두 재배열 방식을 동시에 사용자에게 제공하는 하이브리드 믹스*

강필섭 ° 김종욱 홍만표 이경석
아주대학교 정보통신전문대학원
{kpslen8205, kju, mphong, khlee} @ajou.ac.kr

Hybrid Mix that gives user two reordering method at a time

Pilseob Kang° Jonguk Kim Manpyo Hong Kyungsuk Lhee
Graduate school of information and communication, Ajou university

요 약

지금까지 각 믹스 시스템은 재배열 방식을 한가지씩만 사용했기 때문에 사용자는 믹스 시스템이 제공하는 재배열 방식만을 사용할 수 밖에 없었다. 그러나 믹스 시스템의 일부 사용자들은 메시지의 지연시간보다 익명성을 보장받을 수 있는 재배열 방식을 원하지만, 다른 사용자들은 메시지의 익명성 보다는지연시간을 보장받을 수 있는 재배열 방식을 사용하기 원할 수 있다. 본논문에서는 이러한 믹스 시스템 사용자들을 위해 하이브리드 믹스(hybrid mix)을 제안한다. 그리고 제안한 하이브리드 믹스와 기존의 믹스의 재배열 방식을 비교해 익명성과 메시지 지연시간의 보장 측면에서 어떤 특징을 갖는지 분석하고 마지막으로 하이브리드 믹스를 사용할 때 발생할 수 있는 상황에 대해 제시한다.

1. 서 론

인터넷이 대중에 널리 전파되면서 이를 이용한 서비스가 사용자에게 제공되었다. 하지만 인터넷은 그 구조적 특징으로 인해 사용자가 보낸 메시지가 곧바로 목적지로 전송되는 것이 아니라 네트워크 상의 여러 노드를 거쳐 수신자에게 전송이 된다. 그래서 해당 메시지와 관련 없는 제 3자가 쉽게 메시지를 볼 수 있다. 그뿐만 아니라, IP 헤더에 메시지의 송수신자 주소가 포함되기 때문에 관찰자는 메시지의 출발지와 목적지를 알아낼 수 있다. 인터넷의 이런 구조적 특징은 개인의 프라이버시 보호에서 익명성(anonymity)보장에 악영향을 끼칠 수 있으므로 이를 보호하기 위한 익명통신로에 대한 연구가 진행되었다.

익명통신로는 메시지 관찰자가 메시지의 송신자와 메시지의 대응관계를 알 수 없도록 하는 기술을 의미한다[6]. 익명통신로에 대한 최초의 연구가 이루어진 것은 D. Chaum이 제안한 믹스넷(Mixnet)이다[1]. 믹스넷은 크게 2가지 방법을 사용해 메시지의 익명성을 보장한다. 첫 번째는 믹스로 들어오는 메시지의 복호화를 통해 들어오는 메시지의 모양을 바꿔 노드 안으로 들어온 메시지와 노드 밖으로 내보내는 메시지의 연관성을 관찰자로부터 숨기는 방법이다. 두 번째는 메시지의 재배열을 통해서 들어오고 내보내는 메시지의 순서를 바꿔 관찰자로부터 메시지의 연관성을 숨기는 방법이다.

믹스 시스템에 관한 연구가 진행되면서, 메시지 재배열 방식은 각각 독립적인 특징을 가지고 발전했다. 그러나 기존의 믹스는 재배열 방식을 한가지씩만 사용했기 때문에 사용자들은 해당 믹스 시스템에서 제공해 주는 것 이외의 방식을 사용할 수 없었다.

본 논문에서는 이와 같은 상황을 해결하기 위해서 기존의 timed mix와 timed pool mix를 혼합한 하이브리드 믹스(hybrid mix)를 제안한다. 우선 2장에서는 기존의 믹스 재배열 방식이 어떻게 동작하고 어떤 장점과 단점이 있는지 알아본다. 3장에서는 하이브리드 믹스 시스템의 동작 방법과 기존 timed mix, timed pool mix와 메시지 지연시간과 익명성 보장 측면에서 비교한다. 마지막으로 실제로 하이브리드 믹스를 사용하면서 일어날 수 있는 상황을 분석한다.

2. 관 련 연 구

믹스는 재배열 방식에 따라 심플 믹스(simple mix), 풀 믹스(pool mix) 그리고 컨티뉴어스 믹스(continuous mix)로 분류된다[4]. 그림 1은 믹스 시스템의 종류를 보여준다. 본 논문에서는 timed mix와 풀 믹스를 중심으로 설명해 나간다. 풀 믹스는 크게 다음과 같이 3가지 방식으로 구분할 수 있다[4].

- Threshold pool mix

* "본 연구는 21세기 프론티어 연구개발사업의 일환으로 추진되고 있는 정보통신부의 유비쿼터스컴퓨팅및네트워크 원천기반기술개발사업의 지원에 의한 것임"

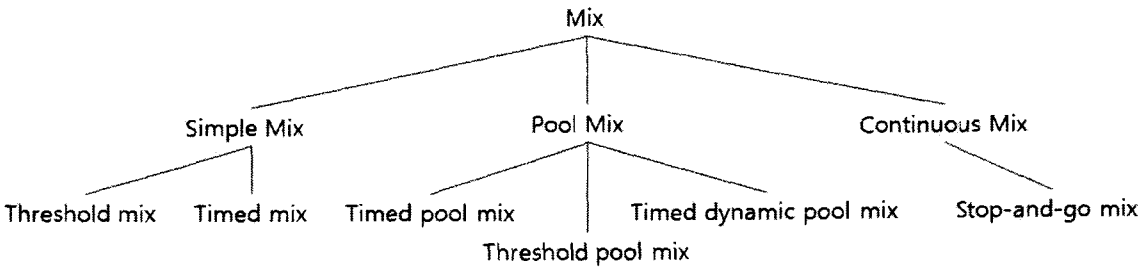

그림 1. 재배열 방식에 따른 믹스의 분류

- Timed pool mix
- Timed dynamic pool mix

풀 믹스에서는 메시지의 익명성과 지연시간의 보장이 트레이드오프(trade-off)관계에 있기 때문에 어떤 특정한 재배열 방법이 최선의 것이라고 말할 수는 없다. 다음부터 각 믹스 재배열 방식이 익명성과 메시지 지연시간의 보장 측면에서 어떤 특징을 갖는지 분석한다.

2.1 Threshold mix and Threshold pool mix

Threshold pool mix는 풀 안에 저장된 메시지의 수가 일정한 수에 도달하면 미리 정한 수(p)의 메시지만 남기고 전부 내보내는 방식이다. D. Chaum이 제안한 믹스넷이 사용한 threshold mix는 p = 0인 threshold pool mix의 한 종류라고 할 수 있다. 항상 일정 수의 메시지가 pool안에 채워진 후 메시지가 내보내지기 때문에 관찰자는 pool의 크기만큼 메시지를 고려해야 한다. 하지만 네트워크 트래픽이 낮을 경우 메시지가 내보내지지 않고 계속 pool 안에 남을 수 있다는 문제가 있다. 게다가 블렌딩 공격(blending attack)에 취약하다는 단점도 존재한다[5]. 그림2는 임계치가 100이고 p는 60인 재배열 알고리즘을 사용하는 믹스의 그래프로 P(n)은 믹스 내 총 메시지 개수에서 메시지를 내보낼 확률이다

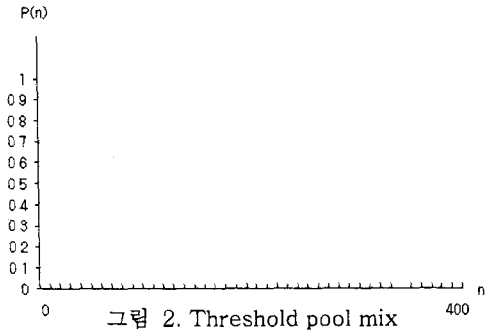

그림 2. Threshold pool mix

2.2 Timed mix

이 믹스는 일정한 시간 주기 동안 저장된 모든 메시지를 내보낸다. Timed mix는 네트워크 트래픽이 낮아 풀 안에 들어가는 메시지가 없을 경우 threshold pool mix가 메시지를 계속 저장하는 상황을 방지하기 위해서 제안된 방식이다. 하지만 Timed mix는 threshold pool mix와 반대로 한 라운드에 적은 양의 메시지가 도착할 경우 익명성 보장에 악영향을 받는다. 메시지가 내보낼 시간이 되었을 때 풀 안에 메시지가 한 개만 있다면 관찰자는 이 메시지가 어디로 가는지 쉽게 알 수 있다. 공격자는 이를 활용해 고의적으로 해당 노드로 보내는 메시지를 줄이는 방식으로 믹스를 공격한다.

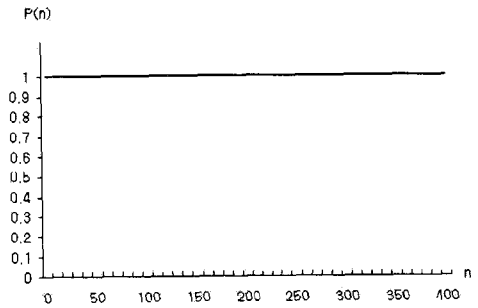

그림 3. Timed mix

2.3 Timed pool mix

이 믹스는 만약 일정한 시간이 된 후 풀 안에 미리 정해놓은 수(p) 이상의 메시지가 존재할 경우 f개의 메시지를 제외한 나머지를 모두 보내고 그렇지 않을 경우 메시지를 내보내지 않고 다음 라운드에 들어오는 메시지를 수신한다. 이 방식은 저장하고 있는 풀의 크기에 따라서 익명성과 메시지 지연시간이 영향을 받는다. 즉 풀의 크기가 크다면 메시지의 익명성 집합(anonymity set)의 크기가 커지므로 익명성이 풀의

크기가 작은 경우보다 더 보장되는 반면 이전 라운드에 들어온 메시지가 풀 안에 남아있을 확률도 높아지기 때문에 메시지 지연시간이 길어지는 단점이 있다.

공격자는 timed pool mix가 많은 메시지를 받는 경우 정해진 수의 메시지만 저장하고 나머지를 내보내는 점을 이용해 플루딩 공격(flooding attack)을 노드에 가할 수 있다[5]. 그림4는 p = 20인 경우의 timed pool mix를 나타낸 그래프이다.

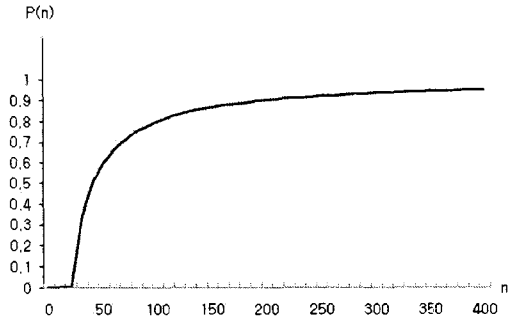

그림 4. Timed pool mix

2.4 Timed dynamic mix

이 방식은 믹스에 들어오는 메시지 수가 n개 이상이 되었을 때부터 일정 시간이 지나면 메시지를 내보내는 방식이다. 그래서 네트워크 트래픽이 높거나 낮은 경우에 따라서 메시지를 내보내는 빈도가 달라진다. 또한 믹스에서 내보내지는 메시지의 수는 timed pool mix와는 달리 믹스 안에 들어온 메시지에서 일정 비율을 선택하는 함수 P(n) 에 의해서 결정된다.

Timed pool mix는 고정된 숫자 만을 남기고 나머지 메시지는 모두 내보내기 때문에 믹스 안으로 들어오는 메시지의 수가 증가할 수록 들어온 메시지에 수에 비해 풀 안에 저장되는 메시지의 수의 비율이 적어진다. 반면에 timed dynamic pool mix는 들어온 메시지와 풀 밖으로 내보내는 메시지의 비율이 P(n)으로 유지되기 때문에 timed pool mix에서 가능한 플루딩 공격(flooding attack)은 timed dynamic pool mix에선 유효하지 않다.

하지만 timed dynamic pool mix는 믹스로 들어오는 메시지의 수가 증가할수록 풀 안에 저장되는 메시지 또한 증가하기 때문에 그에 따른 메시지 저장 비용과 메시지 지연시간이 증가한다. 그림5는 P(n) = 0.7이고 n=20 일 경우의 그래프를 나타낸다.

3. 하이브리드 믹스

2장에서 본 것처럼 믹스 재배열 방식은 각각 서로 다른 특징을 가지고 있다. 기존의 믹스 시스템은 하나의

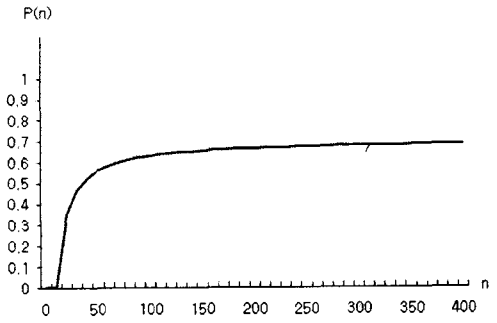

그림 5. Timed dynamic mix

재배열 방식 만을 사용했기 때문에 사용자는 자신의 메시지를 사용자의 메시지는 믹스가 사용하는 재배열 방식을 사용할 수 밖에 없었다. 그래서 이번 장에서 timed mix와 timed pool mix를 혼합한 하이브리드 믹스 시스템을 제안한다. Timed mix는 메시지의 지연시간이 보장되지만, 다른 풀믹스에 비해서 상대적으로 익명성 보호에 취약하다. 반면에 timed pool mix는 timed mix보다 메시지의 익명성이 보장되지만 메시지의 지연시간이 길어질 수 있다. 이와 같이 하이브리드 믹스는 서로 다른 특징을 가지는 믹스를 사용자가 선택할 수 있도록 해준다. 다음 장부터 하이브리드 믹스가 서로 다른 재배열 방식을 사용하는 메시지를 어떻게 처리하는지 믹스가 동작하는 방법에 대해 설명하고, 하이브리드 믹스가 메시지 지연시간과 익명성 측면에서 기존의 timed mix와 timed pool mix와 비교해서 어떤 성질을 가지고 있는지 알아본다.

3.1 하이브리드 믹스의 동작 방법

하이브리드 믹스가 기존 믹스 시스템과 달리 어떻게 동작하는지 알기 위해서 크게 메시지 포맷형식과 재배열 알고리즘 방식으로 나눠 설명한다.

3.1.2 메시지 형식

하이브리드 믹스는 믹스로 들어오는 메시지들이 사용하는 재배열 방식을 관찰자가 구별할 수 없게 timed mix를 사용하는 메시지와 timed pool mix를 사용하는 메시지 둘 다 같은 포맷형식으로 사용한다. 하지만 믹스 시스템은 노드로 들어온 메시지를 구별해서 처리해야 하기 때문에 메시지 포멧에 timed mix와 timed pool mix를 구별하는 플래그(flag)를 설정한다.

3.1.2 재배열 알고리즘

하이브리드 믹스는 timed mix방식을 사용하는 메시지와 timed pool mix 방식을 사용하는 메시지를 각각의 재배열 알고리즘을 사용해서 처리하기 위해 서로 독립적인 2개의 풀을 노드 안에 가진다. 하이브리드 믹스는 각 풀 안에 들어있는 메시지를 처리할 때 기존의 믹스와 마찬가지로 각각의 재배열 알고리즘을 사용한다. 그러나 들어오는 메시지와 나가는 메시지가 어떤 재배열 방식을 사용하는지 구별할 수 없게 메시지 포맷을 동일하게 만든 것처럼 하이브리드 믹스의 재배열 알고리즘은 timed mix와 timed pool mix의 메시지를 내보낸 시간과 메시지를 받아들이는 시간주기를 동일하게 한다. 즉 하이브리드 믹스는 일정한 시간 동안 다른 노드로부터 오는 메시지를 수신하고, 수신된 메시지를 복호화해 각 풀로 구분해서 저장한다. 메시지를 내보낼 시간이 되면 하이브리드 믹스는 각 믹스가 사용하는 재배열 알고리즘을 사용해 동시에 두 종류의 메시지들을 내보낸다.

3.2 적용 효과

믹스를 거쳐가는 메시지는 복호화를 통해 메시지의 모양이 바뀌기 때문에 관찰자는 들어오는 메시지와 나가는 메시지를 서로 구별을 할 수 없다. 하이브리드 믹스도 timed mix를 사용하는 메시지와 timed pool mix를 사용하는 메시지가 동일한 메시지 포맷을 가지고, 두 재배열 방식을 사용하는 메시지들이 동일한 시간 동안에 풀 안에 저장되었다가 모양이 바뀐 후 동시에 믹스 밖으로 내보내지기 때문에 관찰자는 해당 메시지가 어떤 재배열 방식을 사용하는지 알 수 없다. 기존의 믹스 시스템에서 공격자는 해당 믹스 시스템이 사용하는 재배열 알고리즘의 취약점을 이용해 목표로 삼은 메시지를 공격할 수 있었다. 예를 들어 timed mix인 경우, 해당 노드로 자신의 타겟 메시지(target message)만을 보내고 나머지 메시지들은 해당 노드로 들어가지 못하게 막는 방법을 사용해서 공격할 수 있다. 하지만 하이브리드 믹스를 사용하는 메시지는 어떤 재배열 알고리즘을 사용하는지 공격자가 알기 힘들다. 그래서 공격자 측면에서 하이브리드 믹스는 기존의 풀 믹스보다 공격하기 까다롭다.

3.3 Timed mix, Timed pool mix와 비교

하이브리드 믹스가 기존의 timed mix와 timed pool mix와 비교해 익명성과 메시지 지연시간 측면에서 어떤 성질을 가지는지 알아보기 위해서 Andrei Serjantov가 사용한 방법을 이용한다[5].

메시지 지연시간이 최소가 되는 경우는 해당 메시지가 믹스에서 메시지를 내보내기 직전에 들어가는 경우가 되고 이 때 메시지의 지연시간은 ε이 된다. 반면에 메시지의 지연시간이 최대가 되는 경우는 사용자가 timed mix를 사용할 때는 한 라운드 동안의 시간이 되고, timed pool mix를 사용할 경우에는 이론적으로는 무한대의 시간이 된다. 하지만 관찰자는 메시지가 어떤 종류의 믹스를 사용하는지 알 수 없기 때문에 메시지 최대 지연시간은 무한대가 된다[5].

하이브리드 믹스 시스템의 익명성을 측정하기 위해선 우선 3가지 믹스 시스템이 동일한 메시지 저장공간을 가진 믹스 노드를 사용한다고 가정한다. 익명성 집합(anonymity set)이 최소가 되는 경우는 어떤 재배열 방식을 사용하는 메시지가 노드로 들어가는가에 나눠진다. 한 라운드 동안 timed mix방식을 사용하는 하나의 메시지가 하이브리드 믹스로 들어갔을 경우 믹스가 메시지를 보낼 시간이 되면, 메시지가 전송되지만, 관찰자는 해당 메시지가 어떤 재배열 방식을 사용했는지 알 수 없기 때문에 어느 풀에서 나온 메시지인지 알 수 없다. 하이브리드 믹스로 timed pool mix를 사용하는 메시지 하나가 들어갔을 경우에는 timed pool mix의 상태에 따라 메시지를 송신여부가 결정된다. 이때 관찰자가 고려할 메시지의 집합은 timed pool mix의 풀 안에 들어있는 메시지들이 된다.

메시지의 익명성 집합(anonymity set)이 최대가 되는 경우를 보기 위해서 믹스 안의 총 메시지 수 중에서 다음 라운드에 메시지가 남을 확률이 하이브리드 믹스 Timed pool mix와 기존 timed pool mix가 둘 다 같다고 가정한다. 믹스로 들어오는 메시지에서 다음 라운드에 저장할 메시지의 수가 증가하면 사용자의 메시지 지연시간도 길어진다. 하지만 하이브리드 믹스는 두 믹스가 공간을 나눠 쓰기 때문에 하이브리드 믹스에서 다음 라운드에 timed pool mix를 사용하는 메시지가 남아 있는 양은 기존 timed pool mix에서 남아있는 양보다 작다.

익명성은 믹스 안에 있는 총 메시지의 개수보다 다음 라운드에 남아있는 메시지 개수에 더 큰 영향을 받는다[6]. 그래서 하이브리드 믹스의 익명성은 timed pool mix보다는 작지만, 풀을 사용하지 않는 timed mix보다는 크다.

3.4 하이브리드 믹스 내 두 풀의 비율

하이브리드 믹스 내에서 timed pool mix와 timed mix가 서로 차지하는 비율에 따라 하이브리드 믹스의 성질이 바뀐다. 만약 하이브리드 믹스 내 timed pool mix를 사용하는 비율이 높아지면, 다음 라운드에 믹스 안에 남아있는 메시지의 수가 증가되면서 timed pool mix가 가지고 있는 특징에 가까워 진다. 반면에 timed mix의 크기가 커질 경우, 반대로 익명성 정도가 작아진다. 그래서 믹스 네트워크 안에서 익명성 보장을 중요시하는 사용자들이 많을 경우 하이브리드 믹스 내 timed pool mix가 차지하는 비율을 높일 수 있다. 즉

하이브리드 믹스는 이것을 사용하는 네트워크의 성격에 따라서 조정될 수 있다.

3.4 믹스 사용률이 한쪽으로 치우치는 경우

하이브리드 믹스가 사용되면서 사용자의 이용률이 한쪽으로 치우치는 경우가 발생할 수 있다. 이 경우 두 믹스를 모두 사용하지 못하기 때문에 다음과 같은 문제가 발생한다. 공격자는 노드가 어떤 재배열 방식을 사용하는 지 쉽게 추측할 수 있기 때문에 믹스에 대한 공격이 용이해진다. 이러한 문제를 해결하기 위해서 하이브리드 믹스는 더미 메시지를 사용할 수 있다. 사용률이 낮은 믹스에서 더미 메시지를 생성하여 두 믹스 사용률의 균형을 맞춘다. 이 방법은 공격자가 해당 노드가 어떤 재배열 방식을 사용하는지 추측할 수 없게 만든다.

5. 결 론

지금까지 하이브리드 믹스가 어떻게 동작하는지 알아보았다. 또한 하이브리드 믹스를 사용함으로써 나타나는 특징에 대해서 분석했고 하이브리드 믹스를 기존의 timed mix, timed pool mix와 익명성과 메시지 지연시간 보장 측면에서 비교해봤다. 하이브리드 믹스는 재배열 방식이 하나만 사용되는 믹스에 비해서 믹스 사용자들의 선택 범위를 넓혀주기 때문에 많은 메시지들이 송수신되는 인터넷에서 더욱 필요하다. 본 논문에서는 풀 믹스를 중심으로 이야기했지만, 풀 믹스와 컨티뉴어스 믹스(continuous mix)를 혼합한 하이브리드 믹스 시스템에 대한 연구도 이루어져야 한다.

참 고 문 헌

[1]David Chaum. Untraceable electronic mail, return addresses and digital pseudonyms. Communications of the A.C.M., 24(2):84-88, 1981.

[2]Claudia Diaz and Andrei Serjantov. Generalising mixes. In Privacy Enhacing Technologies, LNCS, Dresden, Germany, April 2003.

[3]D. Kesdogan, J. Egner, and R. Buschkes. Stop and-go-MIXes providing probabilistic anonymity in an open system. In Proceedings of the International Information Hiding Workshop, April 1998.

[4]Andrei Serjantov, Roger Dingledine, and Paul Syverson. From a trickle to a flood: Active attacks on several mix types. In F. Petitcolas, editor, Information Hiding Workshop, October 2002.

[5]Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In Paul Syverson and Roger Dingledine, editors, Privacy Enhancing Technologies, LNCS, San Francisco, CA, April 2002.

[6]이현숙,변진욱, 박현아, 이동훈, 임종인. 익명통신로에 관한 최근 연구 동향.한국정보보호학회지 , 1598-3978 , 제14권6호 , pp.53-61 , 2004

[7]Claudia Diaz and Bart Preneel. Taxonomy of Mixes and Dummy Traffic. In the Proceedings of INetSec04:3rd Working Conference on Privacy and Anonymity in Networked and Distributed Systems, Toulouse, France, August 2004.