# Dummy Rate Analysis of Buffer Constrained Chaum Mix

Abhishek Mishra
Lehigh University
PA, USA 18015
abm210@lehigh.edu

Parv Venkitasubramaniam
Lehigh University
PA, USA 18015
parv.v@lehigh.edu

## Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*Security and Protection*; C.2.6 [**Computer-Communication Networks**]: Internetworking—*Routers*

## General Terms

Security

## Keywords

Chaum Mix, Dummy rate, Buffer constraint

## 1. PROBLEM DESCRIPTION

Providing anonymity to network communication refers to the prevention of "networking information" retrieval, not only using the *content* of transmitted data packets but also from their *timing information.* Access to such timing information can reveal the source-destination pairs or path of data flow which is a violation of user privacy, and can further be used to jam a particular flow or create black holes. Chaum Mixes [1] are often used to obfuscate this timing information from malicious eavesdroppers. A Chaum Mix is a relay node or a proxy server that collects packets from different users, then uses layered encryption and packet padding to make outgoing packets appear indistinguishable to eavesdroppers. Furthermore, it also alters the timing information by reordering and batching together packets from different users, so that the probability of any outgoing packet belonging to a particular user is identical for all users, thus achieving perfect anonymity.

A drawback of the ideal Chaum Mix is that if long data streams (eg. media streaming) are transmitted, then it may require a very large buffer capacity to work well. In fact, it has been proven that any batching strategy, under the constraint of a limited buffer size, would eventually reveal the source identities [2]. It is, however, possible to mask the actual pattern of traffic flow through the insertion of **dummy traffic**. For example, consider a network where all nodes transmit packets according to scheduled departure times. If an actual packet is unavailable for transmission at its time of departure, a dummy packet can be transmitted in its place and perfect anonymity is still maintained. But the extensive use of dummy packets can reduce network throughput, which leads to some interesting questions:

1. For a fixed buffer size, what is the optimal mixing strategy to achieve perfect anonymity with minimum possible dummy rate (number of dummy packets per second)?

2. For a fixed buffer size, what is the minimum dummy rate required for a mix to achieve perfect anonymity? Further, how does this minimum dummy rate scale with buffer size?

In this work, the above questions are addressed using an analytical approach.

## 2. APPROACH AND CONTRIBUTION

**Optimal mixing strategy $\Psi$:** For a fixed buffer size $B$, the mix waits until at least one packet from each user arrives or until the buffer is full. If packets from all users arrive before the buffer is full, then one packet of each user is selected and they are randomly ordered and sent in succession. Otherwise, the mix generates one additional dummy packet for each user missing in the buffer, and performs the above operation treating the dummy packets as actual packets of missing users.

THEOREM 1. *The strategy $\Psi$ is optimal for achieving minimum dummy rate.*

Due to the constraint of page limit, we omit the proof here. Using strategy $\Psi$, we can compute the exact rate of dummy transmission for a two source mix.

THEOREM 2. *The minimum dummy rate required by a mix serving two sources of rates $\lambda_1, \lambda_2$ respectively is given by*

$$\mu = \frac{\lambda_1 \rho^{2B+1} - \lambda_1 \rho^{2B} + \lambda_2 \rho - \lambda_2}{\rho^{2B+1} - 1} \text{ where } \rho = \frac{\lambda_1}{\lambda_2}.$$

*and hence when $\lambda_1 = \lambda_2 = \lambda$,*

$$\mu = \frac{2\lambda}{2B+1}.$$

The proof is omitted for brevity. The same proof technique can also be applied when more than two users are present, but it involves the analysis of multidimensional Markov chains which are analytically intractable. We can, however, provide lower and upper bounds on the minimum achievable dummy rate for the general $k$ user case.

*SIGCOMM'11,* August 15–19, 2011, Toronto, Ontario, Canada.
ACM 978-1-4503-0797-0/11/08.

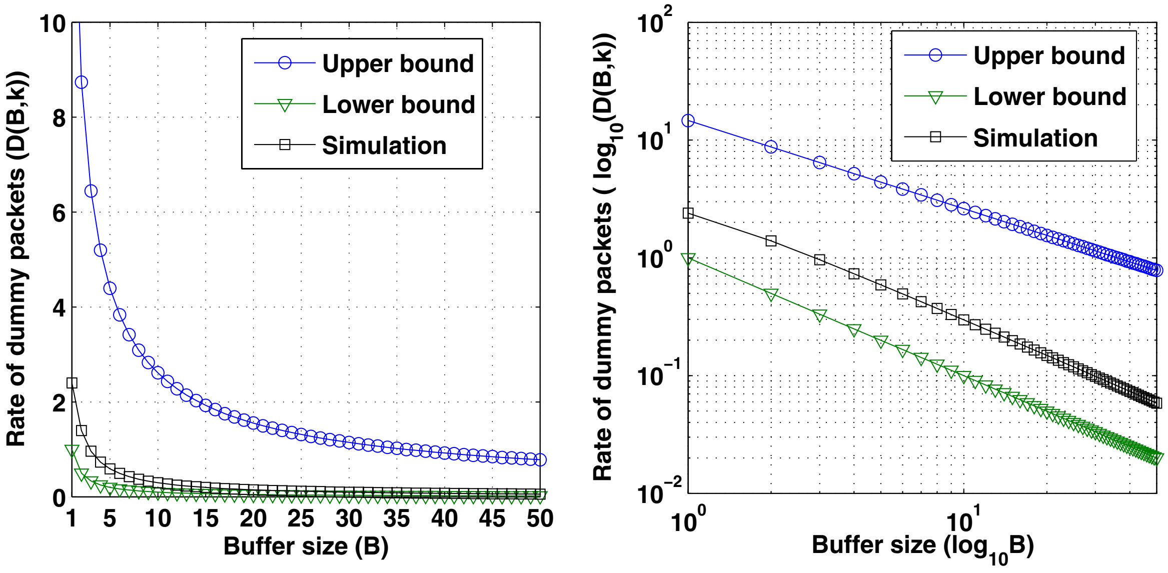

**Figure 1: Bounds on dummy packets as a function of buffer size**

THEOREM 3. *For a $k-$source mix with buffer size $B$, there exists constants $K_1, K_2$ independent of $B$ such that the minimum required dummy rate $D(B,k)$ satisfies:*

$$\frac{K_1}{B} \leq D(B,k) \leq \frac{K_2}{B^{3/4}}.$$

The lower bound on the dummy rate holds true regardless of the mixing strategy employed, thus providing a useful performance benchmark.

To get the upper-bound, we consider a sub-optimal strategy $\overline{\Psi}$, under which the mix first divides the buffer into two equal halves, each of size $B/2$. The mix stores all incoming packets in the first half of the buffer until it is full. Amongst the $B/2$ packets now present in the buffer, the mix transmits the maximum number of packets in a batch, such that the number of packets from each source is equal. The remaining packets are shifted to the second half of the buffer. This process is repeated until the second half of the buffer is full. At this point, all the packets in the second half of the buffer are transmitted in a single batch with enough dummy transmissions such that the number of packets from each source is equal. Thus, by calculating the upper limit on the required dummy rate for strategy $\overline{\Psi}$, we obtain an upper bound on $D(B,k)$ . The detailed proof is omitted due to paucity of space.

In Fig. 1, we plot the numerically computed dummy rate using the multidimensional Markov chain (optimal strategy), and compare it with the bounds. As is evident from the plots, the convergence rate of the optimal strategy matches that of the lower bound. This fact, in conjunction with the result on the convergence rate for the two-source system indicates that the optimal convergence rate is $O\left(\frac{1}{B}\right)$.

## 3. ANONYMITY IN A NETWORK

Thus far, we studied the dummy transmission rate required at a single mix to achieve perfect anonymity. The anonymity, as quantified using the Shannon entropy of packet sources would amount to $\log k$, if there are $k$ sources transmitting to the mix. In a general multihop network, as the packet streams traverse through multiple mixes, the anonymity of packet sources at any given link would be the result of cumulative mixing at previous nodes in the path of the streams. This cumulative anonymity can be computed using the chain rule of entropy [3].

THEOREM 4. *If $A_1, A_2, \cdots, A_k$ denote the entropy of packet sources of $k$ streams that are mixed at a given intermediate node, then the entropy of sources on the outgoing stream of packets is given by:*

$$A = \frac{\sum A_i}{k} + 1.$$

Using the fact that mixes that received packets directly from sources achieve anonymity of $\log k$, we can recursively compute the anonymity of packets on any link.

It is important to note that to achieve this anonymity, it is critical that the generated dummy packets are relayed by subsequent nodes as though they were data packets, and consequently, while anonymity builds up at successive mixes, the true rate of data packets drops, thus resulting in a tradeoff between anonymity and throughput.

## 4. REFERENCES

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

[2] X. Fu, B. Graham, R. Bettati, and W. Zhao, "On countermeasures to traffic analysis attacks," in *Information Assurance Workshop, 2003. IEEE Systems, Man and Cybernetics Society*, pp. 188–195, 18-23 June 2003.

[3] T. Cover and J. Thomas, *Elements of Information Theory.* John Wiley & Sons, Inc., 1991.