Forty-Ninth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 28 - 30, 2011

The Anonymity of an Almost Fair Chaum Mix

Abhishek Mishra
ECE Department
Lehigh University
Email: abm210@lehigh.edu

Parv Venkitasubramaniam
ECE Department
Lehigh University
Email: parv.v@lehigh.edu

***Abstract*—The first-come-first-serve fair queuing algorithm for a router is known to minimize the per packet delay in a single server queue. The policy however, provides no user anonymity to transmitted packets; mere observation of transmission times can reveal the source of every transmitted packet. The information-theoretic analysis of the anonymity of queuing policies under a relaxation of the First-come-first-serve fair queuing is considered in this work. An entropy-based metric of anonymity is proposed to quantify the anonymity of queuing policy under a fairness relaxation where each packet from a user can be transmitted ahead of at most one packet from another user sharing the mix. Inner and outer bounds on the maximum achievable anonymity are characterized as functions of the available memory at the mix.**

I. Introduction

Privacy in networked communication extends beyond the protection of communicated data; it is equally critical to protect the identities of communicating parties. Anonymous communication systems protect the privacy of the users by hiding who is talking to whom and how packets are traversing the network. These systems, several of them deployed on the Internet, support applications with strong privacy requirements such as e-voting protocols, intelligence gathering for law enforcement, military communications, and such like. The importance of such systems is increasing and the largest deployed anonymity network, Tor [1] has attracted an estimated half a million users.

Most anonymity systems such as Tor are based on the concept of Chaum mixes; a mix is special proxy server that uses re-encryption, random bit padding and batching to provide user anonymity to transmitted packets. Commonly deployed mix-networks, while they provide good protection against packet content/length based information retrieval, are vulnerable to timing analysis of packets. The primary reason for the vulnerability is the lack of optimized mix-network protocols under resource limitations of the network nodes in terms of memory and bandwidth, and QoS requirements such as delay and fairness. Guarding against unauthorized *timing analysis* incurs a penalty in network resources and QoS, and it is imperative to optimize the design of anonymity systems under constraints on resources and QoS requirements.

A key barrier to optimizing the design of anonymous networks thus far is the lack of a quantitative metric to measure anonymity in a network that can take into account the different factors that influence anonymity- namely the resource limitations such as bandwidth, memory, the QoS requirements such as delay, throughput and fairness. In recent work [2], [3], we formulate such metrics to quantify anonymity in an system bounded by buffer limitations and strict delay constraints, and demonstrate the tractability of the metric in optimizing the batching strategies of mix nodes to maximize anonymity within the limitations on delay and memory. In this work, our goal is to understand the relationship between fairness on the achievable anonymity of a Chaum mix.

Fairness has been an important criteria in resource allocation problems, particularly in scheduling processor times at intermediate routers serving multiple input flows, and in the fair allocation of bandwidth. The requirement of fairness in scheduling can, however, be detrimental to anonymity; in other words a tradeoff exists between anonymity and fairness. For instance, consider a single mix serving packets from two users. The fair First-come-First-serve scheduling policy would mandate that packets be released strictly in the order of arrival thus rendering the mix incapable of providing any anonymity. If, however, there were no requirement of fairness, then any number of arrived packets can be shuffled prior to transmission, and by increasing the number of packets shuffled, the uncertainty from the eavesdropper's perspective can be increased significantly. In this work, we propose a *relaxation* of the FCFS fairness, so that at most η packets from one user can be scheduled ahead of a packet from the other user that arrived first, then the options available to the mix to reorder packets would increase the achievable anonymity. In this work, we study the maximum achievable anonymity of an *almost fair* mix, where $\eta = 1$.

A. Related Work

Timing analysis to detect traffic information has roots that go back to the early days of World War II. Its extensive usage in compromising privacy in computer networks is well documented [4], [5]. For example, the weaknesses of Internet protocols such as HTTP [4] and SSH [6] have been exposed through timing analysis. On the Internet, sender anonymity is achieved using networks of Chaum mixes [7]. The basic idea is that each sender picks a sequence of Mixes (deterministically [8] or randomly [9]) to route its data to the destination. Subsequent to the original mixing idea by Chaum, many batching strategies have been proposed to deal with resource constraints such as memory and QoS limitations such as delay, the strategies were based on ad hoc ideas rather than a rigorous quantitative analysis. Although metrics have been proposed

for anonymity, using either the size or Shannon entropy of the anonymity set (set of plausible sources of a packet) [10], [11], these definitions do not reflect the complete information available to an adversary. Furthermore, the limited scope of the definition prevents the analysis of the metric with regard to practical networking constraints.

In previous work [2], [3], we quantified anonymity of a Chaum mix under limitations on buffer and delay. In this work, we adapt the framework to quantify the anonymity of a mix under a fairness restriction.

Fairness is a subjective notion, and FCFS fairness is one popular notion. Another well known definition in the context of scheduling algorithms is max-min fairness, where the minimum data rate provided to a flow is maximized. The fair queuing [12] algorithm is the most popular scheduling algorithm that provides max-min fairness. If in a networking model, all packets are of equal length, fair queuing algorithm turns out to be the round-robin strategy for scheduling. The *Proportional method* [12] of allocating resources is based on yet another notion of fairness, where the resources are shared in proportion to demand with additional constraints of memorylessness and demand monotonicity. But it is not usually preferred in the context of packet scheduling, because it can cause a large amount of delay to a user with small demand.

In this work, we consider an almost fair mix by relaxing the FCFS criterion, and characterize the maximum achievable anonymity under no other resource limitations. We show that the optimal strategy for the mix requires waiting indefinitely for all packets to arrive prior to scheduling them. We then consider a practical alternative, where the mix has a limitation on packet storage prior to scheduling. We provide lower and upper bounds on the maximum achievable anonymity as a function of the memory available to the mix for storage.

II. Mathematical Model

Consider a mix receiving packets from 2 users. Let $X_R(t)$ and $X_B(t)$ denote the arrival processes of the two users respectively (referred to as red(R) and blue(B) packets for convenience), which for the purposes of analysis are modeled as two independent Poisson process with equal rates λ. The Poisson assumption is not critical; as long as every new packet that arrives is equally likely to be from either source, the results and techniques in this paper would be valid. Let $Y(t)$ denote the departure process of the packets, as observed by the eavesdropper (Eve). Since Eve cannot identify the sources of packets on the outgoing process, $Y(t)$ is also a point process.

Almost Fair Mix: The mix uses layered encryption and packet padding to obfuscate the contents and lengths of incoming and outgoing packets. In addition, the mix is allowed to reorder the packets subject to a fairness constraint described below.

Fairness constraint η: Under a fairness constraint η, no packet from one user that arrived more than η packets ahead of a packet from another user is transmitted before that packet. The mix has access to private randomness (unknown to the eavesdropper, Eve) and is allowed to randomize the schedule of packets under the fairness constraint η. Let Ψ_η denote the class of all mixing strategies that satisfy the fairness constraint η.

Eve The eavesdropper, Eve, observes three point processes $X_B(t), X_R(t)$ and $Y(t)$. As mentioned above, due to encryption and padding, the sources of packets on the observed outgoing stream $Y(t)$ are unavailable to Eve. Using her observation, and knowledge of the mixing strategy, her goal is to determine the source identities of each outgoing packet on $Y(t)$. Here it is important to note that even though Eve has knowledge of mix's strategy, she does not know the realization of mix's private randomness. Let Φ denote the complete observation of Eve.

A. Anonymity Definition

Given the observation Φ of Eve, and knowledge of the mixing strategy, the uncertainty in the mix's action would induce an a posterior distribution of the sources of packets on the outgoing stream $Y(t)$ from Eve' perspective. In an outgoing sequence of n packets on $Y(t)$ starting from $t = 0$, let $Y_1, \cdots, Y_n$ denote the random variables that are jointly distributed according to this a posterior distribution ($Y_i \in \{R, B\}$ refers to the source of the i^{th} outgoing packet from Eve's perspective). Then the anonymity of a mixing strategy ψ is defined as:

$$\mathcal{A}^\psi = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[H(Y_1 \cdots Y_n)|\Phi],$$

Entropy is computed for every realization of the processes $(X_R(t), X_B(t), Y(t))$ based on the randomness in the the mix's strategy. The expectation is over the randomness in the arrival process.

For a 2 source mix, it is easy to see that:

$$0 \le \mathcal{A}^\psi \le 1.$$

In the absence of a fairness constraint, the maximum value of 1 is achievable in the limit; consider a mix that waits until it receives n packets from both the users, then reorders them into one of $\binom{2n}{n}$ permutations chosen uniformly. Although the arrival process is random, due to lack of any restrictions on the reordering, it is always possible for mix to wait until atleast n packets arrive from each user. Consequently, every transmitted batch will have the entropy $\log\binom{2n}{n}$ and anonymity of the strategy is

$$\mathcal{A}^\psi = \frac{\log\binom{2n}{n}}{2n} \tag{1}$$

which goes to 1 as $n \to \infty$. We are interested in studying the maximum achievable anonymity under a fairness restriction η:

$$\mathcal{A}(\eta) = \sup_{\psi\in\Psi_\eta} \mathcal{A}^\psi.$$

The results in this paper are focused on the maximum achievable anonymity when $\eta = 1$: in other words, no packet from one user can be transmitted ahead of 2 or more packets from the other user. This is the minimum possible relaxation of the fairness constraint and as will be demonstrated, even

with this minimum relaxation of fairness a significant amount of anonymity is achievable.

III. Anonymity under Limited Storage

Consider a mix with no storage limitations, in other words, the mix can store packets indefinitely before deciding to transmit a packet. Intuition may suggest that even though the mix has unlimited storage, the tight fairness restriction would force the mix to transmit a packet, and hence the number of packets required to be stored would be finite. This, however, is not true and the argument below demonstrates why.

Let there is a optimal mixing strategy ψ that requires a finite buffer size k. Since the strategy ψ is optimal, the anonymity achieved by strategy ψ is greater than or equal to that achievable by any other strategy even if the strategy uses a buffer size greater than k. We will now demonstrate that this is not possible by providing a strategy that achieves higher anonymity than ψ thus leading to a contradiction.

Consider the arrival of $3k$ packets to the mix; strategy ψ would have to have transmitted $2k$ packets by the time of arrival of the $3k$th packet. Furthermore the first k packets of these would have been transitted by the time of arrival of the $2k$th packet. Consider a strategy ψ_1 which works with a buffer os size $3k$. When $3k$ packets arrive to the mix, it chooses $2k$ packets as would have been chosen by strategy ψ. Further, it randomly shuffles the $2k$ packets while satisfying the the fairness restriction. Strategy ψ_1 repeats this process with every subsequent set of $2k$ packets. Since shuffling packets will strictly increase the entropy unless all packets are from the same source,

$$\begin{aligned} \mathbb{E}[H(Y_1,\cdots,Y_{2nk})|_{\psi} &< H(Y_1,\cdots,Y_{2nk})|_{\psi_1} \ \forall n \geq 1 \quad (2)\\ \mathcal{A}^{\psi} &< \mathcal{A}^{\psi_1} \quad (3)\end{aligned}$$

which is a contradiction.

Thus, when the mix has unlimited storage, the decision to schedule an arrived packet will be delayed indefinitely. While this represents an impractical scenario, the achievable anonymity obtained under this assumption would serve as a benchmark for memory limited mixing. In the remainder of this paper we will study the anonymity as a function of the buffer size k. Let $\mathcal{A}_k$ denotes the anonymity when buffer size is limited to hold k packets. An upper and lower bound on $\mathcal{A}_k$ is calculated in the next sections.

A. Upper Bound

It is important to note that when the memory of a mix is limited, it is sufficient for Eve to know the sequence of arriving packets in place of the complete timing information of the arrival point process; a packet can wait in the buffer until the next packet arrives regardless of what time it arrives, so all that matters is the source of next arriving packet. Furthermore, the decision to transmit can be made at the time of arrival of a new packet to a full buffer such that there is no uncertainty in the departure process as observed by Eve; a packet leaves the mix if and only if a packet arrives into a full buffer. Without loss of generality, we will assume that Eve's observation is restricted to the incoming sequence of packets.

We define the following notation:

1. As mentioned above, Y_1^n denotes collection of outgoing packets $(Y_1,\ldots,Y_n)$.
2. X_n denotes the random variable corresponding to the source of the n^{th} arrival. X_n is Bernoulli distributed with parameter 0.5 and X_1^n denotes collection of random variables $(X_1,\ldots,X_n)$.
3. $m(n)$ denotes the expected number of permissible ways in which we can permute the first n arrivals under fairness restriction.

The main idea behind calculating the upper bound is as follows:

1. We know that if $\mathcal{X}$ is random variable that takes $\mathcal{M}$ values then

$$\begin{aligned} H(\mathcal{X}) &\leq \log_2(\mathcal{M}) \quad (4)\\ \implies \mathbb{E}[H(\mathcal{X})] &\leq \mathbb{E}[\log_2(\mathcal{M})] \quad (5)\end{aligned}$$

   But, since $\mathcal{M}$ is also a random variable, Jensen inequality implies that

$$\mathbb{E}[\log_2(\mathcal{M})] \leq \log_2(\mathbb{E}[\mathcal{M}]) \quad (6)$$

   (5) and (6) implies that

$$\mathbb{E}[H(\mathcal{X})] \leq \log_2(\mathbb{E}[\mathcal{M}]) \quad (7)$$

   If $p(n)$ denotes the expected possible number of values that the sequence Y_1^n can take, then (7) implies that

$$\mathbb{E}[H(Y_1\cdots Y_n)] \leq \log_2(p(n)) \quad (8)$$

2. If buffer size is limited to k, then Y_1^n can be chosen only from X_1^{n+k+1} which implies that the possible ways in which we can permute X_1^{n+k+1} under buffer and fairness constraint will give us an upper bound on the possible number of values Y_1^n can takes. If $m(n+k+1)$ denotes the expected possible ways in which we can permute X_1^{n+k+1}, then

$$p(n) \leq m(n+k+1) \quad (9)$$

   (8) and (9) implies that

$$\mathbb{E}[H(Y_1\cdots Y_n)] \leq \log_2(m(n+k+1)) \quad (10)$$

   (10) and continuity of $\log(x)$ implies that

$$\mathcal{A} \leq \log_2(\lim_{n\to\infty}(m(n))^{\frac{1}{n}}) \quad (11)$$

Consider the simplest case, when $k = 1$ and $\eta = 1$, the mix can be viewed to be in either one of two states; the new arriving packet is either identical or different in color from the packet in the buffer. When the packets have identical colors, the mix has no choice but to transmit the packet in the buffer. When the packets are from different sources, the mix has a choice to transmit either one of the packets, and can choose to transmit one of them with a specific probability. This represents a classical Markov Decision Process, although

the reward as measured through anonymity is not additive and is a non-linear function of the state transition probabilities.

Theorem 1: Buffer is constrained to hold maximum k arrived packets,
For $k = 1$

$$\mathcal{A}_1 \leq \log_2\left(\frac{1+\sqrt{3}}{2}\right).$$

For $k > 1$

$$\mathcal{A}_k \leq \log_2(C_0)$$

where C_0 is maximum magnitude root of the equation

$$(2r)^{k+2} = 3(2r)^{k+1} - 2$$

Proof: Refer to appendix.

Using the analysis, we can calculate the upper bound for the unlimited buffer case:

Corollary 1: When the mix has unlimited buffer

$$\mathcal{A}_\infty \leq \log_2\left(\frac{3}{2}\right)$$

Proof: Refer to appendix.

B. Unequal rate

The upper bound can be extended to the case when the sources transmit with unequal rates. The analysis technique is very similar to that used in the preceding theorems. Let $X_R(t)$ and $X_B(t)$ be Poisson processes with arrival rates λ_1 and λ_2 respectively. If a packet arrives, it can be red with probability $q = \frac{\lambda_1}{\lambda_1+\lambda_2}$ and blue with probability $1-q$. Using similar techniques discussed previously, the following theorem characterizes the achievable anonymity for the unequal rates case:

Theorem 2: When the buffer size is k,
For $k = 1$

$$\mathcal{A}_1(q) = \log_2\left(\frac{1+\sqrt{1+8q(1-q)}}{2}\right)$$

For $1 < k < \infty$

$$A_k(q) = \log_2(C_0(q))$$

where $C_0(q)$ is the largest magnitude root of

$$r^{k+1} = r^k + q(1-q)r^{k-1}\left(\frac{1-\left(\frac{q}{r}\right)^k}{1-\frac{q}{r}} + \frac{1-\left(\frac{1-q}{r}\right)^k}{1-\frac{1-q}{r}}\right)$$

For $k = \infty$

$$\mathcal{A}_\infty(q) = \log_2\left(1+\sqrt{q(1-q)}\right)$$

Proof: Refer to appendix.

The upper bounds derived above were computed by relaxing the constraints on the mix, in particular, allowing the mix to perform non-causal permutation of packets. Further, the analytical characterization requires the use of Jensen's inequality. As a result, the analytical characterization above are strict upper bounds on the maximum achievable anonymity.

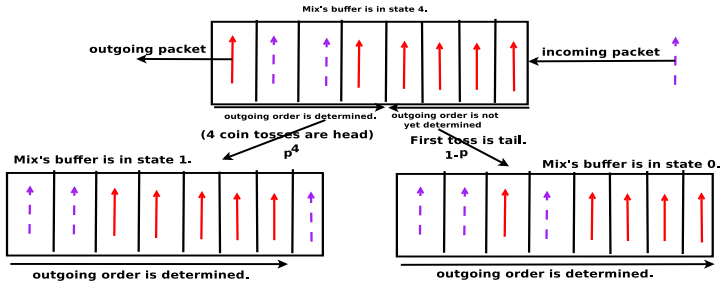


Fig. 1: Strategy $\overline{\psi}$ (discontinuous lines represent blue packet)

IV. Lower bound

The upper bound derived in the previous section provides a benchmark on the achievable anonymity for the almost fair mix. The anonymity of any strategy is less than or equal to the above derived bounds. In this section we provide a lower bound by specifying a strategy for the mix. Under the buffer constraint k, consider the following fixed strategy $\overline{\psi}$. We define the state Z_i of mix's buffer at the time of the i^{th} arrival to be the number of packets whose outgoing order is not yet determined according to strategy $\overline{\psi}$.

Strategy $\overline{\psi}$: As it will be clear from the strategy $\overline{\psi}$, the packets that determined the mix's buffer state (whose outgoing order is not determined) belong to the same source.

1. Mix transmits a packet only if its buffer is full. Consequently, Eve can not get any information by observing the outgoing packets.
2. Mix determines the outgoing order of packets present in its buffer.
3. Without loss of generality, assume that the mix is in state r and all the packets whose order is not yet determined are of red color.
4. If a packet arrives to a empty mix ($r = 0$), then it waits in the mix's buffer for the next packet.
5. If a blue packet (different source) arrives, then mix tosses a coin r times whose probability of showing a head is p. If there are all heads in the coin toss, then the r red packets are ordered as outgoing packets in the buffer. The only packet whose outgoing order is not yet determined is the newly arrived blue packet. Consequently, mix's buffer state turns to 1. Otherwise, the outgoing order of all the $r+1$ packets (including the newly arrived blue packet) is determined. The blue packet occupies the position of the first tail in the coin toss. Consequently, the mix's buffer new state is 0.
6. If a red packet arrives and $r < k$, then mix goes to state $r+1$ in which all the $r+1$ red packets outgoing order is not determined.
7. If $r = k$ and a red packet arrives, then mix transmits a red packet and remain in the state k.

The strategy $\overline{\psi}$ is described in Fig. 1. The states of mix's buffer under the strategy $\overline{\psi}$ can be represented as a Markov process as shown in figure 2. Let $(\mu_0, \cdots, \mu_k)$ represent the stationary

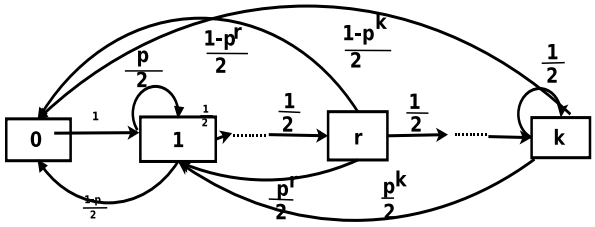
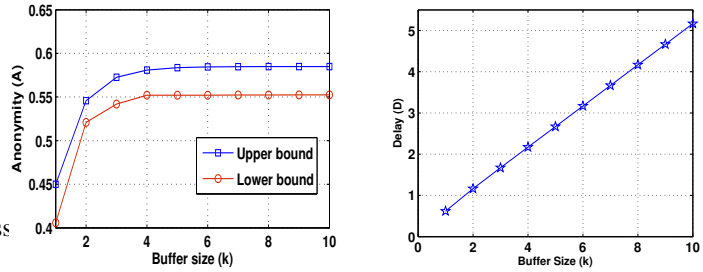


Fig. 2: Mix buffer's states represented as a Markov process under strategy $\overline{\psi}$



(a) $\mathcal{A}$ vs k (b) D vs k

Fig. 3: Anonymity ($\mathcal{A}$) and Delay (D) vs Buffer constraint (k)

distribution of this Markov process.

Theorem 3: When the mix is constrained to hold maximum k packets,

$$\mathcal{A}_k \geq c_1$$

where c_1 is the first coefficient of the equation

$$l(n) = c_1 + c_2 q_2^n + \ldots + c_{k+1} q_{k+1}^n$$

and q_i's are the roots of the equation (except 1 and $1/2$).

$$r^{k+2} - r^{k+1}\left(\frac{1}{2} + \frac{1}{2^k} + p\left(1 - \frac{1}{2^k}\right)\right) + \cdots$$
$$+r^k\left(-\frac{1}{2} + \frac{1}{2^{k+1}} + p\left(1 - \frac{1}{2^{k+1}}\right)\right) + \frac{1-p}{2^{k+1}} = 0$$

c_i's are calculated using the initial conditions

$$\begin{aligned} l(1) &= \left(1 - \frac{1}{2^k}\right) h(p) \\ l(n) &= \left(p + \frac{1-p}{2^k}\right) l(n-1) + \ldots \\ &\quad +(1-p)\sum_{i=2}^{n-1} \frac{l(n-i)}{2^{i-1}} \;\forall\, 2 \leq n \leq k+1 \end{aligned}$$

Proof: Refer to Appendix.

For $k = 1$ and $k = \infty$, the single-signature form for the lower bound is as follows

$$\begin{aligned} \mathcal{A}_1 &\geq \max_{0\leq p\leq 1} \frac{h(p)}{3-p} = 0.4057 \\ \mathcal{A}_\infty &\geq \max_{0\leq p\leq 1} \frac{h(p)}{3-2p} = 0.5515 \end{aligned}$$

V. Anonymity - Delay Tradeoff

As is well known, at a single queue, the FCFS is a delay optimal fairness criterion. Relaxation of the fairness constraint would cause a tangible increase in delay. In this section, we characterize the additional delay (in addition to processing times) incurred by the mix under strategy $\overline{\psi}$.

Theorem 4: Under the strategy $\overline{\psi}$ and buffer constraint k, the average delay D_k of packets is

$$D_k = \sum_{r=0}^{k} \mu_r d(r)$$

where μ_r are as follows:

$$\begin{aligned} \mu_1 &= \frac{2-p}{2\left(1 + (1-p)\left(2 - \left(\frac{p}{2}\right)^k\right)\right)} \\ \mu_0 &= 1 - 2\mu_1 \\ \mu_r &= \frac{1}{2^{r-1}}\mu_1 \;\; 2 \leq r \leq k-1 \\ \mu_k &= \frac{1}{2^{k-2}}\mu_1 \end{aligned}$$

and $d(r)$ are given in the following equation:

$$\begin{aligned} d(r) &= \frac{k}{2\lambda} + \frac{1 - r + p(r+1) - p^r - p^{r+1}(r+1)}{4\lambda(1-p)} + \ldots \\ &\quad + \frac{1 + p^r - 2p^{r+1}}{4\lambda}\left(1 - \frac{1}{2^k}\right) \end{aligned}$$

Proof $d(r)$ in the above theorem refers to the delay faced by a packet when it arrives to a mix when it is in state r. The details of the proof can be found in the appendix.

VI. Numerical Results

Figure 3 plots the lower and upper bounds on the maximum achievable anonymity and average delay faced by the users as a function of the memory k. In general, the state of the mix is the identity of packets in the order of arrival. At every new arrival, the mix has a choice to transmit one of the packets, and can choose to transmit one of them with a specific probability. This represents a classical Markov Decision Process, although the reward as measured through anonymity is not additive and a non-linear function of the state transition probabilities. The lower bound is computed by restricting the mix to a class of stationary strategies, and the upper bound is computed using a restriction on Eve's observation and using Jensen's inequality. Since the conditions for Jensen's inequality to be an equality cannot be satisfied in this case, the upper bound is strict. A trivial strategy to optimize the anonymity asymptotically can be designed as follows: The mix waits for the buffer to get full, and then transmits all packets with the permutation chosen equally likely among all possible fairness satisfying permutations. As the buffer size goes to infinity, it is easy to see that this strategy provides the maximum achievable anonymity; a closed form expression for this strategy does not, however, exist.

Both the plots in Figure 3 suggest that although the buffer size can increase indefinitely, the gain in anonymity saturates quickly under limited buffer conditions thus suggesting that for a fairness constraint k, there exists a suitable buffer size which provides sufficient anonymity at limited resource cost.

VII. Conclusion and future work

In this work, we studied the anonymity of a Chaum mix, under mild relaxation of First-come-First-serve fairness restriction. Although the setup considered was that of a shared router, the problem model is an instance of maximizing privacy in a stochastic control framework. In this setup, however, the action of the mix affects the probability transition matrix. If, alternatively, the problem were modeled as a POMDP, where the state includes the buffer state from the adversarys perspective, the reward would be a non-linear function of the probability of a state. Another instance of such a private stochastic control problem is that of privacy in a demand response smart metering system connected to the electricity grid where electricity costs can be thought of as a reward and scheduling decisions are made to minimize reward.

VIII. Appendix

Proof of theorem 1

Case 1: $k = 1$

It is based on the following observations:

1. For an n length sequence $(\hat{x}_1^n)$ of arrivals in which alternating packets are from different users for any fairness constraint greater than one, we can get total possible permutation of sequence of arrivals $(\mathcal{P}(\hat{x}_1^n))$ equal to
$$\mathcal{P}(\hat{x}_1^n) = F(n+1) \tag{12}$$
where $F(n)$ is the n^{th} Fibonacci number which is equal to $\frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right)$. Since fairness constraint ($\eta = 1$) restricts that first position of any permissible permutation of $\hat{x}_1^n$ can be occupied by X_1 or X_2 and symmetrical structure of $\hat{x}_1^n$ implies that $\mathcal{P}(\hat{x}_1^n) = \mathcal{P}(\hat{x}_1^{n-1}) + \mathcal{P}(\hat{x}_1^{n-2})$.
2. Fairness constraint 1 and buffer constraint 1 ensures that if any n length sequence (x_1^n) has k and $k+1$ packets are from same user, then total possible permutation of this sequence $(\mathcal{P}(x_1^n))$ will be equal to
$$\mathcal{P}(x_1^n) = \mathcal{P}(x_1^k)\mathcal{P}(x_{k+1}^n) \tag{13}$$

Let T_e is the random variable such that $T_e = i$ if i^{th} and $i+1^{th}$ packets are from the same source and $\forall\ k < i$ k^{th} and $k+1^{th}$ packets are from different source. We know

$$\begin{aligned} m(n) &= \mathbb{E}[\mathcal{P}(X_1^n)] && (14)\\ m(n) &= \mathbb{E}[\mathbb{E}[\mathcal{P}(X_1^n)|T_e]] && (15)\\ m(n) &= \sum_{i=1}^{n} P(T_e = i)\mathbb{E}[\mathcal{P}(x_1^n)|Y = i] && (16) \end{aligned}$$

We can see from (12) and (13) that

$$\mathbb{E}[\mathcal{P}(x_1^n)|Y = k]] = F(k+1)m(n-k)\ \ \forall 1 \le k \le n-1$$

and (12) implies that

$$\mathbb{E}[\mathcal{P}(x_1^n)|Y = n] = F(n+1)$$

Using the above equations, we get

$$m(n) = \sum_{k=1}^{n-1} \frac{1}{2^k} F(k+1)m(n-k) + \frac{1}{2^{n-1}}F(n+1) \tag{17}$$

Using (17), we can get

$$m(n+2) = m(n+1) + \frac{1}{2}m(n) \forall\ n \ge 1 \tag{18}$$

(18) is recursive definition for $m(n)$, whose characteristics equation is

$$r^2 = r + \frac{1}{2} \tag{19}$$

which has solutions $\frac{1\pm\sqrt{3}}{2}$. Using initial conditions, $m(1) = 0$ and $m(2) = \frac{1}{2}$, we get

$$m(n) = \frac{1}{\sqrt{3}}\left(\left(\frac{1+\sqrt{3}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{3}}{2}\right)^{n+1}\right)$$

and

$$\begin{aligned} \log_2(\lim_{n\to\infty}(m(n))^{\frac{1}{n}}) &= \log_2\left(\frac{1+\sqrt{3}}{2}\right)\\ &= 0.45 \quad \square \end{aligned}$$

Case 2: $k > 1$

Let P_i is a random variable such that $P_i = n$ if X_n occupies i^{th} position in a permutation. Swapping the packets of the same user do not increases the anonymity so without loss of generality, we can assume following restriction on permutation of arrived packets:

1. If $X_i = X_j$, $P_m = i$ and $P_n = j \implies m < n$.
2. $\forall j > i$, $P_i = j$ only if
$$j \le i+k+1,\ \ X_j \ne X_i \text{ and } X_m = X_i \forall m, i+1 \le m \le j-1$$

Above restriction tells us that if X_i is fixed at position 1, then packets $1, 2, \cdots, i-1$ are necessarily from the same source and have to be transmitted in succession following the first packet. In this case the expected total number of permutation would be $m(n-i)$. Since this occurs with probability $\frac{1}{2^{i-1}}$, we have the following recursive equation on $m(n)$,

$$m(n) = \sum_{i=1}^{k+1} \frac{1}{2^{i-1}} m(n-i) \forall n > k+1 \tag{20}$$

Characteristics equation for the recursive equation (20) is

$$\begin{aligned} \frac{1}{2} &= (2r)^{-1} + (2r)^{-2} + (2r)^{-3} + \ldots + (2r)^{-k-1}\\ (2r)^{k+2} &= 3(2r)^{k+1} - 2 && (21) \end{aligned}$$

Using the theory of finite difference equation, we know that

$$\log_2(\lim_{N\to\infty} m(n)^{\frac{1}{n}}) = \log_2(C_0)$$

where C_0 is the root of (21) with maximum magnitude. This can be numerically calculated and the results are plotted in section VI. □

Proof of corollary 1

For the case, when $k \to \infty$, we can see that characteristics equation (21) becomes

$$\frac{1}{2} = \frac{1}{2r-1} \implies r = \frac{3}{2}$$

which implies that

$$A_\infty \leq \log_2\left(\frac{3}{2}\right) \implies A_\infty \leq 0.5850$$

□

Proof of theorem 2

$k = 1$

Using the similar analysis as in the case of equal rate, we get the recursive equation $m(n) = m(n-1) + 2p(1-p)m(n-2)$ which leads us to the result.

$k = \infty$

In this case the recursive equation is

$$\begin{aligned} m(n) &= m(n-1) + (1-q)\sum_{i=1}^{n-2} q^i m(n-i-1) + \cdots \\ &\quad + q\sum_{i=1}^{n-2}(1-q)^i m(n-i-1) \end{aligned} \tag{22}$$

And the characteristics equation of (22) is $r^2 - 2r + 1 - q(1-q) = 0$ and $1 + \sqrt{q(1-q)}$ is its largest root.

Proof of theorem 3

Lets assume that F is a random variable such that

$$\begin{aligned} F &= i \implies X_1 = X_2 = \cdots = X_i \text{ and } X_1 \neq X_{i+1} \\ &\quad \forall i = 1, \cdots, n+k-1 \\ F &= n+k \implies X_1 = \cdots = X_{n+k} \end{aligned}$$

Consider

$$\begin{aligned} l(n) &= \mathbb{E}[H(Y_n|Y_1^{n-1}, \Phi)] \\ &= \mathbb{E}[\mathbb{E}[H(Y_n|Y_1^{n-1}, \Phi)|F]] \\ &= \sum_{i=1}^{n+k+1} P(F=i)\mathbb{E}[H(Y_n|Y_1^{n-1}, \Phi)|F=i] \end{aligned} \tag{23}$$

We can see that $\overline{\psi}$ ensures that if $F = i$, then Mix will go to state 0 or state 1 on the arrival of $i+1^{th}$ packet. If it goes to state 0, then average anonymity of the packets that leaves Mix after the i^{th} arrival will be independent of the departure before the arrival of the i^{th} packet. Similarly, in case when Mix's goes to state 1, then average anonymity of packets that leaves Mix after the arrival of i^{th} packet can be calculated by assuming that Mix has just one packet when the $i+1^{th}$ packet arrives.This observation can be written in the form of following equations:

$$l(n) = \begin{cases} p^i l(n-i) + (1-p^i)l(n-i-1) & 1 \leq i \leq k \\ p^k l(n-i) + (1-p^k)l(n-i-1) & k+1 \leq i \leq n-1 \\ p^{n+k-i-1}h(p) & n \leq i \leq n+k-1 \\ 0 & i = n+k \end{cases} \tag{24}$$

(23) and (24) implies that

$$l(n) = \left(\frac{1+p}{2}\right) l(n-1) + \left(\frac{1-p}{2}\right) l(n-2) + \sum_{i=2}^{k} \frac{p^i}{2^i}(l(n-i) - l(n-i-1)) \tag{25}$$

For $n \leq k+1$

$$\begin{aligned} l(n) &= \sum_{i=1}^{n-1} \frac{1}{2^i}\left(p^i l(n-i) + (1-p^i)l(n-i-1)\right) \cdots + \\ &\quad + h(p)\sum_{i=n}^{k} \frac{p^{n-1}}{2^i} + h(p)\sum_{i=k+1}^{n+k-1} \frac{p^{n+k-1-i}}{2^i} \\ l(1) &= \left(1 - \frac{1}{2^k}\right) h(p) \end{aligned}$$

The characteristics equations of (25) is

$$\begin{aligned} &r^{k+2} - r^{k+1}\left(\frac{1}{2} + p\right) + \cdots \\ &+ r^k\left(-\frac{1}{2} + p\right) - \frac{p^k(1-p)}{2^{k+1}} r + \frac{p^k(1-p)}{2^{k+1}} = 0 \end{aligned} \tag{26}$$

Clearly 1 and $p/2$ are the roots of (26). If $q_2, \ldots, q_{k+1}$ are the other roots of (26), then $l(n) = c_1 + c_2 q_2^n + \ldots + c_{k+1} q_{k+1}^n$. If magnitude of any of the roots is greater than one or if 1 occurs multiple time as a root, then theory of recursive equations confirms that $\frac{\sum_{n=1}^{N} l(n)}{N}$ would be a diverging sequence. But upper bound on anonymity ensures that $\frac{\sum_{n=1}^{N} l(n)}{N}$ should converge. This implies all the q_i's are strictly less than one. So

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} l(n)}{N} = c_1 \tag{27}$$

Definition of anonymity implies that

$$A_k \geq \lim_{N\to\infty} \frac{\sum_{n=1}^{N} l(n)}{N} \tag{28}$$

Using (28) and (27), we can say that

$$A_k \geq c_1 \square$$

Proof of theorem 4

The stationary distribution of the Markov process as shown in figure 2 satisfies the following equation:

$$\mu_0 = \sum_{r=1}^{k} \frac{1-p^r}{2}\mu_r \tag{29}$$

$$\mu_1 = \mu_0 + \sum_{r=1}^{k} \frac{p^r}{2}\mu_r \tag{30}$$

$$\mu_r = \frac{1}{2^{r-1}}\mu_1 \quad 2 \leq r \leq k-1 \tag{31}$$

$$\mu_k = \frac{1}{2^{k-2}}\mu_1 \tag{32}$$

And

$$\sum_{r=0}^{k} \mu_r = 1 \implies \mu_1 = \frac{2-p}{2\left(1 + (1-p)\left(2 - \left(\frac{p}{2}\right)^k\right)\right)} \tag{33}$$

And

$$\mu_0 = 1 - 2\mu_1 \implies \mu_0 = \frac{(1-p)\left(1-\left(\frac{p}{2}\right)^k\right)}{1+(1-p)\left(2-\left(\frac{p}{2}\right)^k\right)}$$

Lets calculate the average delay faced by the arrived packet (X_n) when mix's buffer is in state 0. Lets assume that X_n is a red packet. It is based on the following observation:

1. If all the next k packets are red, then X_n would leave the mix only on the arrival of $(k+1)^{th}$ packet. In this case, it would have to face $\frac{k}{2\lambda}$ unit delay and the probability of this event is $\frac{1}{2^k}$.
2. Otherwise, in all other cases, $\overline{\psi}$ says that the next arrived blue packet would go before X_n if and only if the outcome of first coin toss is tail. It implies with probability p, X_1 has to face $\frac{k}{2\lambda}$ unit delay and with probability $1-p$, it would face $\frac{k+1}{2\lambda}$ unit delay.

The above observation implies the average delay ($d(0)$) of X_n when it arrives to a mix in state 0 is

$$d(0) = \frac{k}{2\lambda} + \frac{1-p}{2\lambda}\left(1-\frac{1}{2^k}\right)$$

Lets calculate the delay faced by packet X_n under the strategy $\overline{\psi}$ when it arrives to a mix which is in state $r > 0$. Without loss of generality, lets assume that mix is in state r due to r red packets whose outgoing order is not yet determined.

Case1: The arrived packet (X_n) is red

1. If the next k arrivals are red packets, then X_n would face $\frac{k}{2\lambda}$ delay.
2. Otherwise, when a blue packet comes, mix would toss the coin. If the result of initial $r+1$ tosses are head, then X_n would go before the blue packet and its delay would be $\frac{k}{2\lambda}$.
3. If any of the first $r+1$ toss is tail, then blue packet would go before X_n and consequently, X_n would face $\frac{k+1}{2\lambda}$ delay.

Using above observation we can see that, delay faced by X_n when it is red ($d(X_n = R)$)is

$$d(X_n = R) = \frac{k}{2\lambda} + \frac{1-p^{r+1}}{2\lambda}\left(1-\frac{1}{2^k}\right)$$

Case 2: The arrived packet (X_n) is blue.

1. If first i coin tosses give head and $(i+1)^{th}$ coin toss give tail (probability of this event is $p^i(1-p)$), then X_n would be placed at $(k-r+i+1)^{th}$ position in the buffer for transmission. Eventually, it would face the delay $\frac{k-r+i+1}{2\lambda}$.
2. If all the r coin tosses result in head, then it would face the delay same as delay faced by the packet when it arrives to a mix which is in state 0.

Using above observation we can see that, delay faced by X_n when it is blue ($d_n(X_n = B)$)is

$$\begin{aligned} d(X_n = B) &= \frac{k-r}{2\lambda}(1-p^r) \\ &\quad + \frac{1+p-p^r-rp^r-p^{r+1}}{2\lambda(1-p)} + p^r d(0) \end{aligned}$$

Using above results, we can calculate the delay ($d(r)$) suffered by X_n when mix is in state r is

$$d(r) = \frac{1}{2}d(X_n = B) + \frac{1}{2}d(X_n = R)$$

The average delay of the X_n is

$$\begin{aligned} d_n &= \sum_{r=0}^{k} P(Z_n = r)d(r) \\ \lim_{n\to\infty} d_n &= \sum_{r=0}^{k} \mu_r d(r) \end{aligned}$$

But we know that if $a_n \to a$ and $b_n = \frac{\sum_{i=1}^{n} a_i}{n}$ then $b_n \to a$ which implies that

$$D_k = \sum_{r=0}^{k} \mu_r d(r)$$

REFERENCES

[1] "The TOR Project: Anonymity Online." http://www.torproject.org, Feb.

[2] P. Venkitasubramaniam and V. Anantharam, "Anonymity of Mix Networks under Light Traffic Conditions," in *Proceedings of the 36th Allerton Conf. on Communications, Control, and Computing*, (Monticello, IL), October 2008.

[3] P. Venkitasubramaniam and A. Mishra, "Anonymity in Military Networks under Memory Restrictions." submitted to IEEE Journal for Selected Areas in Communication, Nov. 2010.

[4] Q. Sun, D. R. Simon, Y. Wang, W. Russell, V. N. Padmanabhan, and L. Qiu, "Statistical identification of encrypted web browsing traffic," in *Proceedings of the 2002 IEEE Symposium on Security and Privacy*, (Berkeley, California), p. 19, May 2002.

[5] N. Matthewson and R. Dingledine, "Practical traffic analysis: Extending and resisting statistical disclosure," in *Privacy Enhancing Technologies: 4th International Workshop*, May 2004.

[6] D. X. Song, D. Wagner, and X. Tian, "Timing Analysis of Keystrokes and Timing Attacks on SSH," in *Proc. 10th USENIX Security Symposium*, 2001.

[7] D. Chaum, "Untraceable electronic mail, return addresses and digital pseudonyms," *Communications of the ACM*, vol. 24, pp. 84–88, February 1981.

[8] G. Danezis, R. Dingledine, and N. Mathewson, "Mixminion: design of a type iii anonymous remailer protocol," in *Proceedings of 2003 Symposium on Security and Privacy*, pp. 2–15, May 2003.

[9] G. Danezis, "Mix-networks with restricted routes," in *Proceedings of Privacy Enhancing Technologies Workshop (PET 2003)*, Springer-Verlag, LNCS 2760, April 2003.

[10] C. Díaz, S. Seys, J. Claessens, and B. Preneel, "Towards measuring anonymity," in *Proceedings of Privacy Enhancing Technologies Workshop (PET 2002)* (R. Dingledine and P. Syverson, eds.), Springer-Verlag, LNCS 2482, April 2002.

[11] D. Kesdogan, J. Egner, and R. Buschkes, "Stop-and-go MIXes providing probabilistic security in an open system," in *Second International Workshop on Information Hiding (IH'98), Lecture Notes in Computer Science*, vol. 1525, (Portland, Oregon), pp. 83–98, April 1998.

[12] A. V. D. Stiliadis, "Efficient Fair Queuing Algorithms for Packet-Switched Networks," *IEEE Trans. Networking*, vol. 6, pp. 175–185, April 1998.