

## LIFE-CYCLE CONSIDERATIONS FOR WIDE-AREA DISTRIBUTION OF MULTIMEDIA DATA

Carsten Griwodz<sup>1</sup> and Ralf Steinmetz<sup>2</sup>

<sup>1</sup>*Darmstadt University of Technology, Darmstadt, Germany*

<sup>2</sup>*GMD IPSI, Darmstadt, Germany*

**Abstract:** Various applications for world-wide multimedia object distribution arise currently as more and more companies integrate streamed continuous media elements in their web pages and as TV broadcasters add computer programs and data to their video streams. Currently, the streams that can be found on the web have only very small data volume, but as network capacities increase, the data volume that is used by applications will also be increased. TV applications have a lack of networking support and are unable to give feedback to the sender. In the future the integration of computer networks, telephony and broadcast networks will achieve a functional integration of these two worlds.

But as the amount of continuous media data which is offered for everybody to retrieve increases, simple caching mechanisms that work currently well will not be appropriate any more. To address this danger, the introduction of distribution mechanisms which provide more coordination among caches is necessary. As a basis for such a development, we investigate the probability of movies to be requested. This is chosen as a starting point because the current simple caching mechanisms operate very successfully for small-volume data such as text or icons, and high-quality movies are currently the most attractive medium which has the very high requirements in terms of storage as well as bandwidth. In this paper, we review current models on access probabilities and we give strong indication that the current assumptions about hit rates are invalid if the number of users connecting to an individual server differ widely. Since on-demand retrieval of large continuous media streams (especially movies) is currently available only as music CD and movie rental or sale, we take that data as a basis to create a model for access probabilities on this most storage- and bandwidth-demanding multimedia data.

## 1 INTRODUCTION

Although the amount of data for the distributed applications of today is relatively low because the focus is still on information transfer rather than information presentation, this will change as the commercial competition for customers on the network grows.

The commercial web pages have been designed to be user friendly and attractive for some time now, but some consideration is still given to network bandwidth. Typical active elements are currently introduced to web pages by animated GIFs or Java programs, and on Windows machines also by ActiveX. Further multimedia support for video and audio is already offered by server software and hardware vendors, but functional, decentralized distribution systems for this kind do not exist. The typical solution to round-trip or bandwidth problems today is the installation of mirror servers, from which the customer can select.

Among the distributed applications with a growing importance for the real-world are the digital libraries; they provide restricted on-line access to large collections of data. The full services of a digital library are typically only accessible to a limited group of users, which makes issues like data security, the management of access rights and the support of ownership rights an important issues for the provider. Correctness and originality of the data is important for the customer. Current digital library installations have a central administration and a single access point. The data is delivered in proprietary formats directly to the customer, which is a partial solution of the above problems. When digital libraries become accessible on a world-wide scale to large numbers of customer, for example for medical or educational use, the direct connection between the library and the individual customer becomes inefficient. A caching mechanism similar to web caches is necessary, but the additional requirements of digital libraries must be taken into account. For seekable continuous media material, the same caching problems mentioned with the web have to be solved.

Another application which had drawn a lot of attention for some time and which is currently declining from popularity as expensive early implementations become available is the video-on-demand. Its popularity may not increase for quite some time since the restrictions of the existing broadcast infrastructure such as a very limited number of streams in the distribution backbones (in this context I consider 200 or 500 streams very limited) will remain sufficient for quite a while. Broadcasters will be able to add services in sufficient number before they have to consider personal video-on-demand services. New services that are just slightly more personalized but a lot more expensive due to the technical effort are not accepted by the customers when the basic service can be obtained with hardly any costs. For a success, operations like VCR controls and real video-on-demand (as opposed to near video-on-demand) must be provided, probably on an international scale. Because of the various drawbacks of a centralized infrastructure, such as stability problems of the central system and the immense networking costs, a distribution system that comprises various intermediate nodes which act as caches would be affordable earlier. Although such a distribution

system is probably administered centrally and intermediate nodes of the distribution network are controlled by a single entity, distribution mechanisms must be found that are not based on local decisions.

The integration of data services into television services is demonstrated by various broadcasters using the Intericast approach. Although the data transfer is unidirectional and the television service can not be manipulated based on the downloaded data, this is another indication that the broadcasters are not waiting until their services are integrated into computer networks but that they work towards an integration, too. As all applications grow together and mix, the amount of bandwidth that is used up by media contents increases; simple caching mechanisms that work well currently will not be appropriate any more. Especially the caching of movies on every first level cache server through which a user accesses the data is a problem. Either it takes up large amounts of space on the server or the use of a cache does not result in a decrease of network load because the largest objects are not cached at all.

Web caches apply various algorithms which decide on the basis of strictly local information whether a cached object should be kept in the cache or whether it should be removed. Some of these caches are quite elaborate and even prepared for video caching. The simplest decision that is made in each web cache, however, is whether an object's size is below an administrator-defined limit. If it is not, the caching algorithms are simply not applied; the object is never cached. This limit may be considerably higher in second-level or third-level web cache servers. Very large objects, for example movies, are not even cached by these servers. If network and disk space costs continue to decrease at the current rate, even the caching of video clips on the basis of local, static and uncoordinated decisions may become affordable. Still, there will always be objects large enough to make this mechanism unaffordable.

One of the ways to change this individuality is the creation of a designed network with a central administration or at least central administration rules. However, the most extensive collection of computer networks today, the Internet, as well as the collection of all telephone networks, radio or broadcast networks, seem to be proof that a central authority is not to be expected. An ongoing development of the Internet, to which new networks connect and from which other networks disconnect continuously is more likely to remain the reality for the foreseeable future. This implies a few things: First of all, solutions to networking problems must be open for implementation in all connected networks, or freely available. That is not necessarily so, but large parts of the user community may be blocked out otherwise. To protect the users as well as the providers in such networks, copyright and security issues need generic solutions that are not inhibited by the fact that transferred data can be captured in uncontrollable and unintended portions of the network. This means that implementors of applications and data transfer protocols must take into account that for communication between sender and receiver data theft is a reality and that they must not be inhibited by it. The basic reason for the problem is that access to data must be made by the customer through networks that are owned by a third party. It is not necessarily known whether this third

party has a friendly position towards user and provider. In the Internet of today, it is mostly assumed that this is the case. Finally, this heterogeneity implies that network capabilities differ; for the use of video, however, caching must still be possible and handled automatically by transparent mechanisms of the network.

## 2 RELATED WORK

The pragmatism of the Internet shows in the approach taken to address the bandwidth problems which inhibit the use of streamed video: for off-line video, providers are looking for mirror sites. Video caching in web cache servers has been implemented both for research and commercial purposes and a protocol for VCR operations has been implemented by Holfelder ([8]).

In the video server world, the issue has rarely been addressed because server technology is still a sufficiently big problems on its own directly. The idea, however, that cost control can only be done by widely distributed video-on-demand system rather than central-server systems is not new. Among the first to mention that intermediate servers in a hierarchical VoD distribution system can decrease cost are Nussbaumer et al. [11], who provided also simple arguments for this.

The radical approaches towards the establishment of video distribution systems are the server cost-optimizing approach and the network cost-optimizing approach. The lowest server cost is achieved by installing a central server which stores all available movies and requires redundant copies of a movie only because of its limited disk I/O bandwidth. In [13], it is mentioned as a side-note that this approach may finally prove applicable when sufficient user requests are generated that the use of memory instead of disks is feasible. The lowest network cost is achieved by storing all videos at the users' sites. This is obviously never feasible due to storage capacity constraints.

Differences in the access probabilities to individual movies are the factor which makes the positioning of movies in a distribution system an issue. Various approaches towards modeling the probability of access to a specific movie on a video server have been proposed in the literature. Usually, there are no simulation models which consider how much interaction might take place from the point of the user. Rather, an analysis is performed to derive the worst case situation that a server (or network) can cope with. One approach is the modeling of single video streams as they are accessed and played. This is generally done in order to understand how the operation of a single machine or cluster of machines can be optimized. Little and Venkatesh take this approach in [10] with the goal of optimizing disk I/O operations in a single system. Their approach is to build an analytical model for access probabilities based on the work by Ramarao and Ramamoorthy [12]. In [14], Tewari et al. optimize the I/O utilization in a server cluster and use Poisson processes to model the user accesses to the server, with the mean value chosen according to Little's Law [9]. Golubchik et al. investigate in [7] means for sharing video streams in a video-on-demand system that

holds when VCR controls are permitted to the user. Their user model is analytic and assumes a Poisson arrival process. Their analysis does not need long-term movie development.

These approaches are useful for optimizing playout, stacks or disk operations in a system. They do not take user interaction into account once that a movie is playing, and they do not try to model the play time of a single clip in any way.

The modeling of VCR commands requires a model that includes modeling of user habits in applying these VCR operations. Dan and Sitaram analyze in [6] the caching of data in a single server or server cluster under various interactive workloads and models information such as access skew (the distribution of requests on stored clips), the clip length distribution and the viewing time. While they do not consider the aging of individual movies, because they still need a short-term model, they consider the distribution of hits on the available videos and choose the Zipf distribution to model the video popularity. Nussbaumer, aiming at optimizing the cache usage in a single server system or cluster, also assumes the Zipf distribution ([4]) to model video popularity ([11]). The distribution of movies or blocks of single movies over multiple machines for load-levelling or availability purposes has been investigated in [3].

Barnett et al. ([2]) aim at minimizing storage costs in a distributed system and apply caching mechanisms to do this. This requires the kind of long-term analysis we also discuss in this paper. They base their considerations for long-term popularity (in the absence of freely available video-on-demand trial results) on numbers from CD sales. The model they derive is a double exponential curve for the distribution of user accesses on movies and a movie popularity development with only one rising and one declining side. They evaluate various caching strategies. From our findings based on movie rentals, their approach is not able to model the user behavior in video-on-demand systems correctly. Data derived from a video rental store and from video rental magazines which are compared in Figure 3 show that such smoothness in popularity changes is only achieved for large a number of users connecting to a system.

In [5], Dan and Sitaram try to add to the Zipf distribution a long-term dynamic change in time. They take into account that the distribution of movie rentals at a certain date can be approximated by the Zipf distribution but that the ranking among the movies is changing from one day to the next. Their approach for taking this into account is the ordering of the movies. The largest index of a movie is equivalent to the largest index in the Zipf distribution function and the largest probability assignment to this movie. For each day, they apply a permutation function to this ordered list.

In the remainder of this article we discuss the Zipf distribution, the models of Dan and Sitaram, the single-peak model introduced by Barnett et al. ([2]), the experiences we made with our first model and finally, we present our new model.

### 3 INVESTIGATION

We want to find viable approaches to movie distribution and caching in a distributed system that spans a wide area. The investigations we are aiming at comprise the travelling of movies in the system and not only the optimal placement at a certain point in time. Thus, we can not simply subscribe to the *generic possibilities* for movie movement. Instead, we want to model long-term popularity of movies, which might allow for more efficient travelling of movies. We call these changes of movie popularity in time the *life-cycle* of the movie.

We assume that the number of users connecting to a single cache server has an influence on the hit probability of a movie. To verify this, we need to model user interaction on the level of individual users and single movies. Making an analysis and checking the behavior of a system model on this level of detail in the beginning may allow the replacement of the individual user and movie model to a more general one in more complex systems. As a starting point, however, it is necessary to have an accurate model to find inaccuracies in the considerations on the system model.

In order to understand this, we look at various sources of information including but not limited to earlier literature. In extension of the earlier results, we give evidence that the day-to-day changes in rental probability of specific movies results in major exchanges of movie titles in a cache-based algorithm which assumes that low-load hours can be used to update movies according to the findings of the previous high-load period. Thus, we propose a simulation model which takes into account the hit rate changes of movies located in systems with few customers.

#### *Approach*

We want to find out how movie caching in a wide-area distribution network can be done. On the one hand, we need to understand how the probability of a movie to be requested changes. On the other hand, the analysis requires that we model user interaction on the level of individual users. This is necessary because the individual user's request behavior does not directly reflect the development of a movie's overall popularity. Thus, we can not assume how strongly the actual hit number for a movie residing in a cache which serves a small number of users deviates from the overall hit probability of that movie. Making an analysis and checking the behavior of a system model first on this level of detail may allow the replacement of the individual user and movie model to a more general one in more complex systems. As a starting point, however, it is necessary to have a most accurate model to find inaccuracies in the considerations on the system model. To come to a better understanding of this, we consider real-world information as well as proposals from the literature.

*Applicable Real-world Data*

To collect data for an investigation of on-demand movies, we have looked for information from the following sources:

- Reports from video-on-demand trials  
Unfortunately, there are no reports on video-on-demand field trials publicly available that we know of.
- Cinema visitor numbers  
The numbers of cinema visitors provide a source of information to model the initial user interest in new movies. This initial user interest could not be derived from movie rentals (see below) in the same quality because the number of copies for rental is limited. On the other hand the number of cinema visitors are not suitable for long-term modeling. One reason is that only the most favorite movies remain in the cinemas for a long time, and another is that only a limited number of movies is shown in the cinemas at a certain time.
- Movie magazines  
Many of these magazines publish a weekly ranking, e.g. a top10 ranking of the last week's TV, cinema or rental movies. Among the few that provide more information is in Germany the Video Woche [15], which publishes weekly the top 100 movie rentals based upon the Media Control Index. Video Woche is a monthly paper that provides the top 100 list of movie rentals from movie rental stores in Germany. It provides qualitative information on the relative gain and loss in attractiveness of a single movie, as well as quantitative information by giving the order of movie popularity by the chart of the top 100 rentals itself. The Media Control Index is calculated from the rentals of about 1000 movie rental stores nationwide. In that way regional differences are hidden and a broad data base is used. This index has a major disadvantage: The ranking function is not publicly known, there is no guarantee that two movies with the same index value have the same number of rentals. Additionally, since only one data point per movie and week is available, smoothing effects by time-averaging are quite strong. While the Video Woche data provides no information on the user behavior, it is a good indication of movie development. As an example, the development of the movie "Highlander 3" is given in Figure 1 (note that the lowest rank is best in the right figure).
- Movie rental store numbers  
Examining the rental numbers of a single rental store provides information about the probability of access of few users (some thousand) to a large base of movies. It considers completely unpopular movies as well as long-runners, single-day highlights, and all of these mixed with brand new titles. This accounts for a realistic mixture of movies in various states of their developments, and for a realistic pattern of access to the individual titles as it may be the case in an on-demand system. The biggest problem with these numbers is that only popular movies which are available

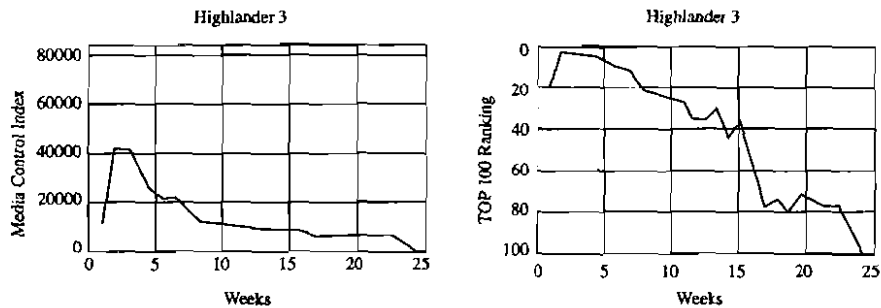


Figure 1: Data from the VideoWoche

in multiple copies are useful for understanding the user behavior. Each copy of a movie on its own has a very unclear development of rentals/day. An additional problem is the limit of the number of rentals that are possible. If a requested movie is not available, a less popular title is rented; this reduces the validity of a comparison with electronic versions.

As the basis of this study, information from the VideoWoche and from a single movie rental store were available.

The movie rental store was one out of a large list of stores that were asked for information and refused to release it. This information did of course not contain customer data, but the indexing scheme allowed to extract the association of the exact dates of rental and return of a movie with its name, it allowed to understand the number of copies of a movie that was available at any time, and the age of the movie itself. The data was sporadically corrupted with incorrect book-keeping on destroyed or missing copies of movies, as well as the chance error in return booking.

To verify the consistency of the two data sources, we checked whether the magazine data and the rental store data have an applicable relation. By the magazine's content we are limited to the movies that are considered the top 100 nation-wide. Figure 2 shows the comparison of the data from VideoWoche magazine (bottom) and the curves of the rental store (top) for the movie "Lion King". We have chosen this movie as an example because it was the one that experienced the largest number of accesses to itself of those movies that remained in the top 100 list of VideoWoche for several weeks during the observation period. If movies with a lower number of accesses to them had been chosen, the similarity in trends with the magazine could have been illustrated only by smoothing the rentals, e.g. by showing the average rentals in three days. By selecting this movie, the similarities become visible without any smoothing. We see the smoothing effects of a large number of users. We conclude from the qualitative similarity of such curves that the co-consideration of these two sources to build a single model would be acceptable.



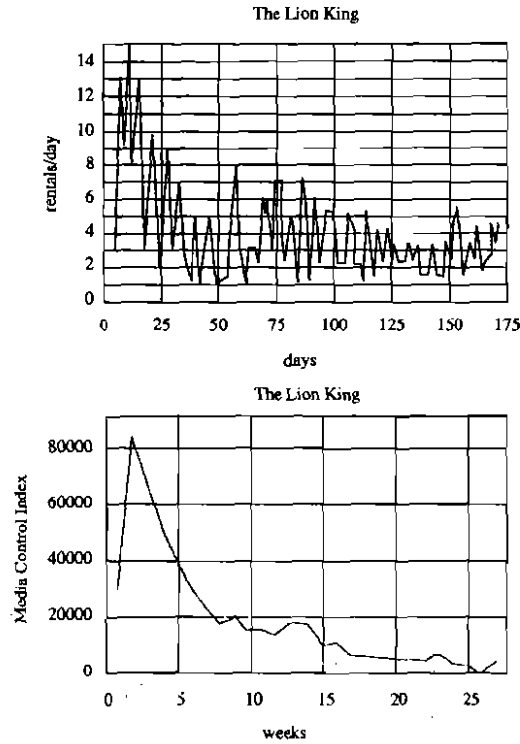


Figure 2: Comparison of rental store (top) and magazine numbers (bottom)

### Literature Discussion

**Zipf Distribution.** Figure 2 demonstrates that the access patterns generated by a small number of users can differ widely from the average user's behavior. The following shows that this affects the usefulness of the Zipf distribution for estimating hit probabilities. The Zipf distribution is defined as

$$z(i) = \frac{C}{i} \quad , C = 1 / \left( \sum_{i=1}^N \frac{1}{i} \right)$$

In this formula,  $N$  is the number of available movie titles.  $i$  is the index of a movie title in the list of  $N$  movies that are sorted in the order of decreasing popularity. It is noteworthy that this distribution, which is typically used as the basis for investigations

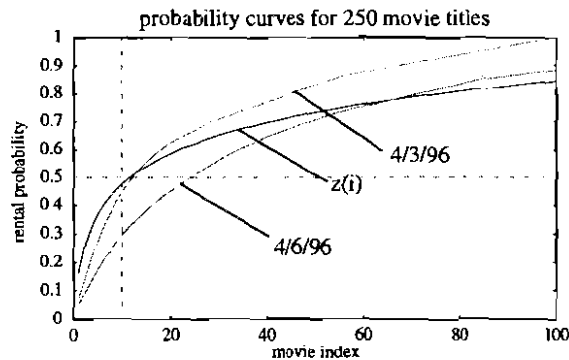


Figure 3: Rental probabilities compared with the Zipf distribution

on video server operations, is completely independent of the number of users that access the set of movies.

To verify the applicability of the distribution to our data, we compare all days of one month in the period covered by the movie data and sort 250 representative movie titles by their popularity at these days. The resulting data is compared with the Zipf distribution for  $N=250$ . Figure 3 displays the first 100 entries of the resulting curves. It presents the curves for the two days with the lowest and highest hit rate on the top 10 movies in one month in comparison with the Zipf distribution for the same number of movies.

It becomes visible that the Zipf distribution, although quite similar to the actual rental probabilities, is somewhat optimistic for small user populations. The upper curve that was derived in this month may be restricted by the number of copies available in the rental store. The lower curve is not affected by this but shows that the diversity in user selection is wider than accounted for by the theoretical function.

This implies that caching algorithms that are designed under the assumption that the Zipf distribution provides a worst-case or at least an average-case boundary for movie hit probability may underestimate the number of cache misses in a server for a small number of users.

**Music CD sales.** Besides the problem of the Zipf distribution that is mentioned above, it is also important that it is a distribution that does not consider temporal change (we call it a static distribution). Consequently, it can not be used to model the life cycle of movies. It has already been shown in [2] that the Zipf distribution in itself is not well suited to simulate real data. Because of this, it is not applicable to investigations that consider temporal changes. Barnett, Anido and Beadle propose a video rental model based on music CD sales. By evaluating these numbers, a life cycle for the rental probabilities of a movie is introduced. The authors' model based on this data assumes that each life cycle has a growth phase and a decline phase. The life cycle of a randomly

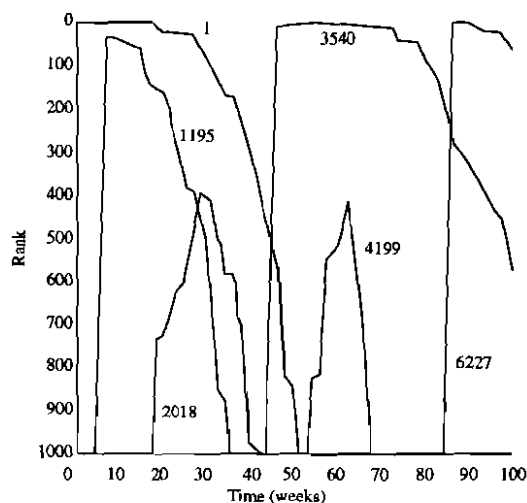


Figure 4: Movie life cycles according to the music CD model

generated movie life cycle is determined by two parameters which control the amount of relevance change and the probability of changing from the growth phase to the decline phase, respectively. Figure 4 shows life cycles of movies that are generated according to this model.

**Rotation model.** Another approach to compensate for the static behavior of the Zipf distribution was taken by Dan and Sitaram [5]. They have created a model based on a modified rotation of movie rental probabilities. Each individual movie title is given an index number, and its current popularity is calculated by the Zipf distribution. After a fixed amount of time, new indices are calculated by rotating the indices. To reduce large jumps in the relevance of a specific movie, the left half of the movie indices is swapped before and after the rotation. By adding the rotation, they try to simulate dynamic changes in the rental probability of individual titles for the cache of a single server. They did not try to create this model for long-term variations.

When movies are sorted according to their relevance, the graphs in Figure 5 display the size of relevance changes from one measurement to the next. Graph 1 in Figure 5 shows this for the rotation model, assuming 150 movies. Obviously, it does not provide sufficient realism for long-term considerations.

Another simple approach would be the permutation of movies' relevances between days. Graph 2 shows the relevance changes for a system with 150 movies that assumes a daily permutation.

In comparison to these two models, graphs 3 and 4 provide two examples of relevance changes for real movies from a movie rental store with a small number of

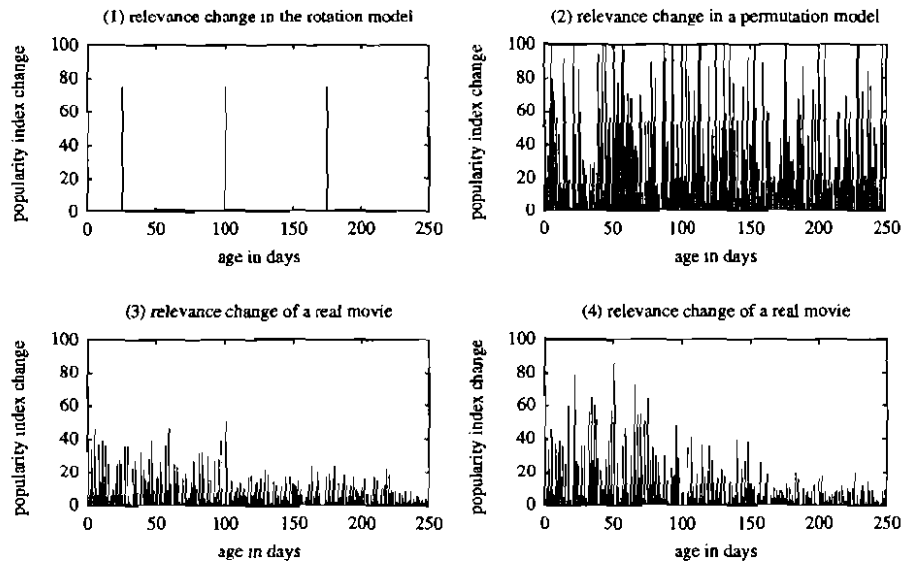


Figure 5: Day-to-day relevance change measurements

customers. The comparison demonstrates that an algorithm which calculates the location of movies' copies in a distribution tree with respect to relevance can not be verified with either of the two models.

**Step Model.** Based on the real-world data from the rental store, we developed the Step Model in a first attempt for a better solution. It copied the approach of [2] which determines a movie's life cycle by a set of parameters. In contrast to that approach, however, we tried to capture the day-to-day relevance changes in small user populations as well as the possibility of comebacks. The outcome of this approach was a model with parameters that was able to copy the behavior of every real-world movie quite well. Unfortunately, as shown in Figure 6, all kinds of absurd behavior were also possible.

We learned from this observation that a model which tries to take user numbers into account needs to separate the movie life cycle from the user's selection of a movie out of the set of available titles. If this separation is not made, the development model itself must be able to separate the long-term behavior from the short-term effects. But even if this is done, applying the model to a distributed system is hardly possible.

In the distribution system, the long-term behavior of the individual movie is supposed to be the same at every node in the system. Actually, it is a development that takes place outside the distribution system. The users' selections, on the contrary, drive the distribution system itself and are an important part of it. A monolithic model that

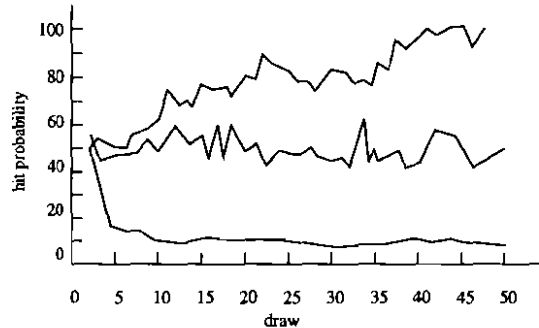


Figure 6: Sketch of parameter effects in the jump model

takes into account the number of users at a specific node would not fit the behavior of users at another node in the system which serves a different number of users.

#### 4 SIMULATION MODEL

**Modeling the Life Cycle.** The data that we have available restricts the realism of our model although we consider it much more realistic for long-term changes than previous models.

We can extract the long-term development of favored movies primarily from the data of the movie magazine, which is based on a large, representative number of movie rental shops and on a large customer base. The raw numbers that we have available from the single rental store provide a verification for these numbers. They can also be used to check the credibility of the model for long-term development of the movies because the rental numbers are continuously collected even after the movie has left the top 100 ranking list. These raw numbers are also the only basis on which we can model the development of movies that never or only pointedly hit the top 100 list and old movies that have reached the bottom line of their development and are still chosen quite frequently.

The access patterns of individual users to movies at their access point (the rental store) can on the other hand not be derived from either of these sources. Rental movies are watched "off-line", i.e. not at the time when the movie is taken away from the store. This is especially true because German rental stores from which both sets of data are derived are usually closed on Sundays, on holidays and usually from midnight to mid-morning, but we assume that the tendency of customers to pick more than one movie at a time is similar in countries without these restrictions. The television viewer would be a better basis to build this model on, but we have not found useful data to model this.

For the reasons given above, we distinguish between movie modeling and user modeling. The behavior of users who want to see a movie and the development of individual movies seem to be decoupled in reality. A user's decision to watch a movie is based mostly on the time he has available rather than the existence of an interesting movie. If a relevant number of movies is inserted into our model at the same time, no equally relevant increase in rentals will be visible. Although all new movies come quickly into the most popular phase of their life cycle the number of accesses to them remains low because the number of users, respectively the time the users have available for viewing, is limited.

The independent modeling is also more convenient for future modifications. This is important to us because we hope to improve the user model in the future by collecting data on daily changes in user interests and viewing frequency. Since we don't have this data yet, a simplification that is applied in our model is the independence of user behavior from a time of day. Especially when video-on-demand is available, we assume that there is no reason for a user to arrange his own schedules according to movie timetables. Thus, the time at which a movie is retrieved is completely independent from the choice which is based on the movie's popularity at the time of retrieval.

**Movie Life Cycle.** Newly published movies exhibit typically, but not always, a steeply rising start peak of user interest. The observation of the rental behavior shows that all movies share a general decrease of user interest in them, but this decrease is not identical for all movies and it is frequently interrupted by increases of user interest later in the life cycle (comebacks). Once a movie has been inserted into the system, its rental probability will never return to zero. On the contrary, the relevance of old movies can be quite high. We did not make a detailed category study, but marginal checks showed that, e.g., the start peak is less relevant (although existent) for movies rated PG-18 but the sustained relevance remains generally high.

As mentioned before, we discarded early models because they needed too many parameters with no explanation for their necessity. Splitting the available data into an underlying curve for long-term behavior of a movie and a random effect that is mainly dependent on the size of the user population led to a more appropriate model. We observed that the underlying curve seems similar to a variation of the exponential curve that is used to describe, e.g., the spread of infectious diseases. We used a version which is parametrized by some constant values to take the quantitative difference in the number of rentals, the steepness of the loss of interest and the remaining interest in a title into account. This function is

$$RP(t) = a \times e^{\frac{2}{\sqrt{10}} + \frac{t}{10 \times b} + \frac{b}{t}} + c$$

We found parameters applicable in the case of movie rentals by least square-fitting the function with the movie rental data.

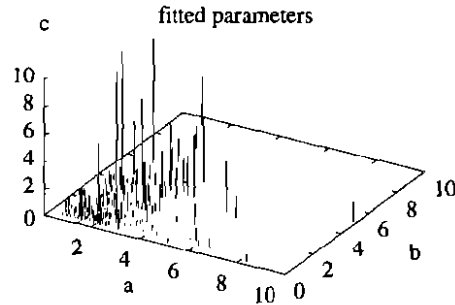


Figure 7: Parameters derived from data

Figure 7 illustrates that the parameters calculated from the least-square fitting are typically small and show no obvious correlation. The check verifies that the coefficient of correlation for each pair of parameters is tiny (the biggest is 0.01 for a and c, where a is interpreted the decline of popularity and c as the remaining popularity).

Because of this observation, we select the parameters for movie modeling independently from each other. For the individual draws that determine the parameters of a newly inserted movie, we use the exponential distribution. That gives us the basic functionality of frequent small and rare big parameters until we have a larger numbers base that allows a more educated decision in this matter.

In Figure 8 we show how daily hits according to our model are distributed and compare them to the real-world rental probabilities that we used in Figure 3 before to illustrate the divergence from the Zipf distribution. From 50 draws, we show the highest and the lowest curve. The behavior of our life cycle function yields a more wide-spread hit distribution when compared with the real world data. It is more pessimistic then the Zipf distribution. We observed also that our model will typically generate curves closer

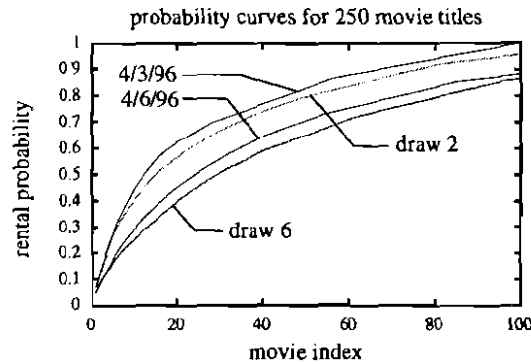


Figure 8: Rental probabilities compared with RP curves

to the lower curve, which implies that a distribution of hits over multiple movies is the typical case, while the generation of higher curves, which implies the dominance of a single movie, is a rare occurrence.

**User Population Size.** The size of the user population in a wide-area distribution system is *not* only relevant as an overall number. The observed increase in the variation of movie popularities for small user groups connecting to a single machine can be relevant for distribution algorithms. It is not reflected by the movie life cycle and must be modeled independently. Figure 9 demonstrates how the model recreates the smoothing effects of increasing user population sizes on the rental probability of a single movie. The life cycle of the example movie and all other movies used for the experiment (initially 150, 1 new movie each day) is defined as by  $RP(t)$ . Each draw is a request for one movie, the number of hits indicates how many of these draws select the example movie. In contrast to the simulation model that we develop, the number of hits per days was fixed for this experiment.

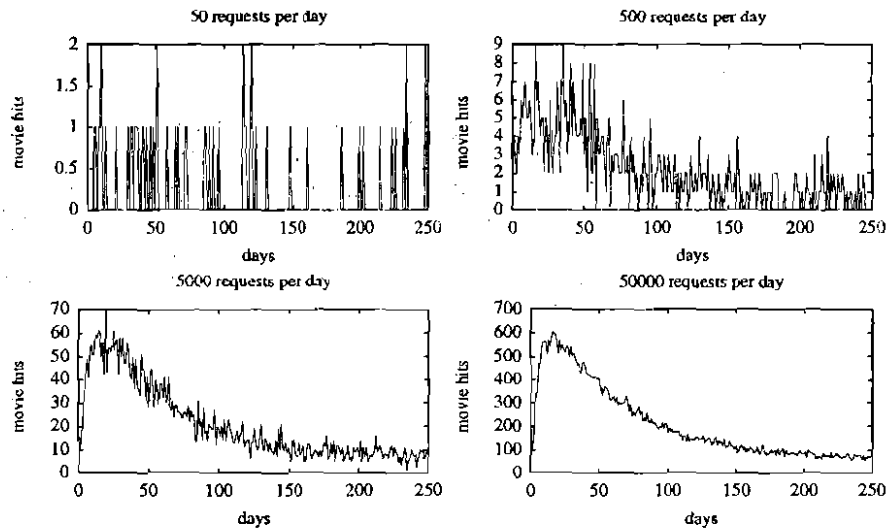


Figure 9: Smoothing effects of growing user populations

## 5 CONCLUSION

In this article, we have presented considerations about the distribution of videos in wide-area distribution systems that are based on caching. We introduced the problem of wide area distribution of video in wide-area distribution networks, under the assumption that a network infrastructure for such a network will be Internet-like.



We have stated why we believe that caching strategies for video data will be necessary in the future, what caching means in the context of wide-area distribution, and deduced that a user model is required to understand the effects of video access in a wide-area system.

To find such a model, we have then studied current models for movie popularities and compared them with real-world data that we have collected. We found that the models from the literature could not be applied easily to the cases that we study and have presented reasons for this opinion. We conclude from our work that a model which operates at the detail level of single user behavior needs to separate the movie life cycle from the selection which a user makes from the set of available movies. Finally, we propose an alternative simple life-cycle function as the basis of our popularity model and demonstrate its features.

In the future, we will use this function for the evaluation of automated distribution mechanisms in wide-area distribution networks. We present a simulation setup for purely hierarchical video-on-demand systems in [1]. We have already applied the model to evaluate some well-known caching algorithms in purely hierarchical distribution networks, and we will extend our evaluations to cover both more caching algorithms and other approaches like prefetching techniques and central decisions. We will also address topologies other than strict hierarchies.

#### ACKNOWLEDGEMENTS

We want to thank Michael Bär, now with SD&M, for his work on this topic which led to the development of the first models, and Ingo Dahm, now with Sun, and our colleague Lars Wolf for various discussions that provided explanations for twists in the data.

#### REFERENCES

- [1] Michael Bär, Carsten Griwodz, and Lars C. Wolf. Long-term Movie Popularity Models in Video-on-Demand Systems or The Life of an on-Demand Movie. In *Proceedings of the Fifth ACM International Multimedia Conference*, Seattle, USA, November 9-13 1997.
- [2] Scott A. Barnett, Gary J. Anido, and H.W. Beadle. Caching Policies in a Distributed Video on-Demand System. In *Proc. Australian Telecommunication Networks and Applications Conference*, Sydney, 1995.
- [3] C. Bernhardt and E. Biersack. The Server Array: A Scalable Video Server Architecture. In *High-Speed Networking for Multimedia Applications*. Kluwer Academic Publishers, 1996.
- [4] Y. S. Chen. Mathematical modelling of empirical laws in computer application: A case study. *Comput. Math. Applicat.*, pages 77-78, October 1992.
- [5] Asit Dan and Dinkar Sitaram. Buffer Management Policy for an On-Demand Video Server. RC 19347, IBM Research Division, 1993.

- [6] Asit Dan and Dinkar Sitaram. A Generalized Interval Caching Policy for Mixed Interactive and Long Video Workloads. RC 20206 (89404), IBM Research Division, September 1995.
- [7] Leana Golubchik, John C. S. Lui, and Richard R. Muntz. Adaptive piggybacking: a novel technique for data sharing in video-on-demand storage servers. *Multimedia Systems*, 4:140-155, 1996.
- [8] Wieland Holfelder. Interactive Remote Recording and Playback of Multicast Videoconferences. In *IDMS 97 Conference Proceedings*, 1997.
- [9] J. D. C. Little. A Proof of the Queuing Formula  $L=IW$ . *Operations Research*, 9:383-387, 1961.
- [10] T. D. C. Little and D. Venkatesh. Popularity-based Assignment of Movies to Storage Devices in a Video-on-Demand System. *ACM Multimedia Systems*, 1994.
- [11] J.P. Nussbaumer, B.V. Patel, F. Schaffa, and J.P.G. Sterbenz. Networking requirements for interactive video on demand. *IEEE Journal on Selected Areas in Communication*, 13(5):779-787, 1995.
- [12] R. Ramarao and V. Ramamoorthy. Architectural Design of On-Demand Video Delivery Systems: The Spatio-Temporal Storage Allocation Problem. In *Proceedings of ICC 91*. IEEE, 1991.
- [13] W. Tetzlaff, M. Kienzle, and D. Sitaram. A Methodology for Evaluating Storage Systems in Distributed and Hierarchical Video Servers. In *COMPCON 94 Conference Proceedings*, pages 430-439, 1994.
- [14] Renu Tewari, Rajat Mukherjee, and Daniel M. Dias. Real-Issues for Clustered Multimedia Servers. Technical Report RC 20108, IBM Research Division, June 1995.
- [15] VideoWoche. Entertainment Media Verlag GmbH & Co.oHG, Stahlgruberring 11a, 81829 München