Dynamic Incentive Mechanism for Industrial Network Congestion Control*

Zhentian WU†(a), Feng YAN††(b), Nonmembers, Zhihua YANG†(c), Member, and Jingya YANG††(d), Nonmember

SUMMARY This paper studies using price incentives to shift bandwidth demand from peak to non-peak periods. In particular, cost discounts decrease as peak monthly usage increases. We take into account the delay sensitivity of different apps: during peak hours, the usage of hard real-time applications (HRAS) is not counted in the user’s monthly data cap, while the usage of other applications (OAS) is counted in the user’s monthly data cap. As a result, users may voluntarily delay or abandon OAS in order to get a higher fee discount. Then, a new data rate control algorithm is proposed. The algorithm allocates the data rate according to the priority of the source, which is determined by two factors: (I) the allocated data rate; and (II) the waiting time.

key words: time dependent pricing, policy control, quality of service (QoS), pricing

1. Introduction

In recent years, multimedia streaming applications, cloud applications that synchronize data among multiple mobile terminals, etc., have been widely deployed and used. These are called capacity intensive applications [1]. Network service providers (NSPs) have to strive to expand their network capacity by investing more infrastructure or new technology. In addition, the fluctuation of all day bandwidth demand has always been a problem for NSP [2], [3]. NSP is under great pressure during the peak period. It needs to provide enough bandwidth for all the traffic in progress and meet its QoS constraints. During the non-peak period, network resources are under utilized, leading to capacity waste. To improve the overall utilization of the network, it is very important for network service providers to transfer the bandwidth demand from peak hour to non-peak hour [4], [5].

2. Related Work

The idea of smart market was first introduced in [6], [7]. In the smart market, prices are adjusted according to congestion. In particular, each user bids for the packets to be transmitted. If the bid value is higher than the congestion cost of the network, the packet will be delivered. Although the dynamic characteristics of smart market are attractive in the sense that it can meet on-demand QoS requirements, there are still many problems to be further studied, including computational complexity, user decision burden and implementation cost [8].

El said et al. [9] provides non-peak discount service. During the peak period, NSPs provide users with preferential charges. Three parameters affecting NSP revenue are studied: (1) the non peak window size during charging discount; (2) the average discount rate of moving a unit amount of traffic load from peak hour to non-peak hour; (3) the percentage of changing peak load due to price incentive. However, due to the lack of mathematical model of demand shift, its contribution is limited. Jiang et al. [10] studies the hourly time-dependent pricing provided by monopolistic and selfish NSPs. On the other hand, if the monopoly NSP has no complete information about the user’s preference for access time, the timeliness pricing of maximizing revenue can also maximize social welfare; on the other hand, if the monopoly NSP has no complete information about the user’s preference for access time. The time-varying pricing of Nash equilibrium is far from the social optimum. Although in [9] congestion is considered in MDE, but the competition among NSPs is not fully studied.

Zhang et al. [11] expands the monopoly scenario to a duopoly one. The competing NSP optimizes its revenue by choosing a context-aware price, while the user considers its time preference, congestion situation and price when choosing NSP. Therefore, the competition among network service providers is also considered. Analysis and simulation results demonstrate that when a single NSP uses time-varying pricing scheme, the other NSP adopts time-varying pricing scheme as a dominant strategy. The sufficient conditions for the existence of Nash equilibrium are established.

Time dependent pricing for the first demo system is named tube [9], where NSP provides the price on the basis of one day ago. The time is divided into 48 periods, each of which is 30 minutes. The unit price is calculated according to the historical load flow. NSP can adjust the price according to the user’s feedback. Based on the analysis of the probability that users transfer the bandwidth demand from peak hour to non-peak hour, the optimal price of bandwidth demand is decided by solving a profit optimization problem. However, the price is selected according to the historical load. When the flow changes rapidly every day, tube can...
3. Questionnaire Survey

In order to estimate user behaviour, this survey collects the following information:

- Basic items: Gender, Age, Region, Occupation, Annual household income
- Usage states: Actual usage states of smart phone, Total communication time per day
- Consciousness: Awareness of communication quality, User behaviour when the communication quality is getting worse
- Behaviour (Time shifting): Communication time that can be shifted in order to get higher communication quality (delay tolerance), Distribution of applications and delay tolerance
- Behaviour (Priority): Applications that users prioritize when the communication quality is poor.

According to the answers, the units of delay tolerance (i.e., second level, minute level, and hour level) are different for different users. Furthermore, there are 92 users, accounting for about 30% of the total, answered that they do not wait. The distribution of the delay tolerance in second level is shown in Fig. 1.

As shown in Fig. 1, the peak occurs at 30 seconds. Furthermore, there are also many users think that their delay tolerance should be within 10 seconds, which is very short.

Figure 2 shows the relationship between the delay tolerance and the total communication time per day. It can be concluded that the delay tolerance is of no correlation to the total communication time per day. Furthermore, the delay tolerance is of no correlation to the characteristics of users such as gender and occupation, either.

By using regression analysis, a logarithm function can be employed to express the quantitative relationship between the cumulative probability and delay tolerance. The contribution ratio (R2) is 0.96, which means the accuracy of the function is good enough. Therefore, this function can be used to model users’ delay tolerance.

From this survey, it can be learnt that LTE has the biggest market share among wireless network services. Furthermore, most people are not satisfied with the current communication quality. Then, for the distribution of delay tolerance, the units of delay tolerance (i.e., second level, minute level, and hour level) are different for different users. In addition, most users think that their delay tolerance should be in second level. The users who said that their delay tolerance should be within 10 minutes account for above 50% of the total. Finally, the applications that users prioritize when congestion happens are in such an order: call, web browsing, e-mail, and SNS. The utilization of streaming service on wireless network is lower than expected. It is supposed that users know the communication quality of wireless network exactly and voluntarily avoid using mobile terminals to watch streaming videos. In this survey, the relationship between the delay tolerance and the characteristics of users is still remained unclear. In other words, the distribution of delay tolerance is independent, and control methods based upon that independent distribution is preferable.

4. Discount Service: Integration of Time-Dependent and App-Based Pricing

4.1 Time-Dependent Pricing Feature

In this letter, time is divided into peak hour and non-peak hour according to congestion state. To utilize the congestion control feature of TDP, fixed rate pricing is used as the benchmark pricing scheme. In particular, users pay a fixed rate on a monthly basis and receive a fee discount if they abandon their use in peak periods. Charging discounts decrease during peak periods as monthly usage increases. Note that in order to improve overall bandwidth utilization, non-peak usage is unlimited. The discount rate of end-user s varies with its monthly usage during peak period. The discount function \( d(v_s) \) expressed by \( v_s \) should meet the following conditions:

If \( v_s = 0 \):

\[
d(v_s) = d_{\text{max}}
\]

If \( 0 < v_s < v_{\text{exp}} \):

\[
d(v_s) = \text{linear function of } v_s
\]
Therefore, users may voluntarily abandon their OAs in order to get higher charge discounts.

5. Resource Management

5.1 Peak-Time Notification

Suppose the logical link of the network is \( L \) (wired or wireless), and the maximum transmission capacity of each link is \( C_l \). The transfer rate of user \( s \) has a range of \([x_s^{\min}, x_s^{\max}]\). Resources are reserved to meet the allocated data rate, and the QoS matrix (for example, throughput or latency) can be limited to a predetermined range [12].

NSPs dynamically monitor traffic on their networks. When congestion occurs, peak hour notifications are sent to the source that will pass through the congestion link. It is expected that users will voluntarily withdraw after receiving peak hour notification to get a higher discount.

5.2 Priority Determination

The time is divided into several periods, and the length of each period is. Each source is characterized by the use of a matrix element \([t_s^0, t_s^1, x_s^{\min}, x_s^{\max}]\). The expected start slot of \( \rho_s^0 \) represents its preferred access time, \( t_s^1 \) is the latest starting position representing its delay tolerance. The minimum transfer rate of \( x_s^{\min} \) and the maximum transfer rate of \( x_s^{\max} \).

On time location \( t \), there is \( S_s(l) \) that will cross link \( l \), where represents the set of HRAs and OAS, respectively. According to the allocated data rate \( x_s \), the priority of the source is given as follows.

\[
p_s^* = \exp(- \frac{x_s - x_s^{\min}}{x_s^{\max} - x_s})
\]

where \( p_s^* \) decreases as a function of the allocated data rate \( x_s \); and varies in the range: (0,1]. The priority of source \( s \in S_s(l) \) base on waiting time is given as follows.

\[
p_s' = 1 - \exp(-\frac{t - \rho_s^0}{t_s^1 - t})
\]

where \( p_s' \) is an increasing function of \( t \). The overall priority of source \( s \) is given as follows.

\[
p_s = p_s^* p_s'
\]

5.3 Transmission Rate Control Algorithm

As shown in Fig. 4, steps involved in bandwidth allocation on each link \( l \in L \) are presented as follows:

Step I: Tentatively, the upper-bound data rate, i.e., \( x_k^{\max} \), is allocated to each source \( s \in S_s(l) \). If \( \sum_{s \in S_s(l)} x_s \leq C_l \), go to Step V; otherwise go to Step II;

Step II: Choose source \( k \) which has the lowest priority in \( S_s(l) \). If \( x_k > x_k^{\min} \), let

\[
x_k = \max\{x_k^{\min}, x_k - \Delta x\}
\]
where $\Delta x$ is the step size. If $\sum_{s \in S(t)} x_s > C_l$, go to Step V; otherwise, repeat Step II for each source in $S(t)$.

**Step III:** Send notification to source $k$ in $S^OA(l)$. If source $k$ voluntarily delay $\Delta t$ for connection, let

$$x_k = 0 \quad (10)$$

If $\sum_{s \in S(l)} x_s \leq C_l$, jump to Step V; otherwise, repeat Step III.

**Step IV:** Find the source $k$ with the lowest priority in $S^OA(l)$. If $t^l_k - t_k > \Delta t$, let

$$x_k = 0 \quad (11)$$

If $\sum_{s \in S(l)} x_s \leq C_l$, jump to Step V; otherwise, repeat Step IV.

**Step V:** Wait until the end of slot $t$, then set

$$t = t + \Delta t \quad (12)$$

and compute the priority of source $s \in S(l)$; jump to Step I.

### 6. Conclusion

In this letter, delay tolerance is considered in such a way that during peak-time, the usage of HRAs is not counted into users’ monthly peak-time usage; while the usage of OAs is counted into the users’ monthly peak-time usage. Therefore, users may voluntarily delay or discard their OAs in order to get higher charge discounts. A novel data rate control algorithm is given. The proposed algorithm allocates transmission rates to sources according to their priorities, which is determined based on two factors: (i) allocated data rate; and (ii) waiting time.

### References

[1] Cisco, “Cisco global cloud index: Forecast and methodology,” Feb. 2018.

[2] Z. Zhou, C. Zhang, J. Wang, B. Gu, S. Mumtaz, J. Rodriguez, and X. Zhao, “Energy-efficient resource allocation for energy harvesting-based cognitive machine-to-machine communications,” IEEE Trans. Cogn. Commun. Netw., vol.5, no.3, pp.595–607, 2019.

[3] B. Gu, X. Zhang, Z. Lin, and M. Alazab, “Deep multi-agent reinforcement-learning-based resource allocation for internet of controllable things,” IEEE Internet Things J., vol.8, no.5, pp.3066–3074, 2021, pp.1–1, 2020.

[4] Z. Zhou, B. Wang, B. Gu, B. Ai, S. Mumtaz, J. Rodriguez, and M. Guizani, “Time-dependent pricing for bandwidth slicing under information asymmetry and price discrimination,” IEEE Trans. Commun., vol.68, no.11, pp.6975–6989, 2020.

[5] B. Gu, X. Yang, Z. Lin, W. Hu, M. Alazab, and R. Kharel, “Multiagent actor-critic network-based incentive mechanism for mobile crowdsensing in industrial systems,” IEEE Trans. Ind. Informat., vol.17, no.9, pp.6182–6191, 2021.

[6] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, “Time-dependent pricing for multimedia data traffic: Analysis, systems, and trials,” IEEE J. Sel. Areas Commun., vol.37, no.7, pp.1504–1517, 2019.

[7] M. Zhang, L. Gao, J. Huang, and M.L. Honig, “Hybrid pricing for mobile collaborative internet access,” IEEE/ACM Trans. Netw., vol.27, no.3, pp.986–999, 2019.

[8] B. Gu and Z. Zhou, “Task offloading in vehicular mobile edge computing: A matching-theoretic framework,” IEEE Veh. Technol. Mag., vol.14, no.3, pp.100–106, 2019.

[9] M. El-Sayed, A. Mukhopadhyay, C. Urrutia-Valdés, and Z.J. Zhao, “Mobile data explosion: Monetizing the opportunity through dynamic policies and qos pipes,” Bell Labs Technical Journal, vol.16, no.2, pp.79–99, 2011.

[10] L. Jiang, S. Parekh, and J. Walrand, “Time-dependent network pricing and bandwidth trading,” NOMS Workshops 2008 - IEEE Network Operations and Management Symposium Workshops, pp.193–200, 2008.

[11] C. Zhang, B. Gu, K. Yamori, S. Xu, and Y. Tanaka, “Oligopoly competition in time-dependent pricing for improving revenue of network service providers with complete and incomplete information,” IEICE Trans. Commun., vol.E98-B, no.1, pp.20–32, 2015.

[12] D. Pan and Y. Yang, “Max-min fair bandwidth allocation algorithms for packet switches,” 2007 IEEE International Parallel and Distributed Processing Symposium, pp.1–10, 2007.