COST TO DEFEAT THE N-1 ATTACK

Neil Chettiar

Abstract

Through this research topic a possible solution to defeat the N-1 active attack on a general MIX is proposed, the basic parameters of the MIX are formulated and checked to see how this MIX fairs under the N-1 attack. The approach used was to add dummy traffic to the output of the MIX. This paper mathematically derives the performance of the proposed MIX under extreme situations and proposes counter measures to mitigate the attack.

1. INTRODUCTION

The Blending attacks have always been a thorn at the side of many MIX networks. Even to this day, the attacks can be successful in reducing the anonymity of a target MIX to zero or close enough to be able to successfully tell if a message coming out of the MIX is the target message or not, with a reasonable probability of success for most MIX networks.

The N-1 attack is a particular kind of blending attack where the attacker blocks all incoming genuine messages to the MIX, fills the MIX with its own N-1 messages and waits for the target message to arrive at the MIX. This triggers the MIX to fire and the attacker can successfully distinguish between its own messages and the target message from the output messages of the MIX. The '*From a Trickle to a Flood: Active Attacks on Several MIX Types'* paper by *Serjantov et al*. talks about the N-1 attack and its many flavors for the different MIXes such as the Threshold, Timed and Pool MIXes. The paper goes into detail explaining how the attack can be exact and certain on the general MIXes and how it can be exact but uncertain on the pool MIXes.

The underlying idea that can be taken from this is that the attacker, in order for the N-1 attack to be successful, must be able to account for the messages that come out of the MIX so that he can identify the target message and follow it to the receiver.

In this paper, a possible solution to the N-1 attack is proposed, the approach is described and the MIX under the N-1 attack is analyzed. Through the study it is seen that the MIX is still susceptible to the N-1 attack to a certain degree and so improvements or counter-measures to the design is suggested in order to help mitigate the attack.

2. THE APPROACH

Adding a random number of dummy messages to the output of the MIX is the general approach followed in-order to obfuscate the attacker. The objective is to reduce the ability of the attacker to be able to definitively account for all the output messages of the MIX, or to make it resourcefully hard to determine which of the output messages is the actual target message.

However, it has been seen that this approach is not very successful for general MIXes such as the Threshold, Timed, Threshold OR Timed and Threshold AND Timed MIXes. For these MIX setups, since the triggering condition is easy to decipher, with a little effort the attacker is able to accurately tell when the MIX is going to fire. Also, the attacker can predict when the target message is going to come out of the MIX as an output. This makes the N-1 attack certain, but due to the addition of random dummy messages at the output of the MIX, the attacker is not able to distinguish between the target messages and

the remaining dummy messages in the output. This makes the attack inexact.

In this way, as a result of an N-1 attack performed on a general MIX, the resulting anonymity set size of the MIX is effectively reduced to k+1 where k is the number of random messages that are added to the output along with the other messages from the MIX. To improve this anonymity set size, the proposed MIX is based on the pool MIX design.

3. THE PROPOSED MIX

The pool MIX design is chosen because it offers a better anonymity metric as the attack cannot know for sure if the message has ever left the MIX or not.

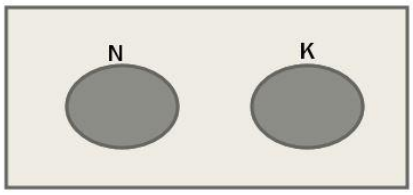


Figure 1: Proposed MIX

The proposed MIX maintains two different pools: one of N genuine messages and another of K dummy messages. It is assumed that the attacker does not have access to the dummy messages, but he can flood the genuine message pool, as he would in the case of an N-1 attack.

To define how this pool MIX works the basic functions according to which it chooses the messages from the pool of N genuine message pool is defined, and the function according to which the number of dummy messages are chosen.

$$n_i = \begin{cases} 0 \text{ if } N_i < N_0 \\ [N_0, N_i/2] \text{ if } N_i > N_0 \text{ and } N_i/2 > N_0] \\ N_0 \text{ if } N_i > N_0 \end{cases}$$

From the formula given above, n_i which is the number of genuine messages chosen at round i is selected depending on the number of messages in the pool of N_i messages at round i in comparison to the threshold value of the MIX given by N_0.

Based on this value of n_i, the number of dummy messages that are chosen are defined as:

$K_i = RAND[n_i /2, n_i]$

According to this definition, the number of dummy messages chosen is a random value based on the number of genuine of messages chosen. Hence, the number of dummy messages generated is limited to a number that is proportional to the number of genuine messages chosen. The limit on the number of dummy messages was set to ensure that the MIX does not add an unnecessarily large number dummy messages at the output, hopefully preventing the network from being flooded with dummy messages.

Based on this design, the probability of selecting a message from the pool of genuine messages can be given as:

$$p(round_i) = \begin{cases} p(n_i) \text{ if } r = i \\ p(n_i) \prod_{j=r}^{i-1} (1 - p(n_j)) \text{ if } r < i \end{cases}$$

$$p(n_i) = \frac{n_i}{N_i}$$

Where $p(round_i)$ is defined as the probability of choosing a message at round i. $p(n_i)$ is defined

as the probability of choosing n_i messages from the pool of N_i messages.

For p(round$_i$), the first line defines the number of n_i messages chosen to be flushed in the same round that they entered the MIX. The second line defines the case when the n_i messages that are chosen entered the MIX at a round r before the current ith round.

From these equations we can generate the equations for the Anonymity set size of the proposed MIX. Anonymity set size is defined as the probability of linking the output messages to the input messages. Hence, the chances of linking the output messages to the input messages can be given as:

$$p(I_i) = \begin{cases} \frac{1}{n_i} \text{ if } i = r \\ \frac{1}{n_i} \prod_{j=r}^{i-1} (1 - p(n_j)) \text{ if } i < r \end{cases}$$

Where, p(I_i) denotes the probability of linking the output messages to the input messages for the current ith round.

Note: this is the anonymity set size of the MIX without the dummy messages, the anonymity due to the dummy messages will be seen after defining the entropy of the system.

We need to consider the effective anonymity set size of the MIX as the AS does not give a clear idea of the actual anonymity that the MIX provides. This mainly stems from the theoretical AS of a pool MIX tending to *infinity* on account of all the messages that ever entered the MIX. We define the Entropy of the MIX as:

$$H_s = -\sum_{i=1}^{r} a_i p(I_i) log_2(p(I_i))$$

Where a_i is defined as the number of messages that have arrived in the MIX at round$_i$. Again, this is the Entropy of the system without the dummy messages.

For the Entropy of the MIX along with the Dummy messages, we can calculate the Entropy of the dummy messages due to the dummy messages separately and add this to the Entropy of the MIX due to the input messages because Entropy is a scalar term hence, the addition of these two non-related components of the MIX is a legal step.

Recall that in the design of the MIX, we added the dummy messages at the output and hence, the probability of selecting the n_i messages from the pool of N_i messages is not dependent on the dummy messages.

Hence, to calculate the Entropy of the MIX along with Dummy messages can be calculated as:

$$p_d = \frac{k_i}{n_i + k_i}$$

$$p(I_i) = (1 - p_d) * p_i$$

$$p_i = \frac{p(n_i) + k_i}{a_i}$$

Where p_d is defined as the probability of selecting a dummy message from the output messages and p_i is defined as the probability of the message being chosen from the input.

Using these terms, we define the Entropy of the proposed MIX with Dummy messages as:

$$H = -p_d log_2(p_d) - \sum_{i=1} (1 - p_d) p_i log_2((1 - p_d) p_i))$$

Source: Ref1

From this we can see that the first term denotes the Entropy due to the Dummy Traffic and the next term is due to the genuine messages chosen from the pool of messages.

Analyzing the output generated by the MIX, the probability of finding the target message from the output is:

$$= \begin{cases} \frac{1}{n_j+k_i} \text{ if } i = r \\ \frac{1}{n_i+k_i} \prod_{j=r}^{i-1} (1 - p(n_j)) \text{ if } i < r \end{cases}$$

Where n_i and k_i represents the messages chosen and the dummy traffic added to the output respectively.

4. UNDER THE N-1 ATTACK

Now that we know how the MIX behaves, let us analyze the performance of the proposed MIX under the N-1 attack.

We can consider the worst possible case when the MIX works at its threshold values, i.e. $n_i = N_0$ and $k_i = n_i/2$. Also, we can consider the worst possible scenario that the attacker was successful in flushing all the genuine messages from the Pool of N messages and replace it with its own poison messages and the one target message that it wants to follow.

In this situation, the anonymity provided by the MIX effectively gets reduced from $1/(n_i + k_i)$ to a value given as $1/(1 + k_i)$ where the value $1 + k_i$ represents the number of unaffected dummy messages and the one target message.

While this value looks surprisingly like the performance of a general Threshold MIX using dummy messages, I would like to highlight the fact that this situation is the most pessimistic worst case scenario of any N-1 attack. This situation assumes that the Pool of messages of the proposed MIX has been completely overrun and the attacker knows when the pool MIX is going to fire. Under these extreme circumstances, the same degree of Anonymity is afforded by both the binomial and Cottrell MIXes.

On a more average scenario, the Anonymity provided by the proposed MIX can be given as $1/(m_i + k_i)$ where m_i can be defined as the number of genuine messages that are chosen from the pool of N_i messages of the pool that contains a large number of attacker's poison messages.

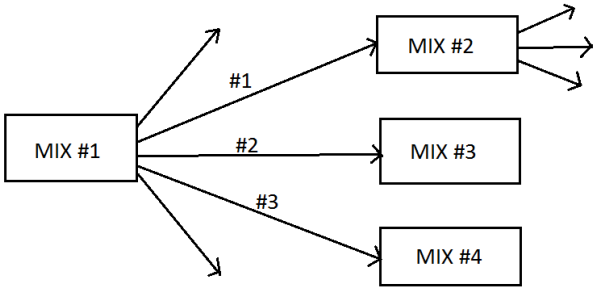
Hence, to put things in perspective, the output of messages from the proposed MIX under the N-1 attack can be shown as:

$$(1 + k_i) < (m_i + k_i) < (n_i + k_i)$$

Again, these terms are generated only considering the situations when the messages are selected in the same round i that they entered the MIX. For the situations when the messages are chosen in a round after they had entered the MIX, we multiply all the previous rounds during which the message was not chosen to be flushed. This further adds to the Anonymity of the system for all scenarios analyzed above.

The success of the MIX under the N-1 attack can be subjective. On one hand, the attacker has been successful in considerably reducing the anonymity of the MIX, but even through his strongest attack the attacker could not reduce the AS to zero. In the same breath, the MIX has been successful in resisting the active N-1 attack by restricting the attack to be uncertain and inexact.

Furthermore, as a result of the attack, the work of the attacker becomes exponential as the attacker still needs to follow each of the un-accounted messages of the output through the network to its receiver.



This is depicted in this short diagram where the MIX under attack is still able to maintain 3

messages that the attacker cannot account for. The attacker then needs to follow these potential target messages to their destinations which may or may not be other MIXes in the network. The attacker has to then attack these new target MIXes which will again launch messages that the attacker cannot account for, and the process repeats for all such MIXes that the current target MIX sends its messages to.

In essence, the success or failure of the attack boils down to the resources that the attacker has at his disposal. If the attacker can easily keep following 50 or 100 messages to the end receiver, then the use of Dummy messages in the network was not beneficial towards the anonymity of the MIX. However, if this makes the attack too expensive for the attacker to pursue the target message, the use of Dummy messages in the proposed MIX setup can be considered as a success.

5. COUNTER MEASURES

To counter the N-1 attack, the following steps are proposed:

- Using Heartbeat traffic introduced by *George Danezis* and *Len Sassaman* in their paper '*Heartbeat Traffic to Counter (n-1) Attacks'*. This paper shows us that dummy traffic can be used to successfully detect an active N-1 attack and thus allow the MIX to take appropriate steps to ensure that the attack is foiled.
- Use Re-group-and-Go traffic to detect an attack. The authors *Jin-Qiao Shi, Bin-Xing Fang* and *Li-Jie Shao* introduce the concept in their paper *'Regroup-And-Go MIXes to counter the (n-1) attack'*. The idea is to segment the message at the sender's side and send the segments of the message to the various MIXes in the MIX network. The segments are scheduled to re-group at a certain MIX in the network, if all the segments fail to re-group within the time-out period, the MIX detects an on-going attack and can take countermeasures.
- Some of the possible steps the MIX can implement once an N-1 attack has been detected are:
    - Wait for the attack to pass: The MIX can wait for an arbitrarily large amount of time till the attacker essentially stops flooding the MIX with poison messages and gives up.
    - Send Crisis messages to the neighboring MIXes in the MIX network to alert them about the N-1 attack. Upon receiving a crisis message from a MIX, the receiving MIX can blacklist it by not sending or drop messages that come from the sending MIX.
    - As a last resort, the MIX under attack can drop all messages from its message pool. This is a lose-lose situation as neither the attacker nor the MIX is succeeding in its tasks. This could be considered as a Denial-of-Service attack.

6. COST OF USING DUMMY MESSAGES

While this paper talks about the theoretical benefits of using Dummy messages to counter the N-1 attack, very little can be said about the practical effects of the same in a communication network.

To obtain useful information about the effects of dummy traffic in the MIX network, a careful implementation of the proposed MIX needs to be carried out in a controlled setting. This could not be accomplished in this paper due to time constraints and due to the following reasons:

- Modelling a MIX network with 4-5 MIXes using the same parameters as the proposed MIX design could have one of two outcomes: either the network supporting the MIX network crashes on account of all the dummy traffic generated or the network works, without any drop in packets transferred. In either situation, analyzing the log files to isolate issues, if they do occur, would be a task because of the multiple points of failure that such a network could have.
- Modelling the MIX network on a small scale would be restricted to a homogenous network, whereas the MIX network might behave very differently on a heterogeneous network which leans more towards a realistic scenario.
- To find the saturation point of the network, the MIX network would have to be simulated in a careful and closely monitored simulation where the number of dummy packets sent by each MIX in the network needs to be slowly incremented until the packet drops begin to become apparent.
- Simulating the N-1 attack on a particular MIX in the network would result in a generation of a large number of dummy messages, and the outcome of the experiment would not provide much insight as the anonymity afforded by the MIX under attack has already been theoretically derived.

7. CONCLUSIONS AND FUTURE WORK

This paper studies the effect of adding dummy messages to the output of a Threshold Pool MIX and delves into the theoretical performance of the MIX under the N-1 attack. While the paper does not provide empirical data to back up the claims made, the experimental simulations of such a network is left as a major chunk of future work for this topic.

Areas of the MIX design that can be further improved are:

- Formulate better parameters according to which the number of Dummy messages are chosen. Incorporate the information from the Heartbeat traffic or the Re-Group-and-Go counter measures to improve the way the number of dummy messages are chosen.
- Design checks into the MIX to ensure that the Heartbeat traffic does not falsely identify a sudden surge of genuine messages as an N-1 attack.
- Incorporate variable thresholds depending on the rate of fire of the MIX: if the MIX does not fire in a long time due to the scarcity of incoming messages, lower the threshold to improve the overall throughput of the MIX.
- Loose Routing to shift or distribute excessive load on a single MIX to other MIXes in the network.

8. REFERENCES

1. ***Reasoning about the anonymity provided by pool mixes that generate dummy traffic. – Claudia Diaz and Bart Preneel . Source: http://freehaven.net/anonbib/cache/pool-dummy04.pdf***
2. ***Heartbeat traffic to counter (n-1) attacks. - George Danezis and Len Sassaman. Source: http://dl.acm.org/citation.cfm?id=1005154***
3. ***From a trickle to a Flood: Active attacks on several MIX types. - Andrei Serjantov and Roger Dingledine and Paul Syverson. Source: http://freehaven.net/doc/batching-taxonomy/taxonomy.pdf***

4. ***Regroup-And-Go mixes to counter the (n-1) attack. - Jin-Qiao Shi, Bin-Xing Fang and Li-Jie Shao. Source: http://www.emeraldinsight.com/doi/abs/10.1108/10662240610656528?journalCode=intr***

5. ***On blending attacks for mixes with Memory extended version. - Luke O'Connor. Source: http://freehaven.net/anonbib/cache/ih05-Luke.pdf***

6. ***Active attacks on Reputable MIX Networks. - Longhai Li, Shaofeng Fu, Xiangquan Che. Source: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5207899&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5207899***

7. ***Untraceable electronic mail, return addresses and digital pseudonyms. - David L. Chaum. Source: http://www.freehaven.net/anonbib/cache/chaum-MIX.pdf***