(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0056889
(43) 공개일자 2021년05월20일

(51) 국제특허분류(Int. Cl.)
H04J 3/06 (2006.01) *H04J 3/14* (2006.01)
(52) CPC특허분류
H04J 3/0658 (2013.01)
H04J 3/14 (2013.01)
(21) 출원번호 10-2020-0080412
(22) 출원일자 2020년06월30일
심사청구일자 2020년06월30일
(30) 우선권주장
1020190143568 2019년11월11일 대한민국(KR)

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
백상헌
서울특별시 서초구 고무래로 35, 109동 1004호
천림
대한민국, 16961, 경기도 용인시 기흥구 기흥로 116번길 60, 507동 502호
임용재
대한민국, 12582, 경기도 양평군 양서면 대심길 52-3
(74) 대리인
김홍석

전체 청구항 수 : 총 7 항

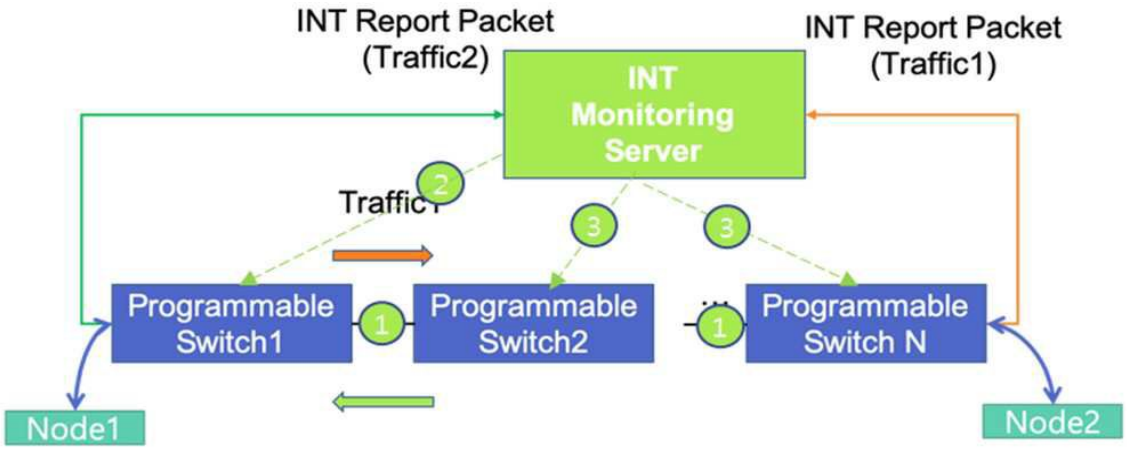
(54) 발명의 명칭 **인-밴드 네트워크 모니터링 방법 및 인-밴드 네트워크 모니터링에서 지연 예측 방법**

(57) 요 약

복수의 스위치들과 모니터 서버를 포함하는 네트워크에서 INT(In-band Network Telemetry) 모니터링 기법을 이용한 로컬 시간 편차 계산 방법이 개시된다. 상기 로컬 시간 편차 계산 방법은 제1 스위치 및 제2 스위치 사이에서 전달되는 패킷에 상기 제1 및 제2 스위치들 각각의 상태 정보가 메타데이터 형태로 추가하는 단계, 상기 제1 스

(뒷면에 계속)

대 표 도 - 도4



위치에서 상기 제2 스위치로 향하는 트래픽에서 상기 제1 스위치에서의 패킷 통과(egress) 타임스탬프(timestamp) 정보(t0)를 획득하는 단계, 상기 제1 스위치에서 상기 제2 스위치로 향하는 트래픽에서 상기 제2 스위치에서 패킷 진입(ingress) 타임스탬프 정보(t1)를 획득하는 단계, 상기 제2 스위치에서 상기 제1 스위치로 향하는 트래픽에서 상기 제2 스위치에서의 패킷 통과 타임스탬프 정보(t2)를 획득하는 단계, 상기 제2 스위치에서 상기 제1 스위치로 향하는 트래픽에서 상기 제1 스위치에서의 패킷 진입 타임스탬프 정보(t3)를 획득하는 단계, 및 상기 타임스탬프 정보들(t0, t1, t2, t3)을 이용하여 상기 제1 및 제2 스위치들 간 로컬 시간 편차(Θ)를 구하는 단계를 포함한다.

이 발명을 지원한 국가연구개발사업

과제고유번호	1711080849
과제번호	2017-0-00195-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	방송통신산업기술개발(R&D)
연구과제명	멀티 서비스를 지원하는 프로그래머블 스위치 제어 기술 개발
기 여 율	1/1
과제수행기관명	고려대학교산학협력단
연구기간	2019.01.01 ~ 2019.12.31

명 세 서

청구범위

청구항 1

복수의 스위치들과 모니터 서버를 포함하는 네트워크에서 INT(In-band Network Telemetry) 모니터링 기법을 이용한 로컬 시간 편차 계산 방법에 있어서,

제1 스위치 및 제2 스위치 사이에서 전달되는 패킷에 상기 제1 및 제2 스위치들 각각의 상태 정보가 메타데이터 형태로 추가하는 단계;

상기 제1 스위치에서 상기 제2 스위치로 향하는 트래픽에서 상기 제1 스위치에서의 패킷 통과(egress) 타임스탬프(timestamp) 정보(t0)를 획득하는 단계;

상기 제1 스위치에서 상기 제2 스위치로 향하는 트래픽에서 상기 제2 스위치에서 패킷 진입(ingress) 타임스탬프 정보(t1)를 획득하는 단계;

상기 제2 스위치에서 상기 제1 스위치로 향하는 트래픽에서 상기 제2 스위치에서의 패킷 통과 타임스탬프 정보(t2)를 획득하는 단계;

상기 제2 스위치에서 상기 제1 스위치로 향하는 트래픽에서 상기 제1 스위치에서의 패킷 진입 타임스탬프 정보(t3)를 획득하는 단계; 및

상기 타임스탬프 정보들(t0, t1, t2, t3)을 이용하여 하기 식에 따라 상기 제1 및 제2 스위치들 간 로컬 시간 편차(Θ)를 구하는 단계를 포함하는 로컬 시간 편차 계산 방법.

$$\theta = \frac{(t1 - t0) + (t2 - t3)}{2}$$

청구항 2

제1항에 있어서,

상기 모니터 서버와 상기 제2 스위치 사이의 로컬 시간 편차(CT)를 계산하는 단계; 및

상기 로컬 시간 편차(CT)와 상기 로컬 시간 편차(Θ)를 더하여 상기 모니터 서버와 상기 제1 스위치 간의 로컬 시간 편차를 구하는 단계를 더 포함하는,

로컬 시간 편차 계산 방법.

청구항 3

제1항에 있어서,

상기 네트워크는 SDN(Software Defined Network)인,

로컬 시간 편차 계산 방법.

청구항 4

제2항에 있어서,

상기 로컬 시간 편차(CT)를 계산하는 단계는,

상기 모니터 서버가 상기 제2 스위치로 시간 동기화 패킷을 전달하는 단계;

상기 모니터 서버가 상기 제2 스위치로부터 상기 제2 스위치가 상기 시간 동기화 패킷을 전달받은 시간(t4)과 상기 전달받은 시간 동기화 패킷을 상기 모니터 서버로 반환할 때 통과한 시간(t5)이 추가된 시간 동기화 패킷을 수신하는 단계;

상기 모니터 서버가 상기 시간 동기화 패킷을 전달한 시간, 상기 제2 스위치가 상기 시간 동기화 패킷을 전달받은 시간(t4), 상기 모니터 서버로 반환할 때 통과한 시간(t5), 및 상기 시간 동기화 패킷을 수신한 시간을 이용하여 상기 모니터 서버와 제2 스위치 사이 로컬 시간 편차(CT)를 구하는 단계를 포함하는,

로컬 시간 편차 계산 방법.

청구항 5

제1항에 있어서,

마지막 수집 시간(T(last))에 계산된 상기 제1 스위치와 상기 제2 스위치 간 로컬 시간 차이 G(last)와, 상기 마지막 수집 시간(T(last)) 전의 시점(T(old))에 계산된 상기 제1 스위치와 상기 제2 스위치 간 로컬 시간 차이 G(old)를 이용하여 현재의 시간(T(now))에서의 상기 제1 스위치와 상기 제2 스위치 간의 로컬 시간 차이(G(now))를 추정하는 단계를 더 포함하는,

로컬 시간 편차 계산 방법.

청구항 6

제5항에 있어서,

상기 로컬 시간 차이(G(now))를 구하는 단계는 하기 식을 이용하여 상기 로컬 시간 차이를 추정하는,

로컬 시간 편차 계산 방법.

$$G(now) = \frac{\{G(last) - G(old)\} \times \{T(now) - T(last)\}}{T(last) - T(old)} + G(last)$$

청구항 7

복수의 스위치들과 모니터 서버를 포함하는 네트워크에서 INT(In-band Network Telemetry) 모니터링 기법을 이용한 로컬 시간 편차 계산 방법에 있어서,

상기 모니터 서버가 엣지 스위치로 시간 동기화 패킷을 전달하고, 전달 시간(t0)을 기록하는 단계;

상기 모니터 서버가 상기 엣지 스위치로부터 상기 엣지 스위치가 상기 시간 동기화 패킷을 전달받은 시간(t1)과 상기 전달받은 시간 동기화 패킷을 상기 모니터 서버로 반환할 때 통과한 시간(t2)이 추가된 시간 동기화 패킷을 수신하는 단계;

상기 모니터 서버가 상기 엣지 스위치에서 반환 패킷을 접수하는 시간(t3)을 기록하는 단계; 및

상기 시간들(t0, t1, t2, t3)을 이용하여 하기 식에 따라 상기 모니터 서버 와 엣지 스위치 사이 로컬 시간 편차(Θ)를 구하는 단계를 포함하는 로컬 시간 편차 계산 방법.

$$\theta = \frac{(t1 - t0) + (t2 - t3)}{2}$$

발명의 설명

기 술 분 야

[0001] 본 발명은 네트워크 모니터링 방법에 관한 것으로, 보다 상세하게는 인-밴드 네트워크의 모니터링 방법 및 인-

밴드 네트워크 모니터링에서 지연 예측 방법에 관한 것이다.

배 경 기 술

[0002] 데이터 평면의 정보를 실시간으로 모니터링하기 위한 수단으로 대중적인 모니터링 방식은 주기적으로 스위치 모니터링 정보 패킷을 생성하여 모니터링 서버에 전달하는 방식이다. 한편, 최근 프로그래머블 스위치 기술이 개발되고 활용되면서 INT(in-band network telemetry)-모니터링 기법이 소개되었다. 이는 데이터 평면에 배포되어 있는 각 스위치에서 패킷을 경유하는 순간, 현재 상태 정보를 경유하는 패킷에 추가 삽입하여 스위치 상태 정보들이 추가된 패킷 정보를 모니터링 서버에 전달하는 방식을 사용한다. INT-모니터링 기법은 기존의 기타 모니터링 방식과 달리, 특정 트래픽의 전달 경로를 실시간으로 파악할 수 있으며 전달 과정에서의 스위치의 상태를 실시간으로 모니터링 하는 장점이 있다.

[0003] INT-모니터링 방법에서, 수집하는 스위치의 정보는 패킷이 진입할 때와 통과할 때의 시간을 포함한다. 하지만, 해당 시간 정보는 각 스위치의 로컬 시간 정보를 참조하기 때문에 각 스위치 간의 시간 정보가 동기화되지 않을 경우, 모니터 서버에서 수집된 시간 정보들을 취합하고 분석하기 어렵다. 따라서 시간 정보를 정확히 활용하기 위해 각 스위치 간의 시간 동기화 기법이 필요하다.

[0004] 기존의 시간 동기화 기법은 스위치 간에 시간 동기화 트래픽을 별도로 필요로 하고, 각 스위치 간의 클록 밀림 현상 차이로 인한 오차범위를 막기 위해 지속적으로 동기화 메시지를 주고 받는다. 따라서, 데이터 평면에 트래픽 과부하가 발생할 수 있다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 대한민국 공개특허 제2016-0081312

발명의 내용

해결하려는 과제

[0006] 본원 발명이 해결하고자 하는 과제는 INT-모니터링을 사용하는 데이터 평면에서 별도의 시간 동기화 프로토콜을 사용하지 않고, INT-모니터링 과정에서 수집된 각 스위치 시간 정보들을 활용하여 스위치 간 시간을 동기화하는 방법을 제공하는 것이다.

과제의 해결 수단

[0007] 본 발명의 일 실시예에 따른 로컬 시간 편차 계산 방법은 복수의 스위치들과 모니터 서버를 포함하는 네트워크에서 INT(In-band Network Telemetry) 모니터링 기법을 이용한 로컬 시간 편차 계산 방법으로써, 제1 스위치 및 제2 스위치 사이에서 전달되는 패킷에 상기 제1 및 제2 스위치들 각각의 상태 정보가 메타데이터 형태로 추가하는 단계, 상기 제1 스위치에서 상기 제2 스위치로 향하는 트래픽에서 상기 제1 스위치에서의 패킷 통과(egress) 타임스탬프(timestamp) 정보(t0)를 획득하는 단계, 상기 제1 스위치에서 상기 제2 스위치로 향하는 트래픽에서 상기 제2 스위치에서 패킷 진입(ingress) 타임스탬프 정보(t1)를 획득하는 단계, 상기 제2 스위치에서 상기 제1 스위치로 향하는 트래픽에서 상기 제2 스위치에서의 패킷 통과 타임스탬프 정보(t2)를 획득하는 단계, 상기 제2 스위치에서 상기 제1 스위치로 향하는 트래픽에서 상기 제1 스위치에서의 패킷 진입 타임스탬프 정보(t3)를 획득하는 단계, 및 상기 타임스탬프 정보들(t0, t1, t2, t3)을 이용하여 하기 식에 따라 상기 제1 및 제2 스위치들 간 로컬 시간 편차(Θ)를 구하는 단계를 포함한다.

[0008] $$\theta = \frac{(t1 - t0) + (t2 - t3)}{2}$$

[0009] 또한, 상기 로컬 시간 편차 계산 방법은 상기 모니터 서버와 상기 제2 스위치 사이의 로컬 시간 편차(CT)를 계산하는 단계, 및 상기 로컬 시간 편차(CT)와 상기 로컬 시간 편차(Θ)를 더하여 상기 모니터 서버와 상기 제1 스위치 간의 로컬 시간 편차를 구하는 단계를 더 포함할 수 있다.

[0010] 또한, 상기 네트워크는 SDN(Software Defined Network)일 수 있다.

[0011] 또한, 상기 로컬 시간 편차(CT)를 계산하는 단계는, 상기 모니터 서버가 상기 제2 스위치로 시간 동기화 패킷을 전달하는 단계, 상기 모니터 서버가 상기 제2 스위치로부터 상기 제2 스위치가 상기 시간 동기화 패킷을 전달받은 시간(t4)과 상기 전달받은 시간 동기화 패킷을 상기 모니터 서버로 반환할 때 통과한 시간(t5)이 추가된 시간 동기화 패킷을 수신하는 단계, 상기 모니터 서버가 상기 시간 동기화 패킷을 전달한 시간, 상기 제2 스위치가 상기 시간 동기화 패킷을 전달받은 시간(t4), 상기 모니터 서버로 반환할 때 통과한 시간(t5), 및 상기 시간 동기화 패킷을 수신한 시간을 이용하여 상기 모니터 서버와 제2 스위치 사이 로컬 시간 편차(CT)를 구하는 단계를 포함할 수 있다.

[0012] 또한, 상기 로컬 시간 편차 계산 방법은 마지막 수집 시간(T(last))에 계산된 상기 제1 스위치와 상기 제2 스위치 간 로컬 시간 차이 G(last)와, 상기 마지막 수집 시간(T(last)) 전의 시점(T(old))에 계산된 상기 제1 스위치와 상기 제2 스위치 간 로컬 시간 차이 G(old)를 이용하여 현재의 시간(T(now))에서의 상기 제1 스위치와 상기 제2 스위치 간의 로컬 시간 차이(G(now))를 추정하는 단계를 더 포함할 수 있다.

[0013] 또한, 상기 로컬 시간 차이(G(now))를 구하는 단계는 하기 식을 이용하여 상기 로컬 시간 차이를 추정할 수 있다.

[0014] $$G(now) = \frac{\{G(last) - G(old)\} \times \{T(now) - T(last)\}}{T(last) - T(old)} + G(last)$$

[0015] 본 발명의 다른 실시예에 따른 로컬 시간 편차 계산 방법은 복수의 스위치들과 모니터 서버를 포함하는 네트워크에서 INT(In-band Network Telemetry) 모니터링 기법을 이용한 로컬 시간 편차 계산 방법으로써, 상기 모니터 서버가 엣지 스위치로 시간 동기화 패킷을 전달하고, 전달 시간(t0)을 기록하는 단계, 상기 모니터 서버가 상기 엣지 스위치로부터 상기 엣지 스위치가 상기 시간 동기화 패킷을 전달받은 시간(t1)과 상기 전달받은 시간 동기화 패킷을 상기 모니터 서버로 반환할 때 통과한 시간(t2)이 추가된 시간 동기화 패킷을 수신하는 단계, 상기 모니터 서버가 상기 엣지 스위치에서 반환 패킷을 접수하는 시간(t3)을 기록하는 단계, 및 상기 시간들(t0, t1, t2, t3)을 이용하여 하기 식에 따라 상기 모니터 서버 와 엣지 스위치 사이 로컬 시간 편차(Θ)를 구하는 단계를 포함한다.

[0016] $$\theta = \frac{(t1 - t0) + (t2 - t3)}{2}$$

발명의 효과

[0017] 본 발명의 실시 예들에 따르면, 스위치 사이에 별도의 시간 동기화 메시지를 생성하지 않고, INT-모니터링 서버에서 수집한 타임스탬프(timestamp) 정보를 활용하여 각 스위치 사이의 시간 오차범위를 계산할 수 있다. 따라서, 트래픽이 흐르는 스위치에서만 작동하기 때문에 데이터 평면의 소모 대역폭을 감소시킬 수 있다. 또한, 스위치에서 의도적으로 프로빙(probing) 패킷 생성을 할 필요가 없어서 스위치 부하를 감소시킬 수 있다.

도면의 간단한 설명

[0018] 도 1은 INT-모니터링 구조를 설명하기 위한 도면이다.

도 2는 도 1의 INT-모니터링 구조의 시간 동기화의 필요성을 설명하기 위한 도면이다.

도 3은 DPTP(precise data-plane time-synchronization protocol)를 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 INT-모니터링을 설명하기 위한 도면이다.

도 5는 본 발명의 일 실시예에 따른 INT-모니터링 방법에서 인접한 스위치 사이 시간 동기화를 설명하기 위한 도면이다.

도 6은 본 발명의 일 실시예에 따른 INT-모니터링 서버와 엣지 스위치 사이 시간 동기화를 설명하기 위한 도면이다.

도 7은 본 발명의 일 실시예에 따른 INT-모니터링 서버와 기타 스위치 사이 시간 동기화를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0019] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시 예들에 대해서 특정한 구조적 또는 기능적 설명들은 단지 본 발명의 개념에 따른 실시 예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시 예들은 다양한 형태들로 실시될 수 있으며 본 명세서에 설명된 실시 예들에 한정되지 않는다.

[0020] 본 발명의 개념에 따른 실시 예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시 예들을 도면에 예시하고 본 명세서에서 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시 예들을 특정한 개시 형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물, 또는 대체물을 포함한다.

[0021] 제1 또는 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로만, 예컨대 본 발명의 개념에 따른 권리 범위로부터 벗어나지 않은 채, 제1 구성 요소는 제2 구성 요소로 명명될 수 있고 유사하게 제2 구성 요소는 제1 구성 요소로도 명명될 수 있다.

[0022] 어떤 구성 요소가 다른 구성 요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성 요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성 요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성 요소가 다른 구성 요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는 중간에 다른 구성 요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성 요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.

[0023] 본 명세서에서 사용한 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로서, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 본 명세서에 기재된 특징, 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0024] 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 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0025] 이하, 본 명세서에 첨부된 도면들을 참조하여 본 발명의 실시 예들을 상세히 설명한다. 그러나, 특허출원의 범위가 이러한 실시 예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.

[0027] 도 1은 INT-모니터링 구조를 설명하기 위한 도면이다.

[0028] 도 1을 참조하면, 모니터링 패킷을 별도로 생성하여 전달하는 기존의 모니터링 방식, 예컨대 Netflow, sFlow 등과는 달리 INT는 스위치의 상태 정보를 메타데이터(metadata)로 정의하여 직접 패킷에 추가하는 방식을 사용한다. 메타데이터는 스위치의 스위치 아이디, 패킷이 도착할 때와 떠날 때의 시간(ingress timestamp, egress timestamp), 홉 처리 지연 시간(hop latency), 큐 점유율(queue occupancy) 등이 포함된다. 모니터링 패킷이 INT를 사용하는 스위치 영역에 들어오면 기존의 패킷에 INT 헤더와 메타데이터들을 추가하고 INT 스위치의 마지막 스위치에서는 INT 패킷을 복사하여 INT-콜렉터(collector) 모니터링 장비에 전달하고, 기존의 패킷은 추가된 정보를 제거하여 복구 작업을 거친 후 INT 스위치 영역 밖으로 내보낸다. INT-콜렉터에서는 실시간으로 패킷을 전달받아 각 스위치의 메타데이터 정보들을 추출하여 모니터링한다. INT-콜렉터 또는 INT 콜렉터 모니터링 장비는 적어도 프로세서(processor) 및/또는 메모리(memory)를 포함하는 컴퓨팅 장치(computing device)를 의미하는 용어로써, 서버, 모니터링 서버, 모니터 서버, INT-모니터링 서버 등으로 명명될 수 있으며, 범용 서버 등으로 구현될 수 있다.

[0030] 도 2는 도 1의 INT-모니터링 구조의 시간 동기화의 필요성을 설명하기 위한 도면이다.

[0031] 도 2를 참조하면, INT-모니터링에서, 스위치 간 시간이 다를 경우, 모니터링 서버에서 수집한 데이터를 이용한 통계 및 분석에 어려움이 따른다. 따라서, 스위치 간 시간 동기화가 필요하다.

[0032] NTP(network time protocol) 동기화 작동은 소프트웨어 계층에서 작동하며 다양한 장비에서 많이 활용되고 있다. NTP는 동기화 서버를 Stratum이라는 여러 계층으로 분리하여 시간을 동기화하는 방법을 사용한다. PTP(precision time protocol)는 하드웨어의 지원을 거쳐 정확도를 높였는데 마스터-슬래브(master-slave) 구조를 사용하여 모든 노드들은 마스터 스위치를 기준으로 동기화 메시지를 주고 받으며 시간을 동기화한다. 하지만, PTP 스위치의 존재 여부를 떠나 범용적으로 시간을 동기화하는 방법이 존재하지 않는다.

[0034] 도 3은 DPTP(precise data-plane time-synchronization protocol)를 설명하기 위한 도면이다.

[0035] 도 3을 참조하면, DPTP는 프로그래머블 스위치 기반의 데이터 평면에서의 시간 동기화 방법이다. 위에서 아래로 마스터-슬래브(master-slave) 구조로 동기화가 진행되며, 동기화할 때 request, response, follow-up 64 바이트의 크기를 갖는 세 개의 메시지가 생성된다. 내부 클럭이 드리프트 시간(drift time)을 보정하기 위해 스위치 사이에서 2,000번의 메시지가 교류된다. 따라서, 스위치 사이에서 3.07mbps 대역폭이 소모되며, 노드가 많을수록 데이터 평면 트래픽에 과부하가 발생한다.

[0036] 이하, 첨부된 도면을 참조하여, 본 발명의 실시예들에 따른 INT-모니터링 방법을 설명한다. 이에 의하면, 스위치 사이에 별도의 시간 동기화 메시지를 생성하지 않고, INT-모니터링 서버에서 수집한 타임스탬프(timestamp) 정보를 활용하여 각 스위치 사이의 시간 오차범위를 계산할 수 있다. 트래픽이 흐르는 스위치에서만 작동하기 때문에 데이터 평면에 소모 대역폭을 감소시킬 수 있다. 또한, 스위치에서 의도적으로 프로빙(probing) 패킷 생성을 할 필요가 없어서 스위치 부하를 감소시킬 수 있다.

[0038] 도 4는 본 발명의 일 실시예에 따른 INT-모니터링을 설명하기 위한 도면이다.

[0039] 도 4를 참조하면, 본 발명의 일 실시예에 따른 INT-모니터링은 크게 세 가지의 동기화 과정으로 나눌 수 있다. INT-모니터링은 인접한 스위치 사이 시간 동기화 과정(①)과 INT-모니터링 서버와 엣지 스위치 사이 시간 동기화 과정(②)과 INT-모니터링 서버와 기타 스위치 사이 시간 동기화 과정(③)으로 나눌 수 있다.

[0041] 우선, 인접한 스위치 사이 시간 동기화 과정(①)을 설명하기로 한다.

[0042] 도 5는 본 발명의 일 실시예에 따른 INT-모니터링 방법에서 인접한 스위치 사이 시간 동기화를 설명하기 위한 도면이다.

[0043] 도 5를 참조하면, 해당 INT-모니터링 영역에서 패킷이 각 스위치에 흐르는 과정에서 스위치의 상태 정보가 메타데이터(metadata) 형태로 패킷에 추가된다. 메타데이터에는 패킷이 스위치를 통과할 때 스위치의 상태 정보가 포함된다. 패킷이 INT-모니터링 영역을 빠져나갈 때 마지막 스위치를 경유하는 과정에서 해당 패킷은 복사과정이 일어나는데 각각 모니터 서버와 목적지(또는 목적 노드(destination node))에 각각 전달된다. 그 중 목적지에 전달되는 패킷은 불필요한 스위치 상태 정보를 제거하는 등 패킷의 복구 작업을 거친 후, 목적지로 전달된다. 반대로, 모니터 서버에 전달되는 패킷은 각 스위치의 상태 정보를 담고 전달된다. 각 스위치의 상태 정보는 패킷이 각 스위치에 진압할 때의 시간과 통과할 때의 시간이 포함되는데, 이를 모니터 서버에서 각 스위치 간의 시간 동기화 해결 수단에 이용할 수 있다.

[0044] 도 5에 도시된 바와 같이, 제1 스위치 및 제2 스위치를 예시적으로 설명하기로 한다.

[0045] 제1 스위치에서 제2 스위치로 향하는 트래픽에서 제1 스위치에서의 패킷 통과(egress) 타임스탬프(timestamp) 정보(t0)를 획득하고, 제1 스위치에서 제2 스위치로 향하는 트래픽에서 제2 스위치에서 패킷 진입(ingress) 타임스탬프 정보(t1)를 획득한다. 제2 스위치에서 제1 스위치로 향하는 트래픽에서 제2 스위치에서의 패킷 통과 타임스탬프 정보(t2)를 획득하고, 제2 스위치에서 제1 스위치로 향하는 트래픽에서 제1 스위치에서의 패킷 진입 타임스탬프 정보(t3)를 획득한다. 여기서, t0는 제1 스위치가 제1 스위치에서 제2 스위치로 향하는 패킷에 추가

한 상태 정보로부터 획득될 수 있고, t1은 제2 스위치가 제1 스위치에서 제2 스위치로 향하는 패킷에 추가한 상태 정보로부터 획득될 수 있고, t2는 제2 스위치가 제2 스위치로부터 제1 스위치로 향하는 패킷에 추가한 상태 정보로부터 획득될 수 있고, t3는 제1 스위치가 제2 스위치로부터 제1 스위치로 향하는 패킷에 추가한 상태 정보로부터 획득될 수 있다. 이어서, 타임스탬프 정보들(t0, t1, t2, t3)을 이용하여 하기 식 1에 따라 제1 및 제2 스위치들 간 로컬 시간 편차(Θ)를 구할 수 있다.

[0046] [식 1]

[0047] $$\theta = \frac{(t1 - t0) + (t2 - t3)}{2}$$

[0049] 이하에서는, 도 3에 도시된 INT-모니터링 서버와 엣지 스위치 사이 시간 동기화 과정(②)을 설명하기로 한다.

[0050] 도 6은 본 발명의 일 실시예에 따른 INT-모니터링 서버와 엣지 스위치 사이 시간 동기화를 설명하기 위한 도면이다.

[0051] 도 6을 참조하면, 서로 다른 INT-모니터링 데이터 평면을 동기화하기 위해서 모니터 서버에서 각 INT-모니터링 영역에서 모니터 서버와의 연결이 제일 안정적인 엣지(edge) 스위치를 선정하여 시간 동기화 패킷을 전달한다. 다만, 본 발명의 권리범위가 이에 제한되는 것은 아니며, 네트워크(예컨대, 소프트웨어 정의 네트워크(Software Defined Network, SDN)를 구성하는 복수의 스위치들 중 적어도 하나의 스위치를 선정하는 것도 가능하다. 이때, 모니터 서버에서 스위치까지의 패킷 전달 시간이 항상 일정하다고 가정하고, 모니터 서버에서 시간 정보 패킷을 처리하기 위한 시간이 항상 일정하다고 가정한다.

[0052] 모니터 서버에서 시간 정보 수집을 위한 패킷을 전달하는 전달 시간(t0)을 획득한다. 모니터링 데이터 평면에서 선정된 엣지 스위치에서는 모니터 서버로부터 오는 패킷을 전달받으면 전달받은 시간(t1)을 기록하고, 전달받은 패킷을 다시 모니터 서버에 반환하게 되는데 패킷을 반환할 때 통과한 시간(t2)을 기록하여 패킷에 해당 정보를 추가한다. 모니터 서버에서는 엣지 스위치에서 반환 패킷을 접수하는 시간(t3)을 획득한다.

[0053] 이때, 모니터 서버와 엣지 스위치의 시간 차이는 하기 식 2와 같다.

[0054] [식 2]

[0055] $$\theta = \frac{(t1 - t0) + (t2 - t3)}{2}$$

[0057] 마지막으로, 도 3에 도시된 INT-모니터링 서버와 기타 스위치 사이 시간 동기화 과정(③)을 설명한다.

[0058] 도 7은 본 발명의 일 실시예에 따른 INT-모니터링 서버와 기타 스위치 사이 시간 동기화를 설명하기 위한 도면이다.

[0059] 도 7을 참조하면, 스위치가 (n+1) 개일 경우(n은 2 이상의 자연수), 스위치 1부터 (n+1)까지 시간 차이 $G_{\theta 1}$부터 $G_{\theta n}$을 식 1을 통해 구한 후, 순차적으로 더하는 방식으로 스위치 1에서부터 간접적으로 연결된 두 스위치 간의 시간 편차를 하기의 알고리즘으로 구할 수 있다. 상술한 과정은 모니터 서버에 의해 수행될 수 있다.

Data: connected device time difference $\theta 1 \ldots n$
Result: time difference between first node to indirectly linked target node n
time difference value initialization;
while *current node i is not target node n* **do**
  time difference value = time difference value + θi;
  move to next node;
end
Algorithm 1: Indirectly connected time difference

[0060]

[0061] 구체적으로, 모니터 서버와 엣지 스위치(제(n+1) 스위치) 사이의 시간 차이(CT)는 앞서 설명한 바와 같이 식 2를 이용하여 계산이 가능하다. 또한, 모니터 서버와 제n 스위치 사이의 시간 차이는 모니터 서버와 엣지 스위치(제(n+1) 스위치) 사이의 시간 차이(CT)에 $G_{\Theta n}$을 더한 값이 된다. 이와 같은 방식으로 각 스위치와 모니터 서버 사이의 시간 차이를 계산할 수 있으며, 제1 스위치와 모니터 서버 사이의 시간 차이는 $G_{\Theta 1}$부터 $G_{\Theta n}$까지 합산할 결과에 CT를 더한 값이 된다.

[0063] 전술한 바와 같이, DPTP 또는 PTP와 같은 기존의 동기화는 1초에 2,000번 이상 시간 동기화 메시지를 전달되어 드리프트 타임(drift time)을 보정하기 위하여 본 발명은 하기의 방법을 제시한다.

[0064] 기존에는 각 스위치 간의 클럭 밀림 현상의 다양함을 극복하기 위해 기존의 시간 동기화 방법에서는 주기적으로 시간 동기화 메시지를 전달하는 방식을 채택하였다. 일정한 시간이 흐르면 스위치 간에 새로운 시간 동기화 메시지를 통해 새로운 시간 차이를 계산하여 대체한다. 하지만, INT-모니터링 방법은 트래픽이 흐르는 경우에 한해서 시간 정보를 수집하기 때문에 트래픽이 발생하지 않을 경우에는 정확한 시간 차이를 계산하기 어렵다. 따라서, 스위치 간 지연시간 예측 방법을 사용한다. 스위치에서의 클럭 밀림 현상이 일정하여 각 스위치 간의 시간 차이가 선형적인 형태로 변경된다고 가정한다면 트래픽이 없는 경우에도 스위치 간의 현재의 시간 차이를 예측할 수 있다.

[0065] 마지막 수집 시간 T(last)에 수집한 두 스위치 간 시간 차이 G(last)를 구하고, 그 전에 수집된 시점 T(old)에 수집한 두 스위치 간 시간 차이 G(old)를 구한다. 현재의 시간을 T(now)라 하면, 현재의 두 스위치 간의 시간 차이 G(now)는 하기의 식 3으로 구할 수 있다.

[0066] [식 3]

[0067]
$$G(now) = \frac{\{G(last) - G(old)\} \times \{T(now) - T(last)\}}{T(last) - T(old)} + G(last)$$

[0068] 즉, 식 3을 이용하면, 적어도 2회의 시간 차이가 이미 계산된 임의의 스위치들 간의 현재 시간 차이를 추정할 수 있다. 또한, 추정된 시간 차이를 앞서 상술한 시간 차이 계산 기법에 이용할 경우, 현재 트래픽이 존재하지 않은 스위치와 모니터 서버 사이의 시간 차이를 계산(또는 추정)할 수 있다. 예컨대, 모니터 서버와 제n 스위치 사이의 시간 차이는 식 3을 이용하여 추정된 제(n+1) 스위치와 제n 스위치 간의 시간 차이에 CT를 더한 값으로 추정할 수 있다.

[0070] 이상에서 설명된 장치는 하드웨어 구성 요소, 소프트웨어 구성 요소, 및/또는 하드웨어 구성 요소 및 소프트웨어 구성 요소의 집합으로 구현될 수 있다. 예를 들어, 실시 예들에서 설명된 장치 및 구성 요소는, 예를 들어, 프로세서, 콘트롤러, ALU(Arithmetic Logic Unit), 디지털 신호 프로세서(Digital Signal Processor), 마이크로컴퓨터, FPA(Field Programmable array), PLU(Programmable Logic Unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(Operation System, OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응

답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술 분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(Processing Element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(Parallel Processor)와 같은, 다른 처리 구성(Processing Configuration)도 가능하다.

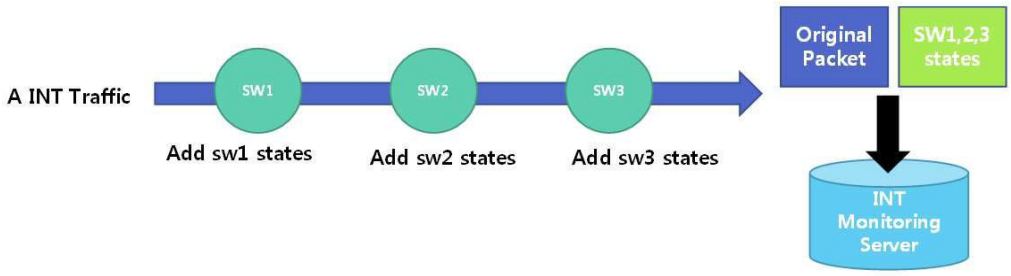
[0071] 소프트웨어는 컴퓨터 프로그램(Computer Program), 코드(Code), 명령(Instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(Collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성 요소(Component), 물리적 장치, 가상 장치(Virtual Equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(Signal Wave)에 영구적으로, 또는 일시적으로 구체화(Embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0072] 실시 예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시 예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 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), 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시 예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

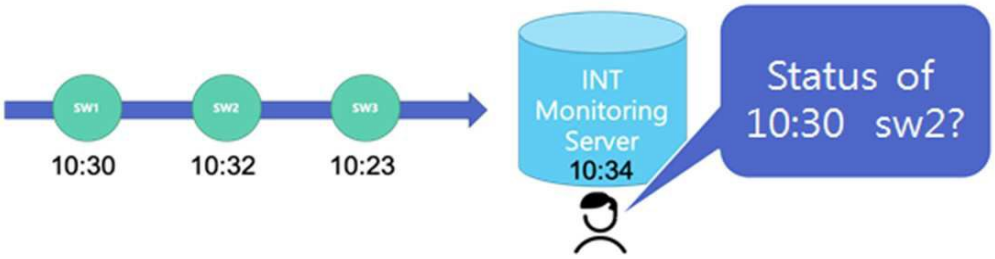
[0073] 본 발명은 도면에 도시된 실시 예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시 예가 가능하다는 점을 이해할 것이다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성 요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성 요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다.

도면

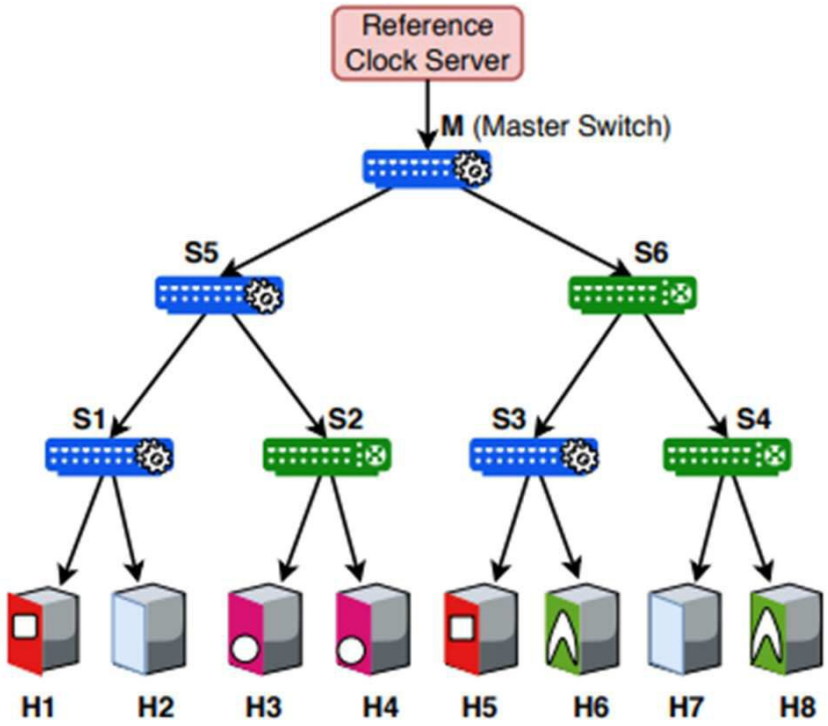
도면1



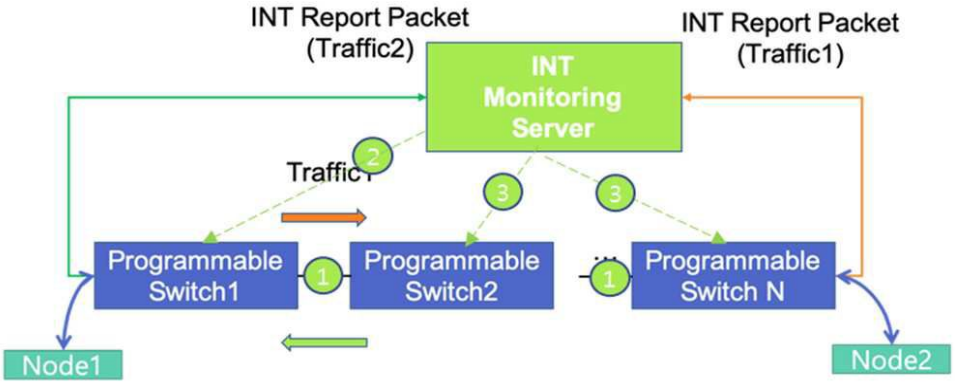
도면2


SW1
SW2
SW3
10:30
10:32
10:23
INT
Monitoring
Server
10:34
Status of
10:30 sw2?

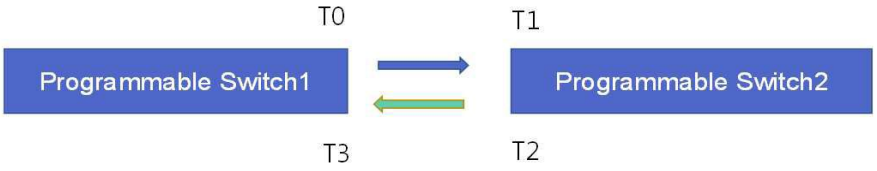
도면3


Reference
Clock Server
M (Master Switch)
S5
S6
S1
S2
S3
S4
H1
H2
H3
H4
H5
H6
H7
H8

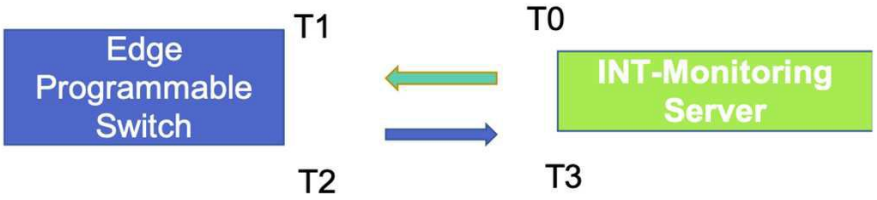
도면4


INT Report Packet
(Traffic2)
INT Report Packet
(Traffic1)
INT
Monitoring
Server
Traffic1
2
3
3
Programmable
Switch1
1
Programmable
Switch2
1
...
Programmable
Switch N
Node1
Node2

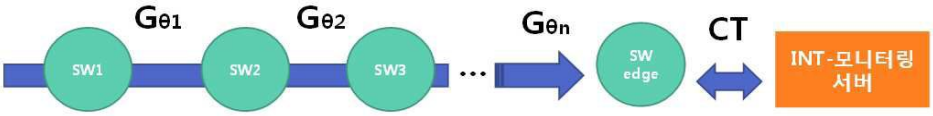
도면5


T0
T1
Programmable Switch1
Programmable Switch2
T3
T2

도면6


T1
T0
Edge Programmable Switch
INT-Monitoring Server
T2
T3

도면7


Gθ1
Gθ2
Gθn
CT
SW1
SW2
SW3
SW edge
INT-모니터링 서버