Baidu Netdisk's File Transfer Strategy

Zhiyi Wang, 121090586, Chenrui Xu, 121090648, Yin Yang, 120090516, Qiyue Zhang, 120090391

***Abstract*—Baidu Netdisk, a major application that dominates the centralized cloud storage market, is most widely known for its particularly slow file download rates for non-VIP users. In 2021, Baidu Netdisk officially launched a P2P network-based file transfer mode, which allows users to accelerate the download rate of some popular files for free by joining the P2P network. This paper aims to analyze the user ecology of Baidu Cloud before and after the addition of P2P network as a new transmission method, and the strategies taken by Baidu, including VIP pricing, bandwidth allocation and the number of P2P users.**

***Index Terms*—Baidu Netdisk, Peer-to-Peer network, bandwidth allocation**

I. Introduction

With the development of Cloud Computing, Cloud Storage technology was born. Cloud storage is a service which provides storage resource and service based on the remote storage servers. Cloud Disks are one of the applications. In China, Baidu Netdisk occupies the dominant market of cloud storage with exceeding 800 million users By 2022. However, unlike the strategies of other companies, Baidu Netdisk does not limit the storage space of the cloud drive but the download rate of users. For non-VIP users, the download rates are generally stable at below 100 KB which means it should take more than 10 hours or even days to download a file of several GB. Around April 2021, Baidu Netdisk launched a user incentive plan: users can choose to join the Peer-to-Peer (P2P) network to increase the file transfer rate, but at the same time, users who join the P2P network also become nodes within the network structure, and their idle bandwidth will be occupied to form the exclusive transfer channel for P2P users. But compared to VIP users, P2P can only accelerate the transmission of those popular resources, because the more popular a resource is, the more likely it is to be held by more nodes, and a P2P network is efficient only when a sufficient number of nodes can transmit a resource together. The addition of this strategy breaks the traditional binary market of VIP and non-VIP users. From Baidu's point of view, this strategy on the one hand alleviates some of the pressure on bandwidth resources, but on the other hand it may also reduce the number of VIP users to some extent.

As a for-profit application, Baidu Netdisk cannot launch new plans that are detrimental to its own interests. Based on the above discussion, we are curious about the impact of the addition of the P2P mechanism on the original user ecology of Baidu Netdisk, as well as the bandwidth allocation between non-VIP users and VIP users and the design of the P2P network based on the motivation to maximize the profit of Baidu Netdisk.

A. Questions

We divide the topic into three main sub-questions which are as follows:

1. Why does Baidu Netdisk allow P2P networks to download files? That is to say, whether and how the P2P mechanism bring more profit to Baidu Netdisk.

2. How does Baidu Netdisk allocate its bandwidth to non-VIP users and VIP users, and how does P2P mechanism affect the total bandwidth allocation among three types of users?

3. What strategy should Baidu Netdisk employ to gain the theoretically highest utility? (For example, how much should Baidu charge for each VIP membership? How many p2p users should be allowed by Baidu?)

B. Assumptions

We also make some assumptions before modeling:

1. Baidu is the single monopolists of the monopoly market.

2. Non-VIP users and VIP users can converted to each other, and the amount of P2P users can be controlled by Baidu; total number of users is fixed.

3. The total amount of bandwidth provided by Baidu Netdisk is fixed.

The first two assumptions limit the context of the problem and provide constraints and known conditions for the model; the third assumption is an initial conjecture about the model, and the specific results will be discussed in detail in Section 5.

II. Literature Review

Baidu Netdisk as a non-open source application, there is not much relevant research. Most of the existing researches are related to the privacy security [2] and traffic

classification [4]. Little research focused on the file transfer mechanism of Baidu Netdisk's itself. Therefore, We can study the architecture of two transmission networks from Baidu Netdisk, one is the traditional server-client architecture and the other is the P2P network.

A. Cloud Storage

The architecture of cloud storage can be organized into five layers and they are network and storage infrastructure, storage management, metadata management, storage overlay, service interface from bottom to upper respectively [3]. For traditional server-client model in cloud storage, users access the data in the central cloud server through service interface layer which is totally centralized. All bandwidth used to transfer files is provided by the service provider.

B. Peer-to-Peer Network

As a decentralized network, there is no central server in P2P network, all users have a consistent identity, both as a server and a client. P2P network can monitor the network traffic and distributed content to ease the load on individual nodes, then possibly locate content closer to points of high demand [1]. For content-sharing P2P networks, the more nodes involved in the network, the faster the transmission efficiency and the more robust the whole network.

III. Method of ODE

In this paper, we consider Baidu to be the single monopolists of monopoly market. Baidu will buy a certain amount of bandwidth and allocate them to its users. All users will respond to both the price of VIP membership and their download speed. Baidu will determine its best strategies (price of VIP and allocation of bandwidth) to gain the maximum revenue.

A. Case 1: Model with Non-VIP and VIP users

Consider the case that P2P mechanism is not utilized. We try to find the equilibrium numbers of VIP users and non-VIP users, then seek for the best strategies that Baidu should apply.

1) Basic Notations and Mechanism of Rate Limitation: Denote $I = \{v, n\}$ as the set of users' strategies, where v stands for VIP and n stands for non-VIP. In this case, users will choose either v or n, which will thus be divided in to 2 group. Let B denote the total bandwidth that Baidu can offer. S_n and S_v are the ratio of bandwidth allocated to non-VIP users and VIP users, respectively. So we have $S_n \in [0, 1]$, $S_v \in [0, 1]$ and

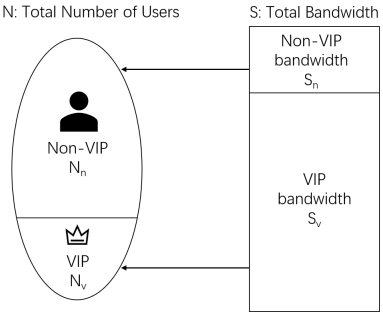

Fig. 1. Bandwidth Allocation Model in Server-client Model

$S_v + S_n \leq 1$. Denote N as the total number of users, which is assumed to be a constant. Denote $N_n(t)$ and $N_v(t)$ as the number of non-VIP and VIP users at time t, respectively (figure 1).

Assume every users in the same $i \in I$ have the same download speed $v_i = k\frac{BS_i}{N_i}, i \in I = \{v, n\}$, which is proportional to allocation ratio and inversely proportional to users' number in same group i.

2) User's Switching Willingness: Denote τ_i as the waiting time of a user in group $i \in I$ to download a unit amount of files. Then $\Delta\tau = \tau_n - \tau_v$ is the time benefit from doing such switching.

Denote P as the price of VIP membership that Baidu designs. Then we need to capture the switching willingness with respect to price, $w(P)$, which is a decreasing function. This function can be identified by experiment on certain society, together with proper regression methods, such as logistic regression.

Then we can define a user's willingness of switching their strategies from non-VIP to VIP, i.e. W_{nv}. Notice that W_{nv} should be positively correlated to time benefit $\Delta\tau$ and negatively correlated to VIP price P. We may assume that

$$\begin{aligned} W_{nv} &= \alpha_1 \Delta\tau - \alpha_2 w(P) = \alpha_1 \left(\frac{N_n}{kBS_n} - \frac{N_v}{kBS_v} \right) - \alpha_2 w(P) \\ &= \frac{\alpha_1}{kB} \left(\frac{N_n}{S_n} - \frac{N_v}{S_v} \right) - \alpha_2 w(P), \end{aligned} \tag{1}$$

where α_1 and α_1 are two constant.

3) Switching Equation: We consider the number of users switching from non-VIP to VIP during small time period Δt, which is exactly one user's willingness of switching multiplied by non-VIP number:

$$\frac{N_v(t + \Delta t) - N_v(t)}{\Delta t} = N_n(t) W_{nv}.$$

Let Δt goes to 0, then we get

$$\frac{\mathrm{d}N_v}{\mathrm{d}t} = N_n W_{nv}.$$

Notice that $\frac{\mathrm{d}N_n}{\mathrm{d}t}$ should be the opposite of $\frac{\mathrm{d}N_v}{\mathrm{d}t}$, so we have

$$\begin{aligned}\frac{\mathrm{d}N_n}{\mathrm{d}t} &= -N_n W_{nv} \\ &= -N_n\left(\frac{\alpha_1}{kB}\left(\frac{N_n}{S_n}-\frac{N_v}{S_v}\right)-\alpha_2 w(P)\right) \\ &= \left(\frac{\alpha_1 N}{kBS_v}+\alpha_2 w(P)\right)N_n - \frac{\alpha_1}{kB}\left(\frac{1}{S_n}+\frac{1}{S_v}\right)N_n^2.\end{aligned} \tag{2}$$

Now we get the differential equation describing the change in the number of non-VIP users.

4) Solving and Analyzing the ODE: We observe that equation(2) is a separable, logistic equation, precisely. We can easily get the solution using integrating factor. Assume the initial value of N_n is $N_n(0) = N_{n0}$. Denote

$$\begin{cases} a = \frac{\alpha_1 N}{kBS_v}+\alpha_2 w(P) \\ b = \frac{\alpha_1}{kB}\left(\frac{1}{S_n}+\frac{1}{S_v}\right)\end{cases} \tag{3}$$

Then the solution is

$$N_n(t) = \frac{aN_{n0}}{bN_{n0}+(a-bN_{n0})e^{-at}}.$$

As $t \to \infty$, we get the equilibrium value of N_n, denoted by M_n:

$$M_n = \frac{a}{b} = \frac{\frac{\alpha_1 N}{kBS_v}+\alpha_2 w(P)}{\frac{\alpha_1}{kB}\left(\frac{1}{S_n}+\frac{1}{S_v}\right)} = \frac{S_n N + S_v S_n \eta w(P)}{S_n+S_v}, \tag{4}$$

where we denote $\eta = \frac{kB\alpha_2}{\alpha_1}$. So the number of VIP users at equilibrium, denoted by M_v, is:

$$M_v = N - M_n = \frac{S_v N - S_v S_n \eta w(P)}{S_n+S_v}. \tag{5}$$

5) The Revenue of Baidu Netdisk: The revenue of Baidu company for a certain time period, denoted by π equals its income subtracted by the cost, where the income equals the price of VIP membership multiplied by number of VIP users.

$$\pi = P\frac{S_v N - S_v S_n \eta w(P)}{S_n+S_v} - hB\left(S_v+S_n\right). \tag{6}$$

Therefore we need to solve the optimization problem:

$$\begin{aligned} max_P \quad & P\frac{S_v N - S_v S_n \eta w(P)}{S_n+S_v} - hB\left(S_v+S_n\right) \\ s.t. \quad & P_- \le P \le P_+ \end{aligned} \tag{7}$$

where P_- and P_+ are the reasonable lower bound and upper bound of price.

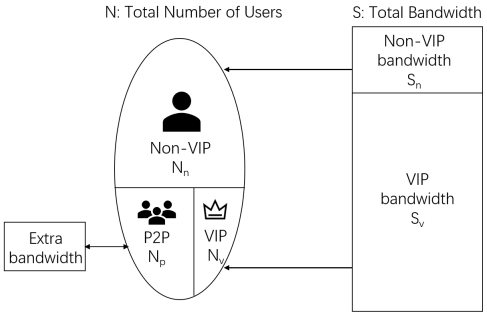

Fig. 2. Bandwidth Allocation Model in Server-client Model with P2P Model

B. Case 2: Model with non-VIP, VIP and P2P Mechanism

In this case, we assume that Baidu will release some quota, denoted by N_p to let some people use P2P method to increase their download rate, according to the mechanism of P2P network. And we can assume that the quota is filled. Now we analyze how the amount of N_p will effect the equilibrium and Baidu Netdisk's revenue. Therefore we can find the best strategy (release how much N_p to maximize the revenue) that the it should take.

1) The New Equilibrium: After the P2P quota is released, there will be $N - N_p$ users still take non-VIP or VIP strategy. According to the concepts of P2P networks, we may assume that those N_p users produce a total bandwidth of $N_p\overline{B_p}$, where $\overline{B_p}$ is the average bandwidth of each users. We consider the equilibrium of them. Modifying the N in equation (4) and (5) by $N - N_p$, we get:

$$M_n = \frac{S_n(N-N_p) + S_v S_n \eta w(P)}{S_n+S_v}, \tag{8}$$

$$M_v = \frac{S_v(N-N_p) - S_v S_n \eta w(P)}{S_n+S_v}. \tag{9}$$

2) The New Revenue of Baidu Netdisk: Notice that with these P2P users, Baidu netdisk only need to provide its bandwidth to $N - N_p$ users (and a very little amount of bandwidth ϵ to P2P users). Therefore we can modified (6) into:

$$\pi = P\frac{S_v(N-N_p) - S_v S_n \eta w(P)}{S_n+S_v} - h\frac{N-N_p}{N}B\left(S_v+S_n\right). \tag{10}$$

Then we need to solve the optimization problem:

$$\begin{aligned} max_{P,N_p} \quad & \pi \\ s.t. \quad & P_- \le P \le P_+ \\ & 0 \le N_p \le N_{pm} \end{aligned} \tag{11}$$

where P_- and P_+ are the reasonable lower bound and upper bound of price, and N_{pm} is the reasonable upper bound of P2P users.

IV. Method of Markov Chain

In this section, we treat the VIP, non-VIP, and P2P users and three distinct states, and analyze the problem from the perspective of the time dependent Markov chain.

A. The definition of transition state matrix

The transition probability matrix Q is defined as follows:

In each term, k_1 represents the likelihood of a VIP user

$$\begin{pmatrix} 1-(\alpha+\gamma)P-\frac{\beta\ L}{1+e^{-k(t-t0)}} & \alpha\ P+\frac{\beta\ L}{1+e^{-k(t-t0)}} & \gamma\ P \\ k_1 P & 1-(k_1+k_2)P & k_2 P \\ k_3 & k_4 P & 1-k_3-k_4 P \end{pmatrix}$$

$k_1,\ k_2,\ k_3, k_4,\ \alpha, \beta, \gamma$ *and the logistic parameters* $L,\ k,\ t_0$ can be estimated by observing historical data and regression.

Fig. 3. The transition state matrix for the markov process

downgrading to a normal user based on the VIP price $P(t)$, with a higher k_1 indicating greater price sensitivity. Similarly, k_2 denotes the likelihood of a VIP user transitioning to a P2P user due to $P(t)$, with a higher k_2 suggesting a higher probability of switching when VIP prices increase. On the other hand, k_3, a constant, signifies the likelihood of a P2P user becoming a normal user, independent of $P(t)$ and marketing efforts, while k_4 indicates the likelihood of a P2P user upgrading to VIP based on $P(t)$, with a higher value suggesting greater price sensitivity. The coefficients α, β, and γ represent the influence of VIP price on the transition probability of Normal to VIP (P_{NV}(t)), marketing campaign impact on P_{NV}(t), and VIP price influence on the transition probability of Normal to P2P (P_{NP}(t)), respectively. We model the price function as a capital P plus a function M of t (P + M(t)), where P represents the constant charging price, and M(t) represents the decaying impact of marketing activities, promotions, and other factors. $L,\ k,\ t_0$ is the logistic parameter for M(t). k_1, k_2, k_3, k_4, α, β, γ and A, B, $L,\ k,\ t_0$ can be estimated by observing historical data and doing regression or optimization.

B. The steady state distribution

The transition state matrix defined earlier is not a constant matrix, so it is impossible to use the traditional way to find the eigenvector of the transition matrix. However, we can use an alternative method to handle it:

$$\lim_{t\to\infty} Qx[t] \ = \ x[t]$$

We take the limit of time as infinity and then solve its eigenvector, and treat it as the steady state distribution of three types of users. To attain the final result, we conducted a survey and acquired related data and fixed the initial percentage of P2P user as 10 percent.

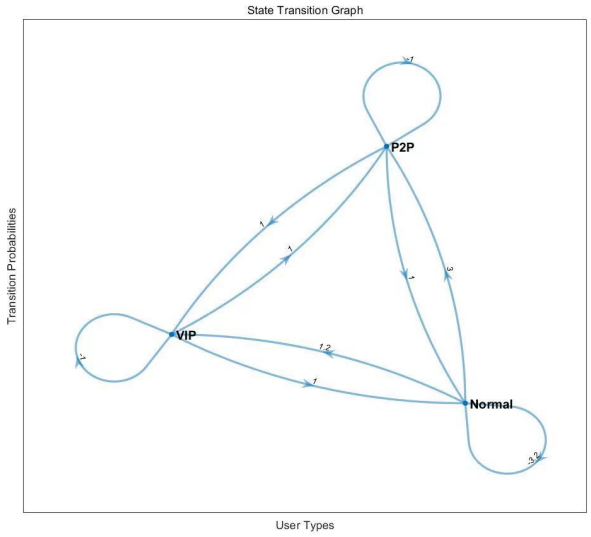

Fig. 4. The state diagram for the markov chain

V. Results and Discussions

A. Case1

1) Data Processing: We have processed the data obtained from the questionnaire survey. Specifically, we have sorted the expected prices of 93 Baidu Netdisk VIP users and potential VIP users from a total of 148 questionnaires. In order to complete the subsequent calculations, a small amount of noise was added to the data, and the blue points on the figure were plotted based on this modified data. We use logistic regression to get the approximated switching willingness. Finally, we obtain the switching willingness function with respect to price:

$$w(P) = \frac{1}{1+e^{(-4.62+0.26P)}},$$

which is depicted by the red line in figure 5.

To calculate the optimal solution for Baidu's revenue by determining the price and the ratio of bandwidth used by VIP users, which are key factors that impact the revenue of Baidu Cloud Drive. We have set up some

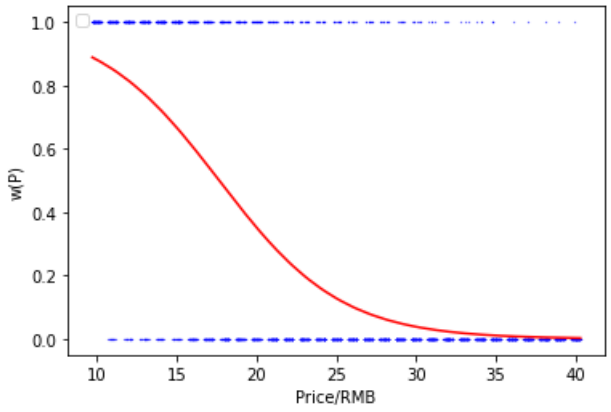

Fig. 5. Willing function with respect to Price

TABLE I
PARAMETER SETTING OF CASE 1

Parameter	Value
N: total number of users (millions)	8.0×10^2
N_nmin: Minimum number of non-VIP users	$0.3 * N$
η: A constant	7.5×10^5
range of S_v	[0.995, 0.999]

parameters for our calculations, which are listed in the Table 1. These parameters come from the report of Baidu Cloud and the data analysis of the questionnaire.

2) Switching and Equilibrium Results: We get the switching behavior and equilibrium of case 1 by iteration (figure 6).

3) Simulation: In view of the fact that Baidu Cloud can control the ratio of bandwidth occupied by non-VIP and VIP, and the Price of VIP, we draw the relationship between S_v, P and Revenue in the model with other parameters as figure 7. According to the simulation, the maximum Revenue is obtained with S_v=0.999 and P=14.8.

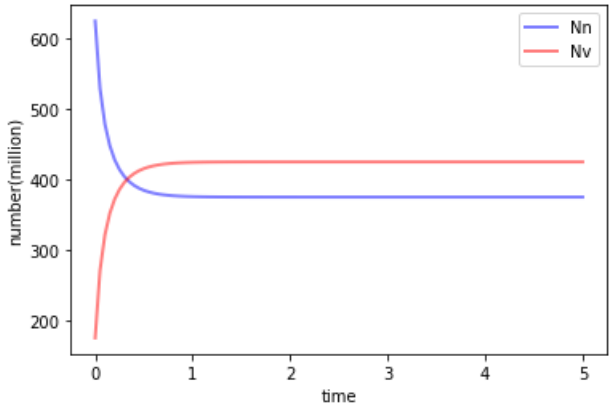

Fig. 6. Equilibrium of users in Case 1

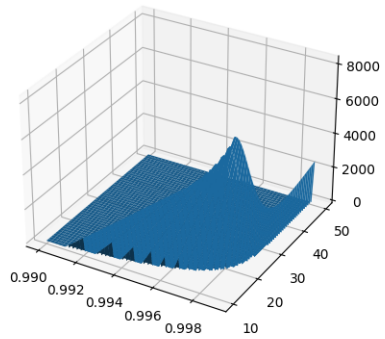

Fig. 7. Surface Plot with respect to S_v, P and Revenue

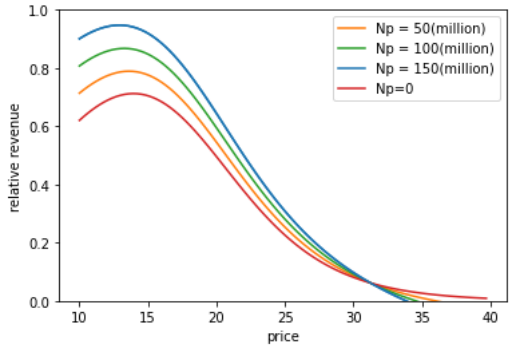

Fig. 8. The relative revenue of Baidu Netdisk with different number of P2P users

Moreover, our analysis indicates that users are more likely to cancel their VIP service subscription when the VIP user's bandwidth ratio is relatively limited or restricted, or when the VIP price is relatively high.

B. Case2

1) Simulation: Based on the analysis in the Method part, with the same data processing method as Case1, we can obtain the relative revenue of Baidu Netdisk with respect to different quota of P2P users.

2) Discussion: According to the figure, we find that, under the condition of the same P and S_v, the revenue of existing P2P users is higher than that of only VIP and non-VIP users.

Moreover, the existence of P2P mechanism can slightly decrease the price of VIP at optimal revenue. So, this is a win-win mechanism for both Baidu and users.

C. Method of Markov Chain

1) Data Processing: To ensure the reliability and validity of our simulation results of Model 2, we carefully

TABLE II
PARAMETER SETTING OF CASE 2

Parameter	Value
k_1	0.4
k_2	0.1
k_3	0.05
k_4	0.1
L	5
k	0.1
t_0	50.0

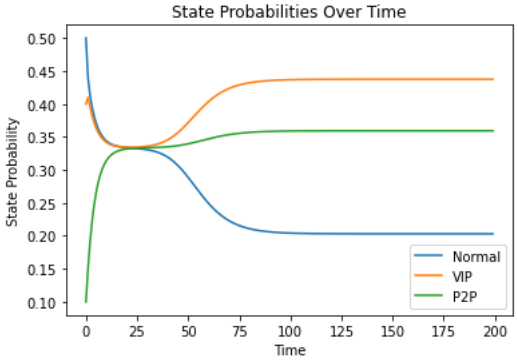

Fig. 9. Equilibrium of Model 2

selected the parameters listed in the table based on their relevance to the research questions and their theoretical significance.

2) Simulation: Based on our simulation results on figure 9, we have discovered that the state probabilities of each user type in the Baidu Netdisk system converge to stable values over time. This indicates that the system reaches a steady-state where the distribution of users across the different types remains relatively constant.

3) Discussion: Our analysis demonstrates that the VIP user type has the highest proportion, followed by P2P users, and finally non-VIP users. This observation suggests that the Baidu Netdisk system may be heavily reliant on VIP users for its revenue stream. Furthermore, the high proportion of P2P users is noteworthy since they typically consume a significant amount of resources in the system, such as network bandwidth and storage space.

VI. CONCLUSION

Our researches have found out the best strategies for Baidu in the case without and with P2P mechanisms. Our findings prove that the existence of P2P users can reduce the bandwidth provided by Baidu Netdisk itself, also show P2P users will not greatly affect the number of VIP users; the application of P2P mechanism will be a win-win strategies for Baidu and users.

Our findings highlight the significance of comprehending the dynamics of user behavior in the Baidu-Netdisk system and the necessity for implementing strategies that can effectively manage user types and maintain a balance between resource consumption and revenue generation.

VII. LIMITATION

A. Limitation of ODE perspective

The current model does not consider the loss caused by the loss of users as we default Baidu as a monopoly company. However, we despise the negative feedback of users after exceeding limit value of price and VIP bandwidth ratio. Moreover, in the reality of the survey, more than 35% of the users indicated that they would not recharge VIP under any circumstances. These phenomena bring changes that current models cannot reflect.

B. Limitation of Markov chain perspective

Utilizing a Markov chain model for the Baidu-Netdisk problem's state transitions between Normal, VIP, and P2P users has its limitations. Firstly, the memoryless property of Markov chains assumes that future states only depend on the current state, not on previous states, which may not accurately represent user behavior influenced by past experiences or cumulative satisfaction. Secondly, the time-dependent transition matrix complicates the model and analysis, which would significantly increase the amount of effort in analyzing. Thirdly, a Markov chain model may oversimplify factors impacting users' decisions to upgrade or switch between user types, as real-world behavior can be influenced by various factors like social network effects, user satisfaction, changing needs, or external events that are not easily captured in the model.

REFERENCES

[1] Manoj Parameswaran, Anjana Susarla, and Andrew B Whinston. P2p networking: an information sharing alternative. *Computer*, 34(7):31–38, 2001.

[2] Jian Tang, Zhihua Xia, Lan Wang, Chengsheng Yuan, and Xueli Zhao. Oppr: An outsourcing privacy-preserving jpeg image retrieval scheme with local histograms in cloud environment. *Journal on Big Data*, 3(1):21, 2021.

[3] Wenying Zeng, Yuelong Zhao, Kairi Ou, and Wei Song. Research on cloud storage architecture and key technologies. In *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, pages 1044–1048, 2009.

[4] Huiyi Zhou, Yong Wang, and Miao Ye. A method of cnn traffic classification based on sppnet. In *2018 14th international conference on computational intelligence and security (CIS)*, pages 390–394. IEEE, 2018.