(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2011-0070464
(43) 공개일자 2011년06월24일

(51) Int. Cl.
H04L 12/26 (2006.01)
(21) 출원번호 10-2009-0127293
(22) 출원일자 2009년12월18일
심사청구일자 2009년12월18일

(71) 출원인
한국전자통신연구원
대전 유성구 가정동 161번지
고려대학교 산학협력단
서울 성북구 안암동5가 1
(72) 발명자
최태상
대전광역시 유성구 관평동 대덕테크노밸리 1011-1204
양선희
대전광역시 유성구 노은동 새미래아파트 814동 801호
김명섭
충청남도 연기군 조치원읍 신흥리 푸르지오1차 107-401
(74) 대리인
특허법인 신지, 유경열, 박진석

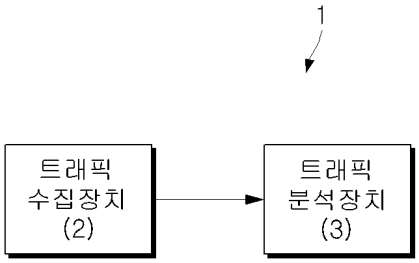
전체 청구항 수 : 총 20 항

(54) 트래픽 수집장치, 트래픽 분석장치, 시스템 및 그 분석방법

(57) 요 약

트래픽 수집장치, 트래픽 분석장치, 시스템 및 그 분석방법이 개시된다. 본 발명의 일 실시예에 따른 트래픽 분석 시스템은 네트워크를 통해 수집된 적어도 하나의 패킷으로부터 생성된 양방향 플로우를, 페이로드가 존재하는 페이로드 패킷의 전송방향 및 크기 정보를 이용하여 응용 프로그램별로 분류한다.

대 표 도 - 도1



이 발명을 지원한 국가연구개발사업
과제고유번호
부처명
연구관리전문기관
연구사업명
연구과제명 미래기술연구부 원천기술개발과제
기여율
주관기관 한국전자통신연구원, 고려대학교 산학협력단
연구기간 2009년 03월 01일 ~ 2010년 02월 28일

특허청구의 범위

청구항 1

네트워크를 통해 적어도 하나의 패킷을 수집하는 패킷 수집부;

상기 수집된 적어도 하나의 패킷으로부터 양방향 플로우를 생성하는 플로우 생성부; 및

상기 생성된 양방향 플로우에서 페이로드가 존재하는 페이로드 패킷을 대상으로 상기 페이로드 패킷의 전송방향 및 크기 정보를 갖는 페이로드 통계정보를 생성하는 페이로드 통계정보 생성부를 포함하는 것을 특징으로 하는 트래픽 수집장치.

청구항 2

제 1 항에 있어서,

상기 페이로드 통계정보는 상기 페이로드 패킷의 전송방향과 페이로드 크기를 나타내는 페이로드 패킷 벡터 및 상기 페이로드 패킷 벡터를 구성하는 페이로드 패킷의 갯수의 조합인 것을 특징으로 하는 트래픽 수집장치.

청구항 3

제 2 항에 있어서,

상기 페이로드 패킷 벡터의 전송방향은 양의 부호(+) 또는 음의 부호(-)로 표현되고,

상기 양의 부호(+)는 상기 페이로드 패킷의 전송방향이 클라이언트에서 서버로 향함을 의미하고, 상기 음의 부호(-)는 상기 페이로드 패킷의 전송방향이 서버에서 클라이언트로 향함을 의미하는 것을 특징으로 하는 트래픽 수집장치.

청구항 4

제 3 항에 있어서,

호스트 간 패킷 전송시에 전송제어절차 프로토콜(TCP)을 사용하는 경우에는 연결요청 패킷을 수신한 호스트를 서버로 지정하고, 사용자 데이터그램 프로토콜(UDP)을 사용하는 경우에는 첫 패킷을 수신한 호스트를 서버로 지정하는 것을 특징으로 하는 트래픽 수집장치.

청구항 5

제 2 항에 있어서,

상기 페이로드 패킷 벡터의 크기는 응용계층 정보를 포함하는 페이로드의 데이터 크기인 것을 특징으로 하는 트래픽 수집장치.

청구항 6

제 1 항에 있어서, 상기 페이로드 통계정보 생성부는,

복수의 페이로드 패킷 중에 먼저 수집된 n개의 페이로드 패킷을 대상으로 상기 페이로드 통계정보를 생성하는 것을 특징으로 하는 트래픽 수집장치.

청구항 7

제 1 항에 있어서,

상기 페이로드 통계정보와 플로우 식별자 및 플로우 기본정보가 포함된 플로우 레코드를 생성하여 저장하는 플로우 레코드 저장부를 더 포함하는 것을 특징으로 하는 트래픽 수집장치.

청구항 8

제 1 항에 있어서,

상기 페이로드 패킷은 응용계층의 정보를 포함하는 페이로드를 갖는 패킷으로서, 컨트롤 패킷을 제외하는 것을

특징으로 하는 트래픽 수집장치.

청구항 9

응용 프로그램별로 페이로드 패킷의 전송방향 및 크기 정보가 상이한 페이로드 통계 시그니쳐를 저장하는 페이로드 통계 시그니쳐 저장부; 및

트래픽을 수집하는 트래픽 수집장치로부터 전송된 양방향 플로우를 상기 페이로드 통계 시그니쳐를 이용하여 응용 프로그램별로 분류하는 트래픽 분류부를 포함하는 것을 특징으로 하는 트래픽 분석장치.

청구항 10

제 9 항에 있어서,

상기 페이로드 통계 시그니쳐는 전송계층 프로토콜, 페이로드 패킷의 전송방향과 페이로드 크기를 나타내는 페이로드 패킷 벡터, 상기 페이로드 패킷 벡터를 구성하는 페이로드 패킷의 갯수, 거리 임계값 및 응용 프로그램 명칭의 조합인 것을 특징으로 하는 트래픽 분석장치.

청구항 11

제 10 항에 있어서,

상기 페이로드 패킷 벡터의 전송방향은 양의 부호(+) 또는 음의 부호(-)로 표현되고,

상기 양의 부호(+)는 상기 페이로드 패킷의 전송방향이 클라이언트에서 서버로 향함을 의미하고, 상기 음의 부호(-)는 상기 페이로드 패킷의 전송방향이 서버에서 클라이언트로 향함을 의미하는 것을 특징으로 하는 트래픽 분석장치.

청구항 12

제 10 항에 있어서,

상기 페이로드 패킷 벡터의 크기는 응용계층 정보를 포함하는 페이로드의 데이터 크기인 것을 특징으로 하는 트래픽 분석장치.

청구항 13

네트워크를 통해 적어도 하나의 패킷을 수집하여 양방향 플로우를 생성하고, 상기 양방향 플로우에서 페이로드가 존재하는 페이로드 패킷을 대상으로 페이로드 패킷의 전송방향 및 크기 정보를 갖는 페이로드 통계정보를 생성하는 트래픽 수집장치; 및

상기 트래픽 수집장치로부터 상기 페이로드 통계정보를 포함하는 양방향 플로우를 수신하여, 상기 양방향 플로우를 응용 프로그램별로 페이로드 패킷의 전송방향 및 크기 정보가 상이한 페이로드 통계 시그니쳐를 이용하여 응용 프로그램별로 분류하는 트래픽 분석장치를 포함하는 것을 특징으로 하는 트래픽 분석 시스템.

청구항 14

응용 프로그램별로 페이로드 패킷의 전송방향 및 크기 정보가 상이한 페이로드 통계 시그니쳐 리스트를 구축하는 단계;

네트워크를 통해 수집된 양방향 플로우에서 페이로드가 존재하는 페이로드 패킷의 전송방향 및 크기 정보를 갖는 페이로드 통계정보를 상기 페이로드 통계 시그니쳐 리스트에 해당되는 페이로드 통계 시그니쳐와 비교하는 단계; 및

상기 비교 결과에 따라 상기 양방향 플로우를 응용 프로그램별로 분류하는 단계를 포함하는 것을 특징으로 하는 트래픽 분석방법.

청구항 15

제 14 항에 있어서,

상기 페이로드 통계정보는 페이로드 패킷의 전송방향과 페이로드 크기를 나타내는 페이로드 패킷 벡터 및 상기

페이로드 패킷 벡터를 구성하는 페이로드 패킷의 갯수의 조합인 것을 특징으로 하는 트래픽 분석방법.

청구항 16

제 14 항에 있어서,

상기 페이로드 통계 시그니쳐는 전송계층 프로토콜, 페이로드 패킷의 전송방향과 페이로드 크기를 나타내는 페이로드 패킷 벡터, 상기 페이로드 패킷 벡터를 구성하는 페이로드 패킷의 갯수, 거리 임계값 및 응용 프로그램 명칭의 조합인 것을 특징으로 하는 트래픽 분석방법.

청구항 17

제 14 항에서, 상기 비교하는 단계는,

상기 페이로드 통계정보의 전송계층 프로토콜 및 상기 페이로드 통계 시그니쳐의 전송계층 프로토콜을 비교하는 단계;

상기 비교 결과, 전송계층 프로토콜이 일치하면 상기 페이로드 통계정보의 페이로드 패킷 벡터를 구성하는 페이로드 패킷의 갯수 및 상기 페이로드 통계 시그니쳐의 페이로드 패킷 벡터를 구성하는 페이로드 패킷의 갯수를 비교하는 단계; 및

상기 비교 결과, 페이로드 패킷의 갯수가 일치하면 상기 페이로드 통계정보의 페이로드 패킷 벡터 및 상기 페이로드 통계 시그니쳐의 페이로드 패킷 벡터 간의 거리가 상기 페이로드 통계 시그니쳐의 거리 임계값 내에 포함되는지를 비교하는 단계를 포함하는 것을 특징으로 하는 트래픽 분석방법.

청구항 18

제 17 항에서, 상기 양방향 플로우를 응용 프로그램별로 분류하는 단계는,

상기 페이로드 통계정보의 페이로드 패킷 벡터 및 상기 페이로드 통계 시그니쳐의 페이로드 패킷 벡터 간의 거리가 상기 페이로드 통계 시그니쳐의 거리 임계값 내에 포함되면, 상기 양방향 플로우를 상기 페이로드 통계 시그니쳐에 해당되는 응용 프로그램으로 분류하는 것을 특징으로 하는 트래픽 분석방법.

청구항 19

제 17 항에 있어서, 상기 페이로드 통계 시그니쳐의 거리 임계값 내에 포함되는지를 비교하는 단계는,

비교 시에 시티 블록 거리(city-block distance) 계산 방식을 이용하는 것을 특징으로 하는 트래픽 분석방법.

청구항 20

제 14 항에서, 상기 비교하는 단계는,

상기 양방향 플로우의 페이로드 통계정보와 상기 페이로드 통계 시그니쳐 리스트에 해당되는 하나의 페이로드 통계 시그니쳐를 비교한 결과 일치하지 않으면, 상기 페이로드 통계 시그니쳐 리스트에 해당되는 다른 페이로드 통계 시그니쳐를 선택하여 상호 비교하는 것을 특징으로 하는 트래픽 분석방법.

명 세 서

발명의 상세한 설명

기 술 분 야

[0001] 본 발명의 일 양상은 네트워크 관리 및 서비스 기술에 관한 것으로, 보다 상세하게는 트래픽 관리 기술에 관한 것이다.

배 경 기 술

[0002] 네트워크 트래픽을 응용 프로그램별로 분류하기 위한 기술들이 있다. 그 중 시그니쳐 기반 분류방법은 응용 프로그램별로 트래픽에서 사용되는 고유한 특징인 시그니쳐를 기준으로 트래픽을 분류하는 방법이다.

[0003] 시그니쳐 기반 분류방법 중 하나는 페이로드 스트링 시그니쳐 기반 분류방법이다. 이는 트래픽을 구성하는 패

킷의 페이로드 내에서, 응용 프로그램의 고유한 스트링이 존재하는지를 판단하여 트래픽을 분류하는 것으로 분류 정확도를 높일 수 있다.

[0004] 그러나, 페이로드 스트링 시그니쳐 기반 분류방법은 페이로드의 내용을 검사해야 하므로 개인의 사생활을 침해할 수 있다. 즉, 패킷의 페이로드에는 개인정보가 포함될 수 있으므로 페이로드의 내용을 검사하는 방법은 사생활 침해라는 법적인 문제를 야기시킬 수 있다.

[0005] 또한, 페이로드 스트링 시그니쳐 기반 분류방법은 패킷의 페이로드를 모두 검사해야 하므로 트래픽 분류 시에 빠른 처리성능이 요구된다. 나아가, 현재의 트래픽 분류는 실시간으로 수행되어야 하므로 네트워크의 많은 양의 트래픽을 동시에 처리하기 위해서는 고성능의 하드웨어가 요구된다. 따라서, 페이로드 스트링 시그니쳐 기반의 분류 방법은 Gpbs 단위 이상의 고속 네트워크에 적용되기가 쉽지 않다.

발명의 내용

해결 하고자하는 과제

[0006] 일 양상에 따라, 고속 네트워크에 적용이 가능하고 개인정보 침해의 문제가 없는 네트워크 트래픽 분류 기술을 제안한다.

과제 해결수단

[0007] 일 양상에 따른 트래픽 수집장치는, 네트워크를 통해 패킷을 수집하는 패킷 수집부, 적어도 하나의 패킷으로부터 양방향 플로우를 생성하는 플로우 생성부 및 양방향 플로우에서 페이로드가 존재하는 페이로드 패킷을 대상으로 페이로드 패킷의 전송방향 및 크기 정보를 갖는 페이로드 통계정보를 생성하는 페이로드 통계정보 생성부를 포함한다.

[0008] 한편 다른 양상에 따른 트래픽 분석장치는, 응용 프로그램별로 페이로드 패킷의 전송방향 및 크기 정보가 상이한 페이로드 통계 시그니쳐를 저장하는 페이로드 통계 시그니쳐 저장부 및 트래픽을 수집하는 트래픽 수집장치로부터 전송된 양방향 플로우를 페이로드 통계 시그니쳐를 이용하여 응용 프로그램별로 분류하는 트래픽 분류부를 포함한다.

[0009] 또 다른 양상에 따른 트래픽 분석 시스템은, 네트워크를 통해 패킷을 수집하여 양방향 플로우를 생성하고, 양방향 플로우에서 페이로드가 존재하는 페이로드 패킷을 대상으로 페이로드 패킷의 전송방향 및 크기 정보를 갖는 페이로드 통계정보를 생성하는 트래픽 수집장치 및 트래픽 수집장치로부터 페이로드 통계정보를 포함하는 양방향 플로우를 수신하여, 양방향 플로우를 응용 프로그램별로 페이로드 패킷의 전송방향 및 크기 정보가 상이한 페이로드 통계 시그니쳐를 이용하여 응용 프로그램별로 분류하는 트래픽 분석장치를 포함한다.

[0010] 또 다른 양상에 따른 트래픽 분석방법은, 응용 프로그램별로 페이로드 패킷의 전송방향 및 크기 정보가 상이한 페이로드 통계 시그니쳐 리스트를 구축하는 단계, 네트워크를 통해 수집된 양방향 플로우에서 페이로드가 존재하는 페이로드 패킷의 전송방향 및 크기 정보를 갖는 페이로드 통계정보를 페이로드 통계 시그니쳐 리스트에 해당되는 페이로드 통계 시그니쳐와 비교하는 단계 및 비교 결과에 따라 양방향 플로우를 응용 프로그램별로 분류하는 단계를 포함한다.

효 과

[0011] 일 실시예에 따르면, 네트워크를 통해 수집되는 적어도 하나의 패킷으로부터 생성된 플로우를, 페이로드 패킷의 내용이 아닌 전송방향 및 크기 정보를 이용하여 응용 프로그램별로 분류함에 따라, 개인정보 침해의 문제가 없으며, 고속 네트워크에 적용 가능하다.

[0012] 나아가 플로우의 응용 프로그램별 분류 시에, 플로우에서 시간의 순서대로 발생하는 n개의 패킷을 대상으로 분류 가능하므로 플로우 발생 초기에 분류가 가능하다. 나아가, 플로우의 응용 프로그램별 분류 시에 시티 블록 거리 계산 방식을 이용함에 따라 간단하면서도 빠르게 플로우를 분류할 수 있다.

발명의 실시를 위한 구체적인 내용

[0013] 이하에서는 첨부한 도면을 참조하여 본 발명의 실시예들을 상세히 설명한다. 본 발명을 설명함에 있어 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략할 것이다. 또한, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서

이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.

[0014] 도 1은 본 발명의 일 실시예에 따른 트래픽 분석 시스템(1)의 구성도이다.

[0015] 도 1을 참조하면, 일 실시예에 따른 트래픽 분석 시스템(1)은 트래픽 수집장치(2) 및 트래픽 분석장치(3)를 포함한다.

[0016] 트래픽 수집장치(2)는 네트워크를 통해 패킷을 수집하여 양방향 플로우(two-way flow)를 생성한다. 그리고, 양방향 플로우에서 페이로드가 존재하는 페이로드 패킷(payload packet)을 대상으로 페이로드 패킷의 전송방향 및 크기 정보를 갖는 페이로드 통계정보(payload statistic information)를 생성한다.

[0017] 양방향 플로우는 두 호스트 간 통신연결에 사용되는 단방향 플로우 두 개를 합친 것으로서, 전송방향 및 크기에 관한 정보를 갖는다. 페이로드 패킷은 응용계층 정보를 나타내는 페이로드를 포함하는 패킷으로서, 컨트롤 패킷(control packet)은 페이로드 패킷에 해당되지 않는다. 예를 들면, 전송제어절차 프로토콜(TCP)의 경우 컨트롤 패킷인 연결 요청(synchronization : SYN) 패킷, 연결종료 요청(finish : FIN) 패킷 또는 재연결 종료(Reset : RST) 패킷 등은 페이로드 패킷에 해당되지 않는다.

[0018] 페이로드 통계정보는 페이로드 패킷의 전송방향과 페이로드 크기를 나타내는 페이로드 패킷 벡터(V) 및 페이로드 패킷 벡터(V)를 구성하는 패킷의 갯수(n)의 조합 정보이다. 페이로드 통계정보는 양방향 플로우에서 시간의 순서대로 발생하는 n개의 패킷을 대상으로 표현될 수 있다.

[0019] 페이로드 패킷 벡터(V)의 전송방향은 양의 부호(+) 또는 음의 부호(-)로 표현될 수 있다. 이때 양의 부호(+)는 페이로드 패킷의 전송방향이 클라이언트에서 서버로 향함을 의미한다. 이에 비해 음의 부호(-)는 페이로드 패킷의 전송방향이 서버에서 클라이언트로 향함을 의미한다.

[0020] 호스트가 클라이언트인지 서버인지를 지정하는 방법은 프로토콜의 종류에 따라 다를 수 있다. 일 예로, 호스트 간 패킷 전송시에 전송제어절차 프로토콜(TCP)을 사용하는 경우에는 연결요청(SYN) 패킷을 수신한 호스트를 서버로 지정한다. 다른 예로, 사용자 데이터그램 프로토콜(UDP)을 사용하는 경우에는 첫 패킷을 수신한 호스트를 서버로 지정한다.

[0021] 페이로드 패킷 벡터의 크기는 응용계층 정보를 포함하는 페이로드의 데이터 크기이다. 즉, 페이로드 패킷 벡터의 크기는 패킷에서 전송계층 프로토콜 헤더, 네트워크 계층 프로토콜 헤더 등을 제외한 응용계층만의 데이터 크기를 의미하며, 단위는 바이트(byte)로 표시될 수 있다.

[0022] 페이로드 패킷의 전송방향과 페이로드 크기는 같이 표현되어, 양의 부호(+) 또는 음의 부호(-)를 갖는 값으로 표현된다. 예를 들면, +20은 클라이언트에서 서버로 향하는 페이로드 크기가 20 바이트인 패킷을 의미하며, -100은 서버에서 클라이언트로 향하는 페이로드 패킷 크기가 100 바이트인 패킷을 의미한다.

[0023] 한편, 트래픽 분석장치(3)는 트래픽 수집장치(2)로부터 페이로드 통계정보를 포함하는 양방향 플로우를 수신한다. 그리고, 양방향 플로우를 응용 프로그램별로 페이로드 패킷의 전송방향 및 크기 정보가 상이한 페이로드 통계 시그니처(payload statistic signature)를 이용하여 응용 프로그램별로 분류한다.

[0024] 통계 시그니처(statistic signature)는 패킷의 헤더나 패킷의 캡처 정보로부터 얻을 수 있는 통계적 특징 중 다른 응용 프로그램과 구별될 수 있는 응용 프로그램 고유의 특징을 말한다. 본 발명에서는 양방향 플로우에 해당하는 페이로드 패킷의 전송방향 및 페이로드 크기를 이용하는 페이로드 통계 시그니처를 정의한다. 하나의 페이로드 통계 시그니처는 하나의 응용 프로그램에 매칭된다. 응용 프로그램은 여러 개의 페이로드 통계 시그니처를 가질 수 있다.

[0025] 페이로드 통계 시그니처는 전송계층 프로토콜(p), 페이로드 패킷의 전송방향과 페이로드 크기를 나타내는 페이로드 패킷 벡터(V), 페이로드 패킷 벡터(V)를 구성하는 패킷의 갯수(n), 거리 임계값(d) 및 응용 프로그램 명칭(n)의 조합 정보이다.

[0026] 페이로드 패킷 벡터(V)의 전송방향은 양의 부호(+) 또는 음의 부호(-)로 표현될 수 있다. 양의 부호(+)는 페이로드 패킷의 전송방향이 클라이언트에서 서버로 향함을 의미하고, 음의 부호(-)는 페이로드 패킷의 전송방향이 서버에서 클라이언트로 향함을 의미한다.

[0027] 페이로드 패킷 벡터의 크기는 응용계층 정보를 포함하는 페이로드의 데이터 크기이다. 즉, 페이로드 패킷 벡터의 크기는 패킷에서 전송계층 프로토콜 헤더, 네트워크 계층 프로토콜 헤더 등을 제외한 응용계층만의 데이터 크기를 의미하며, 바이트(byte) 단위로 표시될 수 있다.

[0028] 페이로드 패킷의 방향과 페이로드 패킷 크기는 같이 표현되는데, 양의 부호(+) 또는 음의 부호(-)를 갖는 값으로 표현될 수 있다. 예를 들면, +20은 클라이언트에서 서버로 향하는 페이로드 크기가 20 바이트인 패킷을 의미하며, -100은 서버에서 클라이언트로 향하는 페이로드 패킷 크기가 100 바이트인 패킷을 의미한다.

[0029] 전술한 바에 따르면, 본 발명의 트래픽 분석 시스템(1)에서의 트래픽 분류는 트래픽 분류를 위해 페이로드 패킷의 내용을 검사하는 것이 아니라, 페이로드 패킷의 형식인 전송방향 및 크기 정보를 이용하므로, 개인정보 침해의 문제가 없으며, 고속 네트워크에 적용 가능하다.

[0030] 한편, 전술한 트래픽 수집장치(2) 및 트래픽 분석장치(3)에서의 모든 과정은 실시간으로 수행된다. 즉, 네트워크를 통해 모든 패킷을 수집하는 과정, 양방향 플로우를 생성하는 과정, 양방향 플로우 내에서 페이로드 통계정보를 추출하는 과정 및 페이로드 통계 시그니쳐와의 비교를 통하여 양방향 플로우를 응용별로 분류하는 과정이 실시간으로 수행된다.

[0031] 도 2는 본 발명의 일 실시예에 따른 트래픽 수집장치(2)의 구성도이다.

[0032] 도 2를 참조하면, 트래픽 수집장치(2)는 패킷 수집부(20), 플로우 생성부(22) 및 페이로드 통계정보 생성부(24)를 포함한다.

[0033] 패킷 수집부(20)는 네트워크를 통해 패킷을 수집한다. 이때 패킷 수집부(20)는 인터넷 망에서의 라우터 또는 스위치를 통해 모든 패킷을 수집할 수 있다.

[0034] 일 실시예에 따르면, 패킷 수집부(20)는 인터넷 망에서 고속의 물리적인 회선을 태핑하거나 스위치 또는 라우터의 포트 미러링(port mirroring) 기능을 이용하여 실시간으로 모든 패킷을 수집하여 이를 플로우 생성부(22)에 제공할 수 있다. 나아가, 네트워크와 연결된 인터넷 회선이 여러 개인 경우, 패킷 수집부(20)가 여러 곳에서의 패킷을 수집하여 이를 한 곳으로 병합하는 작업이 추가적으로 필요하다.

[0035] 한편, 플로우 생성부(22)는 패킷 수집부(20)를 통해 수집된 적어도 하나의 패킷으로부터 양방향 플로우를 생성한다. 이때, 플로우 생성부(22)는 통신 양단의 IP 주소, 포트 번호 및 통신에 사용하는 전송계층 프로토콜을 포함하는 5가지 정보(5-tuple)를 기준으로 적어도 하나의 패킷을 그룹화하여 양방향 플로우를 생성할 수 있다.

[0036] 플로우는 소스 IP(source IP), 목적지 IP(destination IP), 소스 포트(source port), 목적지 포트(destination port) 및 전송계층 프로토콜(transport layer protocol) 중 적어도 일부가 동일한 패킷들의 집합이다. 이는 하나의 통신 연결에서 한쪽 방향으로 전송되는 모든 패킷들의 집합으로 구성된다. 이러한 단방향 플로우(one-way flow)는 하나의 통신연결에서 두 개가 생성된다.

[0037] 본 발명에서는 두 호스트 간 하나의 통신연결에 사용된 모든 패킷들의 집합을 양방향 플로우(two-way flow)로 정의하고 플로우 레코드(flow record)를 생성한다. 왜냐하면 페이로드 통계 시그니쳐는 하나의 통신연결에서 패킷의 방향과 페이로드 패킷 크기가 요구되기 때문이다. 양방향 플로우 레코드는 단방향 플로우 두 개를 합친 형태로써, 두 호스트의 IP와 포트 번호, 그리고 전송계층 프로토콜을 기본적으로 저장하며, 두 방향에 따라 패킷의 수, 바이트의 총 크기 등 다양한 정보들이 각각 저장될 수 있다.

[0038] 한편, 페이로드 통계정보 생성부(24)는 양방향 플로우에서 페이로드가 존재하는 페이로드 패킷을 대상으로 페이로드 패킷의 전송방향 및 크기 정보를 갖는 페이로드 통계정보를 생성한다. 이때 페이로드 통계정보 생성부(24)는 최대 n개의 패킷을 대상으로 각 플로우의 페이로드 통계정보를 생성할 수 있다. n개의 패킷은 먼저 수집된 순서대로 선정될 수 있다. n의 값은 네트워크의 상황에 따라서 달리 결정될 수 있는데, 예를 들면 n은 4~6 사이의 값일 수 있다.

[0039] 한편, 트래픽 수집장치(2)는 플로우 레코드 저장부(26)를 더 포함하는데, 플로우 레코드 저장부(26)는 페이로드 통계정보와 플로우 식별자 및 플로우 기본정보가 포함된 플로우 레코드를 생성하여 저장한다. 전술한 플로우 레코드에 저장되는 페이로드 통계정보(F)는 아래의 수학식 1과 같다.

수학식 1

[0040] $F = \{n, V\}$

[0041] 수학식 1에 따르면, 플로우 레코드에 저장되는 페이로드 통계정보(F)의 구성요소는 n 및 V이다. n은 페이로드 패킷 벡터(V)를 구성하는 패킷의 개수이다. V는 n개의 구성요소인 $\{F_0, F_1, F_2, \cdots, F_{n-1}\}$의 페이로드 패킷 벡터이고, F_k는 정수(k=1, 2, …, n-1)이다.

[0042] 도 3은 본 발명의 일 실시예에 따른 트래픽 분석장치(3)의 구성도이다.

[0043] 도 3을 참조하면, 일 실시예에 따른 트래픽 분석장치(3)는 페이로드 통계 시그니쳐 저장부(30) 및 트래픽 분류부(32)를 포함한다.

[0044] 페이로드 통계 시그니쳐 저장부(30)는 응용 프로그램별로 페이로드 패킷의 전송방향 및 크기 정보가 상이한 페이로드 통계 시그니쳐를 저장한다. 통계 시그니쳐는 패킷의 헤더나 패킷의 캡쳐 정보로부터 얻을 수 있는 통계적 특징 중 다른 응용 프로그램과 구별될 수 있는 응용 프로그램 고유의 특징을 말한다. 그 예로써는 패킷 크기(packet size) 분포, 패킷 수집시간(inter-arrival time) 분포, TCP의 경우 윈도우 크기(window size) 분포 등이 있다. 본 발명에서는 응용 프로그램들이 각각 자신만이 가지는 패킷 크기 분포를 이용한 페이로드 통계 시그니쳐를 사용한다.

[0045] 본 발명에 따른 페이로드 통계 시그니쳐(S)의 구성요소는 아래의 수학식 2와 같다.

수학식 2

[0046] $S = \{p, n, V, d, A\}$

[0047] 즉, 페이로드 통계 시그니쳐(S)는 전송계층 프로토콜(p), 페이로드 패킷의 전송방향과 페이로드 크기를 나타내는 페이로드 패킷 벡터(V), 페이로드 패킷 벡터를 구성하는 패킷의 갯수(n), 거리 임계값(d) 및 응용 프로그램 명칭(A)일 수 있다. V는 n개의 구성요소인 $\{S_0, S_1, S_2, \cdots, S_{n-1}\}$의 페이로드 패킷 벡터이고 S_k는 정수(k=1, 2, …, n-1)이다. 전송계층 프로토콜(p)은 TCP 또는 UDP일 수 있다.

[0048] 응용 프로그램 명칭(A)은 p, n, V의 값을 가지는 응용프로그램의 이름이다. 페이로드 패킷 벡터(V)는 페이로드 패킷 n개의 크기와 방향을 나타내는 n개의 정수로 구성된다. 각 정수에서 부호는 패킷의 방향을, 절대값은 패킷의 페이로드 크기를 나타낸다. 즉 양수는 클라이언트에서 서버로 향하는 패킷이며, 음수는 서버에서 클라이언트로 향하는 패킷이다. 페이로드 패킷 벡터 V는 n차원의 정수를 갖는 다.

[0049] 거리 임계값(d)은 플로우를 분류하기 위해 사용하는 값으로 양의 정수로 표현된다. 거리 임계값(d)은 플로우 레코드의 페이로드 통계정보와 페이로드 통계 시그니쳐의 비교에서 플로우가 어느 응용 프로그램에 속하는지를 판단하는 기준으로 사용된다.

[0050] 한편, 트래픽 분류부(32)는 트래픽을 수집하는 트래픽 수집장치(2)로부터 전송된 양방향 플로우를 페이로드 통계 시그니쳐를 이용하여 응용 프로그램별로 분류한다. 본 발명의 응용 프로그램 분류 방식은 하나의 양방향 플로우마다 모든 통계 시그니쳐를 검사하여 조건에 일치하는 페이로드 통계 시그니쳐가 발견되면 양방향 플로우를 발견된 페이로드 통계 시그니쳐에 해당되는 응용 프로그램으로 분류하는 방식이다.

[0051] 일 실시예에 따르면, 플로우의 응용 프로그램별 분류 방식은 페이로드 통계정보의 페이로드 패킷 벡터 및 페이로드 통계 시그니쳐의 페이로드 패킷 백터 간의 거리를 이용하여 분류한다. 이때, 페이로드 통계 시그니쳐의 거리 임계값(distance threshold)인 d 이하의 거리에 양방향 플로우가 존재하면 해당 플로우는 해당 페이로드 통계 시그니쳐가 가리키는 응용 프로그램에 해당된다.

[0052] 예를 들면, 페이로드 통계 시그니쳐의 페이로드 패킷 백터가 (+30, -100, -200)이고 거리 임계값 d = 10이라고 가정한다. 그런데, 수집된 플로우의 페이로드 패킷이 총 3개이고, 첫 번째 패킷은 +31, 두 번째 패킷은 -98, 세 번째 패킷은 -200 이라는 방향과 크기의 값을 가진다고 가정하면, 페이로드 통계정보의 페이로드 패킷 벡터는 (+31, -98, -200)이다. 이때, 전술한 두 벡터 간의 거리를 측정한 후, 측정 결과가 10 이하의 거리가 나오면 해당 플로우는 이 페이로드 통계 시그니쳐가 가리키는 응용프로그램에 해당된다.

[0053] 본 발명에 따르면, 두 벡터 간의 거리 측정은 시티 블록 거리(city-block distance) 계산 방식을 통해 수행될 수 있다. 전술한 예에서 시티 블록 거리 계산 방식을 이용할 경우에 전술한 두 벡터 간의 거리는 3 [|{+31-(+30)}| + |{-98-(-100)}| + |{-200-(-200)}|]이다. 따라서 d인 10보다 값이 작으므로 해당 플로우는 페이로드 통계 시그니쳐가 가리키는 응용프로그램에 해당된다. 시티 블록 거리 계산 방식에 대한 상세한 설명은 도 6에서 후술한다.

[0054] 도 4는 본 발명의 일 실시예에 따른 페이로드 통계정보를 포함하는 플로우 레코드를 도시한 구조도이다.

[0055] 도 4를 참조하면, 플로우 레코드는 크게 3부분으로 분류될 수 있다. 즉, 플로우 레코드는 플로우 구분에 사용되는 5-튜플(tuple) 정보에 해당되는 플로우 식별자(40), 플로우에 대한 기본정보(42) 및 페이로드 통계정보(44)로 분류될 수 있다.

[0056] 플로우 식별자(40)는 클라이언트 IP 주소(400), 서버 IP 주소(402), 클라이언트 포트(404), 서버 포트(406), 전송계층 프로토콜(408)을 포함한다. 플로우 기본정보(42)는 총 패킷 수(420), 패킷의 총 크기(422), 플로우 시작시간(424) 및 플로우 종료시간(426)을 포함한다. 한편, 페이로드 통계정보(44)는 하나의 플로우마다 먼저 수집된 최대 n개의 페이로드 패킷의 크기에 방향성을 추가하여 벡터의 형태로 저장될 수 있다. 페이로드 통계정보(44)는 도 1 및 도 2에서 전술하였으므로 상세한 설명은 생략한다.

[0057] 도 5는 본 발명의 일 실시예에 따른 트래픽 분석방법을 도시한 흐름도이다.

[0058] 도 5를 참조하면, 본 발명의 트래픽 분석장치는 응용 프로그램별로 페이로드 패킷의 전송방향 및 크기 정보가 상이한 페이로드 통계 시그니처 리스트를 구축하고 이를 초기화(reset)한다(500). 페이로드 통계 시그니처(S)는 전송계층 프로토콜(p), 페이로드 패킷의 전송방향과 페이로드 크기를 나타내는 페이로드 패킷 벡터(V), 페이로드 패킷 벡터(V)를 구성하는 패킷의 갯수(n), 거리 임계값(d) 및 응용 프로그램 명칭(n)의 조합 형태이다.

[0059] 이어서, 트래픽 분석장치는 페이로드 통계정보(F)를 페이로드 통계 시그니처 리스트에 해당되는 페이로드 통계 시그니처(S)와 비교한다(506,508,510). 페이로드 통계정보(F)는 네트워크를 통해 수집된 양방향 플로우에서 페이로드가 존재하는 페이로드 패킷의 전송방향 및 크기 정보를 갖는다. 이때 수집된 양방향 플로우가 존재하지 않으면(514) 패킷 분석 과정이 종료된다.

[0060] 구체적으로, 트래픽 분석장치는 페이로드 통계정보(F)의 전송계층 프로토콜(F(p)) 및 페이로드 통계 시그니처(S)의 전송계층 프로토콜(S(p))을 비교한다(506).

[0061] 비교 결과, 전송계층 프로토콜이 일치(F(p)=S(p))하면 페이로드 통계정보(F)의 페이로드 패킷 벡터를 구성하는 패킷의 갯수(F(n)) 및 페이로드 통계 시그니처(S)의 페이로드 패킷 벡터를 구성하는 패킷의 갯수(S(n))를 비교한다(508). 이에 비하여 전송계층 프로토콜이 일치하지 않으면(F(p)≠S(p)) 페이로드 통계 시그니처 리스트의 다른 페이로드 통계 시그니처를 선택하여 비교한다.

[0062] 패킷의 갯수가 일치(F(n)=S(n))하면, 트래픽 분석장치는 페이로드 통계정보(F)의 페이로드 패킷 벡터 및 페이로드 통계 시그니처(S)의 페이로드 패킷 벡터 간의 거리(D(F,S))가 페이로드 통계 시그니처(S)의 거리 임계값(S(d)) 내에 포함되는지를 비교한다(510). 이에 비하여 패킷의 갯수가 일치하지 않으면(F(n)≠S(n)) 페이로드 통계 시그니처 리스트의 다른 페이로드 통계 시그니처(S)를 선택하여 비교한다.

[0063] 페이로드 통계정보(F)의 페이로드 패킷 벡터 및 페이로드 통계 시그니처(S)의 페이로드 패킷 벡터 간의 거리가 페이로드 통계 시그니처(S)의 거리 임계값 내에 포함(D(F,S)<S(d))되면, 트래픽 분석장치는 양방향 플로우를 페이로드 통계 시그니처(S)에 해당되는 응용 프로그램으로 분류한다(512).

[0064] 플로우 분석장치는 페이로드 통계 시그니처(S)의 거리 임계값 내에 포함되는지를 비교하는 단계(510)에서, 시티 블록 거리(city-block distance) 계산 방식을 이용할 수 있다. 시티 블록 거리 계산 방식은 아래의 수학식 3과 같다.

수학식 3

[0065]
$$d(F,S)=\sum_{i=0}^{n-1}|F(s_i)-S(s_i)|$$

[0066] 수학식 3에 의하면, F 및 S는 n차원의 벡터이며, $F(S_i)$, $S(S_i)$는 각각 F와 S의 페이로드 패킷 벡터를 구성하는 각 원소들이 된다. F, S 두 벡터 간의 거리는 유클리드 거리(Euclidean distance) 계산방식을 이용하여 계산될 수 있지만, 본 발명에서는 수학식 3에 기재된 바와 같이 대용량 트래픽의 빠른 처리를 위해 계산이 간단한 시티 블록 거리 계산 방식을 이용한다. 거리 계산은 동일한 차원의 벡터끼리에서만 수행된다.

[0067] 거리 D(F, S)가 페이로드 통계 시그니처(S)의 거리 임계값보다 작으면 해당되는 양방향 플로우는 페이로드 통계 시그니처(S)가 속한 응용프로그램(S(A))에 해당된다(510). 이에 비하여 거리 D(F, S)가 페이로드 통계 시그니처(S)의 거리 임계값보다 크면 패킷 분석장치는 페이로드 통계 시그니처 리스트에서 다른 페이로드 통계 시그니

처를 선택하여 전술한 과정을 반복한다.

[0068] 만약 페이로드 통계 시그니처 리스트에서 더 이상 비교할 페이로드 통계 시그니처(S)가 존재하지 않는다면(516) 그 플로우의 응용프로그램은 주어진 페이로드 통계 시그니처(S)로 결정할 수 없다(518). 전술한 과정은 모든 플로우에 대하여 독립적으로 수행된다. 모든 플로우에 대하여 응용 프로그램을 찾는 작업을 마치면 본 분석과정이 완료된다.

[0069] 이제까지 본 발명에 대하여 그 실시예들을 중심으로 살펴보았다. 본 발명이 속하는 기술분야에서 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.

도면의 간단한 설명

[0070] 도 1은 본 발명의 일 실시예에 따른 트래픽 분석 시스템의 구성도,

[0071] 도 2는 본 발명의 일 실시예에 따른 트래픽 수집장치의 구성도,

[0072] 도 3은 본 발명의 일 실시예에 따른 트래픽 분석장치의 구성도,

[0073] 도 4는 본 발명의 일 실시예에 따른 페이로드 통계정보를 포함하는 플로우 레코드를 도시한 구조도,

[0074] 도 5는 본 발명의 일 실시예에 따른 트래픽 분석방법을 도시한 흐름도이다.

[0075] <도면의 주요부분에 대한 부호의 설명>

[0076] 1 : 트래픽 분석 시스템 2 : 트래픽 수집장치

[0077] 3 : 트래픽 분석장치 20 : 패킷 수집부

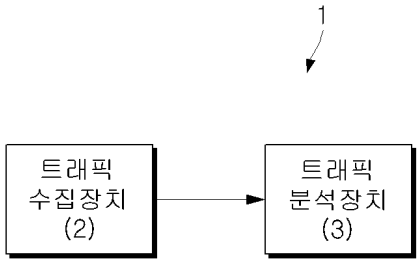
[0078] 22 : 플로우 생성부 24 : 페이로드 통계정보 생성부

[0079] 26 : 플로우 레코드 저장부 30 : 페이로드 통계 시그니처 저장부

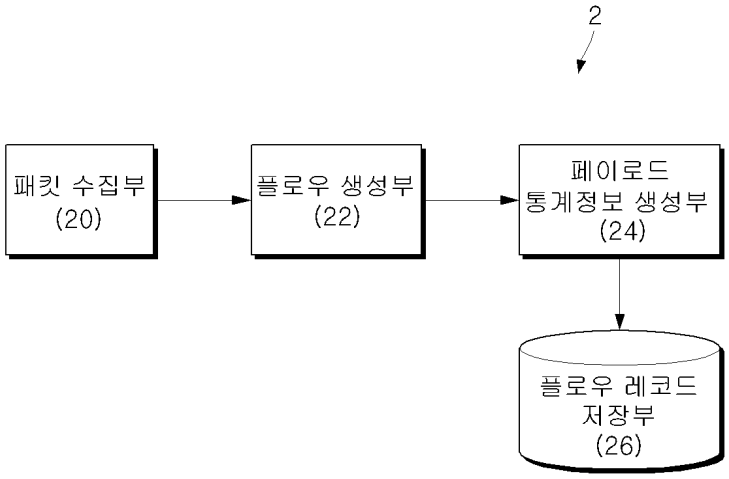
[0080] 32 : 트래픽 분류부

도면

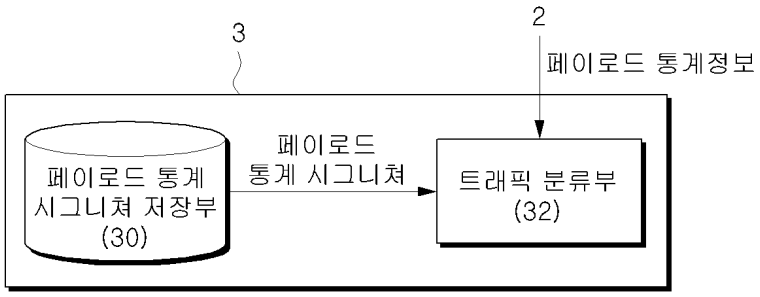
도면1



도면2



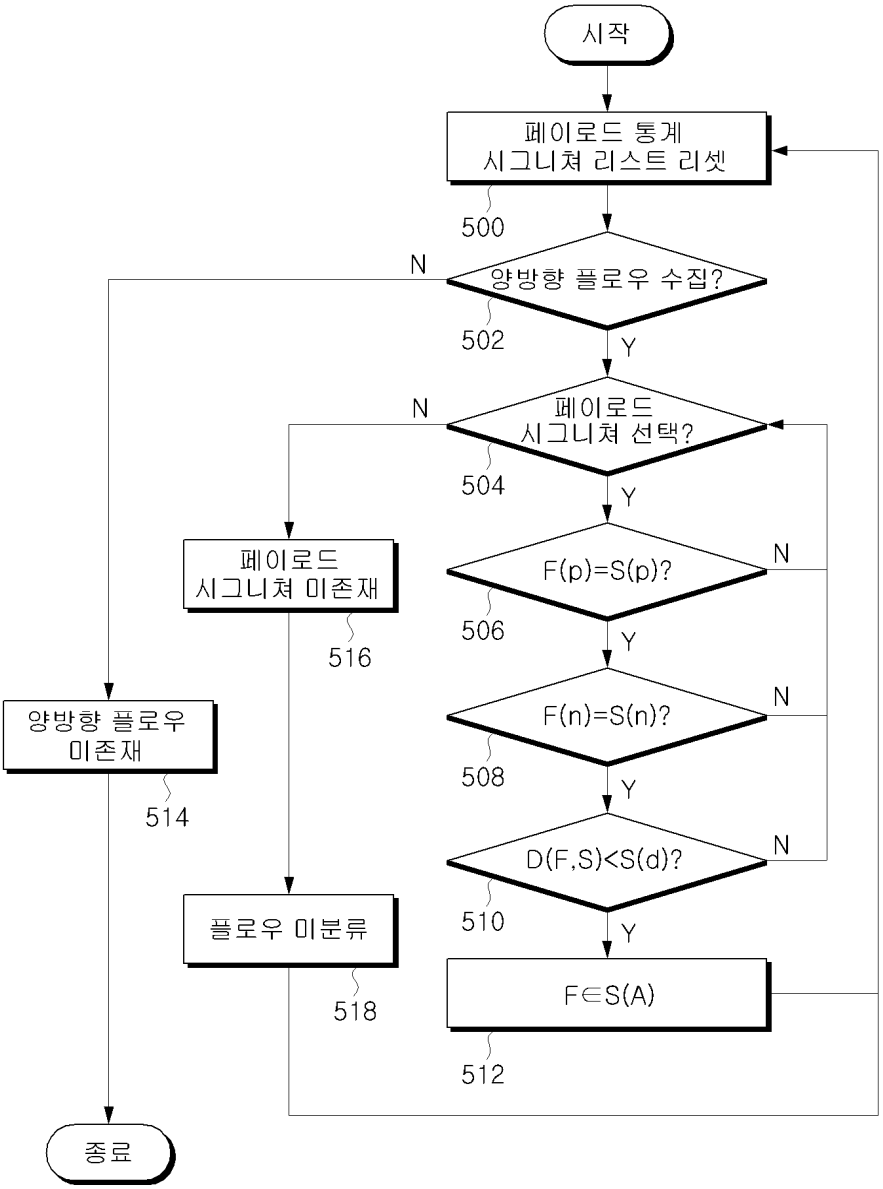
도면3



도면4

클라이언트 IP 주소(400)	클라이언트 포트(404)	플로우 식별자 (40)
서버 IP 주소(402)	서버 포트(406)	
전송 계층 프로토콜(408)		
총 패킷 수(420)	총 크기(422)	플로우 기본정보 (42)
플로우 시작시간(424)	플로우 종료시간(426)	
페이로드 통계정보(44)		

도면5


시작
페이로드 통계
시그니쳐 리스트 리셋
500
양방향 플로우 수집?
502
N
Y
페이로드
시그니쳐 선택?
504
N
Y
F(p)=S(p)?
506
N
Y
F(n)=S(n)?
508
N
Y
D(F,S)<S(d)?
510
N
Y
F∈S(A)
512
페이로드
시그니쳐 미존재
516
양방향 플로우
미존재
514
플로우 미분류
518
종료