본 연구에서는 Kmax 와 dK-2 distance 두 가지 그래프 지표를 이용해 시간에 따른 노드들의 커뮤니케이션 패턴의 변화를 분석하고자 한다. 활용될 수 있는 다른 그래프 변수들은 [5] 에 소개되어 있다. 자세한 비정상 트래픽 탐지 알고리즘은 표 1 에서 확인 가능하다. 비정상 트래픽 판단을 위해 사용되는 static threshold 값은 장기간의 POSTECH 트래픽 정보와 그로부터 얻은 트래픽 분산 그래프들을 분석함으로써 얻을 수 있었다. 알고리즘은 제안한 dK-2 distance 를 연속된 트래픽 분산 그래프 G_i 와 G_{i+1} 로부터 얻기 위해 먼저 joint degree distribution algorithm [12]을 두 그래프 G_i 와 G_{i+1} 의 인접 행렬에 적용한다. 그 후 앞서 얻은 그 결과인 dK(G_i)와 dK-2(G_{i+1}) 사이의 Euclidean distance 를 계산한다. 한 트래픽 분산 그래프의 edge 의 수를 E, vertices 의 수를 V 라고 했을 때 dK-2 알고리즘의 시간복잡도는 O(E+V)이다. Dot format 으로 나타낸 트래픽 분산 그래프를 시각적인 형태로 표현해 비정상 트래픽의 직관적인 탐지를 돕기 위해 Graphviz library[13]가 사용되었다. Graphviz 를 사용해 POSTECH 트래픽을 대상으로 시각화 하는 과정은 2.4HGz Core2 Duo 와 3GB RAM 을 사용하는 PC 에서 수 초 이내에 완료된다.

Joint Degree Distribution 알고리즘

Function: Joint Degree Distribution (A) **returns** joint degree distribution of graph G
Inputs: Adjacency matrix A of graph G
Initialization: n=Number of nodes (G);
K = Maximum (Degree of Graph(G));
JDD[ ] = zeros(K); (Initial matrix JDD with zero elements, size = K)
For i=1 **to** n {
For j=1 **to** n {
if A(i,j) = 1 then{
k1 = Degree (i);
k2 = Degree (j);
JDD (k1,k2) = JDD (k1,k2) + 1;}}}
Return JDD;

dK-2 Distance 알고리즘
JDD1 = Joint Degree Distribution(G_i) JDD2 = Joint Degree Distribution(G_{i+1}) dK-2 distance (G_i, G_{i+1}) = Euclidean distance(JDD1, JDD2); **If** dK-2 distance (G_i, G_{i+1}) < dK-2 distance threshold **then** G_{i+1} is anomaly **Return** status of graph G_{i+1} (anomaly or normal)

표 1. Joint Degree Distribution 알고리즘과 dK-2 Distance 알고리즘

4. 결과 및 검증

4.1. POSTECH DDoS Trace

제안하는 방법이 비정상 트래픽을 탐지할 수 있는지 판단하기 위해 2009 년 7 월 7 일에 전국적으로 발생한 7.7 DDoS 공격 당시의 교내 트래픽을 담고 있는 POSTECH DDoS Trace 를 활용해 검증하였다. POSTECH DDoS Trace 는 1 시간 동안의 POSTECH 네트워크 트래픽을 담고 있는데 당시 POSTECH 내부의 감염된 좀비 PC 들이 국내 또는 해외의 주요 웹사이트들을 공격하는데 사용되었다.

그림 4(a)와 그림 5(a)는 POSTECH DDoS Trace 에 대한 TCP 트래픽의 Kmax 와 dK-2 distance 값의 변화를 각각 보여 준다. 그림에서는 2 분, 22 분 23 분에서 TCP 트래픽의 급격한 변화가 발생했고 이 시각에 비정상 트래픽이 존재함을 확인할 수 있다. 실제로 이 시점에 TCP Syn Flooding 과 HTTP GET/POST Flooding 공격이 이루어진 것으로 확인되었다. 그림 6 에서 볼 수 있듯이 공격을 탐지하기 위해 특정 시점의 패킷 수, 플로우 수, 바이트 크기를 분석하는 간단한 통계적 방법들은 POSTECH DDoS Trace 가 포함하는 공격을 탐지하는데 어려움을 보이지만 dK-2 distance 에 기초한 공격 탐지는 그에 비해 공격 트래픽을 더 정확하게 탐지하는 것이 가능하다

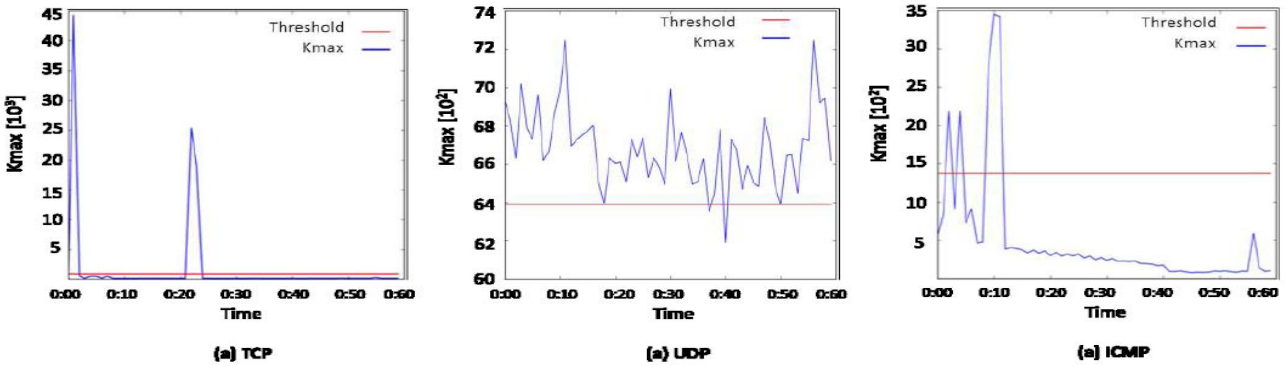

그림 4. POSTECH DDoS Trace 의 Kmax 값 변화

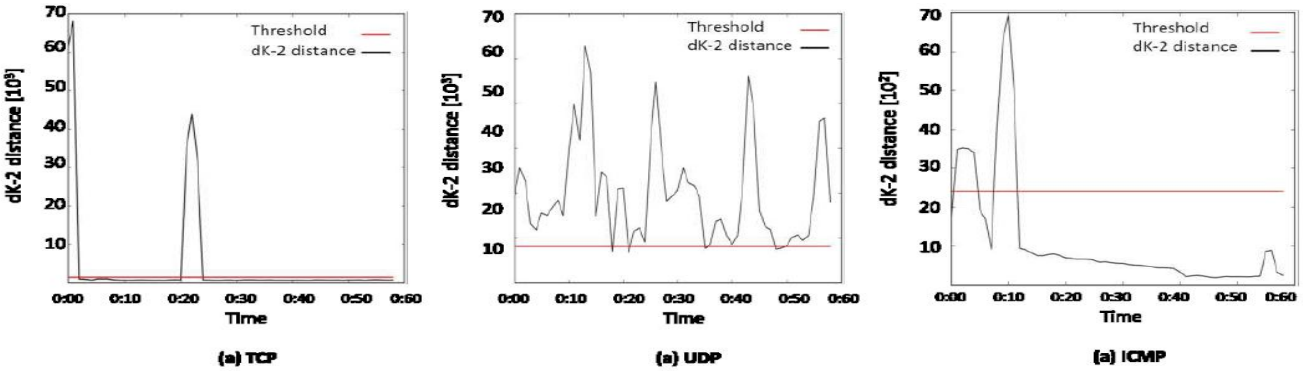

그림 5. POSTECH DDoS Trace 의 dK-2 distance 값 변화

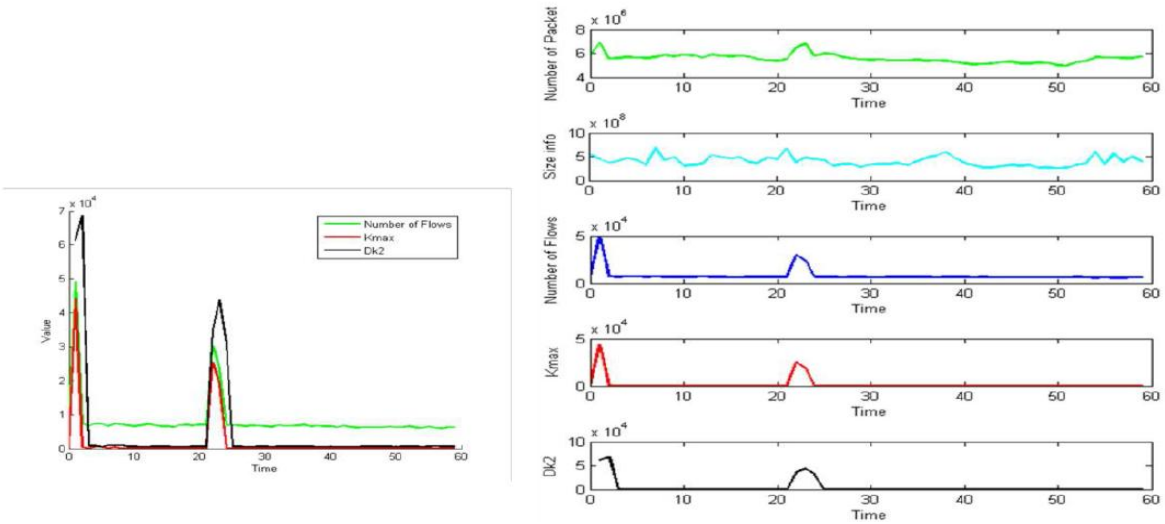

그림 6. POSTECH DDoS Trace 분석

그림 7 은 공격이 발생했을 당시의 POSTECH 네트워크 트래픽을 Graphviz 소프트웨어를 통해 시각화 한 것이다. 그림은 POSTECH DDoS Trace 의 TCP 트래픽 중 22 분에서의 트래픽 분산 그래프이다. 정상적인 트래픽 분산 그래프와 비교했을 때 22 분경에 공격이 발생했음을 쉽게 확인할 수 있다. 해당 DDoS 패턴을 유발한 플로우 정보를 역추적한 결과 그 당시 DDoS 공격이 TCP 포트 6667 번, 즉 Internet Relay Chat(IRC) [14]를 이용한 봇넷에 의한 DDoS 공격이었음을 확인했다. UDP 트래픽의 Kmax(그림 4(b))와 dK-2 distance(그림 5(b)) 값은 1 시간의 POSTECH DDoS Trace 전구간에 걸쳐 threshold 값을 넘어서고 있음을 보여 준다. 이는 UDP 80 번 포트의 Flooding 에 의한 것이다. ICMP Flooding 은 POSTECH DDoS Trace 의 처음 10 분 동안에만 발생했다. ICMP 트래픽의 Kmax 값과 dK-2 distance 값은 각각 그림 4(c)와 그림 5(c) 에서 볼 수 있듯이 처음 10 분 구간에서 급격하게 변화하고 있다.

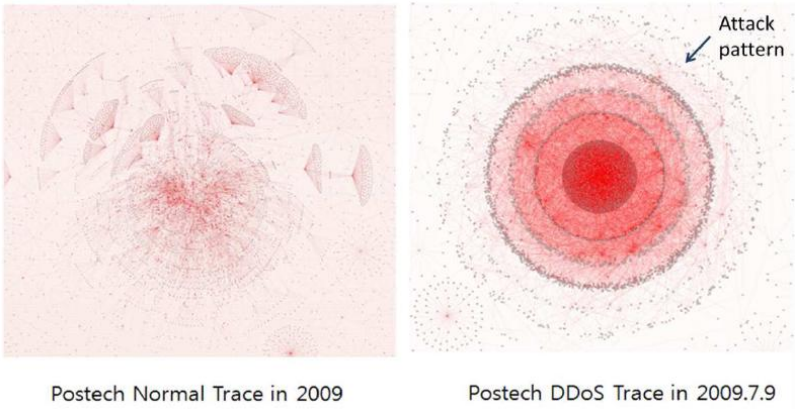

그림 7. POSTECH Trace 의트래픽 분산 그래프

4.2. Synthesized P2P Botnet Trace

본 논문에서 제안하는 비정상 트래픽 탐지 방법이 P2P Botnet 공격의 탐지에도 적합함을 보이기 위해 본 검증 단계에서는 P2P Botnet(Peacomm) Trace 를 자체적으로 생성하여 정상적인 POSTECH Trace 에 삽입하였다. 실제 P2P botnet 트래픽과 유사한 트래픽을 만들고 그것을 검증하기 위해 P2P botnet 의 감염에 사용되는 악성 프로그램인 Trojan Peacomm Binary 파일을 12 개의 호스트로 이루어져 있는 Honeynet 에서 실행하였다. Peacomm botnet 은 Full peer-to-peer(P2P) 네트워크[15,16]를 이루는 botnet 이다. Peacomm 은 Overnet P2P 프로토콜을 이용해 감염된 peer 들이 서로 통신하는데 사용되는데 P2P botnet 이 실제 동작하는 동안의 트래픽을 얻기 위해 Honeynet 에서 감염된 12 개의 호스트들이 서로 통신하는 트래픽을 직접 수집하였다. 그 후 본 논문에서 제안하는 시스템이 P2P botnet 을 탐지할 수 있는지 확인하기 위해 이 과정에서 얻은 botnet trace 를 정상적인 POSTECH trace 의 특정 시간대(29 분)에 삽입하였다.

P2P botnet 들이 서로 통신하고 명령을 내리는 과정에서는 중심이 되는 Central bot 이 존재하지 않기 때문에 일반적인 방법으로는 탐지하기가 쉽지 않다. P2P botnet 내에서 C&C 메시지를 배포하기 위해 각각의 P2P bot 들은 자신만의 작은 Subnet botnet 즉 통신할 Peer list 를 유지하면서 이 peer list 에 등록된 bot 들과 명령을 주고 받는다. 그림 8 에서 볼 수 있듯이 트래픽 크기, 플로우 수, Kmax 값은 P2P botnet 을 탐지하지 못하고 있다. Botnet 들 사이의 통신은 아주 적은 트래픽을 사용하기 때문에 이들과 같은 방법들은 DDoS 공격 전에 이루어지는 P2P botnet 사이의 통신을 탐지할 수 없다. 그림 8 의 dK-2 distance 값은 P2P Botnet Trace 에서 botnet 이 새로운 communication cluster 를 형성하면서 통신을 시작하는 시점을 효과적으로 탐지한다. 그림에서 볼 수 있듯이 dK-2 distance 값은 botnet 의 통신을 시작하는 시점에서 눈에 띄게 증가한 반면 다른 그래프에서는 그 시점을 확실하게 파악할 수 없다. dK-2 distance 는 호스트들의 통신 패턴이 어떻게 변화하는지를 나타내는 지표이기 때문에 다른 통계 기반의 값들과는 달리 botnet 의 공격이 본격적으로 시작되기 전 botnet 들이 cluster 를 형성하고 서로 통신하는 과정에서 botnet 의 움직임을 탐지할 수 있기 때문에 P2P botnet 을 기반으로 한 공격 탐지에도 사용될 수 있다.

5. 실시간 비정상 트래픽 탐지 시스템

5.1. 시스템 아키텍처

본 연구에서 제안한 실시간 비정상 트래픽 탐지 시스템은 기존의 트래픽 모니터링 시스템인 NG-MON2 에 구현되었다. NG-MON2 시스템은 POSTECH 캠퍼스 네트워크 등의 인터넷 트래픽을 실시간으로 수집하고 분석하는 기존의 NG-MON 시스템[10]의 확장이다. NG-MON2 시스템은(그림 9) 세가지 모듈을 포함한다. 첫 번째는 Flow generator 로 인터넷 회선을 지나는 트래픽을 수집하고 그로부터 플로우 정보를 실시간으로 생성한다. 두 번째는 Store 모듈로 이전 단계에서 생성된 플로우 데이터와 패킷 정보를 이후 온라인 또는 오프라인 분석을 위해 저장한다. 세 번째는 Traffic analyzer 모듈로 다양한 기법에 기반한 분석 방법을 통해 수집한 트래픽을 분석한다. 제안한 실시간 비정상 트래픽 탐지 시스템은 NG-MON2 트래픽 모니터링 시스템의 Traffic analyzer 모듈 중 하나로 동작한다.

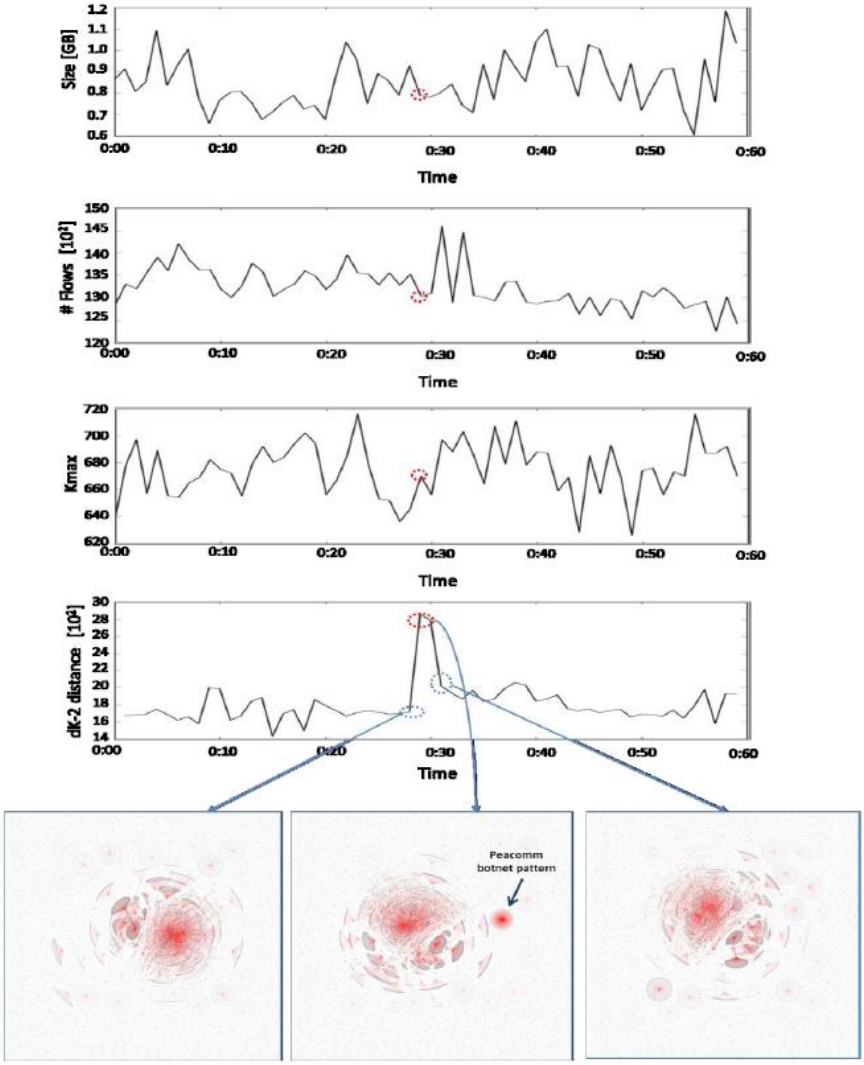

그림 8. Synthesized P2P Botnet Trace 분석

실시간 비정상 트래픽 탐지 시스템의 상세 모듈들은 그림 10 에 소개되어 있다. 먼저 NG-MON2 의 모듈 중 하나인 Flow generation 모듈 에서 생성된 플로우 정보는 트래픽 분산 그래프(Dot format)의 형태로 전환된다. 그 후 Graph metrics analysis 모듈이 분산 그래프의 그래프 지표들을 계산한 후 그 결과를 시스템의 데이터베이스에 저장하고 후에 실시간으로 결과를 출력하기 위해 RRD tool[17]의 데이터베이스에도 그 결과를 저장한다. Anomaly classification 모듈은 데이터베이스에 저장된 분산 그래프의 그래프 변수들이 미리 정해놓은 threshold 값보다 큰지를 비교한 후 현재 시스템의 정상 또는 비정상 여부를 결정한다. 만약 시스템이 비정상 트래픽을 탐지한다면 Notification 모듈이 네트워크 관리자에게 email 과 SMS 를 통해 그 내용을 전달한다. 분석된 모든 결과 값들은 관리자가 쉽게 파악할 수 있도록 User interface 모듈을 통해 웹 상에서 시각화된 형태로 제공된다.

5.1.1. Flow generation

본 연구에서는 플로우를 같은 5-tuple 헤더로 이루어진 패킷들의 집합으로 정의 한다. 5-tuple 은 source IP address, destination IP address, protocol number, source port, destination port[18]로 구성된다. 새로운 패킷이 Flow generator 에 의해 수집되면 Flow generator 는 해당 패킷에 대응되는 기존의 flow data 가 있는지 flow table 을 검색하고 만약 존재한다면 기존의 flow table 에 새로운 패킷의 정보를 추가하여 대응하는 flow 의 총 패킷 수, 바이트 크기 등의 통계 수치들을 업데이트 하게 된다. 만약 새로운 패킷에 대응되는 기존 flow data 가 존재하지 않는다면 Flow generator 는 새로운 flow data 를 생성한다. Flow generator 는 기본적으로 매 분마다 수집한 flow data 를 Flow store 모듈로 전송한다.

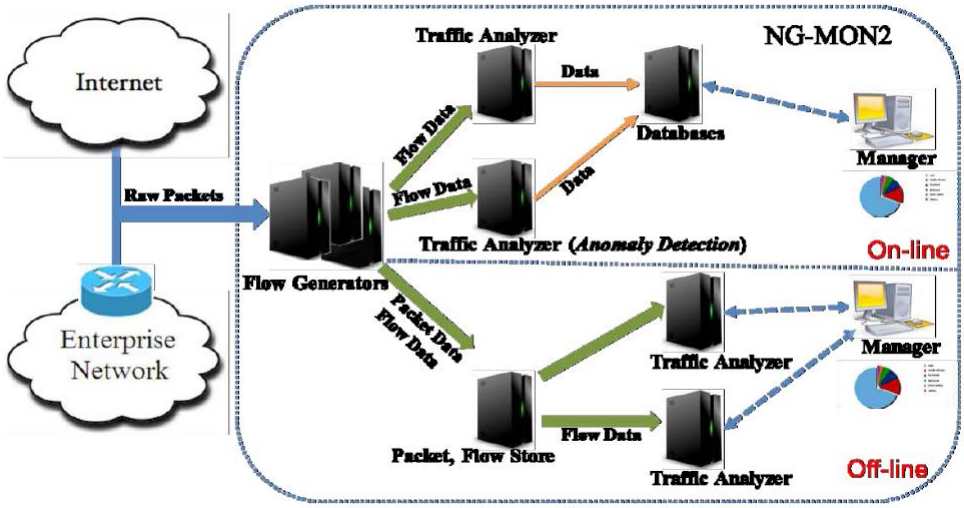

그림 9. NG-MON2 트래픽 모니터링 시스템

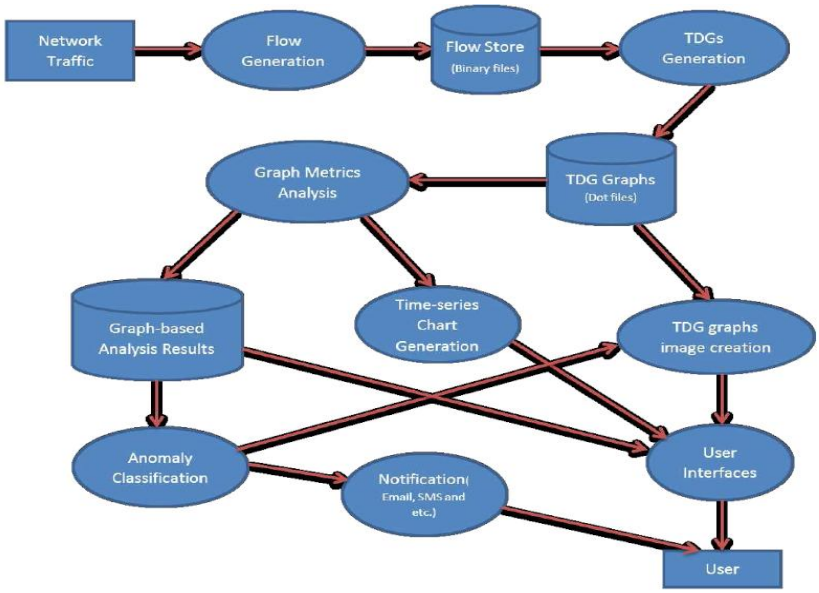

그림 10. 실시간 비정상 트래픽 탐지 시스템의 Function Diagram

5.1.2. Flow store

Flow generator 에서 생성된 flow data 는 로컬 파일 시스템에 매 분마다 새로운 이름으로 저장된다 (예: flow_2012_10_20_09_00). 각각의 flow record 는 48bytes 크기의 binary format 으로 이루어져 있으며 POSTECH 네트워크에서는 매 분마다 평균 6Mbytes 크기의 플로우 데이터가 생성된다. Flow store 는 일정 시간 동안 이 binary flow data 를 유지한 후 이를 삭제 하게 되어 있는데, 예를 들어 NG-MON2 시스템은 flow data 를 저장한 후 72 시간 뒤 자동 삭제함으로써 디스크 용량을 효율적으로 사용할 수 있게 했다.

5.1.3. Traffic dispersion graph generation

TDG generation 모듈은 매 분마다 Flow store 모듈에 flow data file 정보를 요청해 그로부터 dot file 을 생성한다. Dot file 은 각각의 flow data file 에 대응하는 새로운 이름으로 저장된다(예: graph_2012_10_20_09_00). NG-MON2 시스템은 매분 간격으로 트래픽을 수집하고 분석하는 시스템이기 때문에 dot file 생성 작업은 1 분 이내에 이루어져야 한다.

5.1.4. Graph metrics analyzer

Graph metrics analyzer 모듈에서는 먼저 그래프 지표들을 계산하기 위해 트래픽 분산 그래프의 node 와 edge 들의 관계를 인접 행렬(adjacency matrix)로 나타낸다. 매 분마다 얻어 지는 각각의 트래픽 분산 그래프에 대한 그래프 지표 값들은 시스템의 데이터베이스와 RRD tool[17]의 데이터베이스에 저장된다.

5.1.5. Anomaly classification

Anomaly classification 모듈은 Graph metrics analyzer 모듈에서 얻은 결과들을(Kmax 와 dK-2 distance) 정상 네트워크 트래픽의 threshold 와 비교하여 현재 시스템의 상태를 결정한다. 일반적으로 네트워크에서 outlier 를 탐지하는 기법들이 이 과정에서 사용된다[19]. 본 시스템에서는 장기간의 POSTECH 트래픽을 수집 및 분석하고 그로부터 얻은 고정 threshold 를 사용해 비정상 여부를 판단하였다.

5.1.6. Notification

만약 비정상 네트워크 트래픽이 탐지 된다면 Notification 모듈은 시스템 관리자에게 email 과 SMS 를 통해 시스템의 상태를 통보한다. 시스템 관리자에게 전달되는 메시지는 비정상 트래픽 이벤트의 발생 시각, 비정상 트래픽의 종류, 그래프 지표값, 시각화된 트래픽 분산그래프 등을 포함한다.

5.1.7. User interface

User interface 모듈은 지금까지 분석한 트래픽 정보를 출력하는 역할을 담당한다. User interface 는 웹의 형태로 제공되며 매 분마다 실시간으로 업데이트 된다(그림 11). Flow store 에는 이미 모든 패킷들의 헤더 정보가 플로우 별로 저장되어 있기 때문에 사용자는 만약 비정상으로 분류된 플로우가 있다면 해당 트래픽에 대한 구체적인 정보를 역으로 추적 할 수 있다.

5.2. 시스템 테스트

제안한 시스템이 실시간으로 발생하는 비정상 트래픽을 탐지할 수 있는지 검증하기 위해 POSTECH 캠퍼스 네트워크 내부 호스트에서 TCP Port Scanning Tool[20]을 사용해 외부 네트워크로 포트 스캐닝 공격을 실행하였다. 포트 스캐닝 공격은 하루 중 임의의 시각에 실행되도록 미리 설정 되었고 테스트에서는 100 개의 포트 스캐닝 인스턴스를 생성했다. 테스트 결과 시스템은 실시간으로 임의의 시각에 생성된 모든 포트 스캐닝 공격들을 Detection Rate = 100%, False Positive Rate = 0(그림 11) 의 정확도로 탐지 하였다.

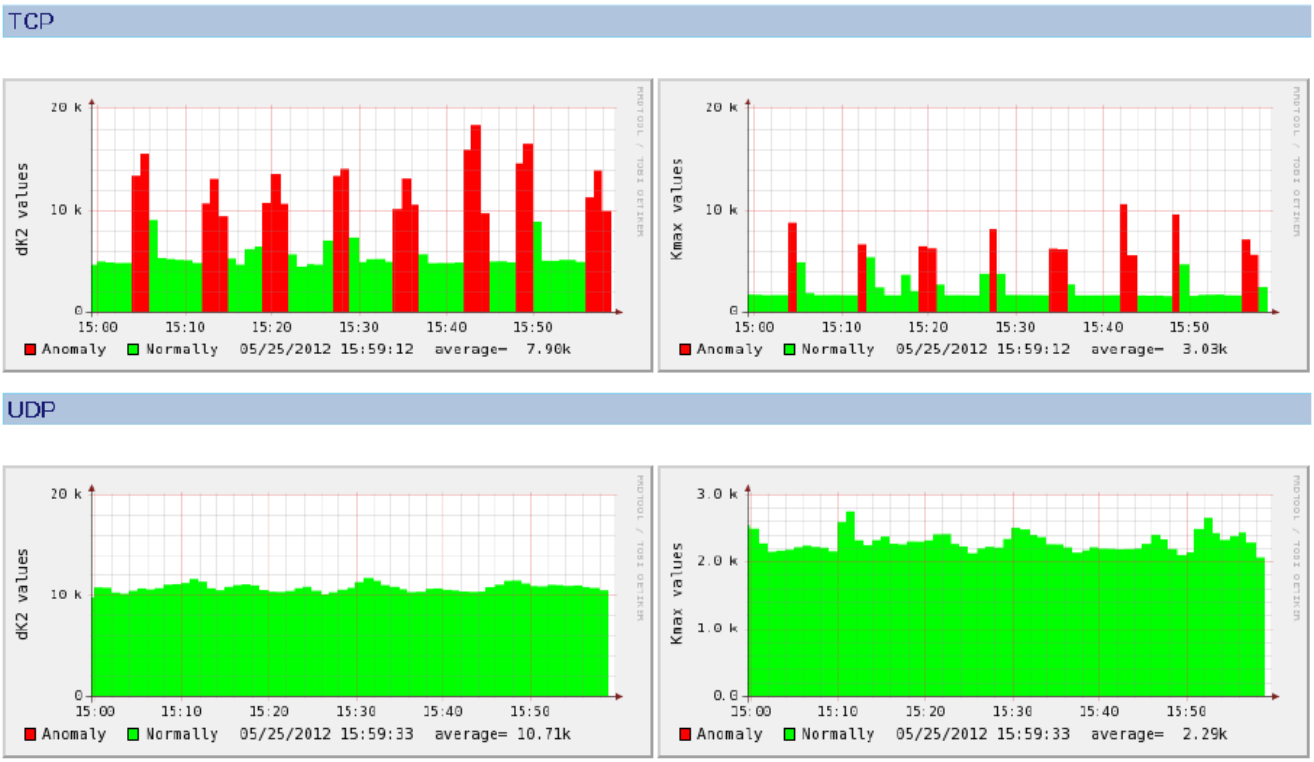

그림 11. TCP 포트 스캐닝 동안의 dK-2 distance 와 Kmax 값

6. 결론

본 연구에서는 시간에 따른 호스트들의 커뮤니케이션 패턴 변화를 그래프 지표를 기반으로 분석함으로써 실시간으로 비정상 트래픽을 탐지하는 방법에 대해 소개하였다. 제안한 방법은 트래픽 분산 그래프로부터 얻은 그래프 지표를 사용하여 네트워크 구조의 변화를 추적해서 네트워크 트래픽의 이상 여부를 확인한다. 제안한 내용을 바탕으로 구현한 시스템은 NG-MON2 시스템을 통해 얻은 POSTECH 네트워크 플로우 정보를 바탕으로 동작하는데 캠퍼스 네트워크에 구현 및 테스트 한 결과 실시간으로 대량의 플로우 정보를 분석하고 비정상 트래픽을 탐지하는 것이 가능했다. 제안한 dK-2 distance 지표는 다른 통계 기반의 비정상 트래픽 탐지 기법과는 달리 네트워크 구조적 특성의 변화를 분석하는 것에 초점을 둔다. 검증 과정에서 2009 년 7.7 DDoS 당시 POSTECH DDoS Trace 를 이용하여 제안한 방법이 DDoS 공격을 탐지할 수 있음을 확인하였다. 또 다른 검증에서 dK-2 distance 지표는 P2P Peacomm Botnet Trace 에서 감염된 봇들의 커뮤니케이션 구조 변화를 쉽게 파악함으로써 제안하는 시스템이 트래픽의 양적 변화를 크게 수반하지 않는 P2P Botnet 탐지에도 적합함을 보였다. 마지막 검증 과정에서는 POSTECH 에서 외부 네트워크로 TCP 포트 스캐팅 공격을 실행했을 때 생기는 트래픽의 변화를 제안한 시스템이 실시간으로 탐지할 수 있음을 확인할 수 있었다. 향후 연구에서는 좀 더 다양한 공격을 포함하고 있는 데이터를 바탕으로 본 시스템과 기존에 제안된 다른 비정상 트래픽 탐지 기법들과의 성능을 비교할 예정이다.

7. 참고 문헌

[1] Sachdeva M, Singh G, Kumar K, Singh K., “DDoS incidents and their impact: A review”, International Arab Journal of Information Technology, vol. 7, pp. 14-19, 2010
[2] Martin R., “Snort - lightweight intrusion detection for networks”, In LISA ’99: Proceedings of the 13th USENIX, pp. 229-238, 1999
[3] Holz T, Steiner M, Dahl F, Biersack E, Freiling F., “Measurements and mitigation of peer-to-peer based Botnets: a case study on storm worm”, In Proceedings of 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats, April 2008.
[4] Iliofotou M, Pappu P., “Network monitoring using traffic dispersion graphs (tdgs)”, In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement (IMC ’07), pp. 315-320, 2007
[5] Iliofotou M, Faloutsos M, Mitzenmacher M., “Exploiting dynamicity in graph-based traffic analysis: techniques and applications”, In Proceedings of the 5th international conference on Emerging networking experiments and technologies (CoNEXT ’09), pp. 241-252, 2009
[6] Le DQ, Jeong T, Roman HE, Hong JWK., “Traffic Dispersion Graph Based Anomaly Detection”, In Proceedings of the Second Symposium on Information and Communication Technology (SoICT), pp. 36-41, 2011
[7] Papadimitriou P, Dasdan A, Garcia-Molina H., “Web graph similarity for anomaly detection”, Journal of Internet Services and Applications, Vol. 1, pp.19-30, 2010
[8] Sala A, Cao L, Wilson C, Zablit R, Zheng H, Zhao B., “Measurement-calibrated graph models for social network experiments”In WWW, pp. 861-870, 2010
[9] Mahadevan P, Hubble C, Krioukov D, Huffaker B, Vahdat A., “Orbis: rescaling degree correlations to generate annotated Internet topologies”, SIGCOMM Computer Communications Review, pp.325-336, 2007
[10] Hong JWK., “Internet traffic monitoring and analysis using NG-MON”, The 6th International Conference, Advanced Communication Technology, Vol.1 , pp. 100-120, 2004
[11] Gansner E, Koutsoos E, North S., “Drawing graphs with dot”, http://www.graphviz.org/Documentation/dotguide.pdf.
[12] Whitney D., “Basic Network Metrics”, Lecture note, http://ocw.mit.edu/courses/engineeringsystems-division/esd-342-network-representations-of-complex-engineering-systems- spring-2010/readings/MITESD 342S10 ntwk metrics.pdf
[13] Graphviz graph visualization software, http://www.graphviz.org/
[14] Kristoff J. Botnets., 32nd Meeting of the North American Network Operators Group, October, 2004.
[15] Grizzard JB, Sharma V, Nunnery C, Kang BB, Dagon D., “Peer-to-peer botnets: Overview and case study”, In HotBots 07 Conference, 2007.
[16] Nunnery C, Kang BB., “Locating Zombie Nodes and Botmasters in Decentralized Peer-to-Peer Botnets”,https://www.os3.nl/media/2007-2008/students/matthew steggink/rp1/p2pdetect conceptpaper.pdf?id=2007-2008.
[17] Oetiker T., RRDTool. http://oss.oetiker.ch/rrdtool/, 2005.
[18] Kim M., “Internet application traffic monitoring and analysis”, PhD Thesis, Dept. of Computer Science and Engineering, Pohang University of Science and Technology (POSTECH), 2004.
[19] Barford P., “A signal analysis of network traffic anomalies”, ACM SIGCOMM Internet Measurement Workshop, pp. 71-82, 2002.
[20] S.K., “Software : TCP Port scanner”, 2011. http://download.cnet.com/Tcp-Port-Scanner/3000-2085 4-75059297.html.

정 태 열
2011 성균관대학교, 컴퓨터공학과 학사
2011 ~ 현재 포항공과대학교, 컴퓨터공학과 통합 과정
<관심분야> 정보 중심 네트워크, 네트워크 보안

Le Quoc Do
2008 Hanoi University of Science and Technology, BSc in Information Technology
2010 ~ 2012 포항공과대학교, 정보전자융합공학부 석사
2012 ~ 현재 다산네트웍스
<관심분야> 비정상 트래픽 탐지, 네트워크 보안

홍 원 기
1983 Univ. of Western Ontario, BSc in Computer Science
1985 Univ. of Western Ontario, MS in Computer Science
1985 ~ 1986 Univ. of Western Ontario, Lecturer
1986 ~ 1991 Univ. of Waterloo, PhD in Computer Science
1991 ~ 1992 Univ. of Waterloo, Post-Doc Fellow
1992 ~ 1995 Univ. of Western Ontario, 연구교수
1995 ~현재 포항공과대학교 컴퓨터공학과 교수
2007~2011 포항공과대학교 정보통신대학원장
2007~2010 포항공과대학교 정보통신연구소 연구소장
2008~2010 포항공과대학교 컴퓨터공학과 주임교수
2008~2012 포항공과대학교 정보전자융합공학부장
2008~현재 포항공과대학교 정보전자융합공학부 교수