(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2016년01월04일
(11) 등록번호 10-1580417
(24) 등록일자 2015년12월18일

(51) 국제특허분류(Int. Cl.)
H04L 29/06 (2006.01)
(21) 출원번호 10-2014-0193573
(22) 출원일자 2014년12월30일
심사청구일자 2014년12월30일
(56) 선행기술조사문헌
KR101229012 B1
KR1020060060671 A

(73) 특허권자
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
김명섭
세종특별자치시 조치원읍 도원1로 16, 111동 902호 (푸르지오1차)
박준상
전라남도 무안군 현경면 수양촌길 88-3
이성호
서울특별시 강남구 삼성로 150, 109동 901호 (대치동, 미도아파트)
(74) 대리인
송인호, 민영준, 최관락

전체 청구항 수 : 총 4 항 심사관 : 문형섭

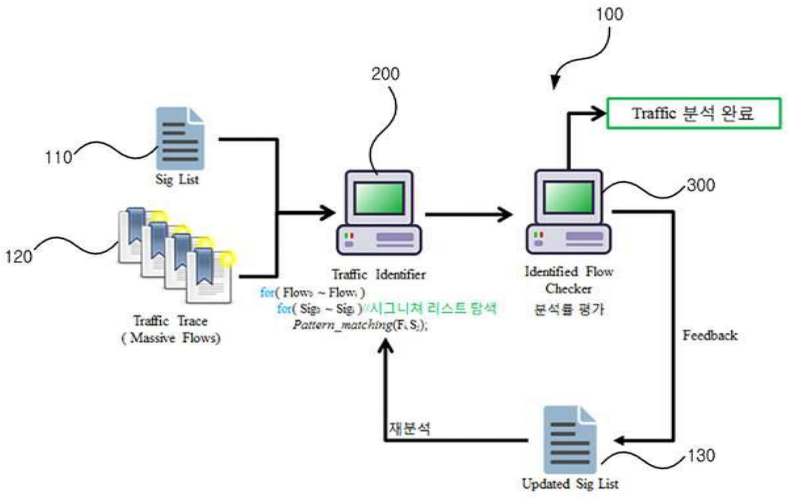
(54) 발명의 명칭 **시그니쳐 리스트 정렬 방법**

(57) 요 약

본 발명은 트래픽 분석을 위한 시그니쳐(signature) 리스트에 존재하는 각 시그니쳐의 가치를 산출하여, 산출된 가치에 따라 각 시그니쳐의 순서를 정하여 정렬하는 시그니쳐 리스트 정렬 방법에 관한 것이다.

본 발명의 일실시예에 의한 시그니쳐 리스트 정렬 방법은 복수의 시그니쳐(signature)에 대해 프리퀀시 밸류(Frequency value), 오프셋 밸류(Offset value) 및 타입 밸류(Type value)를 이용하여 각 시그니쳐에 시그니쳐 가치 값을 산출하는 단계-상기 프리퀀시 밸류는 매칭된 플로우의 빈도에 근거하여 산출되고, 상기 오프셋 밸류는 상기 시그니쳐가 매칭되는 오프셋 위치에 근거하여 산출되고, 상기 타입 밸류는 상기 시그니쳐의 유형에 근거하여 산출됨-; 및 상기 시그니쳐 가치 값에 근거하여 내림차순으로 복수의 시그니쳐를 정렬하는 단계를 포함할 수 있다.

대 표 도 - 도1



명 세 서

청구범위

청구항 1

복수의 시그니쳐(signature)에 대해 프리퀀시 밸류(Frequency value), 오프셋 밸류(Offset value) 및 타입 밸류(Type value)를 이용하여 각 시그니쳐에 시그니쳐 가치 값을 산출하는 단계-상기 프리퀀시 밸류는 매칭된 플로우의 빈도에 근거하여 산출되고, 상기 오프셋 밸류는 상기 시그니쳐가 매칭되는 오프셋 위치에 근거하여 산출되고, 상기 타입 밸류는 상기 시그니쳐의 유형에 근거하여 산출됨-; 및

상기 시그니쳐 가치 값에 근거하여 내림차순으로 복수의 시그니쳐를 정렬하는 단계를 포함하되,

상기 오프셋 밸류는 하기 수학식 1에 의해 산출되는 것을 특징으로 하는 시그니쳐 리스트 정렬 방법.

[수학식 1]

$$Offset\ Value = \frac{(\ Payload_MAX_Length\ ) - Sig_x offset}{Payload_MAX_Length} * 100$$

청구항 2

삭제

청구항 3

제 1 항에 있어서, 상기 시그니쳐의 유형은

분석률 유지를 위해 반드시 필요한 시그니쳐인지 여부에 따라 이센셜 시그니쳐 타입(essential signature type)과 로 밸류 타입(Low Value signature type)으로 구분되는 것을 특징으로 하는 시그니쳐 리스트 정렬 방법.

청구항 4

제 3 항에 있어서, 상기 이센셜 시그니쳐 타입은

항상 플로우에 1대1로 매칭되는 유형의 시그니쳐, 플로우에 1대1로 매칭되면서도 다른 플로우도 분석 가능한 시그니쳐, 및 플로우와 1대1로 매칭되지는 않지만 분석률 유지를 위해 필요한 시그니쳐 중 적어도 하나를 포함하는 것을 특징으로 하는 시그니쳐 리스트 정렬 방법.

청구항 5

제 4 항에 있어서, 상기 산출된 시그니쳐 가치 값은

상기 이센셜 시그니쳐 타입의 시그니쳐 중 최소값을 가지는 시그니쳐의 가치 값이 로 밸류 타입의 시그니쳐 중 최대값을 가지는 시그니쳐의 가치 값보다 큰 것을 특징으로 하는 시그니쳐 리스트 정렬 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 시그니쳐 리스트 정렬 방법에 관한 것으로, 보다 상세하게는 트래픽 분석을 위한 시그니쳐(signature) 리스트에 존재하는 각 시그니쳐의 가치를 산출하여, 산출된 가치에 따라 각 시그니쳐의 순서를 정하여 정렬하는 시그니쳐 리스트 정렬 방법에 관한 것이다.

배 경 기 술

[0002] 네트워크의 고속화와 더불어 다양한 서비스와 응용프로그램이 개발됨에 따라 기업이나 개인들은 인터넷으로 대표되는 네트워크에 대한 의존이 상당히 커져가고 있다. 이와 같은 현실 속에서 네트워크의 효율적 운용과 관리

를 위한 응용 레벨의 트래픽의 모니터링과 분석은 네트워크 사용현황 파악과 확장계획 수립 등의 다양한 분야에서 필요성이 커져가고 있다. 이를 위해서는 다양한 종류의 응용 레벨 트래픽을 정확하게 그리고 빠른 시간 안에 분류할 수 있는 방법이 요구된다.

[0003] 응용 레벨 트래픽 분류 방법에 있어 페이로드 시그니쳐 기반 분석 방법은 패킷의 헤더 정보나 통계 정보를 이용하는 다른 분석 방법들에 비해 상대적으로 높은 분류 정확성과 분석률을 보인다. 하지만 분류 시스템의 처리 속도에 있어 현재의 고속 네트워크상에서 발생하는 대용량 트래픽을 실시간으로 처리하기 위해서는 보다 빠른 방법이 요구 된다. 대용량의 트래픽을 발생시키는 응용수의 증가를 고려했을 때, 페이로드 기반 분석 방법의 처리 속도 향상은 언제나 중요한 과제이다. 이를 해결하기 위해 기존의 다양한 연구에서는 패턴 매칭 알고리즘의 성능 개선 기법에 대한 연구가 주를 이뤘다. 관련된 선행문헌으로 대한민국 공개특허 제10-2010-0116458호가 있다.

[0004] 하지만 매칭 알고리즘의 성능 개선은 제한적이며 현재의 고속 링크의 대용량 트래픽을 수용하는 부분에서 한계를 보이고 있다.

[0005] 따라서 시그니쳐 매칭의 성능을 향상하기 위한 기술에 대한 연구가 필요한 실정이다.

발명의 내용

해결하려는 과제

[0006] 본 발명의 목적은 응용 트래픽 분석 시스템 내에서 페이로드 시그니쳐 리스트를 재정렬시켜 분석 성능을 향상시킬 수 있는 시그니쳐 리스트 정렬 방법을 제공하는 데 있다.

과제의 해결 수단

[0007] 상기 목적을 달성하기 위해 본 발명의 일실시예에 의하면, 복수의 시그니쳐(signature)에 대해 프리퀀시 밸류(Frequency value), 오프셋 밸류(Offset value) 및 타입 밸류(Type value)를 이용하여 각 시그니쳐에 시그니쳐 가치 값을 산출하는 단계-상기 프리퀀시 밸류는 매칭된 플로우의 빈도에 근거하여 산출되고, 상기 오프셋 밸류는 상기 시그니쳐가 매칭되는 오프셋 위치에 근거하여 산출되고, 상기 타입 밸류는 상기 시그니쳐의 유형에 근거하여 산출됨-; 및 상기 시그니쳐 가치 값에 근거하여 내림차순으로 복수의 시그니쳐를 정렬하는 단계를 포함하는 시그니쳐 리스트 정렬 방법이 개시된다.

발명의 효과

[0008] 본 발명의 일실시예에 의한 시그니쳐 리스트 정렬 방법은 비효율적인 매칭 방법을 개선해 불필요한 매칭 횟수는 줄이고 응용 분석 시간은 단축시켜 분석 속도를 높일 수 있다.

도면의 간단한 설명

[0009] 도 1은 본 발명의 일실시예와 관련된 트래픽 분석 시스템을 나타내는 블록도이다.

도 2는 본 발명의 일실시예와 관련된 시그니쳐 중복 매칭 현상을 설명하기 위한 도면이다.

도 3은 본 발명의 일실시예와 관련된 시그니쳐 충돌 매칭 현상을 설명하기 위한 도면이다.

도 4는 본 발명의 일실시예와 관련된 시그니쳐 리스트 정렬 방법을 통해 재정렬된 시그니쳐 리스트와 랜덤 리스트를 통한 성능 비교를 설명하기 위한 도면이다.

도 5는 본 발명의 일실시예와 관련된 시그니쳐 유형을 설명하기 위한 도면이다.

도 6은 본 발명의 일실시예와 관련된 시그니쳐 타입을 결정하는 예를 나타낸다.

도 7은 본 발명의 일실시예와 관련된 시그니쳐 타입을 결정하는 알고리즘을 나타낸다.

도 8은 본 발명의 일실시예와 관련된 시그니쳐 리스트 정렬 방법의 성능과 다른 방법의 성능을 비교하기 위한 그래프이다.

발명을 실시하기 위한 구체적인 내용

[0010] 이하, 본 발명의 일실시예와 관련된 시그니쳐 리스트 정렬 방법에 대해 도면을 참조하여 설명하도록 하겠다.

[0011] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다.

[0012] 도 1은 본 발명의 일실시예와 관련된 트래픽 분석 시스템을 나타내는 블록도이다.

[0013] 도시된 바와 같이, 트래픽 분석 시스템(100)은 트래픽 식별 장치(200) 및 시그니쳐 정렬 장치(300)를 포함할 수 있다. 상기 트래픽 분석 시스템(100)은 플로우 단위 응용 트래픽을 분석할 수 있다.

[0014] 트랙픽 식별 장치(200)는 시그니쳐 리스트(110) 및 트래픽(120)을 입력을 수신할 수 있다. 재정렬되기 전에 입력되는 리스트인 시그니쳐 리스트(110)는 특별한 규칙에 의해 배열된 것이 아닌 랜덤 리스트(random list)일 수 있다. 또한, 시그니쳐 리스트의 순서를 정하기 위해 상기 트래픽(120)은 학습용으로 사용될 수 있다. 상기 트래픽 식별 장치(200)는 수집된 트래픽(Traffic trace)(120)에 시그니쳐 리스트(Sig list)(110)를 매칭시킨 뒤 매칭된 결과를 바탕으로 분석률을 평가하여 결과로서 분석률을 출력할 수 있다. 트랙픽 식별 장치(200)는 시그니쳐 리스트가 재정렬되기 전에는 시그니쳐 리스트에 존재하는 모든 시그니쳐의 풀 매칭(full matching) 작업을 수행할 수 있다.

[0015] 시그니쳐 정렬 장치(300) 상기 시그니쳐의 풀 매칭 작업의 결과인 분석률을 입력으로 수신하고, 상기 분석률에 근거하여 시그니쳐 리스트에 존재하는 각 시그니쳐의 가치 값(signature value)을 산출할 수 있다. 시그니쳐 정렬 장치(300) 산출된 시그니쳐 가치 값에 근거하여 각 시그니쳐의 순서를 결정하고, 결정된 순서에 근거하여 시그니쳐 리스트를 재정렬하여 업데이트 할 수 있다.

[0016] 시그니쳐 리스트가 재정렬된 이후에는 트랙픽 식별 장치(200)는 재정렬된 시그니쳐 리스트(130)를 이용하여 새로 수집된 트래픽과의 매칭 작업을 수행할 수 있다. 이 경우에 트랙픽 식별 장치(200)는 풀 매칭 작업을 하는 것이 아니라, 단위 플로우에 대해 응용 시그니쳐가 매칭되면 매칭 시퀀스를 끝내는 부분 매칭(Partial-matching)방식을 사용한다.

[0017] 한편, 풀 매칭 방식을 사용 할 경우 플로우에 다수의 시그니쳐가 매칭되는 중복과 충돌 매칭 현상이 불가피하게 발생하게 된다.

[0018] 도 2는 본 발명의 일실시예와 관련된 시그니쳐 중복 매칭 현상을 설명하기 위한 도면이다.

[0019] 도 2와 같이, 시그니쳐 중복 매칭은 같은 응용(Torrent)의 시그니쳐가 하나의 플로우에 두 번 이상 매칭 되는 현상이다.

[0020] 도 3은 본 발명의 일실시예와 관련된 시그니쳐 충돌 매칭 현상을 설명하기 위한 도면이다.

[0021] 도 3과 같이, 시그니쳐 충돌은 다른 응용의 시그니쳐가 하나의 플로우에 동시에 매칭되는 현상이다.

[0022] 이하, 실시예에서는 시그니쳐 충돌 매칭이 발생하면 응용 분석이 올바르게 되지 않기 때문에 시그니쳐 자체의 오류로 판단하기로 하겠다.

[0023] 상기 트래픽 분석 시스템(100)에서 풀 매칭 방식을 사용하는 이유는 풀 매칭 분석 과정에서 발생하는 시그니쳐 중복 현상을 통해 모든 시그니쳐의 분석 활용도를 파악할 수 있기 때문이다. 중복 매칭 현상을 분석하면 부분 매칭 방식으로는 알 수 없었던 모든 시그니쳐의 응용 분석 효율을 알 수 있고 이를 바탕으로 시그니쳐 리스트 재정렬이 가능하게 된다.

[0024] 이하에서는 기존의 시그니쳐 리스트처럼 시그니쳐들이 무작위로 정렬된 리스트를 랜덤 리스트(Random list)로 정의하고, 본 발명의 일실시예에 의한 시그니쳐 리스트 정렬 방법을 통해 재정렬 된 리스트를 재정렬 리스트(Reordered list)라고 정의하기로 한다.

[0025] 이 경우, 랜덤 리스트의 문제점은 트래픽의 발생 특징과 리스트 내에 존재하는 각 시그니쳐의 분석 효율을 고려하지 않은 시그니쳐 정렬 방식이다. 그 결과, 리스트 탐색 과정에서 불필요한 추가적 매칭이 발생하고 분석 속도가 감소하게 되는 원인이 된다.

[0026] 도 4는 본 발명의 일실시예와 관련된 시그니쳐 리스트 정렬 방법을 통해 재정렬된 시그니쳐 리스트와 랜덤 리스트를 통한 성능 비교를 설명하기 위한 도면이다. 도 4를 통해 두 리스트 간(랜덤 리스트 및 재정렬 리스트)의

응용 분석 효율을 비교할 수 있다.

[0027] 도 4와 같이 랜덤 리스트는 재정렬 리스트와 비교해 플로우에 시그니쳐를 매칭하는 과정에서 시그니쳐 리스트를 탐색, 매칭하는 과정(matching count)에 불필요한 시간이 추가적으로 소비된다. 랜덤 리스트는 첫 번째 시그니쳐 매칭까지 매칭 횟수가 5번인 반면 재정렬 리스트는 1번의 매칭만으로 플로우 분석이 가능하다. 리스트 탐색에 시간이 추가적으로 소비되면 결국 분석 속도가 떨어지게 된다.

[0028] 결과적으로 보다 빠른 응용 트래픽 분석 속도를 위해 중요한 것은 어떠한 방법을 통해 시그니쳐 리스트를 효율적으로 재정렬시키는가 이다.

[0029] 이하 실시예에서는 효율적 시그니쳐 리스트 재정렬 방법에 대해 보다 구체적으로 설명하기로 한다.

[0030] 본 발명의 일실시예에 의한 시그니쳐 리스트 재정렬 방법은 기본적으로 시그니쳐 리스트에 존재하는 각 시그니쳐들의 가치와 분석 효율을 평가해 수치화 하고 수치화 된 값을 기준으로 높은 값을 갖는 시그니쳐부터 내림차순으로 정렬하는 단계를 포함할 수 있다.

[0031] 이하에서는 시그니쳐 리스트 재정렬을 위한 각 시그니쳐의 가치 값(Signature value)을 계산하는 방법에 대해 설명하기로 한다.

[0032] 시그니쳐 가치 값(Signature value)는 1에서부터 100까지의 값을 갖고, 값이 높을수록 더 가치가 높은 시그니쳐이다. Signature value(이하 SV)는 분석과정에서 발생하는 모든 시그니쳐 매칭 상황을 고려한 3가지 변수를 통해 계산된다.

[0033] 첫 번째 변수는 프리퀀시 밸류(Frequency value)이다. Frequency value(이하 FV)는 수학식 1과 같이 전체 flow 중 특정 시그니쳐에 매칭된 flow의 비율을 나타내는 값으로, 리스트에 존재하는 시그니쳐들이 각각 flow를 얼마나 분석 할 수 있는지를 수치화 한 변수이다. FV는 1에서부터 100까지의 값을 갖고 높은 값을 갖을수록 더 가치가 높은 변수이다.

수학식 1

[0034]
$$Frequency\ Value = \frac{Sig_x\ Matched\ Flows}{Total\ Flows} * 100$$

[0035] 두 번째 변수로 오프셋 밸류(Offset value)가 있다. Offset value(이하 OV)는 각 시그니쳐가 flow내에 존재하는 패킷의 페이로드 offset 중 어느 위치에서 매칭 되는지를 나타낸다. 따라서 앞쪽에서 매칭되는 시그니쳐 일수록 더 높은 OV값을 갖는다. 패킷 페이로드의 앞부분에서 매칭될수록 더 빠르게 매칭된다는 것을 의미하고 빠르게 매칭 된다는 것은 결국 응용 분석 속도 측면에서 또한 빠르다는 것을 의미한다.

[0036] OV는 수학식 2와 같이 계산 된다. 값은 FV와 마찬가지로 1에서부터 100까지의 값을 갖고, 높은 값을 갖을수록 더 빠르게 매칭된다는 것을 의미한다.

수학식 2

[0037]
$$Offset\ Value = \frac{(\ Payload_MAX_Length\ ) - Sig_x\ offset}{Payload_MAX_Length} * 100$$

[0038] 예로 패킷의 50번째 오프셋에서 매칭되는 시그니쳐 X와 200번째 오프셋에서 매칭되는 시그니쳐 Y가 있을 때 시그니쳐 X는 96의 OV를 시그니쳐 Y는 86의 OV를 갖는다. 이 경우, Payload_MAX_Length는 1500이다.

[0039] 세 번째 변수는 타입 밸류(Type value)로 후술할 시그니쳐의 유형에 연관되는 값을 갖는다. Type value(이하 TV)는 시그니쳐의 유형에 따라 고정된 값을 갖거나 유연하게 상황에 맞춰 변화하는 값을 갖는다. TV는 일종의 보정치로 시그니쳐 유형에 따른 각 시그니쳐들 간의 우선순위가 존재 한다고 했을 때 순위가 역전되는 현상을 방지해준다.

[0040] 이하, 실시예에서는 시그니쳐들을 크게 두 가지 유형으로 나누고 각 유형별로 수학식 3과 같이 TV_{ET} 와 TV_{LT}로 정의하기로 한다.

수학식 3

[0041]
$$\begin{aligned} &Essential\ Signature\ Type \ : \ TV_{ET} = 0.5 \\ &Low\ Value\ Signature\ Type \ : \ TV_{LT} = Flexible \end{aligned}$$

[0042] 시그니쳐의 가치 값(Signature value)은 최종적으로 수학식 4와 같이 계산될 수 있다. 시그니처 정렬 장치(300)는 상기 시그니쳐의 가치 값 계산 된 각 시그니쳐들의 SV값을 바탕으로 시그니쳐 리스트를 재정렬할 수 있다.

수학식 4

[0043]
$$Signauture\ Value = TV * (FV + OV)$$

[0044] 도 5는 본 발명의 일실시예와 관련된 시그니쳐 유형을 설명하기 위한 도면이다.

[0045] 도 5를 참조하면, 시그니쳐 유형은 각 리스트에 존재하는 시그니쳐들이 가지고 있는 고유한 특징을 바탕으로 크게 2가지 세부적으로 4가지로 유형으로 정의될 수 있다.

[0046] 시그니쳐 유형을 정의하는 첫 번째 이유는 각 시그니쳐의 직접적인 매칭 효율성과는 별개로 실제 분석 시스템(100)에서 시그니쳐들이 갖는 고유한 특징이나 중요도를 SV값에 적용시키기 위해서이고 두 번째 이유는 이러한 각 시그니쳐의 특성에 따른 중요도를 TV값을 통해 조절하기 위해서이다.

[0047] 이센셜 시그니쳐 타입(Essential signature type)은 분석률 유지를 위해 반드시 필요한 시그니쳐 유형들로 유형 1, 2, 3의 시그니쳐들은 높은 우선순위를 갖는다. 반면 로 밸류 타입(Low Value signature type)은 분석률 측면에서 제외시켜도 되는 시그니쳐 유형으로 유형 4와 같이 flow와 매칭을 따로 해보지 않더라도 이미 해당 flow는 Essential signature type의 시그니쳐에 의해 분석되기 때문에 활용가치가 상대적으로 낮은 시그니쳐 유형이다.

[0048] 이 중 유형 3은 상대적으로 특수한 유형이다. 하나의 플로우에 유형 4의 시그니쳐들 간에 중복 매칭이 발생 할 경우 해당 시그니쳐들 간의 FV+OV 값을 비교 후 더 높은 값을 갖는 시그니쳐는 유형 4에서 유형 3으로 수정해준다.

[0049] 도 6은 본 발명의 일실시예와 관련된 시그니쳐 타입을 결정하는 예를 나타내고, 도 7은 본 발명의 일실시예와 관련된 시그니쳐 타입을 결정하는 알고리즘을 나타낸다.

[0050] 예시로 도 6과 같은 7개의 flow에 매칭 된 6개의 시그니쳐들(0, 9, 11, 16, 17, 19번)에 대해 유형을 정의할 수 있다.

[0051] 도 6에서 주목 할 점은 시그니쳐 0번과 11번, 19번만 있다면 나머지 9, 16, 17번 없이도 7개의 flow를 모두 분석할 수 있다는 것이다. 따라서 도 5의 유형 정의 방법에 의해 시그니쳐 0번과 19번은 유형 1, 11번은 유형 3으로 정의되고 모두 Essential signature type이다. 나머지 9, 16, 17번은 모두 유형 4이고 Low value type으로 정의 할 수 있다.

[0052] 시그니쳐 유형 결정 방법은 도 7의 알고리즘을 바탕으로 구현될 수 있다.

[0053] 각 시그니쳐 들의 유형이 정의되면 유형 별로 TV값이 부여될 수 있다. 수학식 4를 참고하였을 때 FV와 OV값이 합의 최대치가 200이기 때문에 SV의 최대치인 100으로 조정하기 위해서 Essential signature type들(유형 1,2,3)의 시그니쳐들의 TV값(TV_{ET})은 수식 3과 같이 고정적으로 0.5를 부여한다. 반면, 시그니쳐 정렬 장치(300)는 Low value signature type(유형 4)의 시그니쳐들에 대해서는 수학식 5와 같이 계산할 수 있다.

수학식 5

$$TV_{LT} = (SV_{ET_min} / (FV_{LT} + OV_{LT})) - 0.01$$
$$\because SV_{LT_max} < SV_{ET_min}$$
$$TV_{LT} < SV_{ET_min} / (FV_{LT} + OV_{LT})$$

[0054]

[0055] Essential signature type 시그니쳐들이 갖는 SV값들 중 최소값을 SV_{ET_min}이라 정의하고 Low value signature type 시그니쳐가 갖는 SV값들 중 최대값을 을 SV_{LT_max}라 정의했을 때, SVET_min 값은 SVLT_max 값보다 항상 더 커야한다. 따라서 TV_{LT}값은 수학식 5와 같이 계산될 수 있다.

[0056] 모든 시그니쳐의 TV값이 설정되면 SV값을 계산하기 위한 모든 변수들의 값이 정의 되었으므로 각 시그니쳐들의 SV값을 계산 한 뒤 SV값에 따라 높은 값부터 내림차순으로 시그니쳐 리스트를 재정렬 하면 시그니쳐 별 효율을 고려한 최적 시그니쳐 리스트가 완성된다.

[0057] 이하에서 실험예를 통해 본 발명의 일실시예와 관련된 시그니쳐 리스트 정렬 방법의 성능과 다른 방법의 성능을 비교한다. 실험은 다량의 트래픽 분석을 통해 얻어진 포괄적인 재정렬 시그니쳐 리스트를 새로운 트래픽 분석에 적용 시켰을 때 다른 알고리즘의 리스트에 비해 얼마나 더 효율적인지 평가한다.

[0058] 도 8은 본 발명의 일실시예와 관련된 시그니쳐 리스트 정렬 방법의 성능과 다른 방법의 성능을 비교하기 위한 그래프이다.

[0059] 실험예에서 사용되는 트래픽은 재정렬 리스트 제작을 위해 사용되는 Trainig set 트래픽과 각 리스트들의 성능 평가를 위해 사용되는 Testing set 트래픽이 있다. 시그니쳐 리스트로는 전술한 Random list와 본 발명의 일실시예에 의한 Reordered list 그리고 Ideal list가 있다. Ideal list는 Testing set 트래픽에 대한 응용 분석률이 가장 높은 시그니쳐부터 정렬 시킨 리스트이다. Ideal list는 Testing set 트래픽을 분석하기에 가장 이상적인 리스트로 Training set을 바탕으로 제작된 재정렬 리스트의 상대적 성능 비교를 위해 제작되었다.

[0060] 도 8을 참조하면, 전체적으로 Reordered list는 기존의 Random list보다 분석 시간 측면에서는 약 20%의 분석 시간 단축률을 보였고 Ideal list와는 거의 동일한 분석 시간을 보였다. 시그니쳐 매칭 회수 측면에서는 Random list 보다 약 40%의 추가 분석 효율을 보였고 Ideal list 보다는 약 3%정도 분석 효율이 떨어지는 것으로 나타났다

[0061] 전술한 바와 같이, 본 발명의 일실시예에 의한 시그니쳐 리스트 정렬 방법은 비효율적인 매칭 방법을 개선해 불필요한 매칭 횟수는 줄이고 응용 분석 시간은 단축시켜 분석 속도를 높일 수 있다.

[0062] 상술한 시그니쳐 리스트 정렬 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터로 판독 가능한 기록 매체에 기록될 수 있다. 이때, 컴퓨터로 판독 가능한 기록매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 한편, 기록매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.

[0063] 컴퓨터로 판독 가능한 기록매체에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM, DVD와 같은 광기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-Optical Media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다.

[0064] 한편, 이러한 기록매체는 프로그램 명령, 데이터 구조 등을 지정하는 신호를 전송하는 반송파를 포함하는 광 또는 금속선, 도파관 등의 전송 매체일 수도 있다.

[0065] 또한, 프로그램 명령에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0066] 상기와 같이 설명된 시그니쳐 리스트 정렬 방법은 상기 설명된 실시예들의 구성과 방법이 한정되게 적용될 수

있는 것이 아니라, 상기 실시예들은 다양한 변형이 이루어질 수 있도록 각 실시예들의 전부 또는 일부가 선택적으로 조합되어 구성될 수도 있다.

부호의 설명

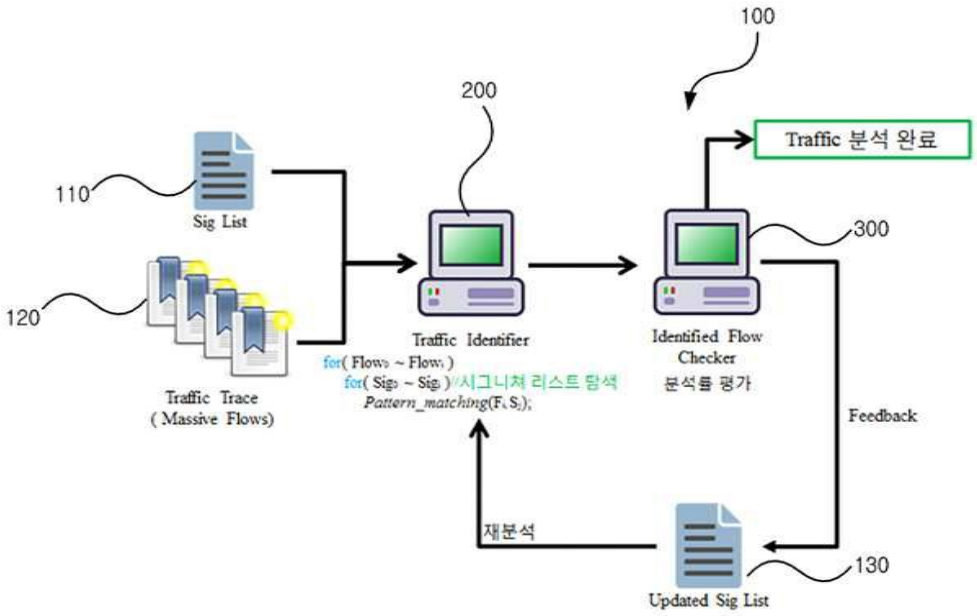
[0067] 100: 트래픽 분류 시스템

200: 트래픽 식별 장치

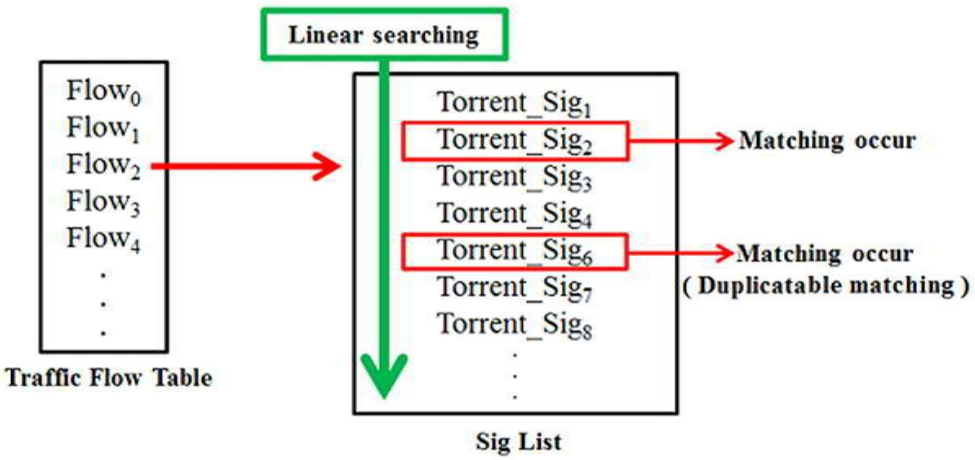
300: 시그니쳐 정렬 장치

도면

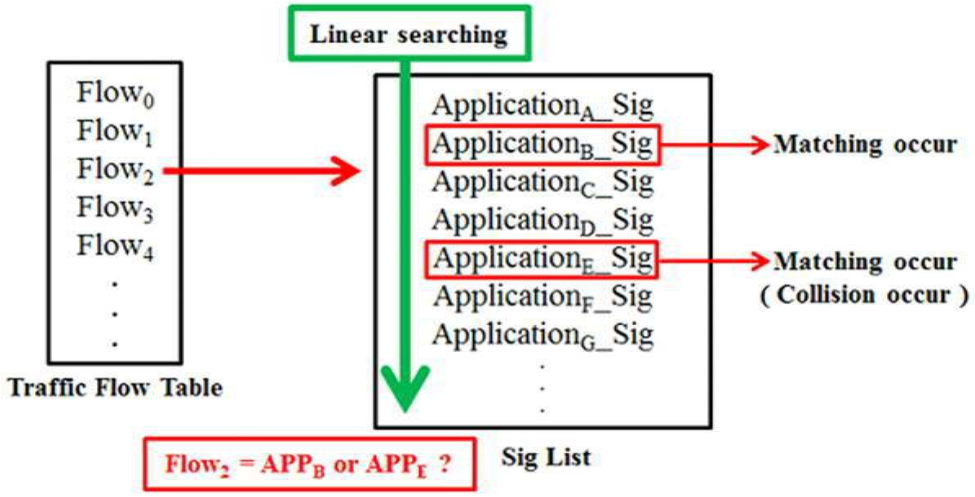
도면1



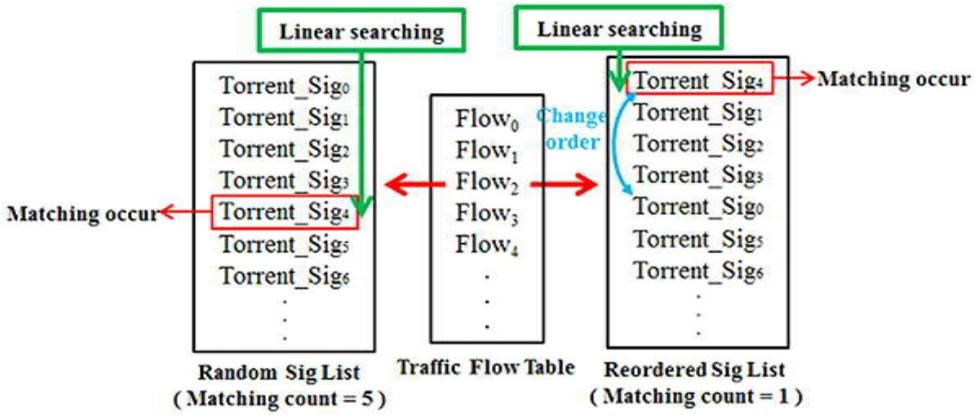
도면2



도면3


Linear searching
Flow0
Flow1
Flow2
Flow3
Flow4
Traffic Flow Table
ApplicationA_Sig
ApplicationB_Sig
ApplicationC_Sig
ApplicationD_Sig
ApplicationE_Sig
ApplicationF_Sig
ApplicationG_Sig
Sig List
Matching occur
Matching occur
(Collision occur)
Flow2 = APPB or APPE ?

도면4


Linear searching
Linear searching
Torrent_Sig0
Torrent_Sig1
Torrent_Sig2
Torrent_Sig3
Torrent_Sig4
Torrent_Sig5
Torrent_Sig6
Matching occur
Random Sig List
(Matching count = 5)
Flow0
Flow1
Flow2
Flow3
Flow4
Traffic Flow Table
Change order
Torrent_Sig4
Torrent_Sig1
Torrent_Sig2
Torrent_Sig3
Torrent_Sig0
Torrent_Sig5
Torrent_Sig6
Matching occur
Reordered Sig List
(Matching count = 1)

도면5



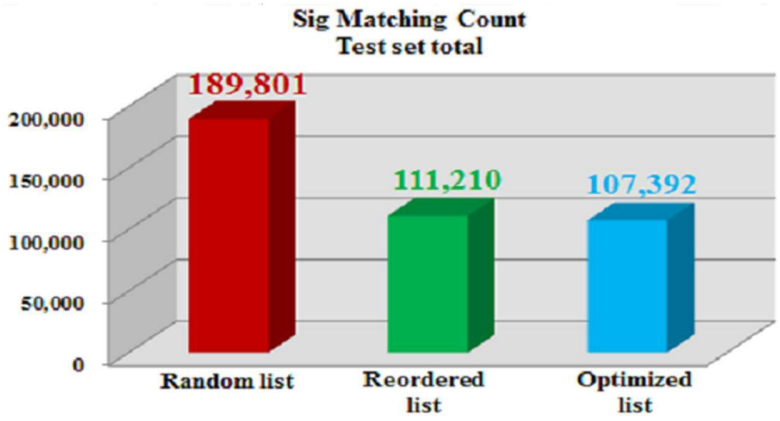
도면6

```
flow ---Matched Sig ID---> [DV:1]: (0th)
flow ---Matched Sig ID---> [DV:2]: (17th) (11th)
flow ---Matched Sig ID---> [DV:2]: (17th) (11th)
flow ---Matched Sig ID---> [DV:1]: (0th)
flow ---Matched Sig ID---> [DV:1]: (19th)
flow ---Matched Sig ID---> [DV:1]: (0th)
flow ---Matched Sig ID---> [DV:3]: (16th) (9th) (11th)
```

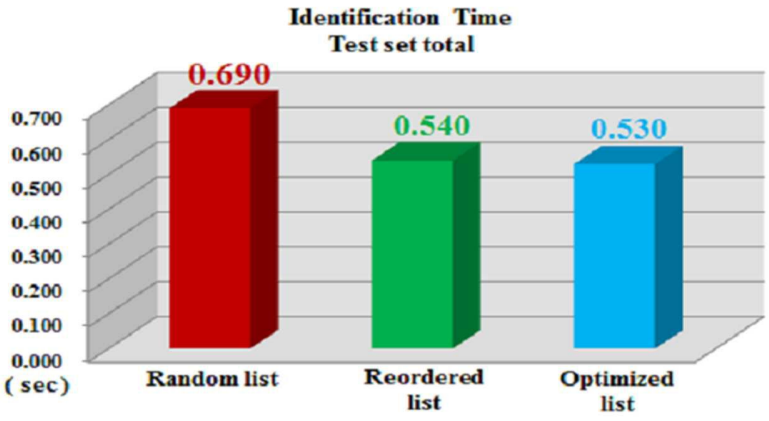
도면7

```
    f : Classified flow
    S : Matched signatures
    T : Signature types
    S(f) = { S | S0, S1, S2 .... Sn }, signatures which match flow f
    T(S) ∈ { T1, T2, T3, T4 }, signature types
1   Give type to each signature which matches flow
2:  Input : all classified flows of a application
3:  for each fi do                                  // Type 1 decision
4:      if ( |S(fi)| == 1 )
5:          T(S0) = T1;
6:  done
7:  for each fi  do                                 // Type 2 , 4 decicison
8:      if( |S(fi)| > 1 )
9:      {
10:         for each Sn in S(fi) do
11:           if ( T(Sn) == T1 )
12:               T(Sn) = T2;
13:           else if ( T(Sn) == None )
14:               T(Sn) = T4;
15:       done
16:     }
17: done
18: for each fi do                                  // Type 3 decision
19:     if ( T(Sn) is all T4 )
20:         find Sk which has maximum FV+OV value among S(fi)
21:         T(Sk) = T3;
22: done
```

도면8



(a)



(b)

【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 4

【변경전】

시그치처

【변경후】

시그니처