(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2015년10월29일
(11) 등록번호 10-1564045
(24) 등록일자 2015년10월22일

(51) 국제특허분류(Int. Cl.)
H04L 12/26 (2006.01) *H04L 12/24* (2006.01)
(21) 출원번호 10-2014-0149230
(22) 출원일자 2014년10월30일
심사청구일자 2014년10월30일
(56) 선행기술조사문헌
JP2007228217 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
김명섭
세종특별자치시 조치원읍 도원1로 16, 111-902 (푸르지오1차)
윤성호
경상북도 포항시 북구 우창동로 157, 103동 702호(창포동, 금호어울림아파트)
(74) 대리인
심경식, 홍성욱

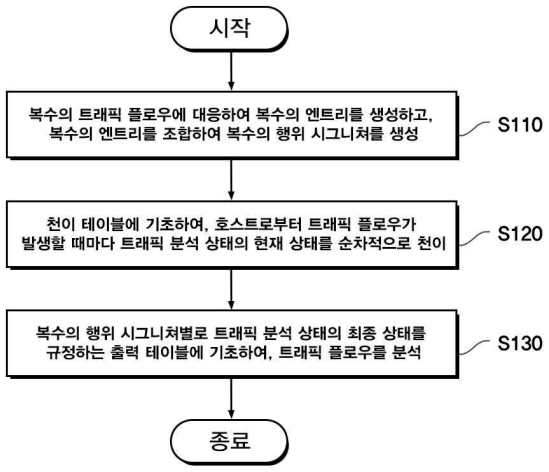
전체 청구항 수 : 총 5 항 심사관 : 전용해

(54) 발명의 명칭 **행위 시그니쳐 기반의 네트워크 트래픽 분석 방법 및 그 장치**

(57) 요 약

네트워크 트래픽 분석 방법이 개시된다. 본 발명의 일 실시예에 따른 네트워크 트래픽 분석 방법은 복수의 트래픽 플로우에 대응하여 상기 복수의 트래픽 플로우 각각을 식별하기 위한 복수의 엔트리를 생성하고, 상기 복수의 엔트리를 조합하여 복수의 응용을 식별하기 위한 복수의 행위 시그니쳐를 생성하는 단계; 상기 엔트리의 생성 순서에 따라 상기 트래픽에 대한 분석 상태의 현재 상태와 다음 상태를 규정하는 천이 테이블에 기초하여, 호스트로부터 트래픽 플로우가 발생할 때마다 상기 트래픽 분석 상태의 현재 상태를 순차적으로 천이시키는 단계; 및 상기 복수의 행위 시그니쳐별로 상기 트래픽 분석 상태의 최종 상태를 규정하는 출력 테이블에 기초하여, 상기 트래픽 플로우를 분석하는 단계를 포함한다.

대 표 도 - 도1



명 세 서

청구범위

청구항 1

복수의 트래픽 플로우에 대응하여 상기 복수의 트래픽 플로우 각각을 식별하기 위한 복수의 엔트리를 생성하고, 상기 복수의 엔트리를 조합하여 복수의 응용을 식별하기 위한 복수의 행위 시그니쳐를 생성하는 단계;

상기 엔트리의 생성 순서에 따라 상기 트래픽에 대한 분석 상태의 현재 상태와 다음 상태를 규정하는 천이 테이블에 기초하여, 호스트로부터 트래픽 플로우가 발생할 때마다 상기 트래픽 분석 상태의 현재 상태를 순차적으로 천이시키는 단계; 및

상기 복수의 행위 시그니쳐별로 상기 트래픽 분석 상태의 최종 상태를 규정하는 출력 테이블에 기초하여, 상기 트래픽 플로우를 분석하는 단계를 포함하고,

상기 트래픽 분석 상태의 현재 상태를 순차적으로 천이시키는 단계는

상기 호스트로부터 트래픽 플로우가 발생하면, 상기 발생한 트래픽 플로우에 대응되는 현재 엔트리를 생성하는 단계;

상기 천이 테이블에 기초하여, 상기 트래픽 분석 상태의 현재 상태를 상기 현재 엔트리에 대응되는 다음 상태로 천이시키는 단계;

상기 트래픽 분석 상태의 상기 천이된 현재 상태에 대한 레코드를 상태 테이블에 저장하는 단계를 포함하고,

상기 트래픽 분석 상태의 상기 천이된 현재 상태에 대한 레코드를 상태 테이블에 저장하는 단계는

상기 현재 엔트리가 상기 상태 테이블에 기 저장된 상기 트래픽 분석 상태의 현재 상태에 대한 레코드인 기존 레코드에 영향을 주는 엔트리이면서, 상기 기존 레코드와 무관한 트래픽 플로우에 대한 엔트리로도 판단되는 경우에는

상기 기존 레코드를 상기 현재 엔트리에 대응하여 상기 기존 레코드의 상기 천이된 현재 상태에 대응되는 상태값으로 갱신하여 저장하면서, 상기 현재 엔트리에 대응되는 새로운 레코드를 상기 상태 테이블에 새롭게 추가하여 저장하는 것을 특징으로 하는 네트워크 트래픽 분석 방법.

청구항 2

삭제

청구항 3

제1항에 있어서,

상기 천이 테이블에 기초하여, 상기 트래픽 분석 상태의 현재 상태를 상기 현재 엔트리에 대응되는 다음 상태로 천이시키는 단계는

상기 현재 엔트리가 최초로 생성된 엔트리인 경우에는 상기 트래픽 분석 상태의 현재 상태를 초기 상태(default state)를 기준으로 상기 현재 엔트리에 대응되는 다음 상태로 천이시키는 것을 특징으로 하는 네트워크 트래픽 분석 방법.

청구항 4

삭제

청구항 5

제1항에 있어서,

상기 출력 테이블에 기초하여 상기 트래픽 플로우를 분석하는 단계는

상기 출력 테이블에 기초하여, 상기 트래픽 분석 상태의 현재 상태가 상기 출력 테이블의 상기 복수의 행위 시그니쳐별 최종 상태 중 하나와 일치하는지 여부를 판단하는 단계;

상기 판단 결과가 일치한다고 판단되면, 상기 복수의 행위 시그니쳐 중에서 상기 트래픽 분석 상태에 대응되는 행위 시그니쳐를 결정하는 단계; 및

상기 결정된 행위 시그니쳐에 대응되는 트래픽 플로우를 식별하는 단계를 포함하는 것을 특징으로 하는 네트워크 트래픽 분석 방법.

청구항 6

제1항에 있어서,

복수의 행위 시그니쳐별 타임아웃 시간을 설정한 타임아웃 테이블에 기초하여, 상기 상태 테이블에서 상기 타임아웃 시간을 초과한 것으로 판단되는 엔트리의 현재 상태에 대한 레코드를 삭제하는 단계를 더 포함하는 것을 특징으로 하는 네트워크 트래픽 분석 방법.

청구항 7

복수의 트래픽 플로우에 대응하여 상기 복수의 트래픽 플로우 각각을 식별하기 위한 복수의 엔트리를 생성하고, 상기 복수의 엔트리를 조합하여 복수의 응용을 식별하기 위한 복수의 행위 시그니쳐를 생성하는 시그니쳐 생성부;

상기 엔트리의 생성 순서에 따라 상기 트래픽에 대한 분석 상태의 현재 상태와 다음 상태를 규정하는 천이 테이블에 기초하여, 호스트로부터 트래픽 플로우가 발생할 때마다 상기 트래픽 분석 상태의 현재 상태를 순차적으로 천이시키는 상태 관리부; 및

상기 복수의 행위 시그니쳐별로 상기 트래픽 분석 상태의 최종 상태를 규정하는 출력 테이블에 기초하여, 상기 트래픽 플로우를 분석하는 트래픽 분석부를 포함하고,

상기 상태 관리부는 상기 호스트로부터 트래픽 플로우가 발생하면, 상기 발생한 트래픽 플로우에 대응되는 현재 엔트리를 생성하고, 상기 천이 테이블에 기초하여 상기 트래픽 분석 상태의 현재 상태를 상기 현재 엔트리에 대응되는 다음 상태로 천이시킨 후, 상기 트래픽 분석 상태의 상기 천이된 현재 상태에 대한 레코드를 상태 테이블에 저장하고,

상기 트래픽 분석 상태의 상기 천이된 현재 상태에 대한 레코드를 상태 테이블에 저장할 때, 상기 현재 엔트리가 상기 상태 테이블에 기 저장된 상기 트래픽 분석 상태의 현재 상태에 대한 레코드인 기존 레코드에 영향을 주는 엔트리이면서, 상기 기존 레코드와 무관한 트래픽 플로우에 대한 엔트리로도 판단되는 경우에는 상기 기존 레코드를 상기 현재 엔트리에 대응하여 상기 기존 레코드의 상기 천이된 현재 상태에 대응되는 상태 값으로 갱신하여 저장하면서, 상기 현재 엔트리에 대응되는 새로운 레코드를 상기 상태 테이블에 새롭게 추가하여 저장하는 것을 특징으로 하는 네트워크 트래픽 분석 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 네트워크 트래픽 분석에 관한 것으로, 특히 행위 시그니쳐 기반의 네트워크 트래픽 분석 방법 및 그 장치에 관한 것이다.

배 경 기 술

[0002] 네트워크 관리의 목적은 네트워크 자원을 최대한 활용하고 비정상 공격으로부터 네트워크 장비를 보호하는 것이다. 이를 위해 네트워크 관리자들은 적절한 네트워크 정책을 수립하여 적용한다.

[0003] 네트워크 정책은 특정 트래픽을 차단하거나 대역폭을 조절하는 방법으로 수행되기 때문에 트래픽의 발생 원천을 알아내는 트래픽 분석 결과를 기반한다. 트래픽 분석은 응용을 대표하는 고유한 특성을 기반으로 트래픽을 발생시킨 응용을 명시하는 것으로써, 분석 결과는 네트워크 정책뿐만 아니라, capacity planning, network provisioning, traffic engineering, fault diagnosis 등과 같은 다양한 분야에서 활용된다.

[0004] 다양한 분석 방법론이 제안되었지만, 점점 복잡해지는 응용의 트래픽 발생 형태로 인해 실제 네트워크 트래픽에 적용하기에는 많은 한계점을 가진다. 포트 기반 분석은 Internet Assigned Number Authority (IANA)가 지정한 잘 알려진 포트 번호를 사용하여 트래픽을 분석한다. 하지만, 지정된 포트 번호 사용을 위반하는 응용들의 출현으로 분석 결과의 정확도를 신뢰하기 어렵다.

[0005] 이러한 한계점을 극복하기 위해 분석 대상 응용을 대표하는 고유한 부분 문자열을 사용하여 트래픽을 분석하는 페이로드 기반 분석 방법이 제안되었다. 실제 발생된 패킷의 페이로드를 검사하기 때문에 높은 분석률과 정확도를 보장하지만, 암호화된 페이로드를 포함하는 트래픽과 터널링 기법을 사용한 트래픽을 분석하지 못하고 사생활 침해 문제를 야기시킨다.

[0006] 최근에는 페이로드 기반 분석 방법의 한계점을 보완하기 위해 통계 기반 분석이 제안되고 있다. 통계 기반 분석은 응용 트래픽이 가지는 고유한 통계 특징을 머신러닝 알고리즘에 적용하여 트래픽을 분석한다. 하지만, 유사한 통계적 특징을 가지는 응용들 간의 구분이 어렵고, 세밀한 기능별 트래픽 분석이 불가능하다는 단점이 존재한다.

[0007] 관련 선행기술로는 등록특허공보 제10-0629756호(발명의 명칭: 네트워크 트래픽 분석 방법 및 시스템, 등록일자: 2006년 9월 22일)가 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명의 목적은 네트워크 트래픽 분석 시간을 단축시키면서도 네트워크 트래픽 분석의 정확성 및 신뢰성을 향상시킬 수 있는 행위 시그니쳐 기반의 네트워크 트래픽 분석 방법 및 그 장치를 제공하는 것이다.

과제의 해결 수단

[0009] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 네트워크 트래픽 분석 방법은 복수의 트래픽 플로우에 대응하여 상기 복수의 트래픽 플로우 각각을 식별하기 위한 복수의 엔트리를 생성하고, 상기 복수의 엔트리를 조합하여 복수의 응용을 식별하기 위한 복수의 행위 시그니쳐를 생성하는 단계; 상기 엔트리의 생성 순서에 따라 상기 트래픽에 대한 분석 상태의 현재 상태와 다음 상태를 규정하는 천이 테이블에 기초하여, 호스트로부터 트래픽 플로우가 발생할 때마다 상기 트래픽 분석 상태의 현재 상태를 순차적으로 천이시키는 단계; 및 상기 복수의 행위 시그니쳐별로 상기 트래픽 분석 상태의 최종 상태를 규정하는 출력 테이블에 기초하여, 상기 트래픽 플로우를 분석하는 단계를 포함한다.

[0010] 바람직하게는, 상기 트래픽 분석 상태의 현재 상태를 순차적으로 천이시키는 단계는 상기 호스트로부터 트래픽 플로우가 발생하면, 상기 발생한 트래픽 플로우에 대응되는 현재 엔트리를 생성하는 단계; 상기 천이 테이블에 기초하여, 상기 트래픽 분석 상태의 현재 상태를 상기 현재 엔트리에 대응되는 다음 상태로 천이시키는 단계; 상기 트래픽 분석 상태의 상기 천이된 현재 상태에 대한 레코드를 상태 테이블에 저장하는 단계를 포함할 수 있다.

[0011] 바람직하게는, 상기 천이 테이블에 기초하여, 상기 트래픽 분석 상태의 현재 상태를 상기 현재 엔트리에 대응되는 다음 상태로 천이시키는 단계는 상기 현재 엔트리가 최초로 생성된 엔트리인 경우에는 상기 트래픽 분석 상태의 현재 상태를 초기 상태(default state)를 기준으로 상기 현재 엔트리에 대응되는 다음 상태로 천이시킬 수 있다.

[0012] 바람직하게는, 상기 트래픽 분석 상태의 현재 상태를 상기 현재 엔트리에 대응되는 다음 상태로 천이시키는 단계는 상기 상태 테이블에 기 저장된 상기 트래픽 분석 상태의 현재 상태에 대한 레코드인 기존 레코드가 존재하는 경우에는 상기 기존 레코드의 현재 상태를 상기 현재 상태를 기준으로 상기 현재 엔트리에 대응되는 다음 상태로 천이시키는 단계를 포함하고, 상기 트래픽 분석 상태의 상기 천이된 현재 상태에 대한 레코드를 상태 테이블에 저장하는 단계는 상기 기존 레코드를 상기 기존 레코드의 상기 천이된 현재 상태에 대응되는 상태 값으로 갱신하여 저장할 수 있다.

[0013] 바람직하게는, 상기 트래픽 분석 상태의 현재 상태를 상기 현재 엔트리에 대응되는 다음 상태로 천이시키는 단계는 상기 현재 엔트리가 생성되면 상기 트래픽 분석 상태의 현재 상태를 초기 상태를 기준으로 상기 현재 엔트리에 대응되는 다음 상태로 천이시키는 단계를 더 포함하고, 상기 트래픽 분석 상태의 상기 천이된 현재 상태에 대한 레코드를 상태 테이블에 저장하는 단계는 상기 초기 상태를 기준으로 천이된 상기 트래픽 분석 상태의 현

재 상태에 대한 레코드를 상기 갱신하여 저장된 상기 기존 레코드 하단에 저장하는 단계를 더 포함할 수 있다.

[0014] 바람직하게는, 상기 출력 테이블에 기초하여 상기 트래픽 플로우를 분석하는 단계는 상기 출력 테이블에 기초하여, 상기 트래픽 분석 상태의 현재 상태가 상기 출력 테이블의 상기 복수의 행위 시그니쳐별 최종 상태 중 하나와 일치하는지 여부를 판단하는 단계; 상기 판단 결과가 일치한다고 판단되면, 상기 복수의 행위 시그니쳐 중에서 상기 트래픽 분석 상태에 대응되는 행위 시그니쳐를 결정하는 단계; 및 상기 결정된 행위 시그니쳐에 대응되는 트래픽 플로우를 식별하는 단계를 포함할 수 있다.

[0015] 바람직하게는, 본 발명의 일 실시예에 따른 네트워크 트래픽 분석 방법은 복수의 행위 시그니쳐별 타임아웃 시간을 설정한 타임아웃 테이블에 기초하여, 상기 상태 테이블에서 상기 타임아웃 시간을 초과한 것으로 판단되는 엔트리의 현재 상태에 대한 레코드를 삭제하는 단계를 더 포함할 수 있다.

[0016] 또한, 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 네트워크 트래픽 분석 장치는 복수의 트래픽 플로우에 대응하여 상기 복수의 트래픽 플로우 각각을 식별하기 위한 복수의 엔트리를 생성하고, 상기 복수의 엔트리를 조합하여 복수의 응용을 식별하기 위한 복수의 행위 시그니쳐를 생성하는 시그니쳐 생성부; 상기 엔트리의 생성 순서에 따라 상기 트래픽에 대한 분석 상태의 현재 상태와 다음 상태를 규정하는 천이 테이블에 기초하여, 호스트로부터 트래픽 플로우가 발생할 때마다 상기 트래픽 분석 상태의 현재 상태를 순차적으로 천이시키는 상태 관리부; 및 상기 복수의 행위 시그니쳐별로 상기 트래픽 분석 상태의 최종 상태를 규정하는 출력 테이블에 기초하여, 상기 트래픽 플로우를 분석하는 트래픽 분석부를 포함한다.

발명의 효과

[0017] 본 발명의 일 실시예에 따르면 행위 시그니쳐 기반으로 네트워크 트래픽을 분석함으로써, 네트워크 트래픽 분석의 정확성 및 신뢰성을 향상시키면서도 분석 시간을 단축할 수 있다.

도면의 간단한 설명

[0018] 도 1은 본 발명의 일 실시예에 따른 행위 시그니쳐 기반의 네트워크 트래픽 분석 방법을 설명하기 위하여 도시한 흐름도이다.

도 2는 본 발명의 일 실시예에 따라 트래픽 분석 상태의 현재 상태를 순차적으로 천이시키는 방법을 설명하기 위하여 도시한 흐름도이다.

도 3은 본 발명의 일 실시예에 따라 출력 테이블에 기초하여 트래픽 플로우를 분석하는 방법을 설명하기 위하여 도시한 흐름도이다.

도 4는 본 발명의 일 실시예에 따른 행위 시그니쳐 기반의 네트워크 트래픽 분석 장치를 설명하기 위하여 도시한 도면이다.

도 5는 본 발명의 일 실시예에 따른 행위 시그니쳐 기반의 네트워크 트래픽 분석 장치를 이용하여 네트워크 트래픽을 분석하는 응용예를 설명하기 위하여 도시한 도면이다.

도 6은 본 발명의 일 실시예에 따라 트랜잭션을 생성할 때 사용되는 스트링 트리의 일례를 도시한 도면이다.

도 7은 본 발명의 일 실시예에 따른 천이 테이블의 예시를 다이어그램으로 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0019] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.

[0020] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0021] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에

직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[0022] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0023] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0024] 이하, 본 발명에 따른 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0025] 도 1은 본 발명의 일 실시예에 따른 행위 시그니쳐 기반의 네트워크 트래픽 분석 방법을 설명하기 위하여 도시한 흐름도이다.

[0026] 단계 110에서는, 네트워크 트래픽 분석 장치가 복수의 트래픽 플로우에 대응하여 복수의 트래픽 플로우 각각을 식별하기 위한 복수의 엔트리를 생성하고, 복수의 엔트리를 조합하여 복수의 응용을 식별하기 위한 복수의 행위 시그니쳐를 생성한다.

[0027] 이때, 복수의 엔트리 각각은 복수의 트래픽 각각에서의 첫번째 질의 패킷(First Request Packet:FRP)에 기초하여 생성되며, 복수의 호스트가 특정한 응용을 사용할 때 나타나는 공통적인 특징을 이용하여 생성될 수 있다.

[0028] 한편, 행위 시그니쳐는 복수의 엔트리를 조합하여 생성되는데, 엔트리의 조합이 동일한 엔트리들로 구성된 경우라도 엔트리 조합 내의 엔트리 생성 순서가 상이하면 상이한 행위 시그니쳐가 된다.

[0029] 엔트리 및 행위 시그니쳐의 생성 방법에 대한 구체적인 설명은 "1. 행위 시그니쳐 생성과 2. 엔트리 생성"에서 후술한다.

[0030] 단계 120에서는, 네트워크 트래픽 분석 장치가 엔트리의 생성 순서에 따라 트래픽에 대한 분석 상태의 현재 상태와 다음 상태를 규정하는 천이 테이블에 기초하여, 호스트로부터 트래픽 플로우가 발생할 때마다 트래픽 분석 상태의 현재 상태를 순차적으로 천이시킨다.

[0031] 이때, 네트워크 트래픽 분석 장치는 호스트에서 발생하는 트래픽 플로우를 수집하는 지점에 위치하는 것으로 가정한다.

[0032] 또한, 네트워크 트래픽 분석 장치가 천이 테이블을 이용하여 트래픽에 대한 분석 상태의 현재 상태를 순차적으로 천이시키는 것은 아호-코라식 알고리즘을 트래픽 분석에 이용하기 위한 것이다.

[0033] 천이 테이블과 현재 상태의 천이 과정에 대한 구체적인 동작에 대해서는 도 2에서 구체적으로 설명한다.

[0034] 단계 130에서는, 네트워크 트래픽 분석 장치가 복수의 행위 시그니쳐별로 트래픽 분석 상태의 최종 상태를 규정하는 출력 테이블에 기초하여, 트래픽 플로우를 분석한다.

[0035] 예컨대, 네트워크 트래픽 분석 장치는 출력 테이블에 기초하여 호스트에서 발생하는 트래픽 플로우가 어떤 응용(application)에 관한 것인지를 분석할 수 있는데, 이에 대해서는 도 3에서 구체적으로 설명한다.

[0036] 이와 같이 본 발명은 행위 시그니쳐에 기초하여 트래픽 플로우를 분석하는데 행위 시그니쳐는 트래픽 플로우의 첫 번째 패킷만을 이용하여 생성된 엔트리의 조합이므로, 트래픽 플로우 전체를 수신한 후에야 트래픽 플로우에 대한 분석이 가능했던 종래 기술에 비하여 훨씬 빠른 시간 내에 트래픽에 대한 분석이 가능하게 된다.

[0037] 또한, 본 발명의 일 실시예는 복수의 호스트가 특정한 응용을 사용할 때 나타나는 공통적인 특징을 이용하여 생성된 엔트리의 조합인 행위 시그니쳐를 이용하여 네트워크 트래픽을 분석하므로, 종래에 비해 트래픽 분석의 정확성과 신뢰도가 향상될 수 있다.

[0038] 도 2는 본 발명의 일 실시예에 따라 트래픽 분석 상태의 현재 상태를 순차적으로 천이시키는 방법을 설명하기

위하여 도시한 흐름도이다.

[0039] 단계 210에서는, 호스트로부터 트래픽 플로우가 발생하면, 네트워크 트래픽 분석 장치가 그 발생한 트래픽 플로우에 대응되는 현재 엔트리를 생성한다.

[0040] 보다 구체적으로는, 네트워크 트래픽 분석 장치는 현재 발생한 트래픽 플로우를 식별하기 위해 현재 발생한 트래픽 플로우의 첫 번째 질의 패킷에 기초하여 현재 엔트리를 생성하게 된다.

[0041] 단계 220에서는, 네트워크 트래픽 분석 장치가 천이 테이블에 기초하여 트래픽 분석 상태의 현재 상태를 현재 엔트리에 대응되는 다음 상태로 천이시킨다.

[0042] 단계 230에서는, 네트워크 트래픽 분석 장치가 트래픽 분석 상태의 천이된 현재 상태에 대한 레코드를 상태 테이블에 저장한다.

[0043] 이때, 레코드를 상태 테이블에 저장하는 방식은 새로운 레코드를 상태 테이블에 새롭게 추가하거나, 기존 레코드를 갱신하는 방법이 있다.

[0044] 먼저, 새로운 레코드를 상태 테이블에 새롭게 추가하는 방법에 대해 설명하면 다음과 같다.

[0045] 현재 엔트리가 최초로 생성된 엔트리인 경우라면 트래픽 분석 상태의 현재 상태를 초기 상태(default state)를 기준으로 현재 엔트리에 대응되는 다음 상태로 천이시킨 상태 값을 상태 테이블에 저장할 수 있다.

[0046] 다음으로, 기존 레코드를 갱신하는 방법에 대해 설명하면 다음과 같다.

[0047] 상태 테이블에 기존 레코드가 존재하는 경우에는 기존 레코드의 현재 상태를 상태 테이블에 저장된 현재 상태를 기준으로 현재 엔트리에 대응되는 다음 상태로 천이시킨 현재 상태에 대응되는 상태 값으로 갱신하여 저장할 수 있다.

[0048] 마지막으로, 현재 엔트리가 기존 레코드에도 영향을 주는 엔트리이고, 기존 레코드와 무관한 트래픽 플로우에 대한 엔트리로 볼 수도 있는 경우라면, 현재 엔트리에 대응하여 기존 레코드를 갱신하면서, 새로운 레코드도 상태 테이블에 새롭게 추가할 수 있다. 이때, 기존 레코드와 무관한 트래픽 플로우에 대한 엔트리로 볼 수 있는 경우는 최초로 생성된 엔트리인 경우와 동일하게 취급하여 새로운 레코드를 상태 테이블에 새롭게 추가하게 된다.

[0049] 한편, 다른 실시예에서는 상태 테이블에 저장된 레코드가 소정 시간을 경과한 경우에는 네트워크 트래픽 분석 장치가 삭제할 수 있다 .

[0050] 예컨대, 네트워크 트래픽 분석 장치는 복수의 행위 시그니처별 타임아웃 시간을 설정한 타임아웃 테이블에 기초하여, 상태 테이블에서 타임아웃 시간을 초과한 것으로 판단되는 엔트리의 현재 상태에 대한 레코드를 삭제할 수 있다.

[0051] 도 3은 본 발명의 일 실시예에 따라 출력 테이블에 기초하여 트래픽 플로우를 분석하는 방법을 설명하기 위하여 도시한 흐름도이다.

[0052] 단계 310에서는, 네트워크 트래픽 분석 장치가 출력 테이블에 기초하여 트래픽 분석 상태의 현재 상태가 출력 테이블의 복수의 행위 시그니처별 최종 상태 중 하나와 일치하는지 여부를 판단한다.

[0053] 단계 320에서는, 판단 결과가 일치한다고 판단되면, 네트워크 트래픽 분석 장치가 복수의 행위 시그니처별 최종 상태 중 하나와 일치하는 것으로 판단된 트래픽 분석 상태의 현재 상태에 대응되는 행위 시그니처를 결정한다.

[0054] 단계 330에서는, 네트워크 트래픽 분석 장치가 그 결정된 행위 시그니처에 대응되는 트래픽 플로우를 식별한다.

[0055] 도 4는 본 발명의 일 실시예에 따른 행위 시그니처 기반의 네트워크 트래픽 분석 장치를 설명하기 위하여 도시한 도면이다.

[0056] 도 4를 참조하면, 본 발명의 일 실시예에 따른 네트워크 트래픽 분석 장치는 시그니처 생성부(410), 상태 관리부(420) 및 트래픽 분석부(430)를 포함한다.

[0057] 시그니처 생성부(410)는 복수의 트래픽 플로우에 대응하여 복수의 트래픽 플로우 각각을 식별하기 위한 복수의 엔트리를 생성하고, 복수의 엔트리를 조합하여 복수의 응용을 식별하기 위한 복수의 행위 시그니처를 생성한다.

[0058] 상태 관리부(420)는 엔트리의 생성 순서에 따라 트래픽에 대한 분석 상태의 현재 상태와 다음 상태를 규정하는

천이 테이블에 기초하여, 호스트로부터 트래픽 플로우가 발생할 때마다 트래픽 분석 상태의 현재 상태를 순차적으로 천이시킨다.

[0059] 트래픽 분석부(430)는 복수의 행위 시그니쳐별로 트래픽 분석 상태의 최종 상태를 규정하는 출력 테이블에 기초하여, 트래픽 플로우를 분석한다.

[0060] 도 5는 본 발명의 일 실시예에 따른 행위 시그니쳐 기반의 네트워크 트래픽 분석 장치를 이용하여 네트워크 트래픽을 분석하는 응용예를 설명하기 위하여 도시한 도면이다.

[0061] 도 5를 참조하면, 타겟 네트워크(510)에서 인터넷(520)으로 향하는 질의 패킷들과 인터넷(520)에서 타겟 네트워크(510)로 향하는 응답 패킷들이 도시되어 있다.

[0062] 플로우 생성기(530)는 타겟 네트워크(510)에서 발생하는 트래픽을 패킷 단위로 수집하며, 그 수집한 패킷들을 이용하여 5-튜플(출발지 IP, 목적지 IP, 출발지 포트, 목적지 포트, 4계층 프로토콜)이 동일한 패킷 집합인 트래픽 플로우를 생성한다.

[0063] 플로우 생성기(530)는 그 생성한 트래픽 플로우 각각의 첫 번째 질의 패킷(FRP)을 식별기(540)에게 전송한다.

[0064] 식별기(540)는 트래픽 플로우들에 대한 첫 번째 질의 패킷들을 본 발명의 일 실시예에 따른 행위 시그니쳐 기반의 네트워크 트래픽 분석 방법에 따라 분석하여 자신에게 입력된 트래픽 플로우들에 해당하는 응용을 식별하게 되고, 분석이 완료된 트래픽 플로우들은 식별된 트래픽이 된다.

[0065] 본 발명의 일 실시예에 따른 행위 시그니쳐에 기초한 네트워크 트래픽 분석 장치는 식별기(540)일 수도 있고, 플로우 생성기(530)와 식별기(540)를 모두 포함할 수도 있다.

[0066] 한편, 도 5에서 플로우 생성기(530)는 트래픽 플로우 각각의 첫 번째 질의 패킷(FRP)을 식별기(540)에게 전송한 후에, 플로우 생성기(530)에 수집된 모든 트래픽 플로우를 식별기(540)에게 전송하는데, 이는 식별기(540)로 하여금 트래픽 플로우의 양을 측정하게 하기 위한 것이다.

[0067] 지금까지는 도 1 내지 도 5를 참조하여, 본 발명의 개략적인 구성에 대하여 설명하였으나, 이하에서는 본 발명의 실시예들에 대해 표 및 수학식 등을 참조하여 보다 구체적으로 설명하고, 그에 따른 실험 및 결과에 대해서도 살펴 보기로 한다.

[0068] **1. 행위 시그니쳐 생성**

[0069] 본 실시예에서는 행위 시그니쳐를 생성하는 방법에 대해 설명한다. 표 1은 3명의 호스트가 특정 인터넷 응용을 사용할 때 발생시킨 트래픽을 나타낸다. 본 추출 알고리즘의 목적은 표 1과 같이 여러 호스트가 발생한 특정 응용의 트래픽으로부터 일정 지지도를 가지는 순차 패턴을 행위 시그니쳐로 추출하는 것이다. 표 2는 표 1의 트래픽에서 최소 지지도 60%를 만족하는(최소 2명의 호스트에서 발생) 행위 시그니쳐를 추출한 결과이다.

표 1

flow ID	srcIP	srcPort	L4Prot	dstIP	dstPort	capture_time	payload
F_1	xxx.xxx.xxx.1	51912	6	173.194.126.130	80	12:30' 11"	pear
F_2	xxx.xxx.xxx.3	21511	6	173.194.126.133	80	12:30' 12"	pea
F_3	xxx.xxx.xxx.2	35691	6	173.194.126.131	80	12:30' 13"	peach
F_4	xxx.xxx.xxx.1	51913	6	173.194.38.71	80	12:30' 14"	lettuce
F_5	xxx.xxx.xxx.3	21513	6	74.125.23.120	80	12:30' 17"	kiwijuice
F_6	xxx.xxx.xxx.2	35692	6	74.125.23.120	80	12:30' 18"	kiwi
F_7	xxx.xxx.xxx.2	35693	17	173.194.38.76	80	12:30' 21"	kiwi
F_8	xxx.xxx.xxx.3	21519	6	173.194.72.95	80	12:30' 22"	leek
F_9	xxx.xxx.xxx.3	21520	6	173.194.126.137	80	12:30' 25"	peachpie
F_{10}	xxx.xxx.xxx.2	35694	6	173.194.72.95	80	12:30' 26"	lemon

[0070]

[0071] 표 1에서 트래픽 플로우 집합 F는 트래픽 플로우 ID(flow ID)로 구분되고, 각각의 트래픽 플로우 ID F_i는 출발지 주소(srcIP), 출발지 포트번호(srcPort), 4계층 프로토콜(L4Prot), 목적지 주소(dstIP), 목적지 포트번호(dstPort), 수집시간(capture_time), 페이로드(payload)를 구성 요소로 가진다.

표 2

행위 시그니쳐 (support > 60%)
$\langle 13000, \{173.194.126.0/24,80,6,pea\}, \{74.125.23.120/32,80,6,kiwi\}, \{173.194.72.95/32,80,6,le\} \rangle$

[0072]

[0073] 표 2의 시그니쳐는 3개의 엔트리와 특정 시간 간격으로 구성된다. 즉, 13,000ms 이내에 제시된 3개의 엔트리가 특정 호스트에서 순차적으로 발생되면 해당 트래픽들을 행위 시그니쳐에 대응되는 것으로 행위 분석한다.

[0074] 행위 시그니쳐는 수학식 1, 2, 3과 같이 정의한다. 행위 시그니쳐 집합 BS는 행위 시그니쳐 BS_i로 구성되며, 각각의 BS_i는 시간 간격(I)와 단일 호스트에서 발생한 트래픽들의 첫 질의 패킷에서 추출한 엔트리(E_i)로 구성된다. 엔트리 E_i는 {ip 주소, 포트, 프로토콜, 페이로드}로 구성된 집합의 영집합(ϕ)을 제외한 멱집합(power set)의 집합 원소로 구성된다.

수학식 1

[0075] $$BS = \{BS_1, BS_2, \ \ldots, BS_b\}$$

수학식 2

[0076] $$BS_i = \left\{ \begin{matrix} I,\ E_1, E_2,\ \ldots, E_e| \\ e \geq 2, Src(E_1) = Src(E_2) = \cdots = Src(E_e) \end{matrix} \right\}$$

수학식 3

[0077] $$E_i = \{2^C | C = \{ip,\ port,\ prot,\ payload\},\ E_i \neq \emptyset\}$$

[0078] **2. 엔트리 생성**

[0079] 네트워크 트래픽 분석 장치는 트래픽 플로우를 입력 받으면 트랜잭션을 생성한다. 입력인 트래픽 플로우는 수학식 4, 5와 같이 정의되고, 출력인 트랜잭션은 수학식 6, 7과 같이 정의된다. 트래픽 플로우 집합 F는 트래픽 플로우 F_i로 구성되며, 각각의 F_i는 수학식 5와 같이 출발지 주소(srcIP), 출발지 포트번호(srcPort), 4계층 프로토콜(L4Prot), 목적지 주소(dstIP), 목적지 포트번호(dstPort), 수집시간(capture_time), 페이로드(payload)를 구성 요소로 가진다. 본 명세서에서 제안하는 행위 시그니쳐는 질의(request) 트래픽만을 대상으로 하기 때문에 내부 호스트에서 외부 서버로 발생하는 트래픽 플로우만을 고려한다. 트랜잭션 집합 T는 트랜잭션 T_i로 구성되며, T_i는 호스트 아이디(host_id), 수집 시간(capture_time), 엔트리(E)로 구성된다.

수학식 4

[0080] $$F = \{F_1, F_2, \ \ldots, F_f\}$$

수학식 5

[0081] $$F_i = \left\{ \begin{matrix} srcIP,\ srcPort,\ L4Prot,\ dstIP,\ dstPort, \\ capture_time,\ payload \mid srcIP = local \end{matrix} \right\}$$

수학식 6

[0082] $$T = \{T_1, T_2, \ \ldots, T_f\}$$

수학식 7

[0083] $$T_i = \left\{ \begin{matrix} host_id,\ capture_time,\ E| \\ E = \{2^C | C = \{ip,\ port,\ prot,\ payload\},\ E \neq \emptyset\} \end{matrix} \right\}$$

[0084] 이하에서는 트래픽 플로우로부터 트랜잭션을 생성하는 엔트리 추출 알고리즘에 대해 설명한다. 트래픽 플로우의 srcIP에 고유한 ID를 부여하여 트랜잭션의 host_id에 저장하고 트래픽 플로우의 capture_time, dstIP, dstPort, L4Prot, payload는 각각 트랜잭션의 capture_time, IP, Port, Prot, payload에 저장된다. dstIP의 경우, 같은 서버군(Server Farm)에서 발생하는 트래픽들을 고유한 엔트리로 추출하기 위해 첫 24bit (C class)만을 추출하여 저장한다. payload의 경우, 트래픽 플로우의 payload에서 첫 N바이트만을 추출하여 저장한다. 이때, 사용하는 N은 도 6과 같은 스트링 트리(String tree)를 사용하여 결정한다. 입력된 모든 트래픽 플로우의 payload들을 하나의 스트링 트리로 구성한다. 스트링 트리의 각 노드는 입력된 payload에 저장된 1byte 문자를 의미한다. 스트링 트리를 구성하고 수학식 8을 사용하여 스트링 트리의 각 노드마다 지지도(support)를 계산한다. 지지도는 트래픽을 발생시킨 전체 호스트 중에 루트 노드부터 해당 노드까지의 문자열을 가지는 트래픽을 발생시킨 호스트의 비율을 의미한다. 예를 들어, 도 6에서 깊이 4의 'c' 노드는 전체 호스트 3명 중에 해당 문자열 "peac"를 가지는 호스트가 2명(표 1의 3행, 9행)이므로 0.67의 지지도를 가진다. 입력된 트래픽 플로우의 payload에서 지지도가 감소하지 않는 최대 길이의 문자열만을 사용함으로써 가변적인 정보(사용자 정보 및 시간 정보)을 포함하는 트래픽으로부터 행위 시그니쳐 추출을 가능하게 한다. 즉, 표 1의 1행, 2행, 3행의 payload인 "pear", "pea", "peach"에서 길이가 3인 "pea"만 추출하여 해당 트래픽 플로우가 동일한 엔트리로 추출되게 한다. 추출 알고리즘 마지막 단계인 패턴 최적화 모듈[도 1의 패턴 최적화부(150)에 대응]에서 간략화된 dstIP, payload를 구체화하는 과정을 수행한다. 표 3은 표 1에 제시된 트래픽을 본 엔트리 추출 모듈에 적용하였을 때 출력되는 결과이다.

수학식 8

[0085] $$Support = \frac{Number\ of\ support\ hosts}{Total\ number\ of\ hosts}$$

표 3

host_id	ip	port	prot	capture_time	payload
1	173.194.126.0	80	6	12:30' 11"	pea
3	173.194.126.0	80	6	12:30' 12"	pea
2	173.194.126.0	80	6	12:30' 13"	pea
1	173.194.38.0	80	6	12:30' 14"	le
3	74.125.23.0	80	6	12:30' 17"	kiwi
2	74.125.23.0	80	6	12:30' 18"	kiwi
2	173.194.38.0	80	17	12:30' 21"	kiwi
3	173.194.72.0	80	6	12:30' 22"	le
3	173.194.126.0	80	6	12:30' 25"	pea
2	173.194.72.0	80	6	12:30' 26"	le

[0086]

[0087] **3. 트래픽 분석**

[0088] 본 발명의 일 실시예에 따른 트래픽 분석 알고리즘은 표 4와 같이 정의될 수 있다.

표 4

$$\textbf{Input}: F = \{F_1, F_2, \ldots, F_f\}, BS = \{BS_1, BS_2, \ldots, BS_b\}$$
$$\textbf{Output}: F' = \{F_1, F_2, \ldots, F_f\}$$
$$\textbf{identifyTraffic}(F, BS)$$

1: $T_{transition}, T_{output}, T_{timeout} \leftarrow constructTable(BS)$

2: **while** true **do**

3: $FRP \leftarrow \textbf{wait}(getFirstRequestPacket(\square))$

4: $T_{status} \leftarrow insertStatus(FPR, T_{status}, T_{transition})$

5: $T_{status} \leftarrow updateStatus(FRP, T_{status}, T_{transition})$

6: $T_{status} \leftarrow timeoutStatus(FRP, T_{status}, T_{timeout})$

7: $F' \leftarrow outputStatus(FRP, T_{status}, T_{output})$

8: **return** F'

*Notation: F : the set of flows, F' : the set of identified flows,
BS : the set of behavior signatures, FRP : first request packet,
$T_{transition}$: transition table, T_{output} : output table,
$T_{timeout}$: timeout table, T_{status} : status table

[0089]

[0090] 트래픽 분석 알고리즘은 1단계의 준비 단계와 4단계의 분석 단계로 구성된다. 알고리즘이 시작하면 준비 단계로써 행위 시그니쳐 BS를 입력으로 테이블을 생성하는 테이블 생성 함수 constructTable(BS)을 이용(line:1)하여 천이 테이블(T_transition), 출력 테이블(T_output), 타임아웃 테이블((T_timeout)을 생성한다. 생성된 테이블들은 트래픽 분석 단계에서 상태 테이블(T_status)을 관리하기 위하여 사용된다.

[0091] 분석 단계는 총 4단계로 구분할 수 있으며 FG로부터 첫 질의 패킷(FRP)이 입력될 때마다 반복적으로 수행(line:2~7)된다. FRP이 입력(line:3)되면 상태 테이블에 트래픽 분석 상태의 현재 상태에 대한 레코드를 추가하기 위하여 insertStatus()함수를 수행(line:4)하고, updateStatus()함수를 통해 상태 테이블에 기존에 저장된 트래픽 분석 상태의 현재 상태에 대한 레코드인 기존 레코드를 갱신한다. 이때, 기존 레코드의 현재 상태를 기준으로 현재 엔트리에 대응되는 다음 상태로 천이시킨 값으로 갱신(line: 5)하게 된다.

[0092] 또한, timeoutStatus()를 통하여 타임아웃 테이블(T_timeout)을 참조하여 상태 테이블(T_status)에서 타임아웃 시간을 초과한 것으로 판단되는 레코드를 삭제(line: 6)한다.

[0093] 마지막으로, outputStatusTable()을 통하여 출력 테이블(T_output)을 참조하여 상태 테이블(T_status)에 저장된 트래픽 분석 상태에 대한 레코드의 현재 상태가 최종 상태(output_state)에 도달하였다고 판단되면 그 최종 상태에 도달한 것으로 판단된 레코드에 대응되는 트래픽 플로우의 목록(history), 즉 해당 행위 시그니쳐에 대

응되는 트래픽 목록을 출력(line:7)한다.

[0094] 한편, 식별기(540)가 플로우 생성기(530)로부터 플로우 생성기(530)가 수집한 모든 트래픽 플로우를 추가적으로 입력 받아 분석된 트래픽의 수량을 측정하는 과정은 비교적 간단한 작업이기 때문에 본 알고리즘 설명에서는 생략한다.

[0095] **3.1 준비 단계**

[0096] 본 실시예는 실제 트래픽 분석의 준비 단계로써 행위 시그니쳐를 통해 천이 테이블(T_transition), 출력 테이블(T_output), 타임아웃 테이블(T_timeout)을 생성하는 과정을 설명한다.

[0097] 본 발명의 일실시예에 따른 트래픽 분석 방법을 설명하기 위해 표 5와 같은 행위 시그니쳐를 사용하여 트래픽을 분석한다고 가정한다. 표 5에는 총 3개의 행위 시그니쳐가 정의되어 있으며 표 5의 오른쪽 열에는 각각의 행위 시그니쳐에 포함되는 첫 엔트리부터 마지막 엔트리까지 생성되는데 걸리는 시간 간격(interval)이 표시되어 있고, 각각의 행위 시그니쳐에 대응되는 엔트리(E1, E2, E3, E4, E5)의 생성 순서가 표시되어 있으며, 가운데 열에는 시간 간격과 엔트리의 세부 구성이 표시되어 있다.

[0098] 표 5의 오른쪽 열에는 BS1이 시간 간격 3000ms 이내에 생성되는 엔트리 조합 E1, E2, E5인 것으로 정의되어 있으며, 가운데 열에는 시간 간격 3000ms, 엔트리 E1, E2, E5 각각에 대한 구성이 {ip 주소, 포트, 프로토콜, 페이로드}로 표시되어 있다.

[0099] 이때, 행위 시그니쳐 각각은 특정한 응용을 식별할 수 있도록 구성된다. 예컨대, BS1은 유튜브를 이용할 때 발생하는 트래픽 플로우들을 나타내는 행위 시그니쳐일 수 있고, BS2는 페이스북을 이용할때 발생하는 트래픽 플로우들을 나타내는 행위 시그니쳐일 수 있고, BS3는 인스타그램을 이용할때 발생하는 트래픽 플로우들을 나타내는 행위 시그니쳐일 수 있다. 즉, 엔트리들이 어떤 행위 시그니쳐에 대한 것인지가 분석되면, 그 엔트리들에 대응되는 트래픽 플로우들이 어떤 응용을 사용할때 발생하는 트래픽 플로우들인지를 알 수 있게 되는 것이다.

표 5

[0100]

ID	BS(original)	BS(symbol)
BS_1	$\langle 3000, \{74.125.23.120/32,80,6,\text{^}kiwi.*\}, \{173.194.72.95/32,80,6,\text{^}le.*\}, \{173.194.38.0/24,80,17,\text{^}ki.*\} \rangle$	$\langle 3000, E_1, E_2, E_5 \rangle$
BS_2	$\langle 13000, \{173.194.126.0/24,80,6,\text{^}pea.*\}, \{74.125.23.120/32,80,6,\text{^}kiwi.*\}, \{174.194.72.95/32,80,6,\text{^}le.*\} \rangle$	$\langle 13000, E_3, E_1, E_2 \rangle$
BS_3	$\langle 3000, \{74.125.23.120/32,80,6,\text{^}kiwi.*\}, \{173.194.38.0/24,80,6,\text{^}let.*\}, \{173.194.1260/24,80,6,\text{^}pea.*\} \rangle$	$\langle 6000, E_1, E_4, E_3 \rangle$

[0101] 표 6은 표 5의 행위 시그니쳐를 이용하여 생성된 천이 테이블(T_transition)이다.

[0102] 또한, 도 7은 본 발명의 일 실시예에 따른 천이 테이블의 예시를 다이어그램으로 도시한 도면으로, 표 6의 천이 테이블에 기초하여 생성된 다이어그램을 나타낸다.

표 6

ID	state	signature	entry	next_state
1	0	BS_1, BS_3	E_1	1
2	1	BS_1	E_2	2
3	2	BS_1	E_5	3
4	0	BS_2	E_3	4
5	4	BS_2	E_1	5
6	5	BS_2	E_2	6
7	1	BS_3	E_4	7
8	7	BS_3	E_3	8

[0103]

[0104] 천이 테이블은 트래픽 분석 상태의 식별자(ID), 현재 상태(state), 매칭 가능 행위 시그니처(signature), 엔트리(entry), 다음 상태(next_state)를 항목으로 가진다.

[0105] 예컨대, 표 6과 도 7에서 아직 어떤 현재 엔트리도 생성되지 않은 경우에 대한 트래픽 분석 상태는 초기 상태인 상태 0으로 설정되어 있다.

[0106] 만일, 초기 상태인 상태 0에서 현재 엔트리 E1이 생성되면 트래픽 분석 상태의 다음 상태는 상태 1로 설정되어 있으며, 해당 트래픽 분석 상태는 행위 시그니처 BS1 또는 BS3와 매칭 가능한 것으로 정의되어 있다. 즉, 표 5에서 BS1 또는 BS3에서의 엔트리 생성 순서를 보면 E1이 가장 먼저 생성되는 것으로 되어 있는데, 현재 엔트리로서 E1이 생성되었으므로 향후 새로운 엔트리가 생성됨에 따라 트래픽 분석 상태의 최종 상태가 BS1 또는 BS3에 해당하는 상태가 될 수 있음을 나타내는 것이다.

[0107] 또한, 초기 상태 상태 0에서 E3가 생성되면 트래픽 분석 상태의 다음 상태는 상태 4로 설정되어 있으며, 해당 트래픽 분석 상태의 현재 상태는 행위 시그니처 BS2와 관련되어 있는 것으로 정의되어 있다.

[0108] 표 7은 표 5의 행위 시그니처를 이용하여 생성된 출력 테이블(T_output)의 예시이다.

표 7

ID	output state	signature
1	3	BS_1
2	6	BS_2
3	8	BS_3

[0109]

[0110] 표 7의 출력 테이블(T_output)은 최종 상태 식별자(ID), 최종 상태(output state), 행위 시그니쳐(signature)를 항목으로 포함한다.

[0111] 예컨대, 상태 테이블(T_status)의 특정 레코드의 현재 상태가 상태 천이 과정을 통해 상태 3에 도달하게 되면, 출력 테이블(T_output)의 ID 1을 참조하여 최종 상태를 상태 3으로 천이시키는데 관련된 트래픽 플로우들(엔트리의 조합)을 행위 시그니쳐 1번(BS1)에 의해 분석된 것으로 판단한다. 즉, 엔트리 E1, E2, E5에 대응되는 트래픽 플로우들은 BS1에 대응되는 트래픽으로 분석하게 된다.

[0112] 표 8은 표 5의 행위 시그니쳐를 이용하여 생성된 타임아웃 테이블(T_timeout)의 예시이다.

표 8

ID	signature	interval
1	BS_1	3000
2	BS_2	13000
3	BS_3	6000

[0113]

[0114] 표 8의 타임아웃 테이블(T_timeout)은 타임아웃 식별자(ID), 행위 시그니쳐(signature), 시간 간격(interval)을 항목으로 포함한다.

[0115] 예컨대, 상태 테이블(T_status)의 특정 레코드가 타임아웃 테이블(T_timeout)에서 명시하고 있는 행위 시그니쳐의 시간 간격(interval)보다 더 오랜 시간 상태 테이블(T_status)에 남아있으면, 해당 레코드는 더 이상 유효하지 않다고 판단하고 해당 레코드를 삭제할 수 있다.

[0116] **3.2 분석 단계**

[0117] 본 실시예에서는 준비 단계에서 구성한 천이 테이블(T_transition), 출력 테이블(T_output), 타임아웃 테이블(T_timeout)을 사용하여 실제 트래픽을 분석하는 방법을 설명한다.

[0118] 여러 호스트에서 발생한 트래픽 플로우가 발생 순서에 따라 네트워크 트래픽 분석 장치에게 입력되기 때문에 호

스트별 상태를 관리할 수 있는 상태 테이블(T_status)을 사용한다.

[0119] 표 9는 본 발명의 일 실시예에 따른 트래픽 분석 방법을 설명하기 위한 첫 질의 패킷(FRP) 테이블의 예시이다.

표 9

FRP ID	srcIP	capture_time	payload	matching entry
FRP_1	xxx.xxx.xxx.1	12:30' 11"	pear	E_3
FRP_2	xxx.xxx.xxx.3	12:30' 12"	pea	E_3
FRP_3	xxx.xxx.xxx.2	12:30' 13"	peach	E_3
FRP_4	xxx.xxx.xxx.1	12:30' 14"	lettuce	E_4
FRP_5	xxx.xxx.xxx.3	12:30' 17"	kiwijuice	E_1
FRP_6	xxx.xxx.xxx.2	12:30' 18"	kiwi	E_1
FRP_7	xxx.xxx.xxx.2	12:30' 21"	kiwi	E_5
FRP_8	xxx.xxx.xxx.3	12:30' 22"	leek	E_2
FRP_9	xxx.xxx.xxx.3	12:30' 25"	peachpie	E_3
FRP_{10}	xxx.xxx.xxx.2	12:30' 26"	lemon	E_2

[0120]

[0121] 표 9에서 FRP ID는 첫 질의 패킷(First Request Packet:FRP)의 식별자이고, srcIP는 첫 질의 패킷가 발생한 호스트의 ip주소이고, capture_time은 첫 질의 패킷이 네트워크 분석 장치에 수집된 시각을 나타내고, payload는 첫 질의 패킷에 포함된 페이로드 값을 나타내고, matching entry는 첫 질의 패킷 각각에 매칭되는 엔트리를 나타낸다.

[0122] 표 10은 FRP1과 FRP2가 수집된 경우에 대한 상태 테이블(T_status)의 예시를 도시한 것이다.

[0123] 표 10에는 FRP1과 FRP2에 대한 레코드가 기록되어 있으며, 레코드 식별자(ID), 호스트 식별자(host), 트래픽 분석 상태의 현재 상태(state), 행위 시그니쳐(signature), 매칭 시각(match_time), 히스토리(history)가 항목으로 포함되어 있다.

[0124] 이때, 레코드 식별자(ID)는 레코드의 식별자를 나타내고, 호스트 식별자(host)는 첫 질의 패킷을 발생시킨 호스트를 나타내고, 트래픽 분석 상태의 현재 상태(state)는 트래픽 분석 상태의 현재 상태를 나타내고, 행위 시그니쳐(signature)는 수집된 첫 질의 패킷과 매칭 가능한 행위 시그니쳐를 나타내고, 매칭 시각(matching time)은 레코드가 추가된 시각을 나타내고, 히스토리(history)는 트래픽 분석 상태의 현재 상태(state)의 천이에 사용된 모든 첫 질의 패킷에 대한 목록을 나타낸다.

[0125] 예컨대, 표 10의 첫 번째 행은 호스트 1에서 FRP1이 발생되었음을 나타내는데 FRP1의 엔트리는 표 9를 참조할때 E3이므로, 표 6의 천이 테이블에 따라 트래픽 분석 상태의 현재 상태는 상태 4가 되며, 매칭 가능한 행위 시그니쳐는 BS2가 되고 레코드가 기록된 시각은 12:30'11''임을 나타낸다.

표 10

ID	host	state	signatures	match_time	history
1	$host_1$	4	BS_2	12:30' 11"	$\langle FRP_1 \rangle$
2	$host_2$	4	BS_2	12:30' 12"	$\langle FRP_2 \rangle$

[0126]

[0127] 표 11은 FRP1과 FRP2에 이어 FRP3가 수집된 경우에 대한 상태 테이블(T_status)의 예시를 도시한 것이다.

[0128] 표 11에서는 insertstatus()함수를 이용해 FRP3가 추가되었으며, 세번째 행에 FRP3에 대한 레코드가 추가된 것을 확인할 수 있다.

표 11

ID	host	state	signatures	match_time	history
1	$host_1$	4	BS_2	12:30' 11"	$\langle FRP_1 \rangle$
2	$host_2$	4	BS_2	12:30' 12"	$\langle FRP_2 \rangle$
3	$host_3$	4	BS_2	12:30' 13"	$\langle FRP_3 \rangle$

[0129]

[0130] 표 12 및 표 13은 FRP5까지 수집된 경우에 대한 상태 테이블(T_status)의 예시를 도시한 것으로, 표 12는 FRP5에 대한 신규 레코드만을 추가한 것을 나타내고, 표 13은 표 12에서의 신규 레코드 추가 후에 FRP5와 관련된 기존 레코드의 상태를 갱신한 것을 나타낸다.

[0131] 표 12를 표 11과 비교하여 설명하면, 표 11의 상황에서 호스트 1로부터 FRP4가 발생되면 표 9에서 FRP4에 해당하는 현재 엔트리는 E4인데, 도 7의 천이 테이블 다이어그램에서 초기 상태 다음에 E4가 발생하거나, E3가 발생(FRP1에 의해 호스트 1에서 E3가 발생)한 다음에 E4가 발생한 경우에 대한 트래픽 분석 상태의 현재 상태가 존재하지 않으므로 상태 테이블(T_status)에 포함된 레코드 중에 변경되는 값은 없다.

[0132] 다음으로, 호스트 3에서 FRP5가 발생하면 표 9에서 FRP5에 해당하는 현재 엔트리는 E1이고, 도 7의 천이 테이블 다이어그램에서 초기 상태인 상태 0에서 엔트리 E1이 발생하면 트래픽 분석 상태의 현재 상태가 상태 1이므로 그와 관련된 레코드를 네 번째 행에 추가한다.

표 12

ID	host	state	signatures	match_time	history
1	$host_1$	4	BS_2	12:30' 11"	$\langle FRP_1 \rangle$
2	$host_2$	4	BS_2	12:30' 12"	$\langle FRP_2 \rangle$
3	$host_3$	4	BS_2	12:30' 13"	$\langle FRP_3 \rangle$
4	$host_3$	1	BS_1, BS_3	12:30' 17"	$\langle FRP_5 \rangle$

[0133]

[0134] 한편, 도 7의 천이 테이블 다이어그램에는 엔트리 E3이 발생한 후에 엔트리 E1이 발생하면 트래픽 분석 상태의 현재 상태가 상태 5로 천이되는 것으로 표시되어 있는데, 표 13에서는 이를 반영하여 세 번째 행의 레코드를 갱신하게 된다.

[0135] 즉, 표 12에서는 세번째 행에서 호스트 3의 트래픽 분석 상태의 현재 상태가 상태 4로 표시되어 있고 히스토리에 FRP3만 표시되어 있는데, 표 13에서는 updatestatus()함수를 이용하여 세번째 행의 트래픽 분석 상태의 현재 상태를 상태 5로 천이시켜 표시하고, 히스토리에는 FRP3과 FRP5를 표시하는 등 세번째 행의 레코드를 갱신한다.

표 13

ID	host	state	signatures	match_time	history
1	$host_1$	4	BS_2	12:30' 11"	$\langle FRP_1 \rangle$
2	$host_2$	4	BS_2	12:30' 12"	$\langle FRP_2 \rangle$
3	$host_3$	5	BS_2	12:30' 13"	$\langle FRP_3, FRP_5 \rangle$
4	$host_3$	1	BS_1, BS_3	12:30' 17"	$\langle FRP_5 \rangle$

[0136]

[0137] 표 14 및 표 15는 FRP7까지 수집된 경우에 대한 상태 테이블(T_status)의 예시를 도시한 것으로, 표 14는 FRP7이 수집될 때까지의 상태 천이를 반영하여 생성된 상태 테이블(T_status)을 나타내고, 표 15는 표 14의 상태 테이블(T_status)에서 타임아웃 테이블에 기초하여 타임아웃 시간을 초과한 레코드를 삭제한 것을 나타낸다.

[0138] 표 14와 표 15를 비교하면, 표 14에서 네 번째 행 및 다섯 번째 행의 행위 시그니쳐(signature)가 BS1, BS3였으나 표 15에서 네 번째 행 및 다섯 번째 행의 행위 시그니쳐(signature)는 BS3만 남아있는 것을 알 수 있다.

[0139] 이는, 표 9에 기초할 때 FRP7이 네트워크 발생 장치에 수집된 시점이 네 번째 행 및 다섯 번째 행의 레코드가 상태 테이블에 기록된지 각각 4000ms, 3000ms가 지난 시점이므로, 표 15에서 네 번째 행 및 다섯 번째 행의 레코드 중에서 행위 시그니쳐 BS1에 대한 레코드를 timeoutstatus()함수를 이용하여 삭제한 것이다.

[0140] 만일, 표 14의 상황에서 추가적인 엔트리가 발생하지 않은 채로 6000ms가 지난 경우라면 표 14의 네 번째 행 및 다섯 번째 행의 레코드 전부가 삭제될 것이다.

표 14

ID	host	state	signature	match_time	history
1	$host_1$	4	BS_2	12:30' 11"	$\langle FRP_1 \rangle$
2	$host_2$	5	BS_2	12:30' 12"	$\langle FRP_2, FRP_6 \rangle$
3	$host_3$	5	BS_2	12:30' 13"	$\langle FRP_3, FRP_5 \rangle$
4	$host_3$	1	BS_1, BS_3	12:30' 17"	$\langle FRP_5 \rangle$
5	$host_2$	1	BS_1, BS_3	12:30' 18"	$\langle FRP_6 \rangle$

[0141]

표 15

ID	host	state	signature	match_time	history
1	$host_1$	4	BS_2	12:30' 11"	$\langle FRP_1 \rangle$
2	$host_2$	5	BS_2	12:30' 12"	$\langle FRP_2, FRP_6 \rangle$
3	$host_3$	5	BS_2	12:30' 13"	$\langle FRP_3, FRP_5 \rangle$
4	$host_3$	1	BS_3	12:30' 17"	$\langle FRP_5 \rangle$
5	$host_2$	1	BS_3	12:30' 18"	$\langle FRP_6 \rangle$

[0142]

[0143] 표 16은 FRP8까지 수집된 경우에 대한 상태 테이블(T_status)의 예시를 도시한 것이다.

[0144] 표 16에서는 전체 6개 행의 레코드 중에서 세번째 행의 레코드의 현재 상태가 6으로 표시되어 있는데, 표 7에 기초할 때 트래픽 분석 상태의 최종 상태에 해당하므로 세번째 행의 레코드와 관련된 트래픽 플로우들은 분석이 완료되었음을 나타낸다. 보다 구체적으로는, FRP3, FRP5, FRP8은 행위시그니처 BS2에 대한 트래픽 플로우인 것으로 분석된다.

표 16

ID	host	state	signature	match_time	history
1	$host_1$	4	BS_2	12:30' 11"	$\langle FRP_1 \rangle$
2	$host_2$	5	BS_2	12:30' 12"	$\langle FRP_2, FRP_6 \rangle$
3	$host_3$	6	BS_2	12:30' 13"	$\langle FRP_3, FRP_5, FRP_8 \rangle$
4	$host_3$	1	BS_3	12:30' 17"	$\langle FRP_5 \rangle$
5	$host_2$	1	BS_3	12:30' 18"	$\langle FRP_6 \rangle$

[0145]

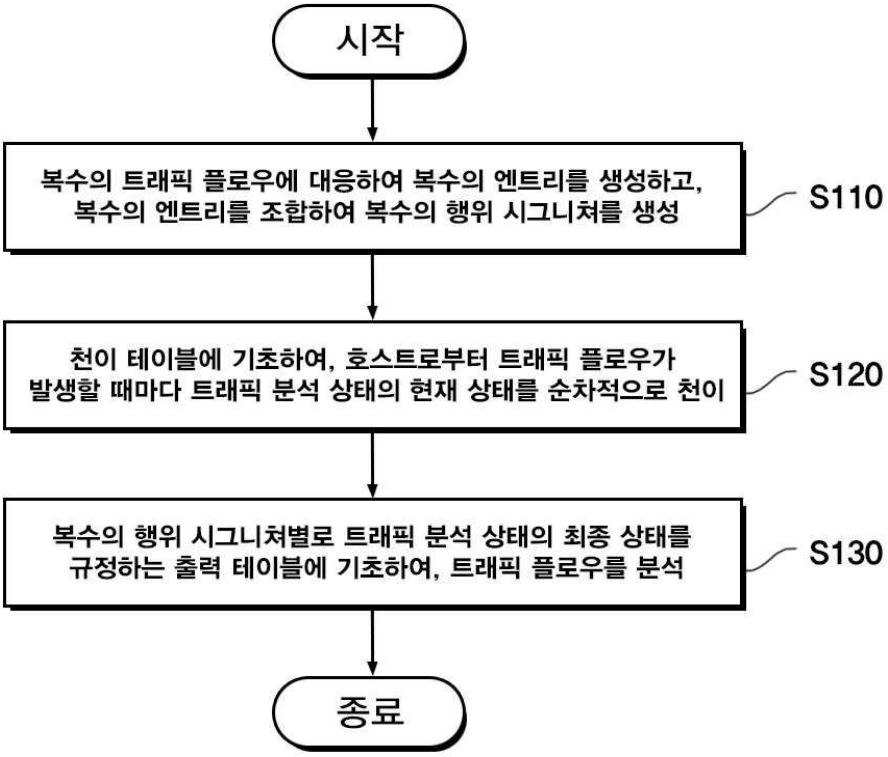
[0146] 한편, 상술한 본 발명의 실시예들은 컴퓨터에서 실행될 수 있는 프로그램으로 작성가능하고, 컴퓨터로 읽을 수 있는 기록매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다.

[0147] 상기 컴퓨터로 읽을 수 있는 기록매체는 마그네틱 저장매체(예를 들면, 롬, 플로피 디스크, 하드디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등) 를 포함한다.

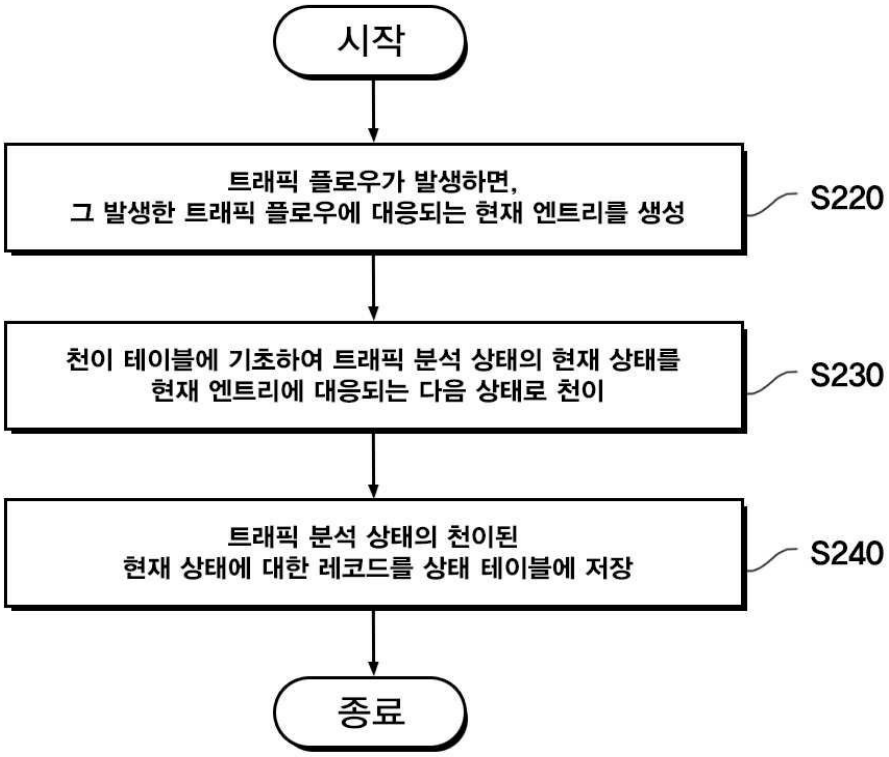
[0148] 이제까지 본 발명에 대하여 그 바람직한 실시예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면

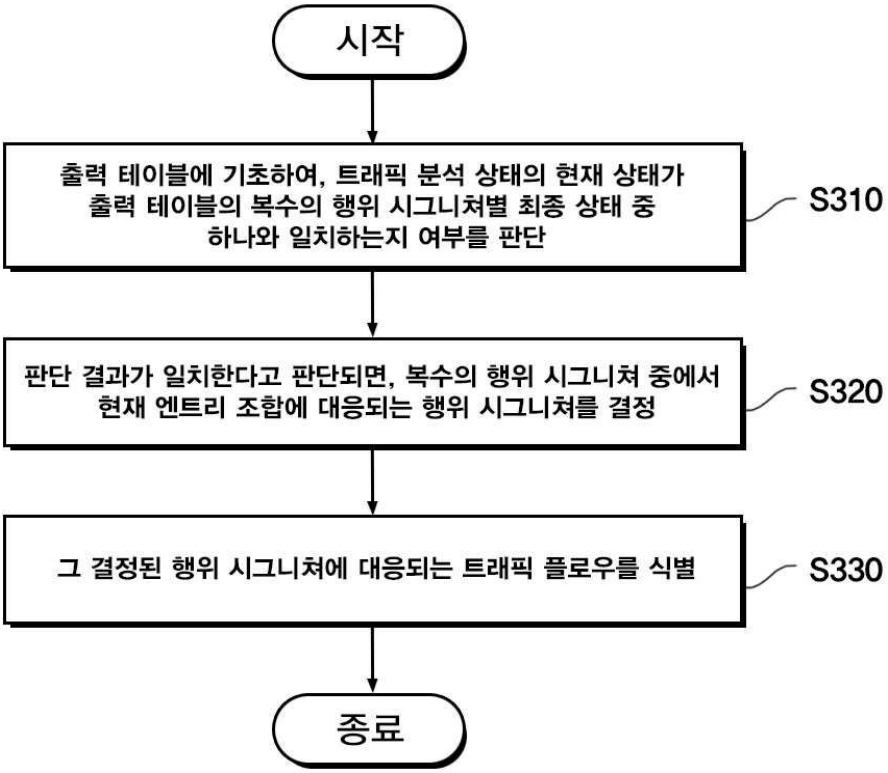
도면1


시작
복수의 트래픽 플로우에 대응하여 복수의 엔트리를 생성하고, 복수의 엔트리를 조합하여 복수의 행위 시그니쳐를 생성
S110
천이 테이블에 기초하여, 호스트로부터 트래픽 플로우가 발생할 때마다 트래픽 분석 상태의 현재 상태를 순차적으로 천이
S120
복수의 행위 시그니쳐별로 트래픽 분석 상태의 최종 상태를 규정하는 출력 테이블에 기초하여, 트래픽 플로우를 분석
S130
종료

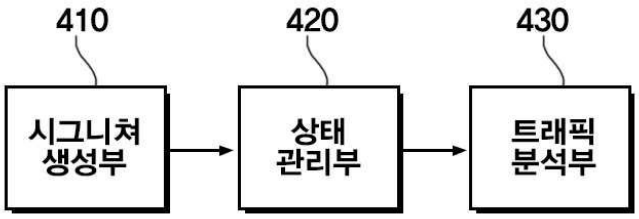
도면2


시작
트래픽 플로우가 발생하면,
그 발생한 트래픽 플로우에 대응되는 현재 엔트리를 생성
S220
천이 테이블에 기초하여 트래픽 분석 상태의 현재 상태를
현재 엔트리에 대응되는 다음 상태로 천이
S230
트래픽 분석 상태의 천이된
현재 상태에 대한 레코드를 상태 테이블에 저장
S240
종료

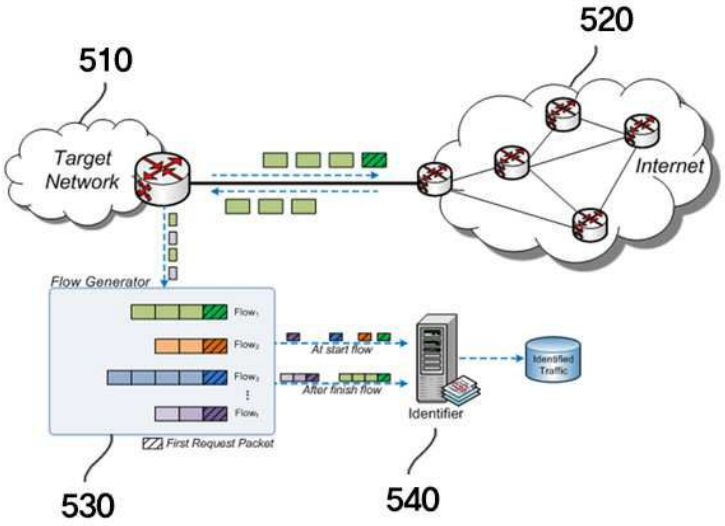
도면3


시작
출력 테이블에 기초하여, 트래픽 분석 상태의 현재 상태가
출력 테이블의 복수의 행위 시그니쳐별 최종 상태 중
하나와 일치하는지 여부를 판단
S310
판단 결과가 일치한다고 판단되면, 복수의 행위 시그니쳐 중에서
현재 엔트리 조합에 대응되는 행위 시그니쳐를 결정
S320
그 결정된 행위 시그니쳐에 대응되는 트래픽 플로우를 식별
S330
종료

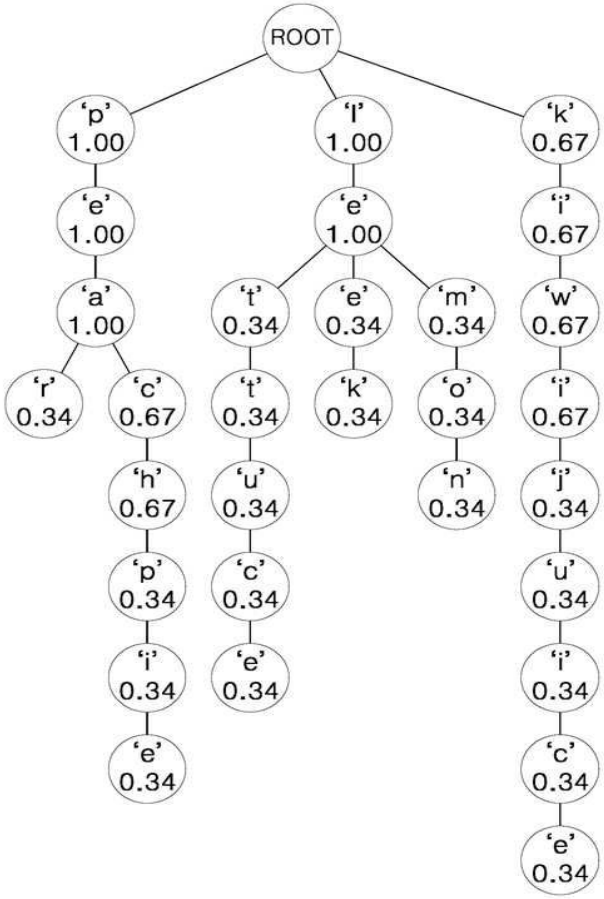
도면4


410
420
430
시그니쳐 생성부
상태 관리부
트래픽 분석부

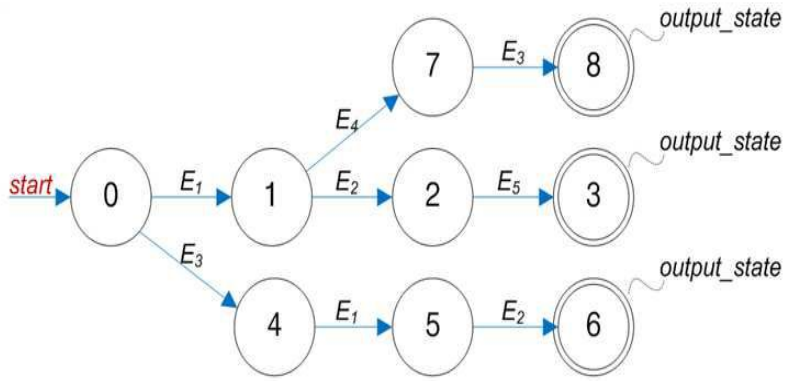
도면5


510
520
Target Network
Internet
Flow Generator
Flow1
Flow2
Flow3
Flowk
At start flow
After finish flow
Identified Traffic
Identifier
First Request Packet
530
540

도면6


ROOT
'p' 1.00
'e' 1.00
'a' 1.00
'r' 0.34
'c' 0.67
'h' 0.67
'p' 0.34
'i' 0.34
'e' 0.34
'l' 1.00
'e' 1.00
't' 0.34
'e' 0.34
'm' 0.34
't' 0.34
'k' 0.34
'o' 0.34
'u' 0.34
'c' 0.34
'e' 0.34
'n' 0.34
'k' 0.67
'i' 0.67
'w' 0.67
'i' 0.67
'j' 0.34
'u' 0.34
'i' 0.34
'c' 0.34
'e' 0.34

도면7


start
0
E1
1
E2
2
E3
3
E4
4
E5
5
E1
6
E2
7
E3
8
output_state
output_state
output_state