Estimating Traffic Latent Due to QoS Deterioration: A Time-Series Causal Inference Approach

Keisuke ISHIBASHI
International Christian University
Tokyo, Japan
ikeiuske@icu.ac.jp

Takumi UCHIDA
ComWorth Co., Ltd.
Tokyo, Japan

***Abstract*—This paper presents a novel method for calculating the impact of Quality of Service (QoS) degradation on traffic reduction. QoS degradation often results in a decrease in Quality of Experience (QoE), prompting users to abandon applications or compelling applications to decrease content size to avoid user abandonment. In both scenarios, overall traffic is diminished. We define to the potential traffic demand before such reductions as latent traffic, which is crucial to estimate for effective capacity dimensioning and traffic engineering.**

Our estimation of latent traffic employs a causal inference approach, perceiving the traffic reduction due to QoS degradation as a causality from QoS degradation to traffic reduction. However, existing causal inference methods assume a one-way relationship between cause-and-effect variables. In our context, a bidirectional relationship exists, as QoS degradation is often triggered by increased traffic. Here, by assuming these two causal links operate on distinct timescales, then we can isolate them into two one-way relationships to estimate causality. Yet, the estimating of the time series is prone to both bias and variance challenges. We tackle these issues through our proposed multi-source iterative causal inference method. We validate our method's effectiveness through simulations, emphasizing the importance of temporal granularity. Moreover, our results affirm that our method can precisely gauge the latent traffic demand.

***Index Terms*—User Engagement, Quality of Experience, Cause Inference**

I. Introduction

Network congestion often leads to Quality of Service (QoS) degradation, such as packet loss or latency, necessitating an upgrade or rerouting of the congested link. Accurate estimation of the traffic demand directed to the link is crucial for mitigating congestion. However, during periods of congestion, incoming traffic can be throttled by TCP congestion control algorithms. Additionally, QoS degradation, as a result of congestion, can cause users and applications to respond to the congestion [1]–[5].

A significant behavioral change involves users abandoning a service. For instance, users often stop watching videos when playback stalls. Additionally, some service providers reduce their content size to prevent abandonment when QoS degrades [6], [7]. In such scenarios, the observed traffic may diminish compared to the original traffic demand, which we term 'latent traffic demand'. If a network link is upgraded based on the observed traffic volume, the improved QoS may actualize latent traffic demand, potentially leading to recurring congestion and QoS degradation. Consequently, latent traffic demand is critical for effective capacity dimensioning and traffic engineering. However, given that latent traffic demand cannot be directly observed or measured, developing a method for its estimation becomes crucial [8]–[12].

Adopting causal inference techniques appears to be a promising approach [13] to estimate the cause-effect relationship from QoS degradation to traffic reduction. However, we encounter a challenge: QoS often degrads due to an increase in traffic, rendering the two variables cyclically interrelated while existing causal inference techniques generally assume acyclicity. Here, we anticipate that the time scales of the aforementioned relationship, traffic increase to QoS degradation and QoS degradation to traffic reduction, are different. The former typically falls within the sub-second range, as suggested by the actual measurement results of round-trip time [14]. The latter, being the sum of human recognition delay and reaction delay, is anticipated to exceed a second [15].

By analyzing the time series with appropriated temporal granularity and categorizing the values of each time slot as variables, we can separate the two causality relationships and identify the QoS degradation effect while assuming acyclicity among traffic and QoS time series variables.

In this paper, we adopt a time-series causal inference approach [17]–[20]. Specifically, we model the interaction among traffic, QoS, and their auto-correlation as a time-series structural model. Then we estimate causality as model coefficients through multivariate time-series regression analysis, which introduce challenges related to bias and variance. To address this, we also propose an iterative estimation method and propose to observe traffic individually by source at the bottleneck for stable causality estimation. We then assess our method using traffic and QoS time series data synthesized with network simulations, finding that the appropriate timescale for detecting QoS degradation depends on TCP flow duration.

Our paper makes the following contributions:

- We introduce the Iterative Multi-Source Estimation (IMSE) method as well as other regression methods to estimate the causal effect of QoS degradation on traffic reduction, aiding in understanding user/application impact and estimating latent traffic demand.
- We establish the spatio-temporal observation conditions and granularity for estimating QoS degradation effects.

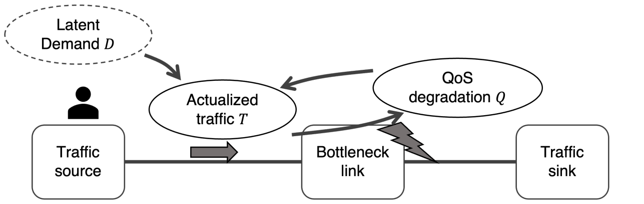

Fig. 1: The underlying system that generates traffic and QoS time series.

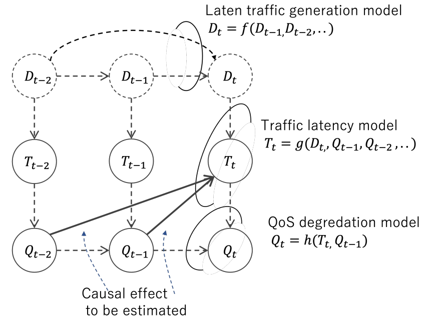

Fig. 2: Structural Causal Model (SCM) among latent traffic demand, actualized traffic and QoS degradation

- We assess our proposed method and conditions using simulations and identify the suitable QoS degradation times.

II. Causal Structure Model for the Causality from QoS Degradation to Traffic Reduction

We introduce a model that explains the causal relationship between QoS (Quality of Service) and traffic observed at a bottleneck link (Fig. 1). In our model, traffic is categorically divided into latent traffic demand and actualized traffic. Latent traffic demand represents the original demand from users or applications as defined in section I. On the other hand, actualized traffic is the traffic that is typically observed on the network.

When QoS deteriorates, latent traffic demand changes. This is due to users or applications reacting to the QoS degradation, resulting in actualized traffic. The causal relationship between actualized traffic and QoS is cyclical. This cycle makes it challenging to estimate the causal effect accurately. To counteract this, our objective is to construct a structural causal model as a multi-variate time series that effectively transforms this cyclic causal effect into an acyclic one.

From here on, we will introduce simplified models to describe the relationships among traffic generation, QoS degradation and user reactions. These models will highlight the possibilities and challenges of estimating causality from QoS degradation to traffic latency using linear regression. Actual relationships among them, specifically loss generation, does not necessarily follow the models and we confirm the applicability of the method in such case in section IV.

A. Structural Causal Model of Traffic / QoS Time Series

We aim to understand the influence of QoS degradation on user traffic by modeling them as discrete time series as follows:

- D_t: Latent traffic demand of users/applications at time step t [byte],
- T_t: Actualized traffic volume flowing to the bottleneck link at time t [byte], and
- Q_t: Degree of QoS degradation at the bottleneck link. In this study, the packet loss ratio serves as the representative metric for QoS degradation.

The temporal granularity of the discrete time series, referred to as δ [s], is assumed to be longer than end-to-end delays, roughly one second. However, δ must not exceed the time scale at which users/applications react to QoS degradation, as will be explained in the next paragraph.

In this paper, we model the causal relationships among these variables as follows (Fig. 2):

$$T_t = f(D_t, Q_{t-l}, \ldots, Q_{t-l-d+1}) + \zeta_t, \tag{1}$$
$$D_t = g(D_{t-1}, \ldots) + \epsilon_t, \tag{2}$$
$$Q_t = h(Q_{t-1}, T_t) \tag{3}$$

Eq. (1) models the generation of T_t as stemming from the latent traffic demand D_t, but is curbed due to the causal effect of QoS degradation Q_{t-k} where $k = l \ldots l + d \quad 1$, l denotes the lag between QoS degradation and user/application response that results in traffic reduction [15], with d representing the duration of this response. Here, ζ_t represents the randomness of the users' response to QoS degradation to actualized traffic. Note that the lag l depends on the temporal granularity δ so that $l\delta$ [s] is the actual lag in seconds and l is the lag counted with δ. Thus, if we change the granularity larger than $l\delta$, the causal effect from QoS degradation to traffic reduction become instantaneous effect, i.e. Eq. (3) includes Q_t, which makes the causal effects Eq. (3) and (2) indistinguishable.

We further model Eq.(1) with a linear model Eq.(4) as follows:

$$T_t = D_t + \sum_{k=l}^{l+d-1} \gamma_k Q_{t-k} + \zeta_t, \tag{4}$$

where γ_k's are the causal effect with coefficients and assumed to be negative, indicating that an increase in QoS degradation leads to a recution in traffic.

Conversely, if we obtain the causal effect γ_k, then we can estimate latent traffic demand D_t with observed actualized traffic demand T_t and the packet loss ratio Q_k as

$$D_t = T_t \quad \sum_{k=l}^{l+d-1} \gamma_k Q_{t-k}, \tag{5}$$

which is the primary goal of this paper.

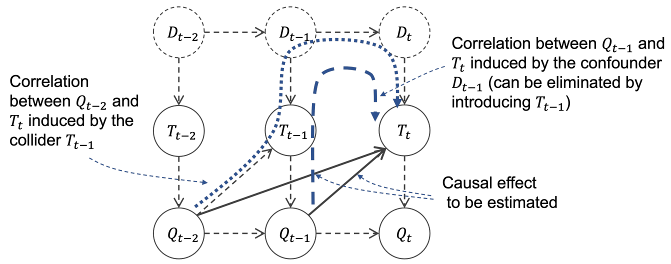

Fig. 3: Bias and variance issues. D_{t-1} is confounder for the causal effect from Q_{t-1} to T_t, whose effect can be eliminated by T_{t-1}. However, T_{t-1} induce a bias in estimating causal effect from Q_{t-2} to T_{t-1}.

Eq. (2) is a time series model for D_t. For example, it can be modeled as the following a linear auto-regressive model with degree p (AR(p) model)

$$D_t = \sum_{k=1}^{p} a_k D_{t-k} + \epsilon_t, \tag{6}$$

where ϵ_t represents the noise term of AR model.

Eq. (3) is the loss generation model. If the packet loss occurs at the single bottleneck queue whose capacity is C [byte/sec] and buffer size is B [byte], Eq. (3) is approximated with Lindley's equation [21] as

$$Q_t = \max(0, Q_{t-1} + T_t \quad C\delta \quad B). \tag{7}$$

III. Estimating Causal Effect

We propose methods to estimate the causal effect parameter γ, considering the relationship among latent traffic demand, actualized traffic generation, and QoS degradation modeled in section II. Our approach uses both linear regression and takes into acolorblackunt potential biases and confounders.

If the QoS degradation model from Eq.(1) can be represented by a linear model from Eq.(4), a straightforward way to estimate γ involves applying linear regression as shown in Eq. (8):

$$T_t = w_0 + \sum_{k=l}^{l+d-1} w_k Q_{t-k} + e_t, \tag{8}$$

where w_0 is the intercept and e_t are the residuals from regression. In this equation, the factor D_k's are not explicitly modeled but are implicitly represented within the w_0 and e_t. Then, the parameter γ_k is estimated as w_k $k = l \dots l + d \quad 1$. We refer to this as Naive Regression (NR) Method.

However, the Naive Regression (NR) Method may lead to biases when estimating the causal effect, when the time series of traffic demand, D_t, demonstrates auto-correlation. As demonstrated in Fig.3, the variable D_{t-1} correlates with D_t and affects both Q_{t-1} and T_t thus acting as a confounding factor. While the ideal solution to remove the bias would involve acolorblackunting for this confounder in the regression model. However, the latent nature of D_{t-1} prevent us from doing so. A possible workaround is to use T_{t-1} as an explanatory variable, given its ability to mitigate the confounding effect by breaking the link between D_{t-1} and T_t in the structural causal model. If we successfully represent Eq. (3), we can estimate the causal effect from Q_{t-1} to T_t as regression coefficient w_k as shown in Eq. (9):

$$T_t = w_0 + \sum_{k=l}^{l+d-1} v_k T_{t-k} + w_k Q_{t-k} + e_t. \tag{9}$$

We label this approach as the Full Regression (FR) Method.

Despite its potential, integrating T_{t-k}'s as explanatory variables introduces two complications. For one, while T_{t-1} proves essential when estimating the causal effect of Q_{t-1} on T_t, it creates a pathway from Q_{t-2} to T_t, rendering T_{t-1} a collider between Q_{t-2} and D_{t-1}, which may reintroduce bias (Fig. 3). Given that T_{t-1} is useful for estimating the causal effect of Q_{t-1} on T_t but detrimental for estimating the causal effect of Q_{t-2} on T_t, we propose an iterative method to separately estimate the causal effect of each $Q_{t-s}, s = l, \dots, l + d \quad 1$ on T_t. In particular, the causal effect from Q_{t-s} is estimated as regression coefficient w_s in the following regression model:

$$T_t = w_0 + v_s T_{t-s} + \sum_{k=l}^{l+d-1} w_k Q_{t-k} + e_t. \tag{10}$$

In the model above, only the regression coefficients w_s is used for inferring the causal effect from Q_{t-s} to T_t. The parameters w_k for $k \notin s$ are not used as they could be biased. Instead, these effects are individually estimated by varing $s = l \dots l + d \quad 1$ as the regression coefficient of the corresponding regression model with the explanatory variable T_{t-s} in Eq. (10).

Another concern involves the potential for increased variance in parameter estimation when introducing T_{t-k}'s as explanatory variables. This variance is notable because the system in Eq. (7) is deterministic, and Q_{t-1} is perfectly determined given $T_{t-k}, k = 1, \dots, r(t)$. Perfect correlations between these variables can lead to multi-collinearity problem, which could make estimation results highly unstable [16].

By analyzing each traffic source as an independent contributor to congestion at the bottleneck, as opposed to treating all traffic as a homogenized flow into the bottleneck link, we can mitigate this multi-collinearity issue. This viewpoint establishes a context where an increment in singular traffic source does not perfectly determine the extent of QoS degradation. If we consider n independent traffic sources flowing through the bottleneck link, then the variables D_t and T_t in Eqs. (4), (6), and (7) are indexed with a superscript $i, i = 1, \dots, n$ to specify the traffic source. Moreover Q_t is determined by the sum of all traffic sources, expressed as:

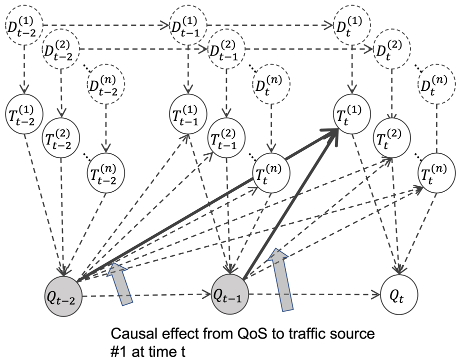

Fig. 4: Structural Causal Model (SCM) for multiple traffic sources. It should be noted that these traffic sources share bottleneck resources and experience the same QoS degradation.

$$D_t^{(i)} = \sum_{k=1}^{p} a_k D_{t-k}^{(i)} + \epsilon_t \tag{11}$$

$$T_t^{(i)} = D_t^{(i)} + \sum_{k=l}^{l+d-1} \gamma_k Q_{t-k} + \zeta_t. \tag{12}$$

$$Q_t = \max(0, Q_{t-1} + \sum_{i=1}^{n} T_t^{(i)} \quad C\delta \quad B) \tag{13}$$

To determine the causal effect, a distinct regression is carried out for each traffic source i and each time slot s, resulting oin w_s as:

$$T_t^{(i)} = w_0^{(i)} + v_s T_{t-s}^{(i)} + \sum_{k=l}^{l+d-1} w_k^{(i)} Q_{t-k} + e_t^{(i)}. \tag{14}$$

Subsequently, the average of w_s^i across all traffic source i is computed, denoted as w_s, and is used as the estimate of the causal effect γ_s. The approache is termed the Iterative Multi-Source Estimation (IMSE) method.

To summarize, using the IMSE method, the latent traffic demand is estimated as:

1) Measure traffic volume $T_t^{(i)}$ and QoS degradation (packet loss ratio) $Q_t^{(i)}$ for traffic source $i, i = 1, \ldots, n$ with interval δ [s].
2) For each $i, i = 1, ..n$ and $s = l \ldots, l + d$, perform regression using Eq. (14) to determine $w_s^{(i)}$ values. Compute their average as w_s
3) Utilizing computed w_k along with measured $T_t^{(i)}$, $Q_t^{(i)}$, estimate the sum of latent traffic demand $\sum_{i=1}^{n} D_t^{(i)}$ using Eq. (15)

$$\sum_{i=1}^{n} D_t^{(i)} = \sum_{i=1}^{n} T_t^{(i)} + n \left(\sum_{k=l}^{l+d-1} w_k Q_{t-k} \right). \tag{15}$$

In this study, we modeled the latent traffic demand generation using the AR(p) process as outlined in Eq.(6). Moreover, the traffic latency resulting from QoS reduction is represented by a linear model, as denoted in Eq.(4). Given these specifications, a linear regression model is anticipated to be effective. However, should these processes deviate from linearity, non-linear regression techniques, such as the Kernel method, might become promising alternatives.

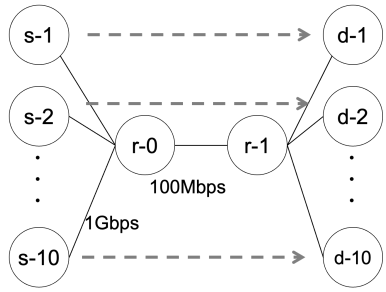

Fig. 5: Network topology for event-driven simulation. The parameters are set so that the average utilization of bottleneck link will be 100 %.

IV. Evaluation

We employ the event-driven simulator, NS-3 [23], to evalute our methods. Firstly, latent traffic demand time series is generated as a Poisson process and their size are calculated acolorblackrding to an AR(1) model, as presented in Eq. (6). Subsequently, actualized traffic was generated as TCP bulk-transfer flows whose sizes are calculated with Eq. (4), that is, if it starts at time t, then its size decreased from the latent traffic demand for the flow acolorblackrding to the packet loss ratio during the time from $t \quad l \quad d + 1$ to $t \quad l$ seconds. In this simulation, although latent traffic demands manifest as flows in the AR(1) process, packets are produced based on TCP congestion control. Packet losses at the bottleneck queue are contingent on the packet arrival patterns and do not strictly adhere to Lindley's equation. We run simulations on a dumbbell topology network, which has source/destination hosts and connects them with a single bottleneck link (Fig. 5). The parameters for simulation are summarized in Table I. We conduct simulations varying the mean of flow interarrival time from 66.7 to 1000 milliseconds so that the ratios of offered traffic compared to the capacity of the bottleneck link are about 10 to 150 %. We set γ_k to be consistent accross all k values and represent the total causal effect $\sum_{k=l}^{l+d-1} \gamma_k$ as $d\gamma$.

Fig 6 shows a sample time series of packet loss ratio, actual traffic volume, and latent traffic demand measure with $\delta = 1$ second time intervals. We can see that when actualized traffic exceeds the link capacity (100 Mbps), then packet loss occurs and after lag $l = 2$ second, the actualized traffic deviates from latent traffic during $d = 8$ seconds. It's worth noting that, in practice, latent traffic demand cannot be directly measured but must be estimated.

We evaluate the possibility of estimating latent traffic demand with observed traffic volume, packet loss ratio, and estimated total causal effect $d\gamma$.

We first compare the accuracy of estimating γ with NR (Naive Regression), FR (Full regression), and IMSE (Iterative

TABLE I: Parameter values specific for TCP simulation.

Parameter	Value(s)
Capacity (access links)	1 Gbps
Capacity (bottleneck link)	100 Mbps
Mean of flow inter arrival time	62.5 - 1000 ms
Mean of TCP flow size	120 KB
Coefficient of variation of TCP flow size	0.5
Degree of AR model p	1
AR parameter α	0.9
Noise for AR model e_t	$\sim \mathcal{N}(0.95C, 0.5C)$
Total Causal effect $d\gamma$	-10
Lag of causal effect l	2
Duration of causal effect d	8
Noise for causal effect η_t	$\sim \mathcal{N}(0, 0.1C)$
Number of traffic sources n	10
Observation period	300 [seconds]

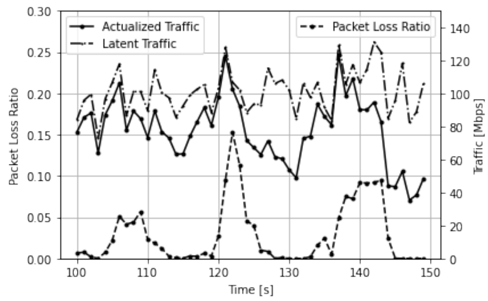

Fig. 6: Sample Time Series of Latent/Actualized Traffic and Packet Loss Ratio

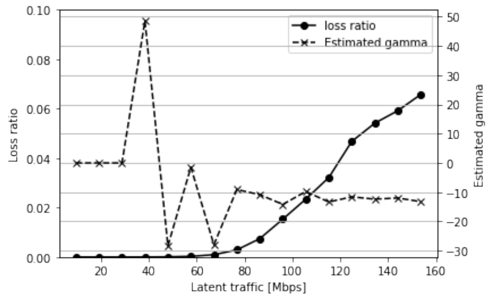

Fig. 8: Loss ratio and estimated . Estimated should be -10 and when latent traffic demand reaches to the link capacity, the estimation will be accurate.

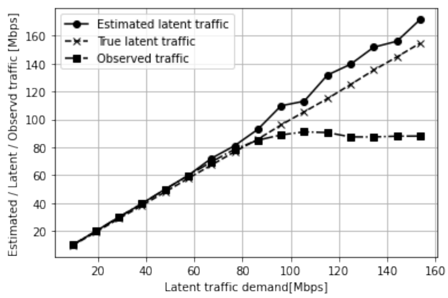

Fig. 9: Estimated latent traffic and observed traffic

Multi-Source Estimation) methods in (14). Fig. 7 shows the normalized mean square errors as Eq. (16) for 10 simulations varying duration from one to eight seconds.

$$\sum_{i=1}^{10} \left(\frac{\hat{\gamma}_i \quad \gamma}{\gamma} \right)^2, \tag{16}$$

where $\hat{\gamma}_i$ is a estimated γ for i-th simulation. As can be seen, for short duration IMSE method achieves higher accuracy compared to NR and FR methods, however, as the duration increases, the errors diminished for all methods. One of the possible reasons is for larger duration, correlation among the flow sizes decreases and confounder effect also decreases. We use IMSE method for the rest of the evaluation.

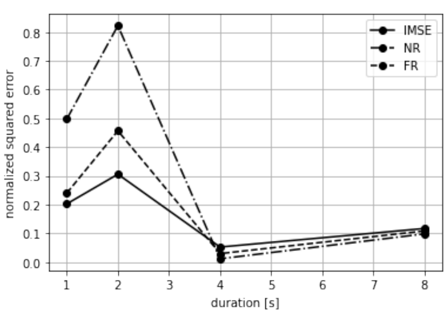

Fig. 7: Normalized mean square errors (NMSE) for estimating with NR, FR, IMSE methods.

Then, we show the results of estimated causal effect by varying the mean flow interarrival time and thus latent traffic arrival rate (Fig. 8). When the rate of latent traffic demand approaches to the link capacity, our method accurately estimate the causal effect ($d\gamma$ = 10). Though in the small traffic rate case, the estimation accuracy degrades, however, in that case, packet loss merely occur and latent traffic and actualized traffic does not deviate and need not to estimate the causal effect.

Figure 9 shows the the estimated latent traffic demand using the estimated causal effect. We can confirm that for a wide range of flow interarrival times, the proposed method can accurately estimate latent traffic demand. An observation is that as the discrepancy between latent traffic and actualized traffic increases, the method somewhat over-estimates the latent traffic demand. We have not yet clarified the reason, but one possible explanation is when the traffic is upper bounded by the link capacity, traffic decrease due to loss can be observed but traffic increase due to the nonexistence of loss cannot be observed. This asymmetry in traffic fluctuation may lead to the overestimation of γ, as well as latent traffic volume.

Note that, even when latent traffic demand exceeds the bottleneck link capacity, observed traffic is upper-bounded with the capacity. This upper-bound is not due to TCP congestion control but due to the impact of QoS degradation. Even in that case, our method can accurately estimate the latent traffic.

We further assess the impact of temporal granularity δ on estimation accuracy, varying it from 100ms to 24 seconds. Figure 10 shows the mean square error of the estimated $d\gamma$.

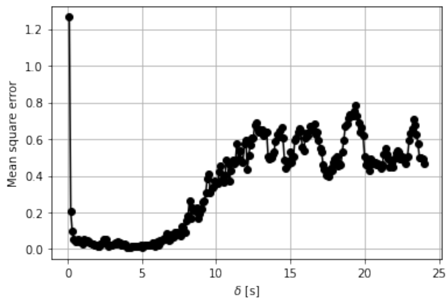

Fig. 10: Mean square errors of inference for different observation time granularity

We can see that with the observation temporal granularity shorter than the QoS degradation lag, we can accurately estimate the QoS degradation parameter $d\gamma$. However, with very short time periods, estimation accuracy is lower. This is because with such small δ, QoS is degraded not only with traffic of the current time slot but also that of past time slots (e.g., an arrow from T_{t-2} to Q_{t-1}). This is due to the time lag between traffic increase and queue length increase, and the past latent traffic again becomes a confounder. Thus, it is safe to set δ larger than the round-trip time, which is the maximum temporal granularity in which traffic increase degrades QoS.

V. CONCLUSION

In this study, we introduced a groundbreaking approach to estimate the causal effect stemming from QoS degradation, leading to traffic reduction as a result of user or application reactions. Given that QoS degradation originates from increased traffic, the inherent causal relationship forms a cycle. Traditional causal inference methods fall short in effectively addressing this cyclical relation. To counter this challenge, we employed a time series causal inference strategy. By adjusting the observation's temporal granularity, we transformed this cyclic relationship into an acyclic one.

Moreover, in recognition of the significant auto-correlation present in both the traffic and QoS time series, we proposed an iterative causal inference approach. Our extensive simulations validated the precision of our technique in discerning the causal effect and latent traffic demand. As we gaze into the future, opportunities arise for extending this methodology to nonlinear models. Integrating Bayesian estimation might also offer a pathway to even more resilient and robust estimations.

REFERENCES

[1] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, "Understanding the Impact of Video Quality on User Engagement," ACM SIGCOMM Computer Communication Review, Vol. 41, Issue 4 pp. 362-–373, August 2011.

[2] Christian Moldovan, Florian Wamser, Tobias Hoßfeld, "User Behavior and Engagement of a Mobile Video Streaming User from Crowdsourced Measurements," QoMEX 2019, Jun. 2019.

[3] S. S. Krishnan and R. K. Sitaraman, "Video Stream Quality Impacts Viewer Behavior," ACM IMC 12, 2012.

[4] M. Katsarakis, R. Teixeira, M. Papadopouli, and V.Christophides, "Towards a causal analysis of Video QoE from Network and Application QoS," Proceedings of the 2016 workshop on QoE-based Analysis and Management of Data Communication Networks, August 2016.

[5] H. Koto, N. Fukumoto, S. Niida, H. Yokota, S. Arakawa, and M. Murata, "Users Reaction to Network Quality During Web Browsing on Smartphones," 6th International Teletraffic Congress (ITC), Sep. 2014.

[6] T. Stockhammer, "Dynamic Adaptive Streaming over HTTP," ACM MMSys'11, pp.133–144, Feb. 2011.

[7] T. Kimura, T. Kimura, A. Matsumoto; K. Yamagishi, "Balancing Quality of Experience and Traffic Volume in Adaptive Bitrate Streaming," IEEE Access, Vol. 9, pp. 15530 – 15547, Jan. 2021.

[8] T. Uchida, K. Ishibashi, and K. Fukuda, "Routing and capacity optimization based on estimated latent od traffic demand, " IEICE Trans. Commun., Vol.E103-B, No. 7, pp.781–790 Jul. 2020.

[9] P. Thakkar, R. Saxena, V. and N. Padmanabhan, "AutoSens: Inferring Latency Sensitivity of User Activity through Natural Experiments," ACM IMC'21, pp.15–21, Nov. 2021.

[10] K. Ishibashi, S. Harada, and R. Kawahara, "Inferring latent traffic demand offered to an overloaded link with modeling QoS-degradation effect," IEICE Trans. Commun., Vol.E102-B, No. 4 Apr. 2019.

[11] K. Ishibashi, S. Harada, and R. Kawahara, "Modeling Upper Layer Reaction to QoS Degradation as a Congestion Avoidance Mechanism," IEICE Transactions on Communications vol. E103-B, No.4, pp.302-311, Apr. 2020.

[12] M. Seufert, S. Wassermann, and P. Casas, "Considering User Behavior in the Quality of Experience Cycle: Towards Proactive QoE-Aware Traffic Management," IEEE Communications Letters, Vol. 23, Issue 7, pp. 1145 – 1148, Jul. 2019.

[13] M.A. Hernán,J.M Robins, *Causal Inference: What If*, Boca Raton: Chapman & Hall/CRC, 2020.

[14] B. Huffaker, M. Fomenkov, D. Moore, and kc claffy, "Macroscopic Analyses of the Infrastructure: Measurement and Visualization of Internet Connectivity and Performance," 2001 USENIX LISA XV, December 2001.

[15] Z. Obrenovic and D. Starcevic, "Modeling Multimodal Human-Computer Interaction," IEEE Computer Vol. 37, Issue 9, September 2004.

[16] Kenneth A. Bollen, *Structural Equations with Latent Variables*, 1989 John Wiley & Sons, Inc.

[17] A. Hyvärinen, K. Zhang, S. Shimizu, and P. O. Hoyer. Estimation of a structural vector autoregression model using non-Gaussianity. Journal of Machine Learning Research, 11:1709-1731, 2010.

[18] J. Peters, D. Janzing, and B. Schölkopf, "Causal Inference on Time Series using Restricted Structural Equation Models" Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS'13), December 2013.

[19] M. Gong, K. Zhang, B. Schölkopf, C. Glymour, D. Tao, "Causal Discovery from Temporally Aggregated Time Series," Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence (UAI), August 2017.

[20] Philipp Geigera, Kun Zhanga,Mingming Gongc, Dominik Janzinga, Bernhard Scholkopf, "Causal Inference by Identification of Vector Autoregressive Processes with Hidden Components,". ICML 2015.

[21] D. V. Lindley, "The theory of queues with a single server," Mathematical Proceedings of the Cambridge Philosophical Society , Vol. 48 , Issue 2, pp. 277 - 289, Apr. 1952.

[22] B. Chen, and J. Pearl, "Graphical Tools for Linear Structural Equation Modeling," Technical Report R-432, Department of Computer Science, University of California Los Angels, CA, 2014.

[23] G. F. Riley and T. R. Henderson, "The ns-3 Network Simulator, " Modeling and Tools for Network Simulation, Springer, Berlin, Heidelberg. pp 15–34, 2010.