The Effects of User Behavior and Internet Provider Policy on the Accessibility of SezamPro On-line System

Jelica Protić, Miloš Prvulović, and Dejan Ristanović*
{jeca@ubbg.etf.bg.ac.yu, prvul@sezampro.yu, dejanr@sezampro.yu}
School of Electrical Engineering, University of Belgrade
Bulevar Revolucije 73, Belgrade, Yugoslavia
*SezamPro On-line, Skadarska 40C, Belgrade, Yugoslavia

Abstract

In this paper we present a model of an on-line system that includes user population, a number of dial-in lines, and specific waiting subsystem. According to this model we perform simulations in order to show the effects of various user behaviors and provider policy on the accessibility of the system. In addition to experiments with assumed values of parameters, we also determine realistic values based on data recorded on the SezamPro, the largest on-line system in Yugoslavia.

1: Introduction

During the last few years connection to the Internet global network became the necessity for all academic computing centers, as well as commercial on-line systems, such as SezamPro [1], the largest on-line system in Yugoslavia. The main obstacle to this process in Eastern European countries is the poor telephone infrastructure, that makes costs of telecom installations very high: for example in Yugoslavia, where the first satellite station for commercial Internet providing was installed in March 1996, a start-up Internet provider should still invest more then $2000 per dial-in telephone line. Even in developed countries such as USA, insufficient number of dial-in lines became a serious problem, especially in academic environment [2]. Therefore, careful planning of the number of dial-in lines and their utilization, suited to specific needs of users, could be essential in performance/cost optimizations of such systems.

2: On-line System Modeling

The goal of this project is to develop a model that describes the behavior of user population, a set of dial-in lines (that may be grouped in cascades), as well as on-line system itself, with its specific policy, that can impose per-session or daily limits on time spent on-line. Few papers that address similar problems deal with the approximate number of 10 users per phone line, effective for 16 or more dial-in lines [3], or concentrate on experimental results comparing the accessibility of existing commercial on-line systems [4]. Unlike the work presented here, these studies do not quantitatively predict non-linear changes of system behavior by changing user population, number of dial-in lines, average time spent on-line, and its probability distribution. Possible methods for such an analysis include analytical and simulation approach. Although

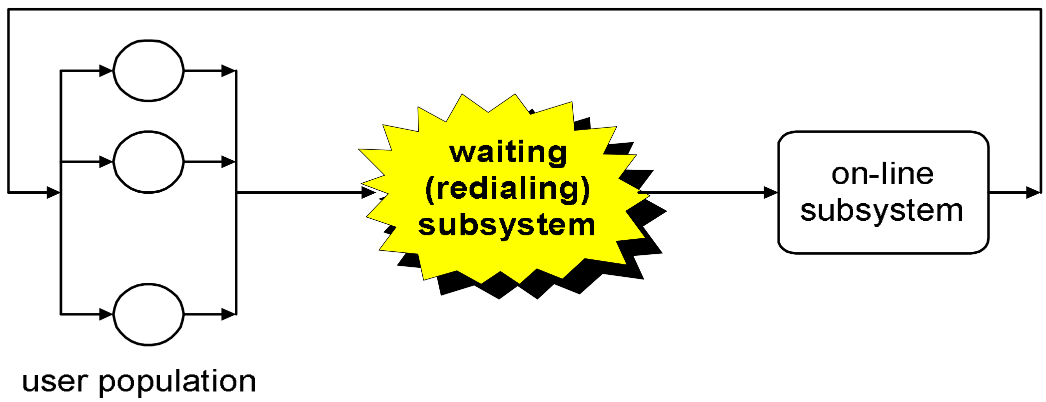
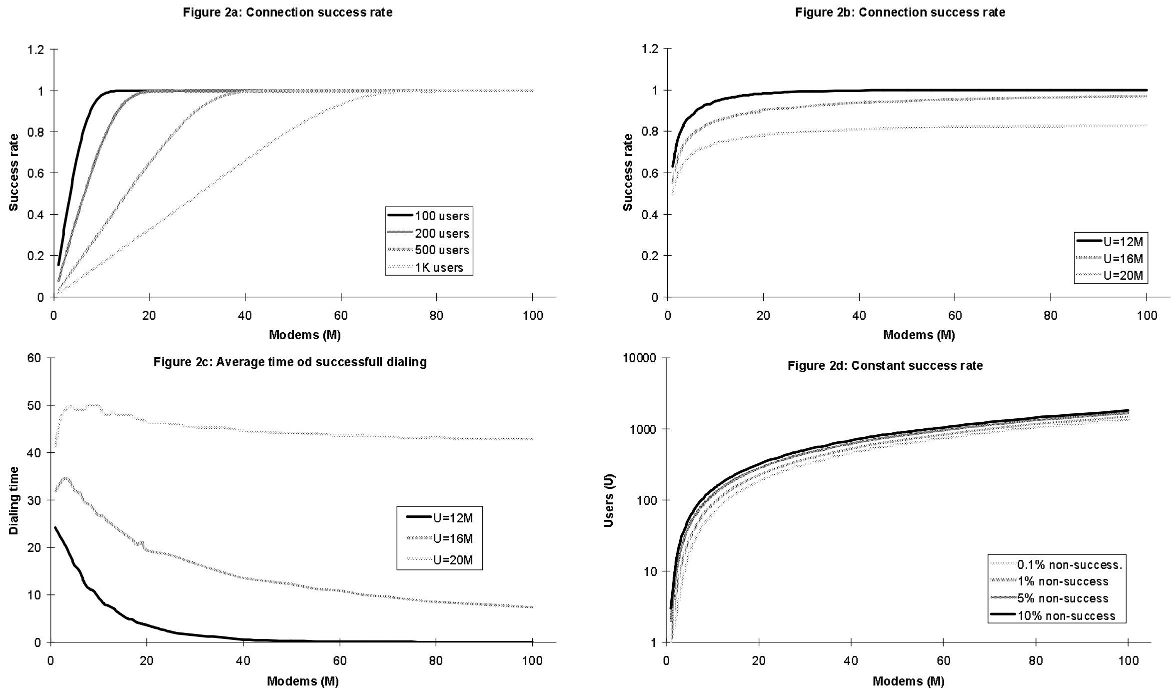


Figure1: Simulated model

the model we consider (Figure 1) resembles an interactive system with a number of terminals and a service center, it differs from it for its specific nature of waiting subsystem, that can not be described as a FIFO queue. Therefore, simulation analysis is more suitable in this case, especially because it makes possible to simulate various types of user behavior, multiple classes of users, per-session time limits etc. The simulator developed and used for experiments presented here was written in C++ and runs on i486 PC. The discrete event simulation is based on the following possible states of the user: *inactive*, *waiting* and *on-line*. A user is *inactive* between the finished session and the next try to get on-line. For the presented set of experiments we assumed that inactivity time has an exponential distribution of probability; in other experiments various distributions are used, considering special behavior of users calling from their offices during working hours, or users taking advantage of low-rate telephone services etc. During the *waiting* state user dials a number, and if provider has any free lines, gets into the *on-line* state. If user gets a busy signal, after a fixed time he dials again, until he gets *on-line*, or until he gives up, when the time interval that user is willing to wait expires. In our experiments, both the duration of this interval, and the time user plans to stay *on-line* have exponential distributions. If the provider does not impose any limits, the user stays *on-line* as long as he planned to; otherwise, his session might end by forced disconnection.

3: Simulation Based on Synthetic Workload without Time Limits

Figures 2 a, b, c, and d present the results of experiments that assume the average inactivity time of 8 hours. The average time a user is willing to wait is 5 minutes, and during that time he redials every 20 seconds. If the user does connect, his session lasts 30 minutes in average (a reasonable duration, considering both "Web surfers" and the other users that login just to check their e-mail), with no time limits imposed by the provider. For such a system, we determined the percentage of successful connections (out of all connection attempts) depending on the number of lines provider has installed. We can see that at some point there are virtually no unsuccessful connections. Before that point, correspondence between the number of lines (modems) and the number of users is close to linear. Assuming the 100% utilization, theoretical maximum of users that can be serviced by a single line is close to 16 (16 users, 3 times a day 30 minutes, equals 24 hours). Therefore, the following experiment presented by Figures 2b (showing the success rate) and 2c (showing the dial time), assumes that the number of users in the system is 12, 16 and 20 times the number of lines. For the small number of lines, there is a considerable probability that all lines are busy, even if the system is underloaded. With the increase of total number of dial-in lines, the success rate approaches its theoretical



maximum, due to statistical distribution of utilization over multiple lines. Naturally, saturation effects are more noticeable for an overloaded system (U=20M), that hardly reaches 80% success rate, than for an underloaded system (U=12M), having the success rate close to theoretical maximum.

Figure 2d presents the results of experiments that determine the number of users an on-line system can serve for a predetermined success rate (0.1, 1, 5 and 10% unsuccessful tries to connect). The scale for users is set to logarithmic to show that, for a given large number of lines, the ratio of numbers of users that can be served with different success rates tends to become constant. In other words, if we have a system with no unsuccessful connections, in order to have 10% more users and keep the number of modems constant, we must allow for an unsuccess rate of 10%, meaning that one out of 10 of all the attempts to establish connection will be unsuccessful.

4: Effects of Per-session Time Limit

When a per-session time limit is introduced, the average service time becomes shorter (the users still request to be serviced for 30 minutes on average, with exponential distribution, but all sessions that would be longer than 45 minutes are truncated, so the average time of service becomes 23 minutes and 19 seconds). Therefore, it becomes possible to service more users with the same rate of successful connections. If we assume that all users are equal and that there are no special rush-hours, introducing a 45-minute per-session limit actually means making a clear tradeoff between forced disconnection rate and the number of users that can be serviced (when the number of lines and the connection success rate are known). In our example, the limit means 22.3% of all sessions are truncated.

However, the assumption that all users behave the same way is not realistic. Consider an example in which there are two groups of users, one with an average requested service time of 1000 seconds, and the other with an average service time of 5000 seconds. If the number of users in the first group is four times the number in the second group, the overall average

Figure 3a: Constant connection success (99%)

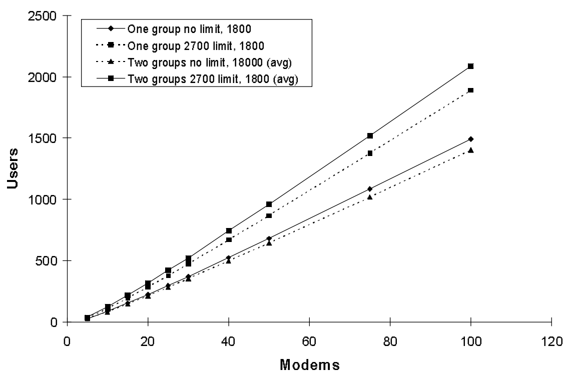
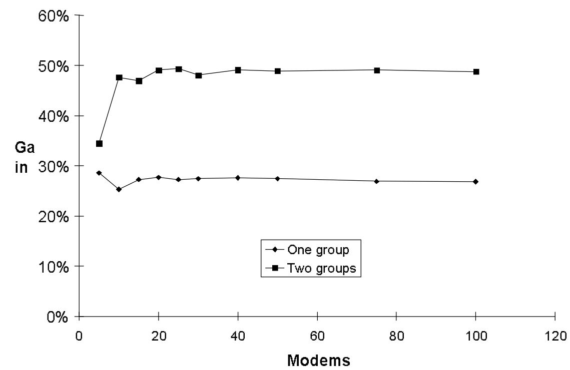


Figure 3b: The gain of per-session limit



requested service time is still 1800 seconds, i.e. 30 minutes. All users call approximately 3 times a day and have their dialing habits just as before.

If there is no limit, the number of users which can be serviced is almost the same as in our previous example (with no limits), because the system load has not changed. However, when the limit of 45 minutes per session is introduced, the forced disconnection rates of the two user groups are different - the first (less demanding) group has a forced disconnection rate of 6.6%, while the second (more demanding) group has a forced disconnection rate of 58.2%. The overall average service time for all users is now around 1182 seconds (19 minutes and 42 seconds), which is less than in the case of a single user group and per-session limit. Consequently, even more users can be served. This shows that the per-session limit becomes more useful if there are multiple classes of users with different requirements. It can be argued that, when a flat rate charging is used, the introduction of the limit essentially means that less-demanding users are preferred over more-demanding ones. Due to the high force-disconnection rates, more-demanding users may want to change the provider and choose one that imposes no time-limits. If the pool of potential users is limited (i.e. there is a fierce competition), loosing demanding customers may be unacceptable, but if the provider has limited resources (on the verge of an overload) and many potential users, introduction of the limit may result in ending up with many low-demand customers instead of the mixture of both more- and less-demanding ones. If we assume that users are charged with the flat-rates, this can even be favorable for the provider.

Figure 3a shows the case when the connection success rate is constant, and it is 99% (1% unsuccessful connections). It is obvious that, when introducing time limits, we gain more when there are two different groups of users than in the case of a single group of users. Figure 3b explicitly shows the gain (in terms of number of users which can be served with a number of modems) of limit introduction in the two scenarios we examined before: 1) only one group of users with an average requested service time of 30 minutes, and 2) two groups, one requesting more and the other less time, but the overall average requested service time being still 30 minutes.

5: Simulation using Parameters Derived from Measurements on SezamPro

Development of simulation tools and postulating user behavior is fine as far as theoretical modeling is concerned. However, our intent is to be able to predict the behavior of real-world systems, such as SezamPro. The first step in that direction would be validation of the assumptions we used when designing our model. For that purpose we collected a log of all connections to SezamPro on-line system over a period of ten days.

Figure 4a: Distribution of time spent on-line

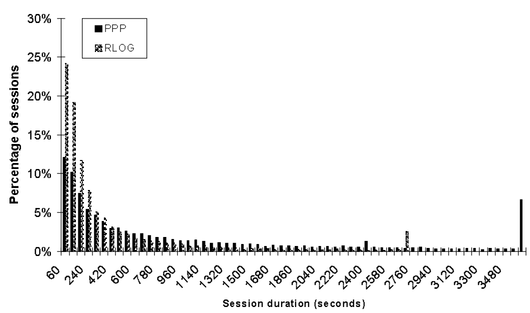
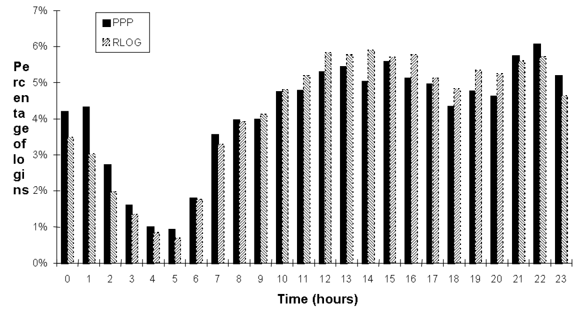


Figure 4b: Login intensity distribution (by time-of day)



SezamPro currently has 41 lines for public access, which are divided into 3 cascades, with 15, 15 and 11 lines, respectively. Three types of public access are possible: telnet, PPP and remote login (RLOG). Telnet connections do not make use of dial-in lines, and are of no interest for this analysis. PPP connections are made to SezamPro via telephone lines using Point-to-Point Protocol, and basically for them SezamPro acts as an Internet Provider. Remote login puts users into a console mode shell for such services as Internet e-mail, local conference system, local chat system etc.

All users of SezamPro system are administratively divided into two groups - those with a full Internet access, and those with local access only. Although users from both groups may use any type of connection, in reality Internet users mainly use PPP access, while users with local access use mainly RLOG access. SezamPro also imposes per-session limit of 60 minutes for users with full Internet access, and 45 minutes for users with local access only.

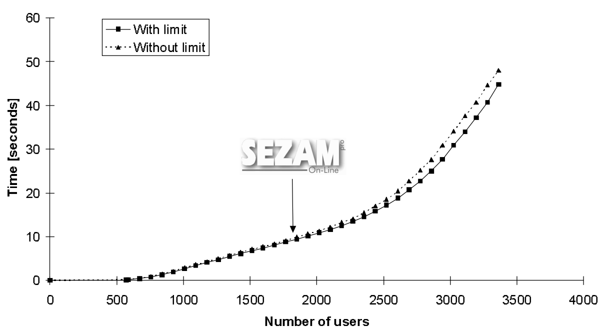
It can be noticed that PPP and RLOG users have different behavior considering both time they spend on-line and the time between successive attempts to get on-line. The distribution of time users spend on-line is shown on Figure 4a. It is obvious that each of these curves closely resembles an exponential distribution (with truncation at 45 and 60 minutes), showing that our assumption about exponential distribution of time users want to stay on-line was realistic. By averaging on-line time, we come up with a mean on-line time of 411 seconds for RLOG, and 979 seconds for PPP connections. These are averages for actual time spent on-line, with truncation. Assuming exponential distribution, it can be calculated that the average time RLOG and PPP users wanted to stay on line was approximately 412 and 1007 seconds, respectively. Figure 4b shows the variations of the login intensity during the day time.

Computing the time between successive attempts of a single user to connect to SezamPro is not so easy. Knowing the number of users U, the time interval T=10 days and the number of successful connections N, we can calculate the average time between successive successful attempts as $T_{avg} = \frac{U \cdot T}{N}$. The formula is exactly the same for average time between successive attempts (successful or not), but the number N must include unsuccessful attempts. Because both PPP and RLOG users compete for the same dial-in lines, the percentage of unsuccessful attempts should be the same for both groups. To find that percentage, we ran the simulation and used successive approximation method to find N. The final result indicates that the percentage of unsuccessful connection attempts is about 2.8%.

The inactivity time, as we defined it, is the time between the end of one connection attempt and start of the next attempt. Therefore, $T_{inact} = T_{avg} - T_{online}$. From the log mentioned above we arrived at the average time of inactivity of 69250 seconds (19 hours and 14 minutes) for RLOG and 43160 seconds (11 hours and 59 minutes) for PPP users.

Having found empirical values for all the parameters needed as input for simulation, we simulated the SezamPro system, varying number of users. It can be seen that the actual number

Figure 5: Average time of (successfull) dialing



of users is such that the system is currently operating at a point almost in the middle of the linear segment between underload and overload, considering the average dialing time (Figure 5). Since it can be expected that a real-world system would actually be somewhere between an overload and underload, we can safely assume that our model describes the real world fairly well.

6: Conclusions

In this paper, we have described a simulation model in order to examine effects of user behavior and the provider policy on the accessibility of an on-line system, concentrating on the assumption that dial-in telephone lines represent the main bottleneck. After performing simulations with a set of assumed parameters, we validated our results using realistic logs of the SezamPro on-line system. However, in all experiments we assumed that users wait for connection 5 minutes on the average (with exponential distribution) and during that time redial every 20 seconds (constant interval). Although we have not collected statistical data to support this assumption, the personal experiences of several users indicate that those numbers are realistic enough. One other assumption was that the system load does not vary during a single day. Figure 4b shows the average daily distribution of successful connections calculated over a ten-day period. It is obvious that the number of connections per hour varies by as much as 50% during the day. However, we have not yet developed a model for such a distribution that is both simple and accurate enough. That remains the topic for further research in this area.

The possibilities for the follow-up research are very promising. Refining the existing model to include more realistic parameters is only one possibility. Another area of particular interest would be modeling and simulation-based analysis of Internet traffic on provider’s link to the rest of the world. This link is another major bottleneck and a huge investment. Therefore, the optimization of its capacity based on the number of users, user habits and provider policy, is very important. The model presented in this paper could be used as a first layer, providing the relation between overall user population and the number of currently connected users, that generate Internet traffic. We plan to continue our efforts in this direction, that represents challenging and possibly fruitful area of research.

7: References

[1] Ristanović, D., “SezamPro: Full Internet Access,” *PC*, PC Press, September 1996.
[2] Horning, K., Eaton, J., “Unprecedented Demand Hampers Dial-in Service,” *Information Technology Digest,* Vol. 4, No. 9, November 13, 1995.
[3] Dennis, D., “Internet Access FAQ,” *http://cgi.amazing.com/internet/faq.html*, December 23, 1996.
[4] Munro, J., “On-line Services: Test Results,” *PC Magazine*, May 14, 1997.