# RPROB - A Family of Binomial-Mix-based Anonymous Communication Systems

Minh-Triet Tran
University of Science
227 Nguyen Van Cu, Dist.5
Hochiminh City, Vietnam
tmtriet@fit.hcmus.edu.vn

Anh-Duc Duong
University of Science
227 Nguyen Van Cu, Dist.5
Hochiminh City, Vietnam
daduc@fit.hcmus.edu.vn

Isao Echizen
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku
Tokyo 101-8430, Japan
iechizen@nii.ac.jp

***Abstract*—We propose RPROB, an infinite family of anonymous communication systems, each of which corresponds to a binomial mix. Any instance of RPROB provides resistance against global active adversary with capabilities to monitor every external activity, to delay and to create messages in the system. Our proposal is to solve the limitation of APROB Channel that concerns only global delaying adversary. Experimental evaluation shows that any instance of RPROB provides higher anonymity than APROB Channel with the same environment and users' behaviors (rate and number of sent messages). Furthermore, because of the randomness provided by a binomial mix, an adversary cannot determine with certainty the probability of a user to be a sender of a delivered message in RPROB system as in Pool-based APROB Channel. The prefix 'R' in RPROB is to emphasize the randomness of our proposal. RPROB also provides flexibility for users to justify their level of anonymity (and speed) and satisfies probabilistic real-time condition which ensures to deliver any message within a predefined duration with high probability.**

***Index Terms*—Anonymity system, binomial-mix-based anonymous communication framework, probabilistic real-time, Global Active Adversary.**

## I. INTRODUCTION

Anonymity has become more and more important in various applications. There is increasing need for privacy preserving solutions not only in traditional domains (such as web surfing, e-mail, social networks,...) but also in pervasive environment, such as location based services [1], [2]... Therefore various anonymous communication systems have been proposed in both theorical models [3], [4], [5] and implementations [6], [7], [8].

Among privacy enhancing technologies (PET), mixes [9] are commonly used to provide unlinkability[10]. A mix receives a number of incoming messages from senders, then transforms and delivers them to next hops or recipients in such a way to prevent matching messages at its input and output with certainty.

G. Tóth and Z. Hornák used an adaptive mix in APROB Channel [11] to provide anonymity with resistance against a global delaying adversary (GDA [11]) with capability to delay any number of incoming messages for arbitrary duration. However this system is vulnerable to blending attacks [12] caused by a global active adversary (GAA [13]) with supplemental capability to monitor all traffic and to generate messages.

To solve this problem, a pool mix is used in Pool-based APROB Channel [13] to replace an adaptive mix of APROB Channel. Although Pool-based APROB Channel is a successful solution for GAA, this system still has some limitations:

- Pool-based APROB Channel is defined in a rigid form. This system keeps a constant number of messages $N_p$ in its pool so that an opponent cannot be certain when a message will be delivered. In fact, it is only required to keep any non-zero number of messages in system. Furthermore, in a high traffic system, the number of incoming messages increases and a system with dynamic $N_p$ can adapt to the new context and keep more messages in its pool to increase anonymity. This is not the case for a system with fixed $N_p$.
- In Pool-based APROB Channel, although a message is delivered only when its anonymity requirement (assigned by its sender) is satisfied, an adversary can calculate with certainty the probability that a user $s_i$ is a sender of an outgoing message as Pool-based APROB Channel is a deterministic system.

In this article, we propose RPROB, a framework that determines a family of binomial-mix based anonymous communication systems. RPROB is not a concrete anonymous system but defines an infinite family of anonymous communication systems, each of which corresponds to a binomial mix to provide anonymity against GAA. In RPROB, the number of messages to be kept in system's pool can be variable. Due to the randomness of a binomial mix, an opponent can no longer determine with certainty the probability that a user $s_i$ is a sender of an outgoing message. The prefix 'R' in RPROB stands for its randomness.

We use the generalized binomial framework [14], [15] to express RPROB in a general specification. Then we derive general formula to evaluate anonymity for an arbitrary sender in RPROB. RPROB satisfies the "probabilistic real-time" condition [13] which ensures to deliver any message within a maximum delay with high probability. Furthermore RPROB also provides flexibility so that users can choose or adjust their own level of anonymity and corresponding delay time.

The rest of the paper is organized as follows: first we review generalized binomial mix framework that we use in

978-1-4244-5113-5/09/$25.00 ©2009 IEEE

the specification of our RPROB. We briefly present and analyse limitations of APROB Channel and Pool-based APROB Channel. Then we present our generalized binomial-mix-based anonymous communication framework RPROB and analyze its properties: security against GAA, probabilistic-real-time property, and flexibility. Conclusions and open questions for future works are in the final section.

## II. Background and Related Works

### A. *Mixes and Generalized Binomial Mix Framework*

The first mix design was introduced by Chaum [9] in 1981. A mix takes a number of input messages from senders, changes the appearance (by encrypting and padding messages) and the flow of messages (by delaying and/or reordering), and delivers them to next hops or recipients in such a way that it is hard to match an output to corresponding input (or an input to corresponding output) with certainty.

To formally specify different deterministic mixes in a unified model, C. Diaz and A. Serjantov presented the generalized mix model [14]. In this generalized mix model, each mix is specified with a function from the number of messages inside the mix to the fraction of messages to be flushed. In this model, the number of messages to be forwarded is deterministic.

Besides the generalized mix models for deterministic mixes, C. Diaz and A. Serjantov also proposed binomial mixes binomial mix [14] and binomial mix framework [15] with randomness.

Let $g : \mathbb{N} \to [0, 1]$ be the probability of forwarding each message and $n$ be the number of messages in the mix right before it flushes messages. Each message in the mix is selected to be forwarded with the probability $g(n)$.

Let $X$ be a random variable corresponding to the number of messages kept in a binomial mix and $P(X = x|n)$ be the conditional distribution of the number of messages kept in the mix. The number of messages kept in a binomial mix follows a binomial distribution $\mathrm{Bin}(n, 1 - g(n))$.

$$P\left(X = x|n\right) = \left(\begin{array}{c} n \\ x \end{array}\right) g\left(n\right)^{(1-x)} \left(1 - g\left(n\right)\right)^{x} \qquad (1)$$

Several mix functions are illustrated in Fig. 1.

- For a binomial mix with $g\left(n\right) = \max\left\{0, 1 - N_p/n\right\}$, the average number of messages kept in its pool is a constant $N_p$ ($N_p = 10$ in this example). This binomial mix originates from a timed pool mix [16].
- For a binomial mix with $g(n) = f$, the mix, on average, outputs a constant fraction $g(n) = f$ of all messages in its pool ($f = 0.7$ in this example). This binomial mix evolves from a timed dynamic mix [16].
- The mix with the mix function based on cummulative normal distribution function [14], [15]

$$g\left(n\right) = f \int_{-\infty}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-\left(x - \mu\right)^2}{2\sigma^2}\right) \qquad (2)$$

This mix is illustrated in Fig. 1 with $f = 0.6$, $\mu = 40$, and $\sigma = 15$.

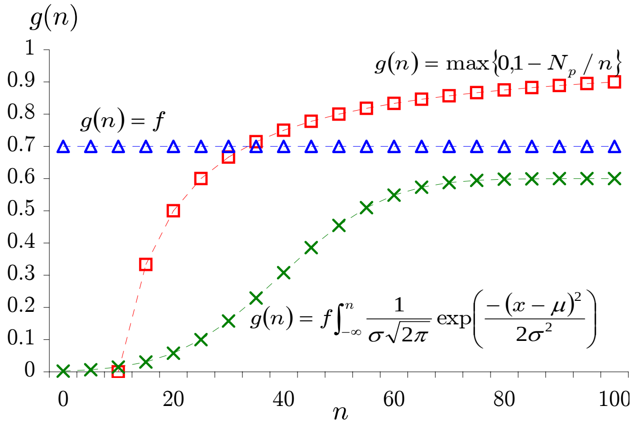

Fig. 1. Mix functions of some binomial mixes.

In this article we use binomial mix framework to specify RPROB in a single general form. Each instance of RPROB is identified by a specific mix function $g(n)$.

### B. *APROB Channel and Global Active Adversary*

To evaluate and analyse the anonymity of a system, we use probability that an opponent calculates for each potential sender $s_i$ in sender list $S$ to be the sender of a message $m$, denoted by $P\left(S\left(m\right) = s_i\right)$. A system is source-hiding [17] with parameter $\Theta$ if an adversary cannot assign any sender to a delivered message with a probability greater than $\Theta$ .

In [11], G. Tóth and Z. Hornák proposed APROB Channel to resist a global delaying adversary (GDA) that can monitor all external activities of the system and delay any number of messages for arbitrary time.

APROB Channel is actually an adaptive mix. Each message $\alpha_k$ arriving in the system has a threshold $\theta(\alpha_k)$ of source-hiding property assigned by its sender $S(\alpha_k)$. The system caches all incoming messages until anonymity requirements of all messages in its buffer can be fulfilled, i.e. $P\left(S\left(\alpha_k\right) = s_i\right) \leq \theta\left(\alpha_k\right)$ for every sender $s_i$. Every message $\alpha_k$ is then transformed into a message $\beta_j$ and delivered to its recipient $R\left(\beta_j\right)$. This is called a round.

Not only APROB Channel but also any anonymous system that flushes all its cached messages when it outputs is vulnerable to blending attacks [12] by a Global Active Adversary (GAA), a stronger yet more practical opponent than GDA. Besides all capabilities of GDA, GAA can generate any number of messages to be accepted by the system. This motivates our proposal of RPROB that uses a binomial mix instead of a regular mix to resist GAA. A binomial mix not only keeps a variable number of messages in its pool when it flushes but also hides the number of messages in its pool.

### C. *Pool-based APROB Channel*

At the kernel of a Pool-based APROB Channel [13] is a pool mix that keeps a constant number $N_p$ of messages in its pool when it flushes.

Similar to APROB Channel, a Pool-based APROB Channel also caches all incoming messages until source-hiding property of every message is satisfied. Then the Pool-based APROB Channel selects a fixed number $N_p$ of messages to be kept in its pool for the next round while other messages are delivered. Thus an opponent cannot be certain when a message is delivered. Besides, due to messages kept from previous rounds, the anonymity in a Pool-based APROB Channel is higher than that in APROB Channel.

A Pool-based APROB Channel achieves the following properties:

- Guaranteed anonymity for senders against GAA: all messages in system are delivered only when the source-hiding property of every message is fulfilled.
- Probabilistic-real-time property [13]: every message is delivered within a maximum delay with high probability if there is no delay caused by any adversary and every sender sends at least one message in a predefined timeslot.
- Flexibility for users: each user can determine his/her threshold of source-hiding property.

There are two issues with Pool-based APROB Channel:

- It is not necessary to keep a constant number of messages in system's pool when it flushes. This leads to the flexibility in designing a system to keep a variable number of messages. With the strategy of keeping a variable number of messages in system's pool, the system can provide higher anonymity in high traffic environment by retaining more messages in its pool.
- An opponent can calculate with certainty the probability that a user $s_i$ is a sender of a delivered message as Pool-based APROB Channel is a deterministic system. This motivates our proposal of RPROB to use randomness in mixes.

In our proposal of RPROB, we reuse the idea in Pool based APROB Channel to retain a number of messages in system's pool when it flushes in order that an opponent cannot be certain when a message will be delivered. However, as our main objective is to propose a family of anonymous communication systems, we need to find an appropriate way to parameterize the system. Besides we aim to add randomness into RPROB to prevent an opponent from calculating with certainty the probability that a user $s_i$ is a sender of a delivered message.

## III. RPROB - Binomial-Mix-based Anonymous Communication Framework

### A. Specification of RPROB

Instead of using a pool mix as in Pool-based APROB Channel, an instance of RPROB uses a binomial mix corresponding to a mix function $g(n)$.

To ensure the expectation of number of messages kept in system's pool is at least 1, we enforce the following constraint on the mix function $g(n)$:

$$n(1-g(n)) \geq 1, \forall n > 0 \tag{3}$$

The processing of RPROB in round $r$ is summarized as follows:

- ***Message collection***: RPROB receives and caches every new message $\alpha_i^{(r)}$ arriving in system. Let $a_r$ be the number of new incoming messages.
- ***Message selection***: Let $N_p^{(r-1)}$ be the number of old messages kept from the previous round $r-1$. Let $n_r = a_r + N_p^{(r-1)}$ be the total number of messages in system's pool (before flushing). When anonymity requirements (the source-hiding property) of all $n_r$ messages can be fulfilled, each message is selected to be forwarded with the probability of $g(n_r)$. Let $b_r$ be the number of selected messages.
- ***Message transformation***: Each selected message $\alpha_i^{(r)}$ is cryptographically transformed into a message $\beta_j^{(r)}$. This procedure aims to prevent an opponent from matching output messages with input ones.
- ***Message delivery***: Each transformed message $\beta_j^{(r)}$ is delivered. $N_p^{(r)} = n_r - b_r$ messages are kept in pool for the next round $(r+1)$.

It should be noted that the number of messages to be forwarded follows the binomial distribution $\mathrm{Bin}(n_r, g(n_r))$ and the number of messages kept in pool follows the binomial distribution $\mathrm{Bin}(n_r, 1-g(n_r))$. Therefore with the same number of messages in system before flusing, the number of messages to be forwarded becomes a random variable.

By observing the number of output messages $b_r$ of round $r$, an adversary cannot be certain about the number of messages $n_r$ in system before it flushes and the number of messages $N_p^{(r)}$ kept in pool for the next round. As a result an opponent cannot calculate with certainty the probability of a sender $s_i$ to be the true sender of a delivered message $\beta_j^{(r)}$.

### B. Analysis and Evaluation

*1) Pool size:* In round $r$, the pool size (the number of messages kept in pool) $N_p^{(r)}$ depends on mix function $g$ and the number of messages $n_r$.

$$N_p^{(r)} \sim \mathrm{Bin}\left(n_r, 1-g\left(n_r\right)\right) \tag{4}$$

The expectation and variance of pool size kept in a binomial mix are as follows:

$$E\left[P\left(N_p^{(r)} = x|n_r\right)\right] = n_r\left(1-g\left(n_r\right)\right) \tag{5}$$

$$Var\left[P\left(N_p^{(r)} = x|n_r\right)\right] = n_r g\left(n_r\right)\left(1-g\left(n_r\right)\right) \tag{6}$$

Fig. 2 illustrates pool sizes of several mixes as functions of the total number of messages $n$ before flusing. Solid lines represent expectation values of pool sizes and dotted lines represent values one standard deviation away from expectation values. We use the same parameters as in Fig. 1.

In this illustration, the variance of the pool size grows as $n$ increases for the mix function based on cummulative normal distribution ($g\left(n\right) = f\int_{-\infty}^{n}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$) and the mix function evolving from a timed dynamic mix ($g(n) = f$). For the mix function originates from a timed pool mix ($g(n) = max\{0, 1-N_p/n\}$), the variance of the pool size approaches zero with increasing $n$. Although either

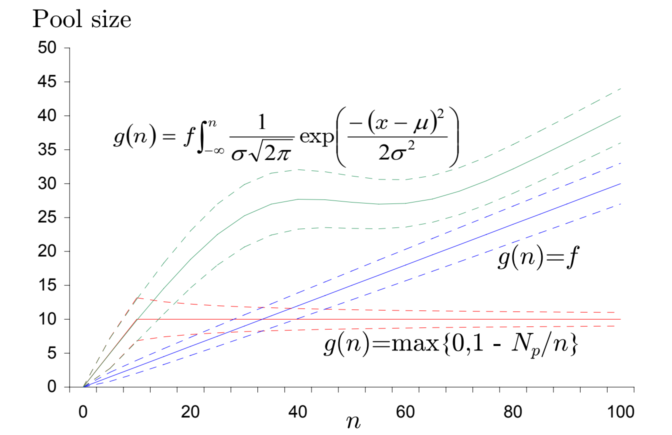

Fig. 2. Expectation value and values one standard deviation away from pool sizes of several binomial mixes.

of these mix functions can be used to construct a binomial mix in RPROB, mix functions with increasing variance when $n$ grows are of preference to increase uncertainty of the system.

*2) Guaranteed anonymity:* For simplicity in presentation, we introduce additional notations: $\varepsilon_S^{(r)} := \left\{ \alpha_k^{(r)} \right\}$ be the collection of all messages sent to the system in round $r$; $A_{s_i}^{(r)}$ be the collection of all messages sent by sender $s_i$ in round $r$. We have $\bigcup_{s_i \in S} A_{s_i}^{(r)} = \varepsilon_S^{(r)}$.

Let $\Phi\left(\beta_j^{(r)}, s_i\right)$ be the probability assigned by an opponent to a sender $s_i$ to send a delivered message $\beta_j^{(r)}$ in round $r$.

$$\Phi\left(\beta_j^{(r)}, s_i\right) = P\left(S\left(\beta_j^{(r)}\right) = s_i\right) \tag{7}$$

In round $r$, there are $N_p^{(r-1)}$ old messages and $a_r$ new messages. Hence the probability of a message $\beta_j^{(r)}$ being a new one is $P\left(\beta_j^{(r)} \in \varepsilon_S^{(r)}\right) = a_r/n_r$ and that of $\beta_j^{(r)}$ being an old message is $P\left(\beta_j^{(r)} \notin \varepsilon_S^{(r)}\right) = N_p^{(r-1)}/n_r$.

The probability that $\beta_j^{(r)}$ is a new message and $s_i$ is its sender is as follow:

$$P\left(S\left(\beta_j^{(r)}\right) = s_i \wedge \beta_j^{(r)} \in \varepsilon_S^{(r)}\right) = \frac{\left|A_{s_i}^{(r)}\right|}{n_r} \tag{8}$$

In case $\beta_j^{(r)}$ is an old message, it arrived in system in some round $k < r$ , then was selected to be kept in system continuously from round $k$ to round $r$–1, finally was selected to be delivered in round $r$. Then the probability that a sender $s_i$ sent $\beta_j^{(r)}$ and $\beta_j^{(r)}$ arrived in system in round $k<r$ is:

$$\begin{aligned} & P\left(S\left(\beta_j^{(r)}\right) = s_i \wedge \beta_j^{(r)} \in \varepsilon_S^{(k)}\right) \\ = & \frac{\left|A_{s_i}^{(k)}\right|}{a_k} \cdot \frac{a_k}{n_k} \cdot \left(\prod_{l=k+1}^{r-1} \frac{N_p^{(l-1)}}{n_l}\right) \cdot \frac{N_p^{(r-1)}}{n_r} \\ = & \frac{\left|A_{s_i}^{(k)}\right|}{n_r} \cdot \left(\prod_{l=k}^{r-1} \frac{N_p^{(l)}}{n_l}\right) \end{aligned} \tag{9}$$

From (8) and (9), the probability that a user $s_i$ is the sender of a delivered message $\beta_j^{(r)}$ in round $r$ can be formulated as follows:

$$\begin{aligned} & \Phi\left(\beta_j^{(r)}, s_i\right) \\ = & \frac{\left|A_{s_i}^{(r)}\right|}{n_r} + \sum_{k=1}^{r-1} \frac{\left|A_{s_i}^{(k)}\right|}{n_r} \cdot \left(\prod_{l=k}^{r-1} \frac{N_p^{(l)}}{n_l}\right) \end{aligned} \tag{10}$$

With RPROB, an opponent cannot calculate with certainty the value of the probability that a sender $s_i$ sent a delivered message $\beta_j^{(r)}$ but only distribution of the value of this probability. As $N_p^{(l)} \sim \text{Bin}\left(n_l, 1 - g\left(n_l\right)\right)$, the expectation value of the probability that a sender $s_i$ sent $\beta_j^{(r)}$ is as follow:

$$\begin{aligned} & E\left[\Phi\left(\beta_j^{(r)}, s_i\right)\right] \\ = & \frac{\left|A_{s_i}^{(r)}\right|}{n_r} + \sum_{k=1}^{r-1} \left|A_{s_i}^{(k)}\right| \frac{\prod_{l=k}^{r-1}\left(1 - g\left(n_l\right)\right)}{n_r} \end{aligned} \tag{11}$$

The anonymity of RPROB are further analysed through examples in the following two scenarios:

**Scenario 1 - The first time sender $s_i$ uses the system**

In case a sender $s_i$ uses an anonymous system for the first time, the only evidence to convince an opponent that $s_i$ is the true sender of a delivered message $\beta_j^{(r)}$ is the activities of $s_i$ in the current round $r$. Thus the probability assigned by adversary to sender $s_i$ to send the message $\beta_j^{(r)}$ is $\left|A_{s_i}^{(r)}\right|/a_r$ and $\left|A_{s_i}^{(r)}\right|/\left(a_r + N_p^{(r-1)}\right)$ for APROB [11] and RPROB, respectively. As $N_p^{(r-1)} \sim \text{Bin}\left(n_{r-1}, 1 - g\left(n_{r-1}\right)\right)$, the propability guessed by an adversary in APROB is always greater than or equal to that in RPROB. Hence RPROB provides better anonymity than APROB in the same context.

In the first experiment of this scenario, we consider an APROB system and three different instances of RPROB. We assume that each system has $n_{r-1} = 100$ messages

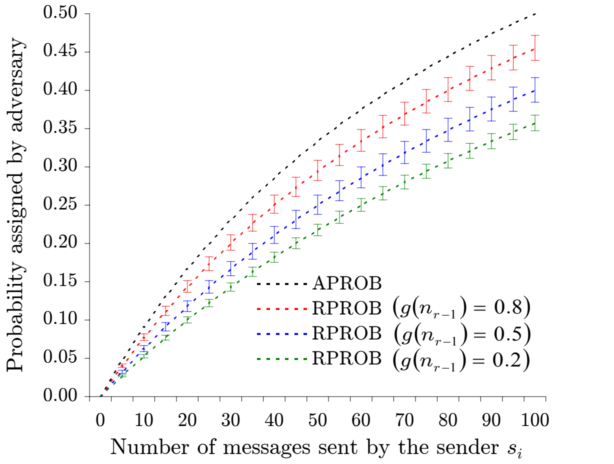

Fig. 3. The probability assigned by an opponent to a sender $s_i$ to send a delivered message $\beta_j^{(r)}$ if $s_i$ uses the system for the first time in APROB and three instances of RPROB.

before it flushes in the previous round $r-1$, and the three RPROB instances correspond to three different functions where $g(n_{r-1}) = g(100) = 0.2$, $0.5$, and $0.8$ respectively. In the current round $r$, a sender $s_i$ begins to send $\left|A_{s_i}^{(r)}\right|$ messages while other senders send 100 messages. Hence each system receives $a_r = 100 + \left|A_{s_i}^{(r)}\right|$ new messages in round $r$.

Fig. 3 illustrates the probability assigned by an opponent to $s_i$ in APROB and the expectation value of the probability guessed by an opponent to $s_i$ in each RPROB instance. The more messages $s_i$ sends in round $r$, the higher the probability assigned to $s_i$ by an adversary. It is clearly shown that any RPROB instance provides higher anonymity than APROB in the same context. The error bars in Fig. 3 represent the variation of the values of the probability corresponding to the values of pool size $N_p^{(r-1)}$ one standard deviation away from expectation values of pool size.

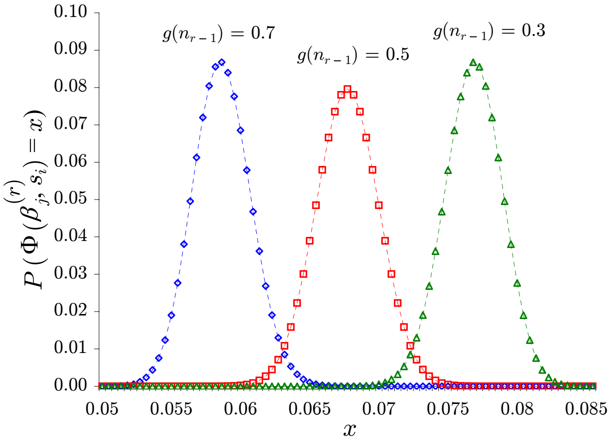

Fig. 4. Distribution of the value of $\Phi\left(\beta_j^{(r)}, s_i\right)$ that a sender $s_i$ to send a delivered message $\beta_j^{(r)}$ if $s_i$ uses the system for the first time.

The second experiment is to illustrate the uncertainty of the probability assigned by an opponent to a sender $s_i$ to send a delivered message $\beta_j^{(r)}$. There are $n_{r-1} = 100$ messages in previous round $r-1$, each of these messages is selected with probability $g(n_{r-1})$. In the current round $r$, a sender $s_i$ begins to use the system for the first time by sending $\left|A_{s_i}^{(r)}\right| = 10$ messages and there are totally $a_r = 100$ new messages.

Fig. 4 illustrates the distribution of the value of $\Phi\left(\beta_j^{(r)}, s_i\right)$ in three instances of RPROB corresponding to three different functions where $g(n_{r-1}) = g(100) = 0.3$,0.5, and 0.7 respectively. The value of $\Phi\left(\beta_j^{(r)}, s_i\right)$ depends on the random variable $N_p^{(r-1)}$, thus the value of $\Phi\left(\beta_j^{(r)}, s_i\right)$ is not deterministic. From the following formula:

$$E\left[\Phi\left(\beta_j^{(r)}, s_i\right)\right] = \frac{\left|A_{s_i}^{(r)}\right|}{a_r + n_{r-1} g\left(n_{r-1}\right)} \tag{12}$$

we have $E[\Phi\left(\beta_j^{(r)}, s_i\right)] = 0.077$, 0.067, and 0.059 corresponding to $g(n_{r-1}) = g(100) = 0.3$, 0.5, and 0.7 respectively.

For a Pool-based APROB Channel with the same environment and pool size $N_p$, we also have $\Phi\left(\beta_j^{(r)}, s_i\right) = 0.077$, 0.067, and 0.059 corresponding to $N_p = 30$, 50, and 70. In this case, an adversary can determine with certainty the value of $\Phi\left(\beta_j^{(r)}, s_i\right)$, while an adversary can only calculate the distribution of $\Phi\left(\beta_j^{(r)}, s_i\right)$ in RPROB systems.

For APROB with the same environment, $\Phi\left(\beta_j^{(r)}, s_i\right)$ is certainly 0.1 as there is no message kept from previous round $r-1$. Clearly any instance of RPROB not only provides higher anonymity than APROB but also prevents an opponent from calculating $\Phi\left(\beta_j^{(r)}, s_i\right)$ with certainty.

**Scenario 2 - Sender $s_i$ uses the system continuously in multiple rounds**

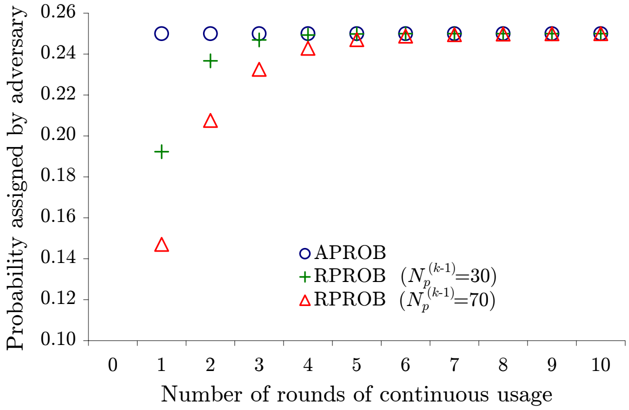

Fig. 5. Expectation value of the probability $P\left(S\left(\beta_j\right) = s_i\right)$ that a sender $s_i$ to send a message $\beta_j$ if $s_i$ uses the system continuously in multiple rounds.

Fig. 5 illustrates expectation value of the probability $P\left(S\left(\beta_j\right) = s_i\right)$ assigned by adversary to sender $s_i$ to send a message $\beta_j$ if sender $s_i$ uses the system continuously in multiple rounds with a constant rate of sending. In this experiment, each system receives a constant $a_k = 100$ new messages per round (including 25 messages sent by $s_i$). For simplicity, each instance of RPROB has a constant pool size ($N_p^{(k)} = 30$ and $N_p^{(k)} = 70$).

The probability assigned by opponent in APROB Channel is a constant (0.25) while the expectation value of this probability in either of the RPROB Channels is always less than 0.25 and converges toward 0.25. This scenario is to illustrate that RPROB system provides higher anonymity than APROB channel even when a user continuously uses a system for multiple rounds.

*3) Probabilistic-real-time property:* Let $\varepsilon_R^{(r)} := \left\{\beta_j^{(r)}\right\}$ be the collection of all messages delivered from the system to next hops or recipients in round $r$. For a message $\alpha_i^{(k)}$ arriving in system in round $k$, the probability that it is forwarded in round $k$ is

$$P\left(\alpha_i^{(k)} \in \varepsilon_R^{(k)}\right) = g\left(n_k\right) \tag{13}$$

and that in consecutive round $k+l$ is:

$$P\left(\alpha_i^{(k)} \in \varepsilon_R^{(k+l)}\right) = g\left(n_{k+l}\right) \prod_{i=0}^{l-1} \left(1 - g\left(n_{k+i}\right)\right) \tag{14}$$

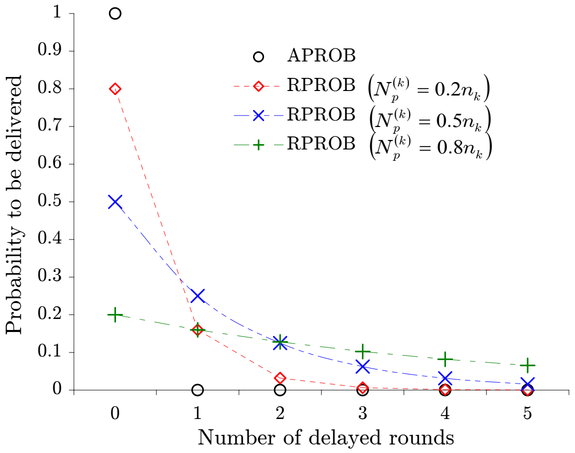

Fig. 6. Probability to deliver a message in APROB Channel and RPROB Channels.

Fig. 6 illustrates probability of delivering a message after its arrival. In this experiment, each system has the same number of new messages per round (i.e. constant traffic). In APROB Channel, it is certain (with probability of 1) that any message will be delivered in the same round. For simplicity, each instance of RPROB has a constant pool size ($0.8n_k$, $0.5n_k$, and $0.2n_k$ for all $k$). As every message is in general delivered within the first few rounds with high probability, RPROB provides probabilistic-real-time property. Note that an adversary cannot be certain when a message is delivered in RPROB.

*4) Flexibility for each user:* Cleary there is a constraint between the maximum number of messages $\varphi(\theta)$ a sender $s_i$ can send in a pre-defined timeslot $\Delta t$ and his frequently used source-hiding property $\theta$. If $s_i$ wants to send a lot of messages in a timeslot, $s_i$ should not assign high source-hiding property to his messages. In case $s_i$ wants to ensure high anonymity for his activities, $s_i$ should send only a small number of messages in a timeslot. RPROB allows flexibility for any user to determine his/her preference of anonymity requirement and the rate of sending messages.

For RPROB Channel, we adopt the strategy [11] to classify sent messages into two categories: ill-timed messages are messages excess $\varphi(\theta)$ and well-timed messages otherwise. The condition for flushing channel's buffer is restricted to only well-timed messages.

## IV. Conclusions and Future Works

The main purpose of our proposal is to define a family of anonymous communication systems (with randomness) to resist global active adversary. RPROB is not a single concrete anonymous communication system but defines an infinite set of anonymous systems based on binomial mixes. Each instance of RPROB uses a binomial mix to preserve variable number of messages in its pool when flusing to prevent an opponent from determining with certainty when a message of interest will be delivered. Furthermore due to the randomness provided by a binomial mix, an adversary can no longer calculate with certainty the probability that a user is a sender of a delivered message as in deterministic systems, including Pool-based APROB Channel.

Due to variable number of messages kept in the pool from previous rounds, RPROB channel, in general, provides higher anonymity than APROB Channel with the same environment (the same external activities and users' rate of sending messages). Especially the expectation probability assigned to a user is considerably small when he/she first uses the system or continues to use the system after a pause for several recent rounds. Besides users can personalize their own source-hiding property under probabilistic-real-time condition.

In fact, some binomial mix frameworks may have some undesirable properties, such as rapidly increase of pool size, low variance of pool size... Thus it is necessary to study the properties of mix functions used to construct binomial mixes to eliminate possible 'weak' functions. Furthermore specific classes of mix functions should be proposed to optimize the performance in some contexts or applications.

## References

[1] B. Gedik and L. Liu, "Protecting location privacy with personalized k-Anonymity: Architecture and algorithms," *IEEE Transactions on Mobile Computing*, vol. 2007, 2007.

[2] C. Bettini, S. Jajodia, P. Samarati, and X. Wang, Eds., *Privacy in Location-Based Applications - Research Issues and Emerging Trends*. Springer, 2009.

[3] D. Kesdogan, J. Egner, and R. Büschkes, "Stop-and-go mixes: Providing probabilistic anonymity in an open system," in *Information Hiding (IH 1998)*, ser. LNCS, vol. 1525, 1998, pp. 83–98.

[4] G. Danezis, "Forward secure mixes," in *7th Nordic Workshop on Secure IT Systems*, 2002, pp. 195–207.

[5] C. Diáz and B. Preneel, "Reasoning about the anonymity provided by pool mixes that generate dummy traffic," in *Information Hiding (IH 2004)*, ser. LNCS, vol. 3200, 2004, pp. 309–325.

[6] P. Syverson, G. Tsudik, M. Reed, and C. Landwehr, "Towards an analysis of onion routing security," in *Designing Privacy Enhancing Technologies*, ser. LNCS, vol. 2009, 2000, pp. 96–114.

[7] U. Moeller, L. Cottrell, P. Palfrader, and L. Sassaman, "Mixmaster protocol," 2002.

[8] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," in *13th USENIX Security Symposium*, 2004, pp. 303–320.

[9] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, vol. 4(2), pp. 84–88, 1981.

[10] G. Danezis and C. Diaz, "A survey of anonymous communication channels," Microsoft Research, Tech. Rep., 2008.

[11] G. Tóth and Z. Hornák, "The APROB channel: Adaptive semi-real-time anonymous communication," in *Security and Privacy in Dynamic Environments*, ser. IFIP, vol. 201, 2006, pp. 483–492.

[12] C. Gulcu and G. Tsudik, "Mixing e-mail with Babel," in *Network and Distributed Security Symposium (NDSS'96)*, 1996, pp. 2–16.

[13] M. T. Tran, T. T. Nguyen, and I. Echizen, "Pool-based APROB channel to provide resistance against global active adversary under probabilistic real-time condition," in *2008 IEEE/IFIP International Symposium on Trust, Security and Privacy for Pervasive Applications (TSP-08)*, 2008, pp. 257–263.

[14] C. Diáz and A. Serjantov, "Generalising mixes," in *3rd Privacy Enhancing Technologies Symposium (PET 2003)*, 2003, pp. 18–31.

[15] A. Serjantov, "A fresh look at the generalised mix framework," in *7th Privacy Enhancing Technologies Symposium (PET 2007)*, ser. LNCS, vol. 4776, 2007, pp. 17–29.

[16] A. Serjantov, R. Dingledine, and P. Syverson, "From a trickle to a flood: Active attacks on several mix types," in *Information Hiding (IH 2002)*, ser. LNCS, vol. 2578, 2002, pp. 36–52.

[17] G. Tóth and Z. Hornák, "Measuring anonymity in a non-adaptive real-time system," in *4th Privacy Enhancing Technologies Symposium (PET 2004)*, ser. LNCS, vol. 3424, 2004, pp. 226–241.