is in promoting the show using the Advertise. In both cases we normalized the curve representing the decision level to fit in the graph with the other curves.

In Figure 10 the initial Quality level is moved toward the Advertise investment because of the great number of spectators already attracted by the quality. When the reduced Quality starts to become a problem in terms of Audience, the change in decision level counters this tendency. Toward the end of the series, given the fact that an high number of N ended in P_0, the Audience seems to have a rising bent. The audience curve of Figure 9 doesn't show the same leaning, suggesting that an
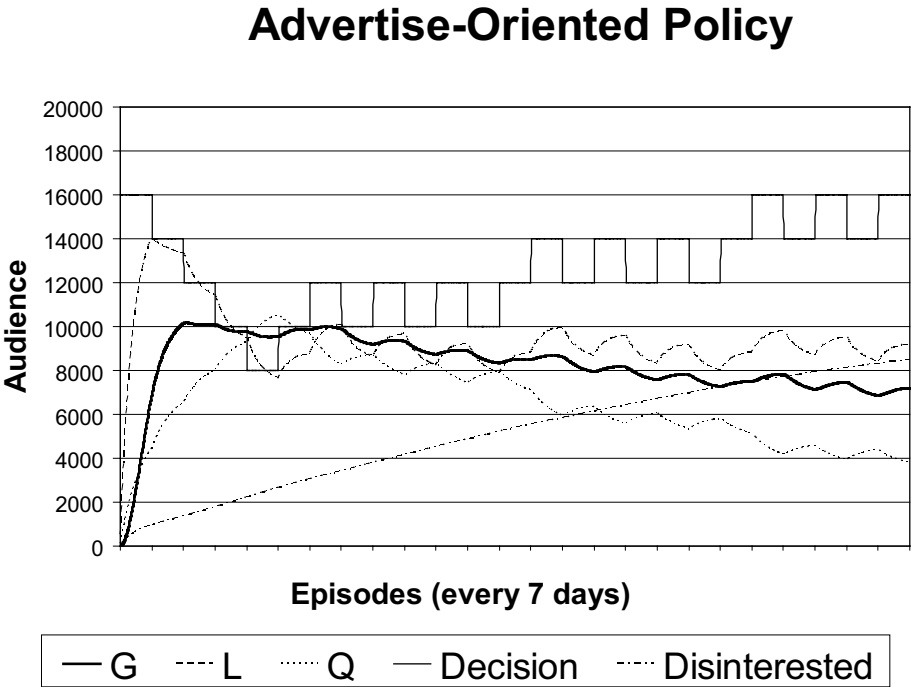
Advertise-Oriented Policy

Fig. 9. Audience and system behavior when priority is given to advertizement.

investment in quality is more convenient in the long period.

4.4 Model Validation

As a validation for our first model, we used audience data taken from the web site[1] of Mediaset, the second greatest TV broadcaster group in Italy. The statistical data displayed on this site were gathered by Auditel, the leading Italian survey agency for TV. The Mediaset site displays daily the audience and share of the group's three channels, divided by different time slots. The age of people involved in the survey is between 15 and 64 years old, permitting a number of different approaches to the use of the data. No gender distinctions were also taken in account.

The data considered in our validation are those showing the audience in the month of February 2009, for Canale 5, the main channel of the Mediaset group. It was impossible for us to collect data for every day, given the fact that their publication on the site is suspended on Sundays.

We choose to focus on the 6.00pm to 8.30pm time frame, just before the starting of the primetime in Italy[2]. Being the primetime a period of strong competition between networks[11][12], the audience data are less dependent on the outcome of a single channel.

In Figure 11 the curve representing the audience evolution of the dataset (labelled "Data") is compared with the curves representing the audience obtained from
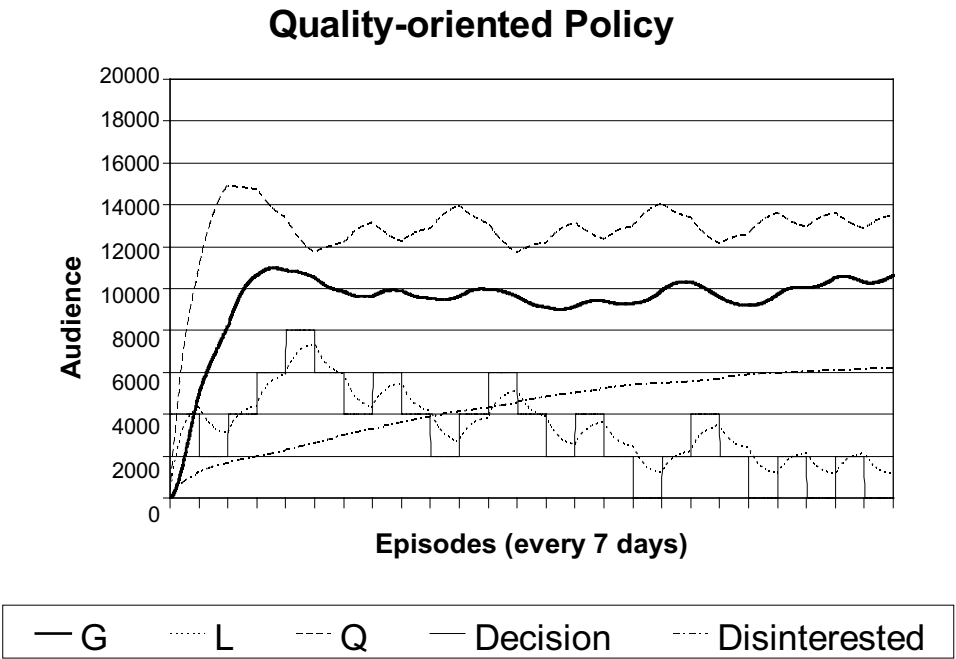
Quality-oriented Policy

Fig. 10. Audience and system behavior when priority is given to quality.

our model, when priority is given either to quality (G_q) or advertisement (G_l). The results provided in the previous section have been adapted to take into account days instead of weeks. In both cases we assumed an evolution time for the system of 28 days.

The curves Data and G in Figure 11 are very similar in tendency, suggesting that our system models the situation properly, if set on the Quality policy. This fact can be explained probably because [8] in that particular time frame people are more likely to watch television attracted by programs not strongly advertised, but matching mainly on their expectations. TV networks prefers to invest strongly in advertising those programs taking place during the prime time when the competition for audience is hardest to win.

5 Conclusions

In this paper we have used Mean Field Analysis to study the behavior of the audience of a movie, or of a television series. Measures defined for Mean Field models have been used in both cases to encode the behavior of a controller acting on the distribution or on the production of the content. Although a proper validation would have been really hard to obtain and outside the scope of this work, we have shown that the model's results are in line with measured data of the audience for both kind of media.
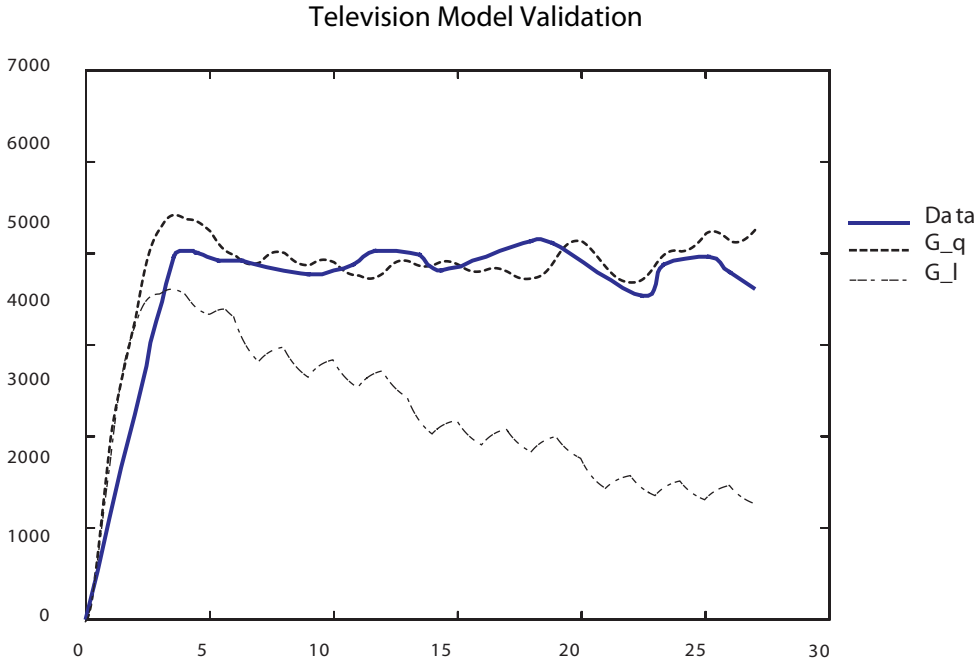
Fig. 11.   Validation of television Data.

References

[1] The *mediaset* web site. http://www.tv.mediaset.it/auditel/ascolti.shtml.

[2] The *prime time* definition on *wikipedia*. http://en.wikipedia.org/wiki/Prime_time.

[3] The *swivel* web site. http://www.swivel.com/data_sets/spreadsheet/1006807.

[4] M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis. *Modelling with Generalized Stochastic Petri Nets*. John Wiley & Sons, 1995.

[5] M. Benaim and J.-Y. Le Boudec. A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 65(11-12):823–838, 2008.

[6] A. Bobbio, M. Gribaudo, and M. Telek. Analysis of large scale interacting systems by mean field method. In *5th International Conference on Quantitative Evaluation of Systems - QEST2008*, St. Malo, 2008.

[7] J.Y. Le Boudec, D. McDonald, and J. Mundinger. A generic mean field convergence result for systems of interacting objects. In *4th International Conference on Quantitative Evaluation of Systems - QEST2007*, Edinburgh, 2007.

[8] F. Di Chio and G. P. Parenti. *Manuale del telespettatore*. Bompiani, Milano, 2008.

[9] M. Gribaudo, D. Cerotti, and A. Bobbio. Analysis of on-off policies in sensor networks using interacting markovian agents. In *4th International Workshop on Sensor Networks and Systems for Pervasive Computing - PerSens 2008*, Hong Kong, 2008.

[10] J. Hillston. Fluid flow approximation of pepa models. In *2nd International Conference on Quantitative Evaluation of Systems - QEST2005*, Turin, 2005.

[11] V.A. *Link Magazine.* RTI, Milano, 2002.

[12] V.A. *Link Magazine.* RTI, Milano, 2005.