플로우의 연관성 모델과 시드 기반의 유동적 혼합 그룹핑을 이용한 악성 트래픽 탐지

박지태, 이성호, 구영훈, 심규석, 김명섭

고려대학교

{pjj5846, gaek5, gyh0808, kusuk007, tmskim}@korea.ac.kr

Attack Traffic Detection using the Correlation of the Flow and Seed based Flexible Combination Grouping Model

Jee-Tae Park, Sung-Ho Lee, Young-Hoon Goo, Kyu-Seok Shim, Myung-Sup Kim

Korea Univ

요 약

오늘날의 네트워크 환경은 인터넷의 고속화와 인터넷 사용자의 증가로 인하여 급격하게 성장하고 있는 추세이고, 이로 인해 악성 트래픽의 종류와 패턴도 날이 갈수록 다양해지고 있다. 이러한 추세에 대응하기 위해서는 악성 트래픽에 대한 정확한 탐지 및 분석이 필요하다. 악성 트래픽을 탐지하는 방법은 여러 가지가 존재하지만, 가장 보편적인 방법은 트래픽의 페이로드 시그니처 기반의 탐지 방법이 있다. 이러한 방법은 높은 정확도와 분석률을 지닌다는 측면의 장점이 있지만 때로는 비효율적이고 패턴 시그니처를 추출 과정이 복잡하다는 단점이 있다. 따라서 최근에는 새로운 악성 트래픽 탐지 방법들이 연구되고 있다. 그 중 한 가지로 플로우의 연관성과 그룹핑을 이용한 악성 트래픽 탐지 방법이 있다. 이는 플로우의 패킷 정보로부터 알 수 있는 통계적 특성 및 헤더 정보를 가지고 그룹핑을 통하여 탐지하는 방법이다. 하지만 이러한 그룹핑 방법은 높은 정확도를 지니는 반면에 분석률 측면에서 어느 정도 한계가 있다. 이에 본 논문에서는 기존의 그룹핑 방법을 개선시키는 방안에 대해서 제안하고 이를 적용하여 실험을 진행한다. 기존의 그룹핑 방법과 개선된 그룹핑 방법을 각각 적용하여 실험하였을 때 기존의 그룹핑 방법의 결과보다 더 높은 정확도 및 분석률의 결과를 얻을 수 있었다.

Ⅰ. 서 론[1)]

오늘날의 네트워크 환경은 과학 기술의 비약적인 발전으로 인한 인터넷의 고속화로 인해 급격하게 성장하고 있다. 이로 인해 응용뿐만 아니라 악성 트래픽의 종류와 패턴도 급격하게 복잡해지고 있다. 이를 해결하기 위해서는 악성 트래픽을 정확하게 탐지 할 수 있어야 한다. 악성 트래픽을 탐지하는 방법에 대해서는 현재까지 많은 연구가 진행되어 왔고, 가장 보편적인 방법이 페이로드 시그니처를 이용하는 방법이다[1,2].

페이로드 시그니처를 이용한 탐지하는 방법은 헤더 등의 시그니처의 패턴을 추출하여 미리 정의하고, 이를 이용하여 악성 트래픽을 탐지하는 방법이다[2]. 이는 정확도 및 분석률 측면에서 높은 성능을 지닌다는 장점이 있지만 트래픽의 시그니처 패턴 추출에 너무 의존적이라는 단점을 가지고 있다.

따라서 요즘에는 페이로드 시그니처를 이용한 탐지 방법 이외의 새로운 탐지 방법들이 시도 되고 있다. 그 중 한 가지로 플로우의 연관성과 그룹핑을 이용한 탐지 방법이 있다. 이 방법은 플로우의 패킷 정보로부터 알 수 있는 통계적 정보와 플로우의 헤더 정보를 이용하여 그룹핑을 진행하는 방법이다[1]. 하지만 이러한 탐지 방법은 높은 정확도로 탐지한다는 측면에서 장점이 있지만 분석률 측면에서 어느 정도 한계가 있다. 따라서 본 논문에서는 기존의 그룹핑을 이용한 탐지 방법을 개선시키는 방안에 대해 제안한다. 그리고 이를 실제 악성 트래픽에 적용하여 기존의 그룹핑 방법보다 그룹핑 결과가 향상되었는지 확인한다.

본 논문의 구성은 본 장의 서론에 이어 2장 본론에서는 기존 그룹핑을 이용한 시스템의 전체 구조와 그룹핑 측면에서 발생하는 문제점 및 개선 방안에 대해 설명한다. 다음 3장에서 제안하는 개선된 시스템으로 실제 악성 트래픽에 적용하여 실험을 진행하고 그 기존의 실험 결과와 개선된 시스템의 실험 결과를 확인한다. 마지막으로 4장에서는 결론 및 향후 연구에 대해 언급한 뒤 본 논문을 마친다.

이 논문은 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.2015R1D1A3A01018057).

Ⅱ. 본론

플로우의 연관성 모델과 시드 기반의 그룹핑을 이용한 악성 트래픽 탐지 시스템의 전체적인 시스템 구조는 4 가지의 모듈로 구성 되어 있다.

먼저 Input으로는 악성 트래픽과 노이즈 트래픽이 섞여있는 트래픽이 있고, Output으로는 분석 결과, 분석 로그 및 그룹핑 정보가 있다. Input으로 들어온 pcap 형태의 트래픽들은 Pre-Processing 모듈을 통해 fwp로 변환 된다. 다음으로 변환된 fwp로 Seed Generation 생성 모듈에서 시드를 생성 한다. 시드란 전 모듈 과정을 거친 트래픽의 5-tuples 정보를 모아서 텍스트 파일로 저장한 것으로 정의한다. 다음으로 Guideline Generation 모듈에서 생성된 시드와 fwp를 가지고 그룹핑을 할 때 기준이 되는 Similarity, Connectivity Threshold 값으로 구성된 가이드라인을 생성한다. 여기서 Similarity 값은 트래픽의 패킷 정보로부터 알 수 있는 통계적 특징을 계산하는 것이고 Connectivity 값은 플로우 헤더 정보를 Euclidean Distance로 계산한 것으로 시간, 주소, 포트번호, 프로토콜 등으로 구성 되어 있다. 다음으로 Sequential Grouping 모듈에서 생성된 가이드라인, 시드 정보, fwp를 이용하여 그룹핑을 진행 하고 마지막에 Output이 나온다.

전체 시스템을 그룹핑 관점에서 보면 기존의 그룹핑 모듈에서 진행 했던 그룹핑 방법은 그룹핑 순서가 고정된 교차 그룹핑 이다. 이는 그림 1과 같이 그룹핑 결과에 상관없이 항상 처음에 Similarity, 다음으로는

Connectivity 순으로 그룹핑이 될 때 까지 진행 한다.

하지만 이 경우 첫 그룹핑에서 Connectivity 그룹핑은 가능하지만 Similarity 그룹핑이 안 되는 경우에는 더 이상 그룹핑이 진행이 되지 않는다는 문제점을 가지고 있다.

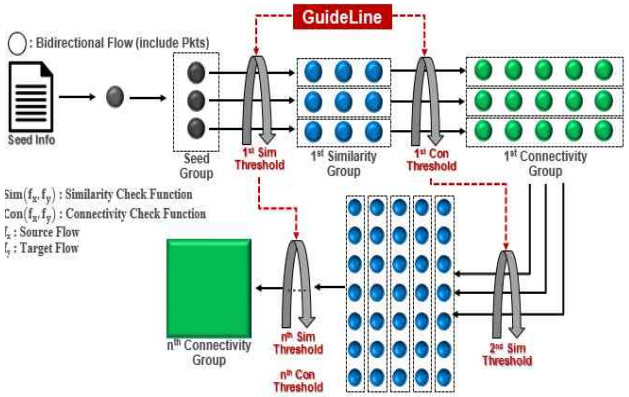

그림 1. 기존의 그룹핑을 적용한 악성 트래픽 탐지 시스템 구조

즉 그룹핑이 충분히 더 될 수 있지만 프로그램이 더 이상 그룹핑을 진행하지 않음으로 써 분석률 측면에서 좋지 않게 나온다. 게다가 두 번째의 그룹핑 경우에도 Connectivity로 그룹핑은 진행되지만 Similarity 그룹핑은 가능하지만 Connectivity 그룹핑은 안되는 경우 위와 같은 문제가 나타난다. 따라서 본 논문에서는 이러한 기존 그룹핑 방법의 문제를 해결하기 위해 유동적 혼합 그룹핑 알고리즘을 제안한다. 제안하는 그룹핑 알고리즘을 개선한 시스템의 구조는 그림 2와 같다.

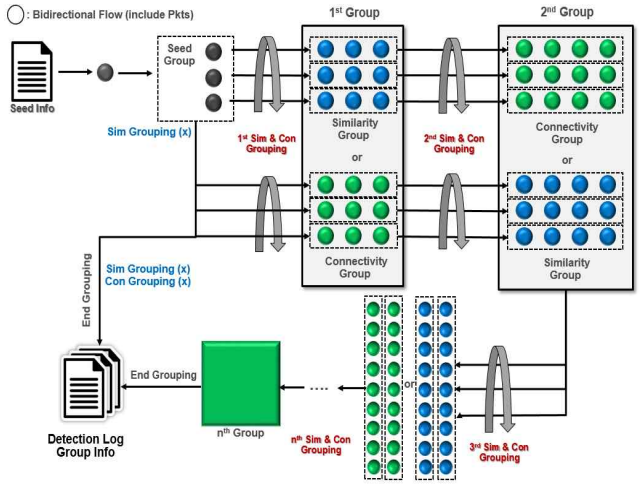

그림 2. 유동적 혼합 그룹핑을 적용한 악성 트래픽 탐지 시스템 구조

유동적 혼합 그룹핑을 적용하여 그룹핑 및 탐지를 진행 하였을 때는 기존의 그룹핑 시스템과 다르게 첫 그룹핑 때 Similarity로 그룹핑이 되지 않는 경우에도 Connectivity 그룹핑을 진행을 하게 된다. 즉 그룹핑이 진행 될 때 Similarity와 Connectivity의 그룹핑 순서가 고정적이 아니라 유동적으로 진행된다. 이를 통해 기존의 방법으로는 그룹핑이 되지 않는 플로우를 탐지할 수 있다.

Ⅲ. 실험

개선된 그룹핑 방법을 기존의 모델에 적용하여 다양한 트래픽으로 실험을 진행 하였다. 실험에 사용 한 악성 트래픽 및 노이즈 트래픽에 대한 정보는 표 1과 같다. 우선 실험에 사용한 악성 트래픽은 Zeprox, Trojan를 사용하였고 노이즈 트래픽은 Skype, Chrome 등의 웹 브라우저 및 메신저 응용 트래픽을 사용하였다. 실험은 총 2 가지 Trace로 나누어 진행 하였고, 각 Trace마다 모든 노이즈 트래픽을 사용하였다. 결과를 판단하는 지표는 여러 탐지 결과 중 가장 좋은 플로우의 탐지 결과를 Recall, Precision으로 나타내었고, 전체 플로우에서 해당 플로우의 비율을 Portion으로 나타내었다.

표 1. 실험에 사용한 악성 및 노이즈 트래픽 정보

ATTACK INPUT TRAFFIC				
Trace #	Attack Type	Size		
		Flow	Pkt	Byte
1	Zeprox	12	483	430,189
2	Trojan	40	1,792	1,607,484
NOISE INPUT TRAFFIC				
Trace #	Application Type	Size		
		Flow	Pkt	Byte
1	Chrome, IE web	491	60,266	46,852,995
2	Skype	576	107,534	95,635,596

기존의 그룹핑 방법과 본 논문에서 새로 제안하는 그룹핑 방법을 각각 적용하여 탐지 실험을 진행한 결과는 표 2와 같다. 전반적으로 보았을 때 두 가지의 악성 트래픽에서의 실험 결과 모두 정확도 면에서는 100%를 보인다. Zeprox에서는 Recall, Precision 수치 모두 100% 탐지가 되지만 Portion 관점에서 보면 수치가 더 향상된 것을 알 수 있다. 그리고 Trojan의 경우에도 Portion 뿐만 아니라 Recall 부분에서도 많이 향상되었음을 알 수 있다.

표 2. 두 가지 그룹핑을 적용한 악성 트래픽 탐지 실험 결과

Cross Grouping Detection Result						
Trace #	Attack Type	Noise Traffic	Measurement	Coverage (%)		
				Flow	Pkt	Byte
1	Zeprox	all	Recall	100	100	100
			Precision	100	100	100
			Portion	75% (9/12)		
2	Trojan	all	Recall	80	73.7	72.4
			Precision	100	100	100
			Portion	17.5% (7/40)		
Flexible Combination Grouping Detection Result						
1	Zeprox	all	Recall	100	100	100
			Precision	100	100	100
			Portion	100% (12/12)		
2	Trojan	all	Recall	100	100	100
			Precision	100	100	100
			Portion	70% (28/40)		

Ⅳ. 결론 및 향후 연구

본 논문에서는 플로우의 연관성 모델과 시드 기반의 그룹핑을 개선하여 악성 트래픽을 탐지하는 방법을 제시하였다. 제시한 방법을 이용하여 실험을 진행한 결과 기존의 그룹핑을 이용하여 탐지 실험을 진행했을 때 보다 분석률과 정확도 측면에서 크게 향상되었다. 향후 연구로는 더 다양한 악성 트래픽을 사용하여 실제로 악성 트래픽을 탐지하는 환경에서 실험을 진행 할 예정이다.

참 고 문 헌

[1] 구영훈, 심규석, 이성호, Baraka D. Sija, and 김명섭 " 네트워크 플로우의 연관성 모델을 이용한 트래픽 분류 방법" 정보과학회 논문지. Vol.44 No.04 April. 2017. pp433-438

[2] 박준상, 윤성호, 안현민, 김명섭, "페이로드 시그니쳐 기반 인터넷 트래픽 분류", May, 15-16, 2014, pp10-14