A Covert Queueing Channel
in FCFS Schedulers

AmirEmad Ghassami and Negar Kiyavash

Abstract
We study covert queueing channels (CQCs), which are a kind of covert timing channel that may be exploited in shared queues across supposedly isolated users. In our system model, a user modulates messages to another user via his pattern of access to the shared resource scheduled in a first-come-first-served (FCFS) manner. One example of such a channel is the cross-virtual network covert channel in data center networks resulting from the queueing effects of the shared resource. First, we study a system comprising a transmitter and a receiver that share a deterministic and work-conserving FCFS scheduler, and we compute the capacity of this channel. We also consider the effect of the presence of other users on the information transmission rate of this channel. The achievable information transmission rates obtained in this study demonstrate the possibility of significant information leakage and great privacy threats brought by CQCs in FCFS schedulers.

Index Terms
Covert Queueing Channel, First-Come-First-Served Scheduler, Capacity Limit.

I. INTRODUCTION
The existence of side and covert channels due to the fragility of isolation mechanisms is an important privacy and security threat in computer networks. Such channels may be created across users which were supposed to be isolated, resulting in information leakage. By definition, a covert channel is a hidden communication channel which is not intended to exist in the system and is created furtively by users [2]. Covert channels may be exploited by a trusted user, or possibly a malware inside a system with access to secret information to leak it to a distrusted user. On the other hand, in a side channel a malicious user attempts to learn private information by observing information not intended for him. In this scenario, there is no collaboration between the source of information and the recipient [3].

AmirEmad Ghassami is with the Department of Electrical and Computer Engineering and, the Coordinated Science Lab, University of Illinois at Urbana-Champaign, Urbana, IL, email: ghassam2@illinois.edu. Negar Kiyavash is with the Department of Electrical and Computer Engineering, Department of Industrial and Enterprise Systems Engineering and, the Coordinated Science Lab, University of Illinois at Urbana-Champaign, Urbana, IL, email: kiyavash@illinois.edu. The results in this paper were presented in part at the 2015 IEEE International Symposium on Information Theory, Hong Kong, June 14 - June 19, 2015 [1].
A special case of covert and side channels is a timing channel in which information is conveyed through timing of occurrence of events (e.g., inter-arrival times of packets). For instance, queueing covert/side timing channels may arise between users who share a packet scheduler in a network.

Packet schedulers serve packets from multiple streams which are queued in a single queue. This causes dependencies between delays observed by users. Particularly, the delay that one user experiences depends on the amount of traffic generated by other streams, as well as his own traffic. Hence, a user can gain information about other users' traffic by observing delays of his own stream. This dependency between the streams can breach private information as well as create hidden communication channels between the users.

One example of a covert/side queueing channel is the cross-virtual network covert channel in data center networks and cloud environments. In recent years, migrating to commercial clouds and data centers is becoming increasingly popular among companies that deal with data. The multi-tenant nature of cloud and sharing infrastructure between several users has made data protection and avoiding information leakage a serious challenge in such environments [4]. In data center networks, software-defined-networks are frequently used for load balancing [5]. This generates logically isolated virtual networks and prevents direct data exchange. However, since packet flows belonging to different VNs inevitably share underlying network infrastructure (such as a router or a physical link), it is possible to transfer data across VNs through timing channels resulting from the queueing effects of the shared resource(s).

In this paper, we study covert queueing channels (CQCs) in a shared deterministic and work-conserving first-come-first-served (FCFS) scheduler. We present an information-theoretic framework to describe and model the data transmission in this channel and calculate its capacity. First, we consider the two users setting and then study the effect of the presence of a third user on the information transmission rate. The approach for analyzing the effect of the presence of the third user may be extended to calculate the capacity of the covert queueing channel serving any number of users.

The rest of this paper is organized as follows: We review related works in Section II. In Section III, we describe the system model. The capacities of the introduced channel for the two and three user cases are calculated in Sections IV and V respectively. Our concluding remarks are presented in Section VI.

II. RELATED WORKS

The existing literature on covert/side timing channels has mainly concentrated on timing channels in which the receiver/adversary has direct access to the timing sequence produced by the transmitter/victim or a noisy version of it. However, in a covert/side queueing channel, the receiver/adversary does the inference based on the timing of his own packets which has been influenced by the original stream.

In a queueing side channel, where a malicious user, called an attacker, attempts to learn another user’s private information, the main approach used by the attacker is traffic analysis. That is, the attacker tries to infer private information from the victim’s traffic pattern. The attacker can have an estimation of the features of the other user’s
stream such as packet size and timing by emitting frequent packets in his own sequence. Previous work shows that through traffic analysis, the attacker can obtain various private information including visited web sites [6], sent keystrokes [7], and even inferring spoken phrases in a voice-over-IP connection [8].

In [9], Gong et al. proposed an attack where a remote attacker learns about a legitimate user’s browser activity by sampling the queue sizes in the downstream buffer of the user’s DSL link. The information leakage of a queuing side channel in an FCFS scheduler is analyzed in [10]. The analysis of more general work-conserving policies has been done in [11] and [12]. The authors in [12] present an analytical framework for modeling information leakage in queuing side channels and quantify the leakage for several common scheduling policies.

Most of the work in covert timing channels is devoted to the case in which two users communicate by modulating the timings, and the receiver sees a noisy version of the transmitter’s inputs [13]–[21]. Also, there are many works devoted to the detection of such channels [14], [22], [23]. The setup of CQC is new in the field of covert channels and as far as the authors are aware, there are very few works on this setup [24], [25].

### III. System Description

Consider the architecture depicted in Figure 1. In this model, a scheduler serves packets from 2 users: $U_e$ and $U_d$. Each user, $U_i$, $i \in \{e, d\}$, is modeled by a transmitter and a receiver node, denoted by $U_i^T$ and $U_i^R$, respectively. $U_i^R$ is the node which receives $U_i^T$’s packet stream. Note that $U_i^T$ and $U_i^R$ could correspond to the uplink and downlink of the same entity. $U_e$ intends to send a message to $U_d$, but there is no direct channel between them. However, since $U_e^T$ and $U_d^T$’s packets share the same queue, $U_e^T$ can encode messages in the arrival times of its packets, which are passed onto $U_d$ via queueing delays. Therefore, a timing channel is created between users via the delays experienced through the coupling of their traffic due to the shared scheduler.

To receive the messages from $U_e$, user $U_d$ sends a packet stream from the node $U_d^T$. He then uses the delays he experiences by receiving the packet stream at $U_d^R$ to decode the message. Therefore, effectively, the nodes $U_e^T$
Fig. 2: An example of the input and output streams of the FCFS scheduler serving two users. Red packets belong to $U_e$ and blue packets belong to $U_d$. We assume that one packet is buffered in the queue at time $A_i$.

and $U^R_e$ are on the encoder side and the nodes $U^T_d$ and $U^R_d$ are on the decoder side of the channel of our interest. Throughout the paper, we call $U_d$’s sent stream the *probe stream*.

We consider an FCFS scheduler, which is commonly used in DSL routers. We assume this scheduler is deterministic and work-conserving. Time is discretized into slots, and the scheduler is capable of processing at most one packet per time slot. At each time slot, each user either issues one packet or remains idle. Furthermore, we assume that all packets are the same size.

Figure 2 shows an example of the input and output streams of the system depicted in Figure 1 with an FCFS scheduler. In this figure, the first stream is the arrival stream i.e., arrivals from both $U^T_e$ and $U^T_d$, depicted by red and blue, respectively. The second stream is the output stream of user $U_e$ (received by $U^R_e$), and the third one is the output stream of user $U_d$ (received by $U^R_d$). In this example, we assume that one packet is buffered in the queue at time $A_i$, where a packet arrives from both $U^T_d$ and $U^T_e$. If user $U_e$ had not sent the two packets (depicted in red), the second packet of user $U_d$ which arrives at time $A_{i+1}$ could have departed one time slot earlier. Therefore, $U_d$ knows that $U_e$ has issued two packets.

Throughout the paper, we assume that the priorities of the users are known. Particularly, without loss of generality, we assume that $U_d$ has the highest priority among all users; i.e., in the case of simultaneous arrivals, $U_d$’s packet will be served first.

As mentioned earlier, at each time slot, each user is allowed to either send one packet or none; hence, the input and output packet sequences of each user could be viewed as a binary bitstream, where ‘1’ and ‘0’ indicates whether a packet was sent or not in the corresponding time slot.
Assume message $W$ drawn uniformly from the message set $\{1, 2, ..., M\}$ is transmitted by $U^T_e$, and $\hat{W}$ is $U_d$'s estimate of the sent message. Our performance metric is the average error probability, defined as follows:

$$P_e \triangleq P(W \neq \hat{W}) = \sum_{m=1}^{M} \frac{1}{M} P(\hat{W} \neq m | W = m).$$

$U_e$ encodes each message into a binary sequence of length $n$, $\Delta^n$, to create the codebook, which is known at the decoder, $U_d$.

In order to send a message, $U^T_e$ emits a packet in the $i$-th time slot if $\Delta_i = 1$ and remains idle otherwise, i.e.,

$$\Delta_i = \begin{cases} 
1 & \Rightarrow U^T_e \text{ issues a packet in time slot } i. \\
0 & \Rightarrow U^T_e \text{ remains idle in time slot } i. 
\end{cases}$$

To decode this message, $U^R_d$ sends a binary length $n$ stream (the probe stream) to the scheduler during the same length $n$ time period. User $U_d$ will use this stream and the response stream received at node $U^R_d$ to decode the sent message.

We define the code, rate of the code, and the channel capacity similar to the definitions in [13], [26] and [27], as follows:

**Definition 1.** An $(n, M, \epsilon)$-code consists of a codebook of $M$ equiprobable binary codewords, where messages take on average $n$ time slots to be received, and the error probability satisfies $P_e \leq \epsilon$.

**Definition 2.** The information transmission rate, $R$, of a code is the amount of conveyed information (logarithm of the codebook size) normalized by the average number of used time slots for the message to be received, i.e.,

$$R = \frac{\log M}{n}.$$ 

Rate may be interpreted as the average amount of information conveyed in a single time slot.

**Definition 3. (Channel Capacity)** The Shannon capacity, $C$, for a channel is the maximum achievable rate at which one can communicate through the channel when the average probability of error goes to zero. In other words, $C$ is the suprimum of rates, $R$, which satisfies the following property [27]:

$$\forall \delta > 0, \; \exists (n, M, \epsilon_n)$-code

s.t. \quad \begin{cases} 
\frac{\log M}{n} > R - \delta \\
\epsilon_n \to 0 
\end{cases} \quad \text{as } n \to \infty.$$

The following notations will be used throughout the paper:
- \( r_i \): \( U_d^T \)'s packet rate.
- \( A_i \): Arrival time of the \( i \)-th packet in the probe stream.
- \( D_i \): Departure time of the \( i \)-th packet of the probe stream.

We assume \( m \) packets are sent by \( U_d \) during \( n \) time slots and we have: \( r_d = \lim_{n \to \infty} \frac{m}{n} \).

- \( X_i \): Number of \( U_e \)'s packets sent in the interval \([A_i, A_{i+1})\). Note that \( X_i = \sum_{j=A_i}^{A_{i+1}-1} \Delta_j \).
- \( T_i = A_{i+1} - A_i \): inter-arrival time between \( i \)-th and \((i + 1)\)-th packet of the probe stream. We denote a realization of \( T \) by \( \tau \).
- \( Y_i = D_{i+1} - D_i - 1 \)
- \( \hat{X}_i \): estimation of \( X_i \) by decoder.
- \( \hat{W} \): decoded message.

In an FCFS scheduler, \( U_d \) can have an estimation of the number of the packets of other users between any of his own consecutive packets. The estimation of the number of packets in the interval \([A_i, A_{i+1})\) is accurate if the scheduler is deterministic and work-conserving and a sufficient number of packets is buffered in the queue at time \( A_i \). In that case, the number of other users’ packets arriving in the interval \([A_i, A_{i+1})\) could be simply calculated by \( D_{i+1} - D_i - 1 \). Note that \( U_d \) cannot pinpoint the location of the sent packets; that is, if the inter-arrival time is \( \tau \), \( U_d \) can distinguish between \( \tau + 1 \) different sets of bit streams sent during this time. Therefore, we look at any probe stream sent during \( n \) time slots as a combination of different inter-arrival times.

If the sum of the packet rates of the users used during sending a message of length \( n \) is on average larger than 1, then the message will be arrived on average during more than \( n \) time slots. Also, this will destabilize the input queue of the scheduler. For example, for a system with two users \( U_d \) and \( U_e \), if \( U_d^T \) sends packets in every time slot, then sending a packet by \( U_e^T \) in any time slot would cause a delay in the serving of the next packet of \( U_d^T \) and hence could be detected. Therefore, in each time slot, \( U_e^T \) could simply idle to signal a bit ‘0’ or send a packet to signal a bit ‘1’, resulting in the information rate of \( \frac{1}{10} \) bit per time slot in the case that bits are equiprobable. But, this scheme is not feasible in practice as it would destabilize the queue and result in severe packet drops.

In order to have queue stability, it suffices that the total packet arrival rate does not exceed the service rate, which for a deterministic and work-conserving scheduler is equal to 1 (see Appendix A for the proof of stability which is based on a Lyapunov stability argument for the general case that the serving rate is assumed to be \( 0 \leq \rho \leq 1 \)).

\[ q(A_i) = D_i - A_i - 1 \]

where \( q(A_i) \) denotes the queue length at the time that the \( i \)-th packet in the probe stream arrives at the queue. The extra \( 1 \) in the formula is the time needed for the \( i \)-th packet of the probe stream to be served. Therefore, user \( U_d \) should always be aware of the queue length and keep it sufficiently large by sending extra packets when needed.
and arbitrary number of users is considered. Specifically, for the case of two users we need

\[ r_e + r_d < 1. \] (1)

On the other hand, if the sum of the packet rates of the users used during sending a message of length \( n \) is on average less than 1, the length of the input queue may go to zero, and consequently \( U_d \) may not be able to count the number of packets of other users correctly. Note that increasing \( r_d \) increases the resolution available for user \( U_d \) and hence this user can have a better estimation of the number of other users’ sent packets. Therefore, in the case of two users, in order to achieve the highest information rate, the operation point should tend to the line \( r_e + r_d = 1 \). Therefore, we focus on the coding schemes where the sum of the rates is held at 1, with considering a preamble stage in our communication to guarantee sufficient queue length.

IV. TWO-USER CASE

In this section, using achievability and converse arguments, the capacity of the introduced system is calculated for a system with a deterministic and work-conserving FCFS scheduler serving packets from two users.

As depicted in Figure 1, user \( U_e \) is attempting to send a message to \( U_d \) through the covert queueing channel between them. Note that since we have considered service rate of 1 for the FCFS scheduler and users can agree on the packet stream sent by \( U_d^T \) ahead of time, the feedback \( U_e^R \) is already available at the encoder. Therefore, the following Markov chain holds

\[ W \rightarrow X^m \rightarrow Y^m \rightarrow \hat{W}. \] (2)

Note that as mentioned earlier, if there is a sufficient number of packets buffered in the shared queue, \( \hat{X}_i \) could be accurately estimated as \( Y_i \).

The main result of this section is the following theorem, the proof of which is developed in the rest of the section.

**Theorem 1.** The capacity of the timing channel in a shared FCFS scheduler with service rate 1 depicted in Figure 7 is equal to 0.8114 bits per time slot, which can be obtained by solving the following optimization problem:

\[
C = \sup_{\alpha, \gamma_1, \gamma_2} \alpha \tilde{H}(\gamma_1, 1) + (1 - \alpha) \tilde{H}(\gamma_2, \frac{1}{2})
\]

s.t.

\[
\alpha(\gamma_1 + 1) + (1 - \alpha)(\gamma_2 + \frac{1}{2}) = 1,
\]

where \( 0 \leq \alpha \leq 1 \) and \( 0 \leq \gamma_1, \gamma_2 \leq \frac{1}{2} \) and the function \( \tilde{H} : [0, 1] \times \left\{ \frac{1}{k} : k \in \mathbb{N} \right\} \rightarrow [0, 1] \) is defined as:

\[
\tilde{H}(\gamma, \frac{1}{k}) = \frac{1}{k} \sup_{X \in \{0, 1, \ldots, k\}} H(X), \quad k \in \mathbb{N}, 0 \leq \gamma \leq 1.
\] (4)

We first investigate some of the properties of the function \( \tilde{H} \).
**Lemma 1.** Let \( U_k \sim \text{unif}(\{0, 1, \ldots, k\}) \). The distribution which achieves the optimum value in (4) is the tilted version of \( \text{unif}(\{0, 1, \ldots, k\}) \) with parameter \( \lambda \):

\[
P_X(i) = \frac{e^{i\lambda}}{\sum_{i=0}^k e^{i\lambda}}, \quad i \in \{0, 1, \ldots, k\},
\]

where \( \lambda = (\psi'_{U_k})^{-1}(k\gamma) \), where the function \( \psi'_{U_k}(\cdot) \) is the derivative of the log-moment generating function of \( U_k \).

See Appendix B for the proof of Lemma 1.

**Lemma 2.** The function \( \tilde{H} \) could be computed using the following expression:

\[
\tilde{H}(\gamma, \frac{1}{k}) = \frac{1}{k} [\log_2 (k+1) - \psi^*_{U_k}(k\gamma) \log_2 e],
\]

where \( U_k \sim \text{unif}(\{0, 1, \ldots, k\}) \), and the function \( \psi^*_{U_k}(\cdot) \) is the rate function given by the Legendre-Fenchel transform of the log-moment generating function,

\[
\psi^*_{U_k}(\gamma) = \sup_{\lambda \in \mathbb{R}} \{\lambda \gamma - \psi_{U_k}(\lambda)\}.
\]

In order to prove this lemma, first we note that for any random variable \( X \) defined over the set \( \{0, 1, \ldots, k\} \),

\[
H(X) = \sum_{i=0}^k P_X(i) \log \frac{1}{P_X(i)}
\]

\[
= \sum_{i=0}^k P_X(i) \log (k+1) - \sum_{i=0}^k P_X(i) \log \frac{P_X(i)}{k+1}
\]

\[
= \log (k+1) - D(P_X||U_k),
\]

where \( D(P_X||U_k) \) denotes the KL-divergence between \( P_X \) and \( U_k \). Therefore, in order to maximize \( H(X) \), we need to minimize \( D(P_X||U_k) \). Using the following well-known fact [28] concludes the lemma.

\[
\min_{\mathbb{E}[X]=k\gamma} D(P_X||U_k) = \psi^*_{U_k}(k\gamma) \log_2 e.
\]

Figure 3 shows the function \( \tilde{H}(\gamma, \frac{1}{k}) \) for different values of \( \gamma \) and \( k \in \{1, 2, 3\} \).

**Lemma 3.** The function \( \tilde{H}(\cdot, \cdot) \) is concave in pair \( (\gamma, \frac{1}{k}) \) in the sense that for integers \( k_1, k_2, k_3 \), and for values \( 0 \leq \gamma_1, \gamma_2, \gamma_3 \leq 1 \), and for \( \alpha \in [0, 1] \) such that \( \alpha(\gamma_1, \frac{1}{k_1}) + (1-\alpha)(\gamma_3, \frac{1}{k_3}) = (\gamma_2, \frac{1}{k_2}) \), we have:

\[
\alpha \tilde{H}(\gamma_1, \frac{1}{k_1}) + (1-\alpha) \tilde{H}(\gamma_3, \frac{1}{k_3}) \leq \tilde{H}(\gamma_2, \frac{1}{k_2}).
\]

See Appendix C for the proof of Lemma 3.

Substituting (5) in (3) and solving it, the capacity of the timing channel in the shared FCFS scheduler with service rate 1 depicted in Figure 1 is equal to 0.8114 bits per time slot, achieved by \( \alpha = 0.177 \), \( \gamma_1 = 0.43 \) and \( \gamma_2 = 0.407 \).
Fig. 3: $\tilde{H}(\gamma, \frac{1}{k})$ for different values of $\gamma$ and $k \in \{1, 2, 3\}$.

**Lemma 4.** For all $\gamma \in [0, 1]$ and $k \in \mathbb{N}$, we have $\tilde{H}(\gamma, \frac{1}{k}) = \tilde{H}(1 - \gamma, \frac{1}{k})$.

See Appendix D for the proof of Lemma 4.

In the following, the proof of Theorem 1 is given. The proof is based on converse and achievability arguments.

**A. Converse**

In the converse side, the ultimate goal is to prove that

$$C \leq \sup_{\alpha, \gamma_1, \gamma_2} \alpha \tilde{H}(\gamma_1, 1) + (1 - \alpha)\tilde{H}(\gamma_2, \frac{1}{2})$$

s.t.

$$\alpha(\gamma_1 + 1) + (1 - \alpha)(\gamma_2 + \frac{1}{2}) = 1,$$

where $0 \leq \alpha \leq 1$ and $0 \leq \gamma_1, \gamma_2 \leq \frac{1}{2}$. We break the proof into two lemmas.

**Lemma 5.** For the timing channel in a shared FCFS scheduler with service rate 1 depicted in Figure 1, any code consisting of a codebook of $M$ equiprobable binary codewords, where messages take on average $n$ time slots to be received, satisfies

$$\frac{1}{n} \log M \leq \sum_{\tau=1}^{n} [\pi_\tau \tilde{H}(\mu_\tau, \frac{1}{\tau})] + \epsilon_n$$

(9)

where $\sum_{\tau=1}^{n} \pi_\tau (\mu_\tau + \frac{1}{\tau}) = 1$ and for all $\tau$, $0 \leq \mu_\tau \leq \frac{1}{2}$. In this expression, $\epsilon_n = \frac{1}{n} (H(P_e) + P_e \log_2 (M - 1))$, $\pi_\tau$ is the portion of time that user $U_d$ sends packets with inter-arrival time equal to $\tau$ in the probe stream, and $\mu_\tau$ is $U_e^T$'s average packet rate when the inter-arrival time is equal to $\tau$.  

9
Proof. We first we note that
\[
\frac{1}{n} \log M \stackrel{(a)}{=} \frac{1}{n} H(W)
\]
\[
\stackrel{(b)}{=} \frac{1}{n} H(W | \tau^m)
\]
\[
= \frac{1}{n} I(W; \tilde{W} | \tau^m) + \frac{1}{n} H(W | \tilde{W}, \tau^m)
\]
\[
\stackrel{(c)}{\leq} \frac{1}{n} I(W; \tilde{W} | \tau^m) + \epsilon_n
\]
\[
\leq \frac{1}{n} I(X^m; Y^m | \tau^m) + \epsilon_n,
\]
where (a) holds because \( W \) is a uniform random variable over the set of messages \( \{1, \ldots, M\} \), (b) follows from the fact that the chosen message is independent of the inter-arrival time of decoder’s packets, (c) follows from Fano’s inequality with \( \epsilon_n = \frac{1}{n} (H(P_e) + P_e \log_2 (M - 1)) \), and (d) follows from data processing inequality in Markov chain in (2). Therefore
\[
\frac{1}{n} \log M \leq \frac{1}{n} \left[ H(X^m | \tau^m) - H(X^m | Y^m, \tau^m) \right] + \epsilon_n
\]
\[
\leq \frac{1}{n} H(X^m | \tau^m) + \epsilon_n
\]
\[
\leq \frac{1}{n} \sum_{j=1}^{m} H(X_j | \tau^m) + \epsilon_n
\]
\[
\leq \frac{1}{n} \sum_{j=1}^{m} \max_{P_{X_j | \tau^m}} H(X_j | \tau^m) + \epsilon_n.
\]
In the maximization above, the mean of the distribution \( P_{X_j | \tau^m} \) is \( \mathbb{E}[X_j | \tau^m] \). As mentioned in Section III in order to find the maximum information rate while having stability, we are interested in the asymptotic regime in which the operating point is converging to the line \( r_e + r_d = 1 \). Therefore, the information rate is upper bounded by having the set of means, \( \{\mathbb{E}[X_1 | \tau^m], \mathbb{E}[X_2 | \tau^m], \ldots, \mathbb{E}[X_m | \tau^m]\} \) satisfying the constraint \( \frac{1}{n} \sum_{j=1}^{m} \mathbb{E}[X_j | \tau^m] + r_d = 1 \).

Let \( \xi_j = \frac{\mathbb{E}[X_j | \tau^m]}{\tau_j} \). Using (4), we have
\[
\max_{P_{X_j | \tau^m} \mid \mathbb{E}[X_j | \tau^m] = \tau_j \xi_j} H(X_j | \tau^m) = \tau_j \hat{H}(\xi_j, \frac{1}{\tau_j}),
\]
where as mentioned in Lemma 1, the distribution for each \( X_j \) which achieves the maximum value in (10) is the tilted distribution of \( U_{\tau_j} \) with parameter \( \lambda \), such that \( \lambda = \left( \psi'_{U_{\tau_j}} \right)^{-1}(\tau_j \xi_j) \). Therefore, we will have:
\[
\frac{1}{n} \log M \leq \frac{1}{n} \sum_{j=1}^{m} \tau_j \hat{H}(\xi_j, \frac{1}{\tau_j}) + \epsilon_n,
\]
such that the set \( \{\xi_1, \xi_2, \ldots, \xi_m\} \) satisfies the constraint \( \frac{1}{n} \sum_{j=1}^{m} \tau_j \xi_j + r_d = 1 \). The inter-arrival times take values in the set \( \{1, 2, \ldots, n\} \). Therefore, in the summation above we can fix the value of inter-arrival time on the value \( \tau \).
and count the number of times that $\tau_j$ has that value. Defining $m_\tau$ as the number of times that the inter-arrival time is equal to $\tau$ (note that $n = \sum_{\tau=1}^{n} \tau \cdot m_\tau$), we can break the summation above as follows

$$\frac{1}{n} \log M \leq \frac{1}{n} \sum_{\tau=1}^{n} \sum_{k=1}^{m_\tau} \tau \tilde{H}(\mu_{\tau,k}, \frac{1}{\tau}) + \epsilon_n$$

$$= \frac{1}{n} \sum_{\tau=1}^{n} \tau \sum_{k=1}^{m_\tau} \tilde{H}(\mu_{\tau,k}, \frac{1}{\tau}) + \epsilon_n$$

$$= \frac{1}{n} \sum_{\tau=1}^{n} \tau \cdot m_\tau \sum_{k=1}^{m_\tau} \frac{1}{m_\tau} \tilde{H}(\mu_{\tau,k}, \frac{1}{\tau}) + \epsilon_n,$$

where $\mu_{\tau,k}$ is equal to the $k$-th $\xi_j$ which has $\tau_j = \tau$.

By Lemma 3, the function $\tilde{H}(\cdot, \cdot)$ is a concave function of its first argument. Therefore, by Jensen’s inequality,

$$\sum_{k=1}^{m_\tau} \frac{1}{m_\tau} \tilde{H}(\mu_{\tau,k}, \frac{1}{\tau}) \leq \tilde{H}(\mu_{\tau}, \frac{1}{\tau}),$$

(11)

where $\mu_{\tau} = \frac{1}{m_\tau} \sum_{k=1}^{m_\tau} \mu_{\tau,k}$. Using (11) and the equation $n = \sum_{\tau=1}^{n} \tau \cdot m_\tau$, we have:

$$\frac{1}{n} \log M \leq \sum_{\tau=1}^{n} \left[ \frac{\tau \cdot m_\tau}{m_\tau} \tilde{H}(\mu_{\tau}, \frac{1}{\tau}) \right] + \epsilon_n$$

(12)

where $\pi_\tau = \frac{\tau \cdot m_\tau}{m_\tau \sum_{\tau=1}^{n} \tau \cdot m_\tau}$.

The packet rates of the users could be written as follows:

$$r_e = \frac{1}{n} \sum_{j=1}^{m} \tau_j \xi_j = \frac{1}{n} \sum_{\tau=1}^{n} \tau \mu_{\tau,k}$$

$$= \frac{1}{n} \sum_{\tau=1}^{n} \tau m_\tau \frac{1}{m_\tau} \sum_{k=1}^{m_\tau} \mu_{\tau,k} = \frac{1}{n} \sum_{\tau=1}^{n} \tau m_\tau \mu_{\tau}$$

$$= \sum_{\tau=1}^{n} \frac{\tau \cdot m_\tau}{m_\tau \sum_{\tau=1}^{n} \tau \cdot m_\tau} \mu_{\tau} = \sum_{\tau=1}^{n} \pi_\tau \mu_{\tau},$$

and

$$r_d = \sum_{\tau=1}^{n} \pi_\tau \frac{1}{\tau}. $$

Therefore, the constraint could be written as follows:

$$\sum_{\tau=1}^{n} \pi_\tau (\mu_{\tau} + \frac{1}{\tau}) = 1.$$ 

(13)

Suppose the set of pairs $\{ (\mu_{\tau}, \frac{1}{\tau}) \}_{\tau=1}^{n}$ satisfies (13) and maximizes the right hand side of (12). By Lemma 4 there exists another set of pairs $\{ (\hat{\mu}_{\tau}, \frac{1}{\tau}) \}_{\tau=1}^{n}$ with $\hat{\mu}_{\tau}$ defined as

$$\hat{\mu}_{\tau} = \begin{cases} 
\mu_{\tau} & \text{if } 0 \leq \mu_{\tau} \leq \frac{1}{2}, \\
1 - \mu_{\tau} & \text{if } \frac{1}{2} \leq \mu_{\tau} \leq 1,
\end{cases}$$

and

$$\sum_{\tau=1}^{n} \pi_\tau (\hat{\mu}_{\tau} + \frac{1}{\tau}) = 1.$$
that gives the same value for the right-hand side of (12), but it has \( \sum_{\tau=1}^{n} \pi_{\tau} (\hat{\mu}_{\tau} + \frac{1}{\tau}) \leq \sum_{\tau=1}^{n} \pi_{\tau} (\mu_{\tau} + \frac{1}{\tau}) \). Therefore, \( U_d \) can increase his packet rate and increase the information rate using the values \( \hat{\mu}_{\tau} \). Hence, in the maximizing set, for all \( \tau \), we have \( 0 \leq \mu_{\tau} \leq \frac{1}{2} \). Therefore, the optimal operating point will be on the line \( r_e + r_d = 1 \), with \( 0 \leq r_e \leq \frac{1}{2} \) and \( \frac{1}{2} \leq r_d \leq 1 \).

\[ \square \]

Applying Lemma 3 we can replace all pairs of form \((\mu_{\tau}, \frac{1}{\tau})\), \( \tau \geq 2 \), with a single pair of form \((\mu, \frac{1}{2})\):

Lemma 6. For any set of pairs \( S = \{(\mu_{\tau}, \frac{1}{\tau}), \tau \in [n]\} \) where for all \( \tau \), \( 0 \leq \mu_{\tau} \leq \frac{1}{2} \), with weights \( \{\pi_{\tau}, \tau \in [n]\} \) with operating point on the line \( 0 \leq r_e \leq \frac{1}{2} \) and \( \frac{1}{2} \leq r_d \leq 1 \), there exists \( 0 \leq \alpha \leq 1 \) and \( 0 \leq \gamma_1, \gamma_2 \leq \frac{1}{2} \) such that

\[ \alpha (\gamma_1 + 1) + (1 - \alpha) (\gamma_2 + \frac{1}{2}) = 1, \]

and

\[ \sum_{\tau=1}^{n} (\pi_{\tau} \tilde{H}(\mu_{\tau}, \frac{1}{\tau})) \leq \alpha \tilde{H}(\gamma_1, 1) + (1 - \alpha) \tilde{H}(\gamma_2, \frac{1}{2}). \]

Proof. For all \( \tau \in \{3, \ldots, n\} \), there exists \( \beta_{\tau} \in [0,1] \), such that

\[ \beta_{\tau}(\mu_{\tau}, 1) + (1 - \beta_{\tau})(\mu_{\tau}, \frac{1}{\tau}) = (\mu_{\tau}^2, \frac{1}{2}), \]

Clearly, the set \( \{(\mu_1, 1), (\mu_2, \frac{1}{2}), (\mu_3^2, \frac{1}{2}), \ldots, (\mu_n^2, \frac{1}{2})\} \) can also give the same operating point as \( S \) does. By Lemma 3

\[ \beta_{\tau} \tilde{H}(\mu_{1}, 1) + (1 - \beta_{\tau}) \tilde{H}(\mu_{\tau}, \frac{1}{\tau}) \leq \tilde{H}(\mu_{\tau}^2, \frac{1}{2}), \quad \forall \tau \in \{3, \ldots, n\}, \]

Therefore,

\[ \sum_{\tau=1}^{n} \pi_{\tau} \tilde{H}(\mu_{\tau}, \frac{1}{\tau}) = \zeta_1 \tilde{H}(\mu_1, 1) + \zeta_2 \tilde{H}(\mu_2, \frac{1}{2}) + \sum_{\tau=3}^{n} \zeta_{\tau} (\beta_{\tau} \tilde{H}(\mu_1, 1) + (1 - \beta_{\tau}) \tilde{H}(\mu_{\tau}, \frac{1}{\tau})) \]

\[ \leq \zeta_1 \tilde{H}(\mu_1, 1) + \zeta_2 \tilde{H}(\mu_2, \frac{1}{2}) + \sum_{\tau=3}^{n} \zeta_{\tau} \tilde{H}(\mu_{\tau}^2, \frac{1}{2}) \]

\[ \leq \zeta_1 \tilde{H}(\mu_1, 1) + (1 - \zeta_1) \tilde{H}(\frac{\zeta_2 \mu_2 + \sum_{\tau=3}^{n} \zeta_{\tau} \mu_{\tau}^2}{1 - \zeta_1}, \frac{1}{2}), \quad (14) \]

where \( \pi_1 = \zeta_1 + \sum_{\tau=3}^{n} \zeta_{\tau} \beta_{\tau}, \pi_2 = \zeta_2 \) and \( \pi_{\tau} = \zeta_{\tau}(1 - \beta_{\tau}) \) for \( 3 \leq \tau \leq n \) and we have used Lemma 3 again in the last inequality.

\[ \square \]

From Lemmas 5 and 6 we have:

\[ \frac{1}{n} \log M \leq \alpha \tilde{H}(\gamma_1, 1) + (1 - \alpha) \tilde{H}(\gamma_2, \frac{1}{2}) + \epsilon \]

\[ \leq \sup_{\alpha, \gamma_1, \gamma_2} \alpha \tilde{H}(\gamma_1, 1) + (1 - \alpha) \tilde{H}(\gamma_2, \frac{1}{2}) + \epsilon. \]
Letting $n \to \infty$, $\epsilon_n$ goes to zero and we have

$$C \leq \sup_{\alpha, \gamma_1, \gamma_2} \alpha \tilde{H}(\gamma_1, 1) + (1 - \alpha) \tilde{H}(\gamma_2, \frac{1}{2})$$

s.t.

$$\alpha(\gamma_1 + 1) + (1 - \alpha)(\gamma_2 + \frac{1}{2}) = 1,$$

where $0 \leq \alpha \leq 1$ and $0 \leq \gamma_1, \gamma_2 \leq \frac{1}{2}$. This completes the proof of the converse part.

**B. Achievability**

The sequence of steps in our achievable scheme is as follows:

- Set $\alpha = 0.177 - \delta$, for a small and positive value of $\delta$.
- Fix a binary distribution $P_1$ such that $P_1(1) = 0.43$ and $P_1(0) = 0.57$. Generate a binary codebook $C_1$ containing $2^{\alpha n R_1}$ sequences of length $\alpha n$ of i.i.d. entries according to $P_1$.
- Fix a ternary distribution $P_2$ over set of symbols $\{a_0, a_1, a_2\}$ such that $P_2(a_0) = 0.43$, $P_2(a_1) = 0.325$ and $P_2(a_2) = 0.245$. Generate a ternary codebook $C_2$ containing $2^{(1 - \alpha)n R_2}$ sequences of length $\frac{1}{2}(1 - \alpha)n$ of i.i.d. entries according to $P_2$. Substitute $a_0$ with 00, $a_1$ with 10 and $a_2$ with 11, so we will have $2^{(1 - \alpha)n R_2}$ binary sequences of length $(1 - \alpha)n$.
- Combine $C_1$ and $C_2$ to get $C$, such that $C$ has $2^{n(\alpha R_1 + (1 - \alpha) R_2)}$ binary sequences of length $n$ where we concatenate $i$-th row of $C_1$ with $j$-th row of $C_2$ to make the $((i - 1)(2^{(1 - \alpha)n R_2}) + j)$-th row of $C$ (note that $2^{(1 - \alpha)n R_2}$ is the number of rows in $C_2$). Rows of $C$ are our codewords. In above, $n$ should be chosen such that $\alpha n R_1$, $\alpha n$, $(1 - \alpha)n R_2$ and $\frac{1}{2}(1 - \alpha)n$ are all integers.
- **Encoding:** $U_d^T$ sends the stream of all ones (one packet in each time slot) in the first $\alpha n$ time slots and sends bit stream of concatenated 10’s for the rest of $(1 - \alpha)n$ time slots.
- To send message $m$, $U_e^T$ sends the corresponding row of $C$, that is, it sends the corresponding part of $m$ from $C_1$ in the first $\alpha n$ time slots and the corresponding part of $m$ from $C_2$ in the rest of $(1 - \alpha)n$ time slots.
- **Decoding:** Assuming the queue is not empty since there is no noise in the system, the decoder can always learn the exact sequence sent by $U_e$.

Consequently, we will have:

$$C \geq \log_2 \frac{2^{n(\alpha R_1 + (1 - \alpha)R_2)}}{n} = \alpha R_1 + (1 - \alpha) R_2,$$

Since in our achievable scheme, $U_d^T$’s packets are spaced by either one or two time slots, it is enough to have one packet buffered in the queue, where since we are working in the heavy traffic regime, it will not be a problem.
In infinite block-length regime, where $n \to \infty$, we can choose $R_1 = H(P_1)$, $R_2 = \frac{1}{2}H(P_2)$ and find codebooks $C_1$ and $C_2$ such that this scheme satisfies the rate constraint. Therefore,

$$C \geq \alpha H(P_1) + (1 - \alpha)\frac{1}{2}H(P_2).$$

Substituting the values in the expression above, and letting $\delta$ go to zero, we see that the rate 0.8114 bits per time slot is achievable.

V. THREE-USER CASE

As an extension to the basic problem, in this section we consider the case that a third user is also using the shared scheduler. We add user $U_p$ to our basic system model. This user has nodes $U_p^T$ and $U_p^R$ as his transmitter and receiver nodes, respectively (Figure 4). We assume that the node $U_p^T$ sends packets according to a Bernoulli process with rate $r_p$ to the shared scheduler. The shared scheduler is again assumed to be FCFS with service rate 1 and we analyze the capacity for coding schemes satisfying queueing stability condition in the asymptotic regime where the operating point is converging to the line $r_e + r_p + r_d = 1$. Also, in this section we consider the extra assumption that the inter-arrival time of the packets in the probe stream is upper bounded by the value $\tau_{max}$.

Assuming that a sufficient number of packets are buffered in the shared queue, user $U_d$ can still count the number of packets sent by the other two users between any of his own consecutive packets, yet he cannot distinguish between packets sent by user $U_e$ and the packets sent by user $U_p$. Hence, user $U_d$ has uncertainty in estimating the values of $X$. We model this uncertainty as a noise in receiving $X$. Suppose $U_d^T$ sends two packets with $\tau_i = 2$. Each of the other users can possibly send at most 2 packets in the interval $[A_i, A_{i+1})$ and hence, $Y \in \{0, 1, 2, 3, 4\}$. Therefore, we have the channel shown in Figure 5 for this example. In the general case, for the inter-arrival time
\( (1 - r_p)^2 \)

\( (1 - r_p)^2 \)

\( (1 - r_p)^2 \)

Fig. 5: Channel between the encoder and the decoder of the system for the case that the inter-arrival time of two packets of the probe stream is 2.

Fig. 6: The graphical model representing the statistical relation between \( W, X_m, Y_m \) and \( \hat{W} \).

\( \tau \), given \( X = x \), we have \( Y \in \{x + 0, ..., x + \tau\} \) such that

\[
Pr(Y = i + x | X = x) = \binom{\tau}{i} (r_p)^i (1 - r_p)^{\tau - i}, \quad i \in \{0, ..., \tau\},
\]

which is a binomial distribution \( Bin(\tau, r_p) \). Therefore, the support of the random variable \( Y \) is \( \{0, 1, ..., 2\tau\} \). For the mean of \( Y \), we have:

\[
E[Y | \tau] = E[E[Y | X, \tau] | \tau] = E[X + \tau r_p | \tau] = \tau (r_e + r_p).
\]

Because of user \( U_p \)'s stream, the encoder is not aware of the stream received at node \( U_e^R \) beforehand and this output can provide information to the encoder about \( U_e^T \)'s stream. The more packets node \( U_e^T \) sends to the scheduler, the more information this stream contains about \( U_e^T \)'s stream. Using this information, the encoder can have an estimation of the output of the channel at the decoder’s side and hence it could be considered as a noisy feedback to the encoder. Figure 6 shows the graphical model for random variables in our system.
The main result of this section is evaluation of the capacity of the introduced channel, presented in Theorem 2. In the following, the subscript $r_p$ denotes that the calculation is done when the rate of $U_p$ is $r_p$.

**Theorem 2.** If the rate of $U_p$ is $r_p$, the capacity of the timing channel in a shared FCFS scheduler with service rate 1 depicted in Figure 4 is given by

$$C(r_p) = \sup_{\alpha, \gamma_1, \gamma_2, \tau} \alpha \tilde{I}_{r_p}(\gamma_1, \frac{1}{\tau}) + (1 - \alpha) \tilde{I}_{r_p}(\gamma_2, \frac{1}{\tau + 1})$$

s.t.

$$\alpha(\gamma_1 + \frac{1}{\tau}) + (1 - \alpha)(\gamma_2 + \frac{1}{\tau + 1}) = 1 - r_p,$$

where $0 \leq \alpha \leq 1$ and $0 \leq \gamma_1, \gamma_2 \leq 1$ and $1 \leq \tau \leq \tau_{\text{max}} - 1$. The function $\tilde{I}_{r_p} : [0, 1] \times \{\frac{1}{k} : k \in \mathbb{N}\} \mapsto [0, 1]$ is defined as:

$$\tilde{I}_{r_p}(\gamma, \frac{1}{k}) = \frac{1}{k} \sup_{X \in \{0, 1, \ldots, k\}} I_{r_p}(X; Y), \quad k \in \mathbb{N}, 0 \leq \gamma \leq 1.$$  

(17)

The proof is based on converse and achievability arguments. Before giving the proof, we first investigate some of the properties of the function $\tilde{I}$.

**Lemma 7.** The function $\tilde{I}_{r_p}$ could be computed using the following expression:

$$\tilde{I}_{r_p}(\gamma, \frac{1}{k}) = \frac{1}{k} \tilde{H}_{r_p}(\gamma, \frac{1}{k}) - \frac{1}{k} H(\text{Bin}(k, r_p)),$$

(19)

where $\tilde{H}_{r_p}(\gamma, \frac{1}{k}) = \sup_{X \in \{0, 1, \ldots, k\}} H_{r_p}(Y)$ and the second term is the entropy of the binomial distribution with parameters $k$ and $r_p$.

See Appendix C for the proof of Lemma 7.

In order to calculate $\tilde{H}_{r_p}(\gamma, \frac{1}{k})$, the following optimization problem should be solved:

$$\max_{P_X \geq 0} \log_2 e \sum_{i=0}^{2k} P_Y(i) \ln \left( \frac{1}{P_Y(i)} \right)$$

s.t.

$$\begin{cases} \sum_{i=0}^{k} i P_X(i) = \mathbb{E}[X] = k\gamma, \\ \sum_{i=0}^{k} P_X(i) = 1, \end{cases}$$

(20)

where $P_Y = P_X * P_{\text{Bin}}(k, r_p)$, that is,

$$P_Y(i) = \sum_{j=0}^{k} P_X(j) P_{\text{Bin}}(k, r_p)(i-j), \quad i \in \{0, 1, \ldots, 2k\}.$$  

(21)

Figure 4 shows the functions $\tilde{I}_0(\gamma, \frac{1}{k}), \tilde{I}_{0.1}(\gamma, \frac{1}{k})$ and $\tilde{I}_{0.2}(\gamma, \frac{1}{k})$ for different values of $\gamma$ and $k \in \{1, 2, 3\}$.
Fig. 7: $\hat{I}_{r_p}(\gamma, \frac{1}{k})$ for different values of $\gamma$ and $k \in \{1, 2, 3\}$ and $r_p \in \{0, 0.1, 0.2\}$.

Fig. 8: capacity of the timing channel in the shared FCFS scheduler of Figure 4 for different values of $r_p$.

**Lemma 8.** For all $0 \leq r_p \leq 1$, integers $1 \leq k_1, k_2, k_3 \leq \tau_{max}$, values $0 \leq \gamma_1, \gamma_2, \gamma_3 \leq 1$, and $\alpha \in [0, 1]$ such that $\alpha(\gamma_1, \frac{1}{k_1}) + (1 - \alpha)(\gamma_3, \frac{1}{k_3}) = (\gamma_2, \frac{1}{k_2})$, we have:

$$\alpha\hat{I}_{r_p}(\gamma_1, \frac{1}{k_1}) + (1 - \alpha)\hat{I}_{r_p}(\gamma_3, \frac{1}{k_3}) \leq \hat{I}_{r_p}(\gamma_2, \frac{1}{k_2})$$  \hspace{1cm} (22)

See Appendix F for the proof of Lemma 8.

Using the mentioned properties, the capacity of the timing channel in the shared FCFS scheduler of Figure 4 for different values of $r_p$ can be calculated. Figure 8 shows the value of the capacity with respect to $r_p$.

The following proof of Theorem 2 is based on converse and achievability arguments.

**A. Converse**

Suppose the rate of $U_p$ is $r_p$. Similar to the proof of Lemma 5 for the timing channel in a shared FCFS scheduler with service rate 1 depicted in Figure 4, any code consisting of a codebook of $M$ equiprobable binary codewords,
where messages take on average \( n \) time slots to be received, satisfies

\[
\frac{1}{n} \log M \leq \frac{1}{n} I_{r_p}(W; \hat{W}|\tau^m) + \epsilon_n
\]

\[
\leq \frac{1}{n} I_{r_p}(W; Y^m|\tau^m) + \epsilon_n,
\]

where \( \epsilon_n = \frac{1}{n}(H(P_e) + P_e \log_2 (M - 1)) \) and \( (a) \) follows from data processing inequality in the model in Figure 6. Therefore,

\[
\frac{1}{n} \log M \leq \frac{1}{n} \sum_{j=1}^{m} I_{r_p}(W; Y_{j}|Y^{j-1}|\tau^m) + \epsilon_n
\]

\[
\leq \frac{1}{n} \sum_{j=1}^{m} I_{r_p}(W, Y^{j-1}; Y_{j}|\tau^m) + \epsilon_n
\]

\[
\leq \frac{1}{n} \sum_{j=1}^{m} \max_{\pi_{X_j}|\tau^m} I_{r_p}(X_j; Y_{j}|\tau^m) + \epsilon_n,
\]

where \( (a) \) again follows from data processing inequality in the model in Figure 6. In the maximization above, the mean of the distribution \( P_{X_j|\tau^m} \) is \( \mathbb{E}[X_j|\tau^m] \) and in order to find the maximum information rate, the set of means, \( \{\mathbb{E}[X_1|\tau^m], \mathbb{E}[X_2|\tau^m], ..., \mathbb{E}[X_m|\tau^m]\} \), should satisfy the constraint \( r_e + r_p + r_d = 1 \), that is, \( \frac{1}{n} \sum_{j=1}^{m} \mathbb{E}[X_j|\tau^m] + r_p + r_d = 1 \). Let \( \xi_j = \frac{\mathbb{E}[X_j|\tau^m]}{\tau_j} \). Using (18), we have:

\[
\max_{\pi_{X_j|\tau^m}=\tau_j \xi_j} I_{r_p}(X_j; Y_{j}|\tau^m) = \tau_j \tilde{I}_{r_p}(\xi_j, \frac{1}{\tau_j}).
\]

(23)

This implies that

\[
\frac{1}{n} \log M \leq \frac{1}{n} \sum_{j=1}^{m} \tau_j \tilde{I}(\xi_j, \frac{1}{\tau_j}) + \epsilon_n.
\]

Next, using Lemma 8 and similar to the proof of Lemma 5 by breaking the summation, using Jensen’s inequality and the equation \( n = \sum_{\tau=1}^{\tau_{max}} \tau \cdot m_{\tau} \), we will have:

\[
\frac{1}{n} \log M \leq \sum_{\tau=1}^{\tau_{max}} \pi_{\tau} \tilde{I}_{r_p}(\mu_{\tau}, \frac{1}{\tau}) + \epsilon_n,
\]

(24)

where \( \mu_{\tau} \) is the average of \( \xi_j \)'s which have \( \tau_j = \tau \) and \( \pi_{\tau} = \frac{\tau \cdot m_{\tau}}{\sum_{\tau=1}^{\tau_{max}} \tau \cdot m_{\tau}} \). In this expression, \( \pi_{\tau} \) could be interpreted as the portion of time that user \( U_d \) sends packets with inter-arrival time equal to \( \tau \).

Also, using the same approach as the one used in the proof of Lemma 5, the constraint of the problem could be written as follows:

\[
\sum_{\tau=1}^{\tau_{max}} \pi_{\tau} (\mu_{\tau} + \frac{1}{\tau}) = 1 - r_p.
\]

(25)
Suppose the set of pairs \( S = \{(\mu_\tau, \frac{1}{\tau}), \tau \in [\tau_{\text{max}}]\} \) with weights \( \{\pi_\tau, \tau \in [\tau_{\text{max}}]\} \) gives \( \sum_{\tau=1}^{\tau_{\text{max}}} \pi_\tau \tilde{I}_r(\mu_\tau, \frac{1}{\tau}) \) and has its operating point on the line \( r_c + r_d = 1 - r_p \), and we have \( \frac{1}{\tau} \leq \sum_{\tau=1}^{\tau_{\text{max}}} \pi_\tau \frac{1}{\tau} \leq \frac{1}{\tau+1} \), for some \( 1 \leq \tau^* \leq \tau_{\text{max}} - 1 \). We have

\[
\begin{align*}
\left\{ \begin{array}{l}
\beta_\tau(\mu_{\tau+1}, \frac{1}{\tau+1}) + (1 - \beta_\tau)(\mu_\tau, \frac{1}{\tau}) = (\mu_{\tau+1}^*, \frac{1}{\tau+1}) & \tau \leq \tau^* - 1, \\
\beta_\tau(\mu_\tau, \frac{1}{\tau}) + (1 - \beta_\tau)(\mu_\tau, \frac{1}{\tau}) = (\mu_{\tau+1}^*, \frac{1}{\tau+1}) & \tau \geq \tau^* + 2,
\end{array} \right.
\end{align*}
\]

for some \( \beta_\tau \in [0, 1] \). Clearly, the set

\[
\{ (\mu_\tau^*, \frac{1}{\tau}), \ldots, (\mu_{\tau-1}^*, \frac{1}{\tau}), (\mu_{\tau+1}^*, \frac{1}{\tau+1}), (\mu_{\tau+2}^*, \frac{1}{\tau+2}), \ldots, (\mu_{\tau_{\text{max}}+1}^*, \frac{1}{\tau_{\text{max}}+1}) \}
\]

can give the same operating point as \( S \) does. Therefore, using the technique presented in (14) and twice use of Lemma 8 we have

\[
\frac{1}{n} \log M \leq \alpha \tilde{I}_r(p, \gamma_1, \frac{1}{\tau^*}) + (1 - \alpha) \tilde{I}_r(p, \gamma_2, \frac{1}{\tau^* + 1}) + \epsilon_n
\]

\[
\leq \sup_{\alpha, \gamma_1, \gamma_2, \tau} \alpha \tilde{I}_r(p, \gamma_1, \frac{1}{\tau}) + (1 - \alpha) \tilde{I}_r(p, \gamma_2, \frac{1}{\tau + 1}) + \epsilon_n.
\]

Letting \( n \to \infty \), \( \epsilon_n \) goes to zero and we get the desired result.

**B. Achievability**

Achieving the proposed upper bound could be done by a method exactly like the one used in Subsection [IV-B]. We need to solve the optimization problem (17) to find parameters \( \alpha, \gamma_1, \gamma_2 \) and \( \tau \). Because of Lemma 8 in order to find the optimal \( \tau \), we can start with \( \tau = 1 \) and optimize other parameters and then calculate \( \alpha \tilde{I}_r(p, \gamma_1, \frac{1}{\tau}) + (1 - \alpha) \tilde{I}_r(p, \gamma_2, \frac{1}{\tau + 1}) \) and, in each step, increase the value of \( \tau \) by 1, stopping whenever the obtained value is decreased compared to the previous step. For instance, for \( r_p \leq 0.1 \), the optimal \( \tau \) is 1, and hence the procedure stops after checking two steps. After calculating parameters \( \gamma_1, \gamma_2 \) and \( \tau \), the optimal input distribution could be obtained using optimization problem (20).

**VI. Conclusion**

We studied covert queueing channels (CQCs) that can occur through delays experienced by users who are sharing a scheduler. As the scheduling policy plays a crucial role in the possible information transmission rate in this type of channel, we focused on deterministic and work-conserving FCFS scheduling policy. An information-theoretic framework was proposed to derive the capacity of the CQC under this scheduling policy. First, we studied a system comprising a transmitter and a receiver that share a deterministic and work-conserving FCFS scheduler. We obtained the maximum information transmission rate in this CQC and showed that an information leakage rate as high as 0.8114 bits per time slot is possible. We also considered the effect of the presence of other users on the information transmission rate of this channel. We extended the model to include a third user who also uses the shared resource and studied the effect of the presence of this user on the information transmission rate. The solution approach presented in this extension may be applied to calculate the capacity of the covert queueing channel.
among any number of users. The achievable information transmission rates obtained from this study demonstrate the possibility of significant information leakage and great privacy threats brought by CQCs in FCFS schedulers. Based on this result, special attention must be paid to CQCs in high security systems. Finding the capacity of CQCs under other scheduling policies, especially non-deterministic policies, remains to be done in the research area of covert communications and is considered as the main direction for future work. Furthermore, a comprehensive study is required to design suitable scheduling policies that can simultaneously guarantee adequate levels of both security and throughput.

APPENDIX A

PROOF OF STABILITY

For the system model with \( M \) users and service rate \( \rho \), the arrival process has a Poisson binomial distribution with probability mass function

\[
P(K = k) = \sum_{A \in F_k} \prod_{i \in A} r_i \prod_{j \in A^c} (1 - r_j),
\]

with support \( k \in \{0, 1, ..., M\} \), where \( F_k \) is the set of all subsets of \( k \) integers that can be selected from \( \{1, 2, 3, ..., M\} \). The mean of this distribution is \( \mu = \sum_{i=1}^{M} r_i \).

We denote arrival, service and queue length at time \( k \), with \( a(k) \), \( s(k) \) and \( q(k) \), respectively, and we have

\[
q(k + 1) = (q(k) + a(k) - s(k))^+.
\]

Using Foster-Lyapunov theorem with Lyapunov function \( V(q(k)) = (q(k))^2 \) and calculating the drift, we have

\[
\mathbb{E}[q^2(k + 1) - q^2(k)|q(k) = q] \leq \mathbb{E}[(q + a - s)^2 - q^2]
\]

\[
= \mathbb{E}[2q(a - s)] + \mathbb{E}[(a - s)^2],
\]

where \( \mathbb{E}[(a - s)^2] \) is a constant and we denote it by \( K \). Therefore, for some \( \epsilon > 0 \), if \( \mu < \rho \), for large enough value of \( q \), we have

\[
\mathbb{E}[q^2(k + 1) - q^2(k)|q(k) = q] \leq 2q(\mu - \rho) + K \leq -\epsilon,
\]

which implies the stability.

APPENDIX B

PROOF OF LEMMA \(^1\)

In order to find the optimum distribution, \( P_X \), the optimization problem could be written as follows:

\[
\max_{P_X \geq 0} \log_2 \sum_{i=0}^{k} P_X(i) \ln\left(\frac{1}{P_X(i)}\right)
\]

\[
s.t. \quad \left\{ \begin{array}{l}
\sum_{i=0}^{k} i P_X(i) = \mathbb{E}[X] = k\gamma, \\
\sum_{i=0}^{k} P_X(i) = 1,
\end{array} \right. 
\]

\[
(26)
\]
which could be solved using the Lagrange multipliers method. The Lagrangian function would be as follows:

$$\sum_{i=0}^{k} P_X(i) \ln\left(\frac{1}{P_X(i)}\right) + \lambda\left(\sum_{i=0}^{k} iP_X(i) - k\gamma\right) + \rho\left(\sum_{i=0}^{k} P_X(i) - 1\right).$$

Setting the derivative with respect to $P_X(i)$ equal to zero, we get

$$\ln\left(\frac{1}{P_X(i)}\right) - 1 + i\lambda + \rho = 0,$$

which implies that

$$P_X(i) = e^{\rho - 1} \cdot e^{i\lambda}.$$  \hfill (27)

Also, from the second constraint we have

$$\sum_{i=0}^{k} e^{\rho - 1} \cdot e^{i\lambda} = 1 \Rightarrow e^{\rho - 1} = \frac{1}{\sum_{i=0}^{k} e^{i\lambda}}.$$  \hfill (28)

Combining (27) and (28), we have:

$$P_X(i) = \frac{e^{i\lambda}}{\sum_{i=0}^{k} e^{i\lambda}},$$

which is the tilted distribution of $U_k$ with parameter $\lambda$.

In order to calculate $\lambda$, from the first constraint:

$$k\gamma = \sum_{i=0}^{k} iP_X(i) = \sum_{i=0}^{k} i \cdot e^{i\lambda} = \frac{\sum_{i=0}^{k} ie^{i\lambda}}{\sum_{i=0}^{k} e^{i\lambda}} = \frac{\mathbb{E}[U_k e^{U_k\lambda}]}{\mathbb{E}[e^{U_k\lambda}]} = \frac{d}{d\lambda}(\ln \mathbb{E}[e^{U_k\lambda}])$$

$$= \psi'_{U_k}(\lambda).$$

**APPENDIX C**

**PROOF OF LEMMA 3**

First we note that

$$\tilde{H}(\gamma, \frac{1}{k}) = \frac{1}{k}\left[\log_2(k+1) - \psi'_{U_k}(k\gamma) \log_2 e\right]$$

$$= \frac{1}{k}\log_2(k+1) - \frac{1}{k}\sup_{\lambda}\{k\gamma\lambda - \log(\frac{\sum_{i=0}^{k} e^{i\lambda}}{k+1})\} \log_2 e$$

$$= -\sup_{\lambda}\{\gamma\lambda \log_2 e - \frac{1}{k}\log_2(\sum_{i=0}^{k} e^{i\lambda})\}.$$ 

Therefore, if we can show that for any given $\lambda$ the function $h(\gamma, \frac{1}{k}) = \gamma\lambda \log_2 e - \frac{1}{k}\log_2(\sum_{i=0}^{k} e^{i\lambda})$ is convex, then since the supremum of convex functions is convex, we can conclude the desired concavity of the function $\tilde{H}(\cdot, \cdot)$.

Noting that $\frac{1}{k}\log_2(\sum_{i=0}^{k} e^{i\lambda}) = \frac{1}{k}\log_2\left(\frac{1 - e^{(k+1)\lambda}}{1 - e^{\lambda}}\right)$, to prove the convexity of $h(\cdot, \cdot)$, it suffices to prove that the function $g(x) = x \log(\frac{1 - e^{(x+1)\lambda}}{1 - e^{\lambda}})$, $0 < x \leq 1$, is concave. This is true from the concavity of the function $\tilde{g}(x) = \log(\frac{1 - e^{(x+1)\lambda}}{1 - e^{\lambda}})$, and the fact that for any function $f$, $xf(\frac{1}{x})$ is concave if $f(x)$ is concave. The concavity of the function $\tilde{g}(\cdot)$ can be easily seen by taking its second derivative.
APPENDIX D

PROOF OF LEMMA 4

For a given $\gamma$ and support set $\{0, 1, ..., k\}$, suppose the distribution $P^*_X$ is defined over $\{0, 1, ..., k\}$ and has mean $E_{P^*_X}[X] = k\gamma$ and

$$\sup_{X \in \{0, 1, ..., k\} \atop E[X] = k\gamma} \frac{1}{k} H(X) = \frac{1}{k} H(P^*_X).$$

Define distribution $Q_X$ as follows:

$$Q_X(i) = P^*_X(k - i), \quad 0 \leq i \leq k,$$

Therefore the entropy of $Q_X$ will be the same as the entropy of $P^*_X$ and we have

$$\mathbb{E}_{Q}[X] = \sum_{i=0}^{k} i Q_X(i) = \sum_{i=0}^{k} i P^*_X(k - i) = -\sum_{i=0}^{k} (-k + (k - i)) P^*_X(k - i)$$

$$= k - \sum_{i=0}^{k} (k - i) P^*_X(k - i) = k - k\gamma = k(1 - \gamma).$$

Hence, we have

$$\tilde{H}(\gamma, \frac{1}{k}) = \sup_{X \in \{0, 1, ..., k\} \atop E[X] = k\gamma} \frac{1}{k} H(X) = \frac{1}{k} H(P^*_X) = \frac{1}{k} H(Q_X) \leq \sup_{X \in \{0, 1, ..., k\} \atop E[X] = k(1 - \gamma)} \frac{1}{k} H(X) = \tilde{H}(1 - \gamma, \frac{1}{k}). \quad (29)$$

Similarly, suppose for the distribution $Q^*_X$, defined over $\{0, 1, ..., k\}$ and with mean $\mathbb{E}_{Q^*}[X] = k(1 - \gamma)$,

$$\sup_{X \in \{0, 1, ..., k\} \atop E[X] = k(1 - \gamma)} \frac{1}{k} H(X) = \frac{1}{k} H(Q^*_X).$$

Define distribution $P_X$ as follows:

$$P_X(i) = Q^*_X(k - i), \quad 0 \leq i \leq k,$$

Therefore the entropy of $P_X$ will be the same as the entropy of $Q^*_X$ and we have

$$\mathbb{E}_{P}[X] = \sum_{i=0}^{k} i P_X(i) = \sum_{i=0}^{k} i Q^*_X(k - i) = -\sum_{i=0}^{k} (-k + (k - i)) Q^*_X(k - i)$$

$$= k - \sum_{i=0}^{k} (k - i) Q^*_X(k - i) = k - k(1 - \gamma) = k\gamma.$$

Hence, we have

$$\tilde{H}(1 - \gamma, \frac{1}{k}) = \sup_{X \in \{0, 1, ..., k\} \atop E[X] = k(1 - \gamma)} \frac{1}{k} H(X) = \frac{1}{k} H(Q^*_X) = \frac{1}{k} H(P_X) \leq \sup_{X \in \{0, 1, ..., k\} \atop E[X] = k\gamma} \frac{1}{k} H(X) = \tilde{H}(\gamma, \frac{1}{k}). \quad (30)$$

Comparing (29) and (30) gives the desired result.
APPENDIX E

PROOF OF LEMMA 7

\[
\tilde{I}_r(p, \frac{1}{k}) = \sup_{X \in \{0,1,\ldots,k\}} \frac{1}{k} I_r(X;Y) \\
= \sup_{X \in \{0,1,\ldots,k\}} \frac{1}{k} [H_r(Y) - H_r(Y|X)] \\
= \sup_{X \in \{0,1,\ldots,k\}} \frac{1}{k} [H_r(Y) - \sum_{x=0}^{k} P_X(x) H_r(Y|X=x)] \\
\overset{(a)}{=} \sup_{X \in \{0,1,\ldots,k\}} \frac{1}{k} [H_r(Y) - \sum_{x=0}^{k} P_X(x) H(Bin(k,r_p))] \\
= \sup_{X \in \{0,1,\ldots,k\}} \frac{1}{k} H_r(Y) - \frac{1}{k} H(Bin(k,r_p)) \\
= \frac{1}{k} H_r(p, \frac{1}{k}) - \frac{1}{k} H(Bin(k,r_p)),
\]

where (a) follows from (15).

APPENDIX F

PROOF OF LEMMA 8

We first prove that the function \( \tilde{I}(\cdot, \cdot) \) is concave in its first argument. Let \( P^*_X \) and \( P^*_Y \) be the optimum distributions resulted from optimization problem (20) for parameters \((\gamma_1, \frac{1}{k})\) and \((\gamma_3, \frac{1}{k})\), respectively. Therefore for any \( 0 \leq \alpha \leq 1 \),

\[
\alpha \tilde{I}_r(p, \frac{1}{k}) + (1-\alpha) \tilde{I}_r(p, \frac{1}{k}) \\
\overset{(a)}{=} \frac{1}{k} \alpha H(P^*_X \ast Bin(k,r_p)) + \frac{1}{k} (1-\alpha) H(P^*_Y \ast Bin(k,r_p)) - \frac{1}{k} H(Bin(k,r_p)) \\
\overset{(b)}{\leq} \frac{1}{k} H(\alpha(P^*_X \ast Bin(k,r_p)) + (1-\alpha)(P^*_Y \ast Bin(k,r_p)) - \frac{1}{k} H(Bin(k,r_p)) \\
\leq \frac{1}{k} \sup_{X \in \{0,1,\ldots,k\}} \frac{1}{k} H(P_X \ast Bin(k,r_p)) - \frac{1}{k} H(Bin(k,r_p)) \\
= \tilde{I}_r(\alpha \gamma_1 + (1-\alpha) \gamma_3, \frac{1}{k}),
\]

where (a) follows from Lemma 7 and (b) follows from the concavity of the entropy function.
Because of the complexity and lack of symmetry or structure in the function $\tilde{I}$, there is no straightforward analytic method for proving its concavity. But we notice that it is sufficient to show that for all $2 \leq k \leq \tau_{\text{max}} - 1$, and $\alpha$ such that
\[
\alpha \frac{1}{k-1} + (1 - \alpha) \frac{1}{k+1} = \frac{1}{k},
\] we have
\[
\alpha \tilde{I}_r(\gamma_1, \frac{1}{k-1}) + (1 - \alpha) \tilde{I}_r(\gamma_3, \frac{1}{k+1}) \leq \tilde{I}_r(\alpha \gamma_1 + (1 - \alpha) \gamma_3, \frac{1}{k}).
\] (32)

From (31) we have $\alpha = \frac{k - 1}{2k}$, hence using Lemma 7, (32) reduces to
\[
2H_r(\gamma_1, \frac{1}{k}) - H_r(\gamma_1, \frac{1}{k-1}) - H_r(\gamma_3, \frac{1}{k+1}) + f(k, r_p) \geq 0,
\] (33)

where $f(k, r_p) = H(\text{Bin}(k-1, r_p)) + H(\text{Bin}(k+1, r_p)) - 2H(\text{Bin}(k, r_p))$. Noting that the left-hand side is a Lipschitz continuous function of $\gamma_1, \gamma_3$ (away from zero), and $r_p$ and the fact that $k$ takes finitely many values, the validation of inequality (33) can be done numerically.

REFERENCES

[1] A. Ghassami, X. Gong, and N. Kiyavash, “Capacity limit of queueing timing channel in shared FCFS schedulers,” in 2015 IEEE International Symposium on Information Theory (ISIT), pp. 789–793. IEEE, 2015.
[2] V. Gligor, “Covert channel analysis of trusted systems. a guide to understanding,” tech. rep., DTIC, 1993.
[3] S. Kadloor, X. Gong, N. Kiyavash, T. Tezcan, and N. Borisov, “Low-cost side channel remote traffic analysis attack in packet networks,” in Communications, 2010 IEEE International Conference on, IEEE, 2010.
[4] D. Wakabayashi, “Breach complicates Sony’s network ambitions,” The Wall Street Journal, April 28, 2011.
[5] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, “Openflow: enabling innovation in campus networks,” ACM SIGCOMM Computer Communication Review, vol. 38, no. 2, pp. 69–74, 2008.
[6] M. Liberatore and B. N. Levine, “Inferring the source of encrypted http connections,” in Proceedings of the 13th ACM conference on Computer and communications security, pp. 255–263, ACM, 2006.
[7] D. X. Song, D. Wagner, and X. Tian, “Timing analysis of keystrokes and timing attacks on SSH,” in USENIX Security Symposium, vol. 2001, 2001.
[8] C. V. Wright, L. Ballard, S. E. Coull, F. Monrose, and G. M. Masson, “Uncovering spoken phrases in encrypted voice over ip conversations,” ACM Transactions on Information and System Security (TISSEC), vol. 13, no. 4, p. 35, 2010.
[9] X. Gong, N. Borisov, N. Kiyavash, and N. Scheir, “Website detection using remote traffic analysis,” in International Symposium on Privacy Enhancing Technologies Symposium, pp. 58–78, Springer, 2012.
[10] X. Gong, N. Kiyavash, and P. Venkitasubramaniam, “Information theoretic analysis of side channel information leakage in FCFS schedulers,” in Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on, pp. 1255–1259, IEEE, 2011.
[11] S. Kadloor, N. Kiyavash, and P. Venkitasubramaniam, “Mitigating timing based information leakage in shared schedulers,” in INFOCOM, 2012 Proceedings IEEE, pp. 1044–1052, IEEE, 2012.
[12] X. Gong and N. Kiyavash, “Quantifying the information leakage in timing side channels in deterministic work-conserving schedulers,” arXiv preprint arXiv:1403.1276, 2014.
[13] V. Anantharam and S. Verdu, “Bits through queues,” IEEE Transactions on Information Theory, vol. 42, no. 1, pp. 4–18, 1996.
[14] S. Cabuk, C. E. Brodley, and C. Shields, “Ip covert timing channels: design and detection,” in Proceedings of the 11th ACM Conference on Computer and Communications Security, pp. 178–187, ACM, 2004.
[15] S. J. Murdoch and S. Lewis, “Embedding covert channels into TCP/IP,” in International Workshop on Information Hiding, 2005.
[16] D. Llamas, A. Miller, and C. Allison, “An evaluation framework for the analysis of covert channels in the TCP/IP protocol suite,” in ECIS, pp. 205–214, 2005.

[17] M. H. Kang, I. S. Moskowitz, and D. C. Lee, “A network pump,” IEEE Transactions on Software Engineering, vol. 22, no. 5, pp. 329–338, 1996.

[18] S. K. Gorantla, S. Kadloor, N. Kiyavash, T. P. Coleman, I. S. Moskowitz, and M. H. Kang, “Characterizing the efficacy of the nrl network pump in mitigating covert timing channels,” IEEE Transactions on Information Forensics and Security, vol. 7, no. 1, pp. 64–75, 2012.

[19] R. Soltani, D. Goeckel, D. Towsley, and A. Houmansadr, “Covert communications on poisson packet channels,” in Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on, pp. 1046–1052, IEEE, 2015.

[20] R. Soltani, D. Goeckel, D. Towsley, and A. Houmansadr, “Covert communications on renewal packet channels,” in Communication, Control, and Computing (Allerton), 2016 54th Annual Allerton Conference on, pp. 548–555, IEEE, 2016.

[21] P. Mukherjee and S. Ulukus, “Covert bits through queues,” in Communications and Network Security (CNS), 2016 IEEE Conference on, pp. 626–630, IEEE, 2016.

[22] V. Berk, A. Giani, G. Cybenko, and N. Hanover, “Detection of covert channel encoding in network packet delays,” Rapport Technique TR536, de l'Université de Dartmouth, p. 19, 2005.

[23] S. Gianvecchio and H. Wang, “Detecting covert timing channels: an entropy-based approach,” in Proceedings of the 14th ACM conference on Computer and communications security, pp. 307–316, ACM, 2007.

[24] R. Tahir, M. T. Khan, X. Gong, A. Ahmed, A. Ghassami, H. Kazmi, M. Caesar, F. Zaffar, and N. Kiyavash, “Sneak-peek: High speed covert channels in data center networks,” in IEEE International Conference on Computer Communications (INFOCOM), IEEE, 2016.

[25] A. Ghassami, A. Yekkehkhany, N. Kiyavash, and Y. Lu, “A covert queueing channel in round robin schedulers,” arXiv preprint arXiv:1701.08883, 2017.

[26] T. M. Cover and J. A. Thomas, Elements of Information Theory. John Wiley & Sons, 2012.

[27] I. Csiszar and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems. Cambridge University Press, 2011.

[28] Y. Polyanskiy and Y. Wu, “Lecture notes on information theory, MIT (6.441), UIUC (ECE 563),” 2014.