# 인터넷 트래픽 분석을 위한 TCP 패킷 역전 및 중복 문제 해결 알고리즘

이수강, 안현민, 김명섭
고려대학교 컴퓨터정보학과

{sukanglee, queen26, tmskim}@korea.ac.kr

# Solution of Out-of-order and Retransmission problems for analysis Internet Traffic

Su-Kang Lee, Hyun-Min An, and Myung-Sup Kim

Korea Univ.

## 요 약

급격한 인터넷의 발전으로 효율적인 네트워크 관리를 위해 트래픽 데이터 분석의 중요성이 강조되고 있다. 발생된 트래픽 데이터를 분석하기 위해 해당 트래픽을 발생시킨 응용을 탐지할 수 있어야 한다. 응용을 탐지하기 위한 방법들 중 하나인 통계정보 트래픽 분류방법을 사용하여 트래픽을 분류할 수 있는데, 이러한 통계정보를 그대로 사용하여 분류하기에는 TCP 세션에서 발생하는 Out-of-order, Retransmission 과 같은 문제점들이 있다. 본 논문에서는 기존의 Out-of-order, Retransmission 해결 알고리즘의 한계점을 보완하기 위해 새로운 Out-of-order, Retransmission 해결 알고리즘을 제안한다.

## 1. 서 론

초고속 인터넷의 보급과 다양한 인터넷 기반 응용의 등장으로 인해 트래픽이 복잡 다양해지고 있다. 이러한 상황 속에서 효과적인 네트워크의 관리를 위해 패킷을 수집하여 해당 패킷을 발생시킨 응용을 탐지할 수 있어야 한다. 트래픽 분석을 위한 응용을 탐지하기 위한 방법들 중 하나인 통계정보 트래픽 분류방법[2]은 패킷의 크기와 전송 방향, 전송 순서, 그리고 캡쳐 시간등을 사용한다. 하지만 통계 정보를 그대로 사용하여 트래픽을 분류 하는데는 한계가 있는데 바로 TCP 세션에서 발생하는 Out-of-order 와 Retransmission 문제이다.

기존 연구[1]에서는 학내망에서 발생된 트래픽을 이용하여 문제들을 조사하고 이를 해결하기 위한 알고리즘을 제시하였다. 기존 연구의 실험 결과에서 완전히 제거되지 않은 Retransmission 패킷에 대해 조사한 결과 SEQ 번호와 ACK 번호가 맞지 않는 쌍을 발견 하였다. 기존 연구에서는 Retransmission 패킷이 발생하게 되면 뒤에 탐지된 패킷만을 저장하였다. 이는 패킷 전송 중 오류로 인해 패킷이 버려진 뒤 Retransmission 된 경우 먼저 전송된 패킷에 있을 오류를 제거하기 위해서이다. 하지만 먼저 전송된 패킷과 후에 발생한 재전송된 패킷은 재패킷화 될 경우 패킷의 크기가 서로 달라지게 된다. 따라서 기존 방법을 사용하여 저장할 경우 두 패킷중 어떤 패킷이 정상적으로 상대방에게 전송되는지 확인을 하지 않고 저장하기 때문에 SEQ 번호와 ACK 번호가 맞지 않게 된다. 본 논문에서는 해당 문제를 조사하고, 이를 해결하기 위한 알고리즘을 제안한다.

본 논문은 다음과 같은 순서로 구성된다. 2 장에서는 기존 연구에 대한 문제점을 분석하고 해결 방법을 제안한다. 3 장에서는 제안한 알고리즘을 바탕으로 실험한 결과를 제시하고 끝으로 4 장에서는 결론 및 향후 연구에 대해 언급한다.

이 논문은 2012 년 정부(교육과학기술부)의 재원으로 한국연구재단 (2012R1A1A2007483) 및 2013 년도 정부(미래창조과학부)의 재원으로 한국연구재단-차세대정보.컴퓨팅기술개발사업(2010-0020728)의 지원을 받아 수행된 연구임.

## 2. 본 론

본 장에서는 Out-of-order, Retransmission 발생시 기존의 해결 방법에 대한 문제점을 분석하고, 개선된 대안을 제시한다.

**Out-of-order** 는 패킷들이 여러 경로를 거치면서 발생할 수 있는 지연현상을 이유로 수집지점에서 원래의 순서가 아닌 다른 순서로 수집되는 문제를 말한다. 이는 각 패킷들의 SEQ, ACK 번호를 비교하여 재정렬 하는 방법으로 문제를 해결할 수 있다.

**Retransmission** 은 전송된 패킷에서 오류를 발견하였거나 일정 시간이 지나도 응답 패킷을 받지 못할 경우 동일한 패킷 또는 재패킷화를 통하여 새로 만들어진 패킷을 재전송하는 문제이다.

**재패킷화(repacketization)**는 성능을 향상시킬 목적으로 다시 전송할 때 좀 더 큰 세그먼트를 전송하는 방법이다.

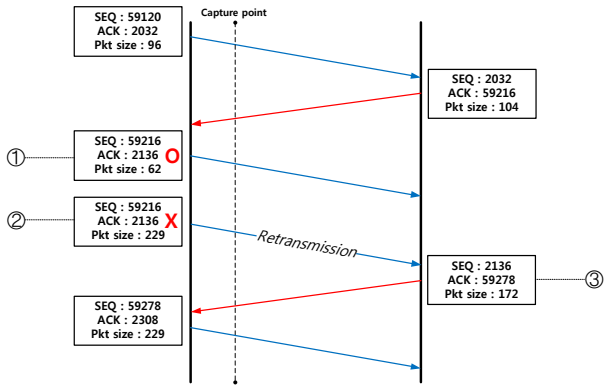


그림 1. Retransmission 패킷의 재패킷화

그림 1 은 Retransmission 패킷의 재패킷화 발생시 원래의 패킷과 재전송된 패킷의 크기가 달라지는 경우를 그림으로 나타낸 것이다. 기존 연구에서는 Retransmission 패킷 발생시 먼저전송된 패킷보다 나중에 전송 된 패킷을 저장하였다. 기존 방법으로는 1 번 패킷은 지워지고 2 번 패킷이 저장된다. 하지만 3 번 패킷의 ACK 번호는 59278 인데 2 번 패킷의 SEQ 번호와 SIZE 의 합은 59445 이므로 응답패킷과 올바르게 대응하는 1 번 패킷이 저장되어야 한다. 따라서 재패킷화 되는 경우엔 Retransmission 패킷과 뒤에 나오는 응답패킷과의 SEQ, ACK 번호를 비교하여 패킷을 저장해야 한다.

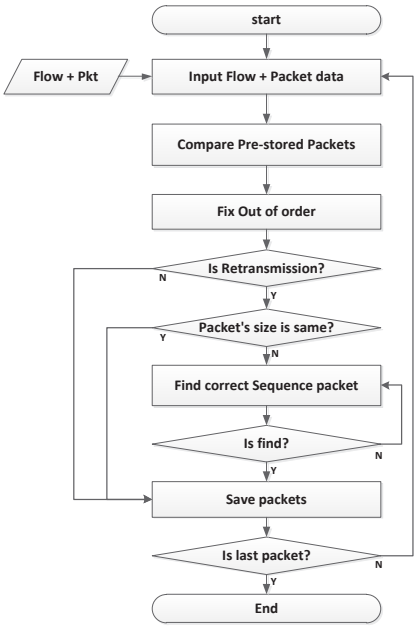


그림 2. TCP 세션의 이상동작 개선 순서도

그림 2 는 Retransmission 발생시 재패킷화 되어 재전송되는 경우 올바른 응답패킷을 선택하여 저장하는 알고리즘의 순서도이다. 해당 알고리즘은 패킷을 포함한 플로우 데이터를 입력으로 받아 Out-of-order 와 Retransmission 을 해결하는 방법이다. Pre-stored Packet 은 플로우에 이미 저장되어 있는 다른 패킷들을 의미하고 Direction 은 패킷의 전송 방향을 의미한다.

Retransmission 을 해결하기 전에 먼저 Out-of-order 를 해결한다. Out-of-order 는 방향이 같으면서 먼저 저장된 패킷의 SEQ 번호보다 나중에 저장된 패킷의 SEQ 번호가 작은 경우이다. 이는 패킷의 위치를 서로 바꿔 SEQ 번호 순으로 재정렬하여 해결한다. Retransmission 은 데이터 전송에서 오류를 발견하였거나 일정 시간 경과해도 응답패킷을 받지 못할 경우 발생하게 된다. 이 때 패킷은 동일한 SEQ 번호와 방향을 갖게 되며 동일한 패킷 또는 재패킷화 된 크기가 다른 패킷을 재전송하게 된다.

Retransmission 이 발생하면 원본 패킷과 재전송된 패킷의 크기를 비교하여 같으면 두 패킷은 동일한 패킷이므로 뒤에 잡힌 패킷을 저장하게 된다. 하지만 크기가 다를 경우 플로우 내에서 재전송된 패킷과 대응하는 ACK 번호를 갖고 있는 응답 패킷을 찾을 경우 해당 재전송 패킷을 저장한다. 마지막 패킷까지 검사를 마치면 해당 플로우는 Out-of-order 와 Retransmission 문제를 해결한 플로우다.

## 3. 실험 결과

본 장에서는 2 장에서 제안된 알고리즘의 이상동작 처리방법의 성능 평가를 위해 학내망에서 수집한 트래픽을 대상으로 실험하였다. 실험에 사용된 트래픽 데이터는 2013 년 5 월 15 일 13 시부터 15 시까지 총 2 시간의 트래픽 데이터를 사용하였다. 데이터 총 양은 표 1 과 같다.

| State | Packet | Flow | Byte(MB) |
|---|---|---|---|
| Total | 113,574,298 | 3,233,978 | 141,398.7MB |
| Normal | 104,486,509 | 2,661,783 | 129,052MB |

표 1. 실험 데이터

본 논문에서 제시한 해결 알고리즘을 바탕으로 학내망에서 정상적인 Flow 는 84.97%, 비정상적인 Flow 는 15%를 차지하였다.

| State | Packet(%) | Flow(%) | Byte(MB) |
|---|---|---|---|
| Normal | 90.18% | 84.97% | 129,052MB |
| Out-of-order | 7.73% | 9.30% | 12,338MB |
| Retransmission | 2.45% | 7.70% | 5,125MB |
| Retransmission & Out-of-order | 0.32% | 2.13% | 507MB |

표 2. 제안 된 방법론 적용 후 실험 결과

표 2 는 TCP 세션의 이상동작 처리를 실제 트래픽에 적용하여 분석된 이상동작 비율과 각각의 데이터 양이다. Out-of-order 와 Retransmission 의 발생 비율은 Flow 기준으로 9.3%, 7.70%, Packet 기준으로 7.73%, 2.45%이며 Byte 로는 12,338MB, 5,125MB 를 차지하고 있었다. 차지하고 있는 비율은 미미해 보이지만 용량으로는 많은 양의 이상동작이 있음을 보였다.

## 4. 결론 및 향후 과제

본 논문에서는 TCP 세션의 이상동작 발생 비율을 조사하고 기존 연구에서 제시한 해결방법의 문제를 개선한 알고리즘에 대해 기술하였다. 그리고 실제 트래픽을 대상으로 개선된 알고리즘을 사용하여 이상동작 개선 실험을 하였고, 이를 통해 본 논문에서 제시한 방법론의 필요성을 입증하였다.

향후 연구과제로는 본 논문에서 제시한 TCP 세션의 이상동작 해결 알고리즘을 적용한 모듈을 보완할 예정이다. 본 논문에서 제시한 알고리즘은 오프라인에서 구동되는 모듈에 적용하였다. 실시간 모니터링 시스템에서 해당 알고리즘을 적용하여 구동하기 위해 모듈의 데이터 처리및 저장 방법에 대한 보완작업을 수행하고자 한다.

## 참 고 문 헌

[1] 안현민, 최지혁, 함재현, 김명섭, "TCP 세션의 이상동작으로 인한 트래픽 분석 방법론의 한계와 해결 방안, KNOM Review, Vol. 15, No. 1, Dec. 2012, pp. 31-39.

[2] Ying-Dar Lina, Chun-Nan Lua, Yuan-Cheng Laib, Wei-Hao Penga and Po-Ching Lina, "Application classification using packet size distribution and port association" Proc. of the Journal of Network and Computer Applications, In Press, Corrected Proof, Available online, March, 20. 2009.