제21회 한국정보처리학회 춘계학술발표대회 논문집 제11권 제1호 (2004. 5)

네트워크 트래픽 분석과 기계학습에 의한 DDoS 공격의 탐지

이철호, 김은영, 오형근, 이진석
국가보안기술연구소
e-mail : chlee@etri.re.kr

Detection of DDoS Attacks through Network Traffic Analysis and Machine Learning

Cheol-Ho Lee, Eun-Young Kim, Hyung-Geun Oh, Jin-Seok Lee
National Security Research Institute, Daejeon, Korea

요 약

본 논문에서는 분산 서비스거부 공격(DDoS)이 발생할 때 네트워크 트래픽의 특성을 분석하기 위해서 트래픽 비율분석법(TRA: Traffic Rate Analysis)을 제안하고 트래픽 비율분석법을 통해서 분석된 다양한 유형의 DDoS 공격의 특성을 기계학습(Machine Learning)을 이용해서 DDoS 공격의 탐지규칙을 생성하고 그 성능을 측정하였다. 트래픽 비율분석법은 감시대상 네트워크 트래픽에서 특정한 유형의 트래픽의 발생비율을 나타내며 TCP flag rate 와 Protocol rate 로 구분된다. 트래픽 비율분석법을 적용한 결과 각각의 DDoS 공격 유형에 따라서 매우 독특한 특성을 가짐을 발견하였다. 그리고, 분석된 데이터를 대상으로 세 개의 기계학습 방법(C4.5, CN2, Naïve Bayesian Classifier)을 이용해서 DDoS 공격의 탐지규칙을 생성하여 DDoS 공격의 탐지에 적용했다. 실험결과, 본 논문에서 제안된 트래픽 비율분석법과 기계학습을 통한 DDoS 공격의 탐지방법은 매우 높은 수준의 성능을 나타냈다.

1. 서론

Yahoo, eBay, E*Trade 와 같은 웹사이트에 대한 분산 서비스 거부(Distributed Denial of Service: DDoS) 공격으로 인한 피해사례에서 보듯이 인터넷에 개방되어 있는 시스템들은 DDoS 공격에 매우 취약하다[10]. DDoS 공격은 짧은 시간 내에 대량의 패킷을 공격대상 호스트 또는 네트워크에 보냄으로써 그들이 제공하는 서비스를 일시적으로 중단시키거나 심지어 시스템의 고장으로까지 유도하게 된다.

이들 DDoS 공격은 인터넷에 개방되어 있으면서 동시에 한정된 자원(네트워크의 대역폭, 시스템의 패킷 처리 용량, 시스템에 도착한 패킷의 처리를 위하여 이용되는 시스템의 제반 자원 등)을 가진 모든 시스템들을 쉽게 공격의 대상으로 하여 피해를 입힌다는 점에서 매우 심각하게 여겨지고 있으며, 이에 대한 다양한 연구가 진행되고 있다[1,5,7,10,12].

본 논문에서는 DDoS 공격 트래픽의 특성을 효과적으로 분석하기 위해서 트래픽 비율분석법(TRA: Traffic Rate Analysis)을 제안한다. 트래픽 비율 분석법은 전체 트래픽에서 특정한 유형의 트래픽이 차지하는 비율(Rate)을 의미하며 TCP flag rate 와 Protocol Rate 로 구분된다. 이 논문에서는 트래픽 비율 분석법을 이용해서 웹서비스 트래픽과 DDoS 공격 트래픽을 분석했으며 그 결과 웹서버에 대한 DDoS 공격이 발생했을 때 정상적인 웹서비스 트래픽과 DDoS 공격 트래픽의 뚜렷한 차이점을 발견할 수 있었다. 또한, 발견된 DDoS 공격 트래픽의 차이점을 세 개의 기계학습 알고리즘(CN2, C4.5, Naïve Bayesian Classifier)에 적용함으로써 DDoS 공격 탐지규칙을 생성해냈다[13,14,15]. 생성된 DDoS 공격 탐지규칙은 다양한 유형의 DDoS 공격의 탐지에서 매우 높은 성능을 나타냈다.

본 논문의 2 장에서는 DDoS 공격의 탐지에 관련된 연구들에 대해서 소개하며, 3 장에서는 본 논문에서 제안하는 트래픽 비율분석법에 대해 소개하고, 4 장에서는 실험환경에 대해서 설명한다. 5 장에서는 실험을 통해서 웹서버에 대한 DDoS 공격이 발생했을 때 웹서비스 트래픽과 DDoS 공격 트래픽의 차이점을 보이고

기계학습을 적용한 DDoS 공격의 탐지성능에 관해서 설명한다. 마지막으로 6 장에서는 결론 및 앞으로의 연구방향에 대해 언급한다.

2. 관련연구

대부분의 DDoS 공격 도구들은 공격자의 위치가 드러나는 것을 피하기 위해서 DDoS 공격 패킷의 Source IP Address 를 Spoofing 해서 사용한다. 여기에서 Spoofing 되는 Source IP Address 는 무작위로 생성되므로 이러한 특성을 이용하면 DDoS 공격을 탐지할 수 있을 것이다. 이 방법에 기반하여 Kolmogorov Complexity 를 사용하는 DDoS 공격을 탐지하는 방법이 소개되었다[6]. 이 방법은 일반적인 상황에서의 네트워크 서비스는 패킷의 Source IP Address 가 지역성(Locality)을 가지고 있으나, DDoS 공격이 발생하면 Flooding 패킷의 Source IP Address 가 무작위로 spoofing 되기 때문에, Kolmogorov Complexity 의 값이 서로 다르게 측정되는 성질을 이용한다.

그리고, 특정 subnet 내부로 들어가는 패킷 비율(to-rate) 과 밖으로 나가는 패킷 비율 (from-rate) 의 불균형 현상(disproportion)을 이용해서 DDoS 공격을 탐지하는 방법도 있다[8,9]. 임의의 Subnet P 에 대해서 R(P)를 to-rate 와 from-rate 의 비율로 가정하면, Rmin < R(P) < Rmax 인 경우를 정상적인 상태, 그 외의 경우를 DDoS 공격 상태로 인식한다. 이 방법 역시, DDoS 공격에 사용되는 공격 패킷의 Source IP Address 가 무작위로 Spoofing 되는 현상을 이용하며 DDoS 공격 패킷에 대한 Victim 의 응답 패킷이 DDoS 공격이 시작된 방향이 아닌 다른 방향으로 빠져나가는 것을 모니터링함으로써, DDoS 공격을 탐지할 수 있다.

하지만, 이런 종류의 탐지방법은 공격자가 Souce IP Address 에 대한 Spoofing 의 정도를 변경하거나 실제의 주소를 사용하는 경우, 제대로 동작하지 못할 것임이 분명하다.

3. 트래픽 비율분석법

트래픽 비율 분석법이란 전체 트래픽에서 특정한 형태를 가진 트래픽의 비율(Rate)을 이용해서 트래픽을 분석하는 방법을 말하며 TCP 플래그 비율(TCP flag rate)과 프로토콜 비율(Protocol rate)로 구분된다. [수식(1)] 은 TCP 플래그 비율(TCP flag rate)을 정의하고 있다.

$$R_{td}[F\,i\,|\,o] = \frac{\sum flag(F)\ in\ a\,TCP\ \ header}{\sum TCP\ \ packets} \quad (1)$$

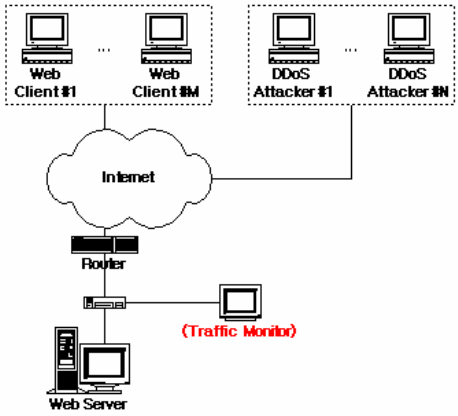
TCP 플래그 비율(TCP Flag Rate)은 TCP 패킷만을 대상으로 하며, 특정한 TCP 플래그를 가진 패킷의 개수를 전체 TCP 패킷의 개수로 나눠서 구할 수 있다[1]. 수식에서 사용되는 TCP 플래그는 S, F, R, A, P, U, N 이며, 각각 SYN, FIN, RST, ACK, PSH, URG, NULL[2]을 의미한다. 그리고, td 는 트래픽 비율을 측정하기 위한 시간 간격을 의미한다. 예를 들어서, R_1[Si]는 Inbound 방향성을 갖는 트래픽, 즉 외부에서부터 측정 대상 호스트로 향하는 트래픽을 대상으로 1 초 단위로 측정했을 때, 전체 TCP 패킷에서 SYN 플래그를 갖는 TCP 패킷의 발생비율을 의미한다.

$$R_{td}[[TCP\,|\,UDP\,|\,ICMP]\,i\,|\,o] = \frac{\sum[TCP\,|\,UDP\,|\,ICMP]\,packets}{\sum IP\ \ packets} \quad (2)$$

[수식(2)]에서 정의하는 프로토콜 비율(Protocol Rate)은 특정한 프로토콜(4 계층 프로토콜: TCP, UDP, 또는 ICMP)을 갖는 패킷의 개수를 전체 IP 패킷의 개수로 나눠서 구할 수 있다. 마찬가지로, td 는 측정단위 시간을 의미한다. 예를 들어서, R_1[TCPo]는 Outbound 방향성을 갖는 전체 IP 패킷에서 TCP 패킷의 발생비율을 1 초에 1 회씩 측정한 결과값을 의미한다.

4. 실험환경

[그림 1]은 실험을 위해서 사용한 네트워크의 구성을 나타낸 것이다.



[그림 1] 실험을 위한 네트워크의 구성

실험에서 사용된 트래픽 모니터(Traffic Monitor)는 libpcap 을 이용해서 개발했으며 Promiscuous 모드로 동작해서 네트워크 트래픽을 모두 캡쳐하여 트래픽 비율 분석법에 의해서 통계치를 계산하는 패킷 수집모듈과 기계학습 방법에 의해서 생성된 DDoS 공격 탐지 규칙을 가지고 DDoS 공격의 여부를 판단하는 DDoS 공격 탐지모듈로 구성된다. 웹 클라이언트(Web Client)는 4 개의 호스트로 구성하여 SPECweb99 를 사용했고, DDoS 공격도구는 2 개의 호스트로 구성하여 TFN2K 를 사용했다[2,4,11]. 그리고, 모든 호스트는 리눅스(Linux 2.4.18)를 사용했고 웹 서버는 아파치(Apache)를 사용했다. SPECweb99 는 보통의 웹 브라우저처럼 웹 서버를 향해서 다양한 형태의 HTTP 요청을 전송하고 웹 서버로부터 HTTP 응답을 전달 받는 역할을 한다[4].

5. 실험결과

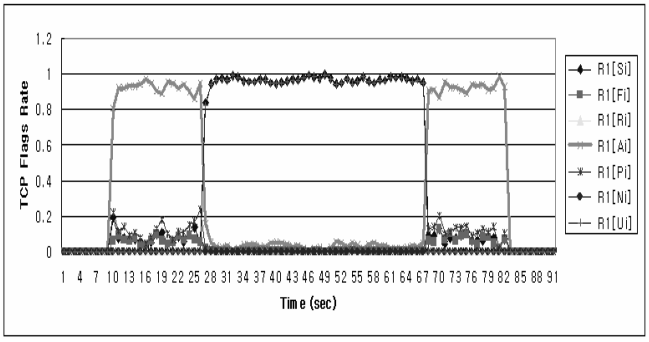
일반적인 네트워크 트래픽이 DDoS 공격으로 인해서 변화하는 모습을 분석하기 위해서는 실제의 인터

[1] TCP 패킷은 복수개의 flag를 가질 수 있으므로 TCP flag rate의 합은 1보다 큰 값이 될 수도 있다.

[2] NULL은 어떤 flag도 set되지 않은 상태를 의미한다.

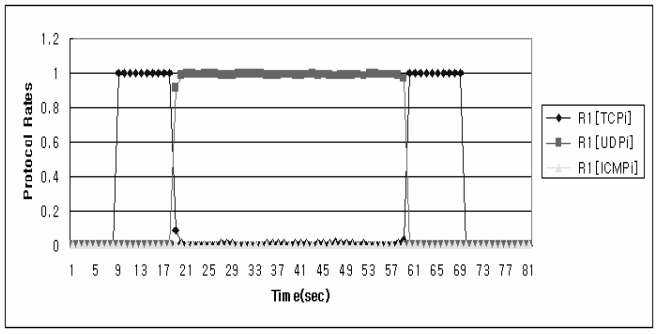
넷 환경과 거의 동일한 웹 환경을 구성해야 할 것이다. 웹 트래픽은 TCP 연결을 기반으로 이루어 진다[3]. 예를 들어서, 사용자가 웹 브라우저에서 특정 웹 주소를 입력하면 웹 브라우저와 웹 서버는 TCP 연결을 맺고, 입력된 웹 주소가 HTTP 의 GET 메소드를 통해서 대상 웹 서버에게 전달되고 그에 대한 응답으로 웹 서버는 요청 받은 문서의 내용을 클라이언트에게 전송하고 전송이 완료된 후에는 TCP 연결을 종료한다.

이처럼, 웹 서비스에 사용되는 HTTP 는 TCP 를 기반으로 한 프로토콜이므로, 하나의 TCP 연결에 수반되는 HTTP 요청의 개수(R/C: Requests per Connection)와 동시 연결 수(SC: Simultaneous Connections) 등의 값을 달리함에 따라서, 다양한 형태의 웹 서비스 환경이 만들어 질 수 있을 것이다. R/C 의 값은 1,2,5,10 을 사용했고, SC 는 5,10,50,100,150,200 을 사용했다. 따라서, 총 24 가지의 다양한 웹서비스 트래픽 환경을 사용해서 DDoS 공격이 발생할 때와 발생하지 않았을 때를 비교해서 실험했다.



[그림 2] SYN flooding attack 시의 네트워크 트래픽

[그림 2]는 정상적인 웹 서비스 트래픽이 발생되는 상황에서(9 초 ~ 83 초) SYN flooding attack 이 발생할 때(26 초 ~ 67 초)의 네트워크 트래픽의 변화를 보여주고 있다. R_1[Si]과 R_1[Ui]의 값이 급격하게 증가한다. 이것은 SYN 과 URG 플래그를 가진 패킷의 유입이 급격히 증가하기 때문에 생기는 현상이다.



[그림 3] UDP flooding attack 시의 네트워크 트래픽

[그림 3]은 19 초부터 59 초 사이에서 UDP flooding attack 이 발생한 모습을 보여준다. R_1[UDPi]의 값이 0 에서 1 로 급격하게 증가함을 볼 수 있다. 이것은 UDP flooding attack 에 사용되는 대량의 UDP 패킷이 유입되기 때문에 발생되는 현상이다.

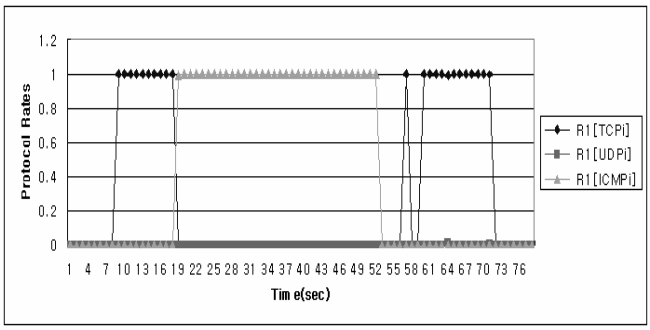


그림 4. ICMP flooding attack 시의 네트워크 트래픽

또, ICMP flooding attack 이 발생하면 위의 그림 4 와 같이 대량의 ICMP echo request 패킷이 발생하기 때문에 R_1[ICMPi]의 값이 0 에서 갑자기 1 로 증가한다.

살펴본 것과 같이, 다양한 형태의 DDoS 공격이 발생함에 따라서 네트워크 트래픽은 독특한 특성을 가지고 변화한다. 본 논문에서는 이렇게 수집된 다양한 분석데이터를 3 개의 기계학습 알고리즘(CN2, C4.5, Naïve Bayesian Classifier)에 적용함으로써 DDoS 공격 탐지규칙을 생성해냈다. [표 1]은 기계학습 결과로 생성된 DDoS 공격 탐지규칙을 보여준다.

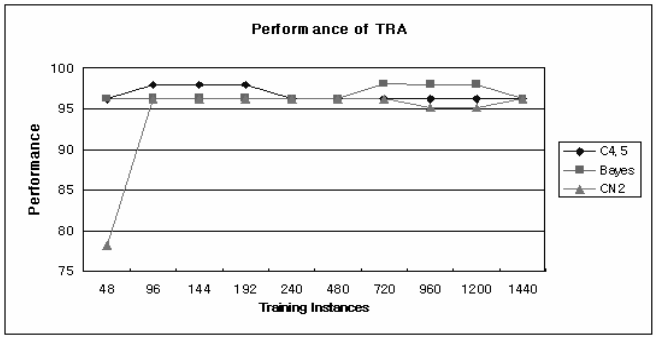
알고리즘	학습결과
C4.5	C4.5 Trees: Decision Tree: R[Si] <= 0.4 : WEB (361.0/1.0) R[Si] > 0.4 : SYN (359.0) Errors = 1/720
CN2	CN2 Decision List: IF "R[Si]" > 0.48 THEN class = "SYN" [0 359] ELSE IF "R[Si]" > 0.02 THEN class = "WEB" [332 0] ELSE IF "R[Si]" < 0.02 THEN class = "WEB" [28 0] ELSE (DEFAULT) class = "SYN" [0 1] Errors = 0/720
Naïve Bayesian Classifier	Bayes Classifier: P(class=WEB) = 0.500000 P(R[Si]|class=WEB) = G(0.071971,0.051441) P(R[Ai]|class=WEB) = G(0.928013,0.051446) P(class=SYN) = 0.500000 P(R[Si]|class=SYN) = G(0.982168,0.067275) P(R[Ai]|class=SYN) = G(0.017831,0.067275) Errors = 1/720

[표 1] SYN flooding attack 분석 데이터의 학습결과

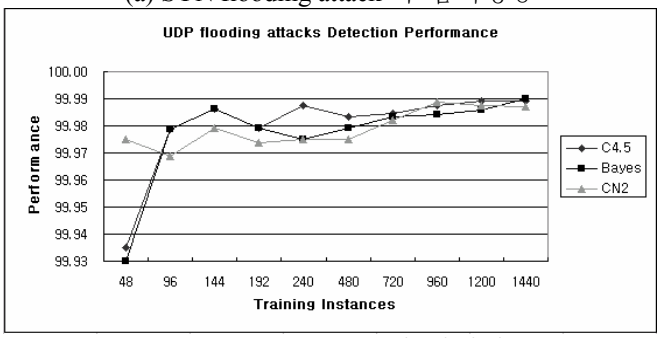
위의 표 1 에서 보듯이 학습에 사용한 3 가지 기계학습 알고리즘은 각각 DDoS 공격을 탐지할 수 있는 탐지규칙을 생성해냈다.

그리고, [그림 5]는 기계학습을 통해서 생성된 DDoS 공격 탐지규칙을 성능을 측정한 것이다. (a)에서 보듯이, SYN flooding attack 의 탐지성능은 학습데이터의 크기에 따라서 다소 차이를 보이기는 하지만 대체적으로 95% 이상의 높은 탐지성능을 보였다. 그리고, (b)와 (c)에서 보듯이, UDP flooding attack 과 ICMP flooding attack 의 탐지성능은 99.9%의 매우 높은 탐지성능을

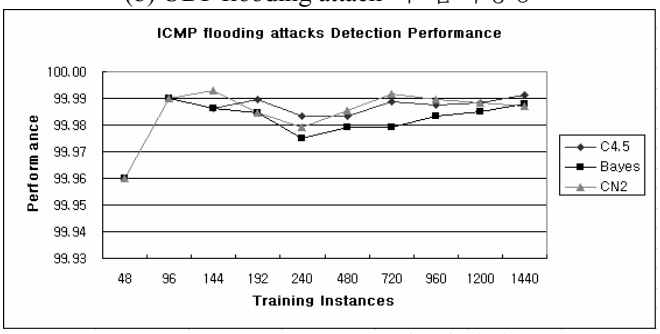
보였다. 이처럼 UDP flooding attack 과 ICMP flooding attack 의 탐지성능이 SYN flooding attack 의 탐지성능보다 월등하다. 이것은 정상적인 네트워크 트래픽으로 사용했던 웹 서비스 트래픽이 TCP 를 기반으로 하기 때문에 동일하게 TCP 를 사용하는 SYN flooding attack 의 경우보다 TCP 가 아닌 UDP 나 ICMP 를 사용하는 UDP flooding attack 과 ICMP flooding attack 의 탐지성능이 더 높게 나타난 것이다.



(a) SYN flooding attack 의 탐지성능



(b) UDP flooding attack 의 탐지성능



(c) ICMP flooding attack 의 탐지성능

[그림 5] DDoS 공격의 탐지성능

6. 결론 및 향후과제

본 논문에서는 DDoS 공격 트래픽을 분석하기 위한 방법으로 전체 패킷에서 특정한 유형의 패킷이 차지하는 비율을 분석하는 트래픽 비율분석법(Traffic Rate Analysis)을 제안했으며, 웹 서비스에 대해서 DDoS 공격이 발생하지 않은 경우와 DDoS 공격이 발생할 경우로 각각 나누어서 주기적으로 트래픽 비율(TCP Flag Rate, Protocol Rate)을 모니터링 하였다. 그 결과, DDoS 공격은 특징적인 패턴이 있음을 밝혔다. 그리고, 밝혀진 DDoS 공격의 패턴을 기계학습의 학습데이터로 사용하여 세 개의 기계학습 알고리즘(C4.5, CN2, Naïve Bayesian Classifier)에 적용하여 DDoS 공격 탐지규칙을 생성하였고 생성된 탐지규칙은 매우 높은 수준의 탐지성능을 보임을 입증하였다. 하지만, 본 논문에서 제시한 방법은 웹 서버를 대상으로만 실험되었고 네트워크 트래픽 생성기로 만들어낸 인위적인 환경을 통한 실험이라는 점에 그 한계가 있다고 하겠다.

향후 과제로는 웹 서버 이외에도 SMTP 또는 FTP 서버와 같은 보다 다양한 트래픽 환경에 대해서도 본 논문에서 제안된 방법을 적용함으로써 다양한 형태의 네트워크에 대한 DDoS 공격의 패턴을 규명하고 DDoS 공격을 탐지하는 방법을 연구할 것이다.

참고문헌

[1] Kevin J. Houle, George M. Weaver, “Trends in Denial of Service Attack Technology”, CERT Coordination Center, October 2001.

[2] Packet Storm, “DDoS Attack Tools”, http://packetstorm.widexs.nl/distributed/indexdate.shtml, 2002.

[3] Fielding, R., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and Bernerslee, T., 1999. “Hypertext Transfer Protocol - HTTP/1.1”, Tech. Rep. RFC 2616 (June), IETF. http://www.ietf.org/rfc/rfc2616.txt

[4] Standard Performance Evaluation Corporation, “SPECweb99 benchmark”, http://www.spec.org/osg/web99, August 2000.

[5] Haining Wang, Danlu Zhang, and Kang G. Shin, “Detecting SYN Flooding Attacks”, In Proceedings of IEEE INFOCOM 2002, 2002.

[6] A.B. Kulkarni, S.F. Bush, and S.C. Evans, “Detecting Distributed Denial-of-Service Attacks Using Kolmogorov Complexity Metrics”, GE Research and Development Center, December 2001.

[7] Allen Householder, Art Manion, Linda Pesante, and George M. Weaver, “Managing the Threat of Denial-of-Service Attacks”, CERT Coordination Center, October 2001.

[8] Thomer M. Gil and Massimiliano Poletto, “MULTOPS: a data-structure for bandwidth attack detection”, In Proceedings of the 10th USENIX Security Symposium, pp. 23-38, August 2001.

[9] Thomer M. Gil, “MULTOPS: a data structure for denial-of-service attack detection”, Master thesis, Division of Mathematics and Computer Science, VRIJE University, December 2000.

[10] L. Garber, “Denial-of-Service Attacks Rip the Internet”, IEEE Computer, pp. 12-17, April 2000.

[11] The Tcpdump Group, “LIBPCAP 0.6.2”, http://www.tcpdump.org, June 2001.

[12] Pars Mutaf, “Defending against a Denial-of-Service Attack on TCP”, In Proceedings of the 2nd International Workshop on Recent Advances in Intrusion Detection(RAID’99), 1999.

[13] Clark, P. and Niblett, T. “The CN2 Induction Algorithm”, Machine Learning Journal 3(4), pp. 261-283, 1989.

[14] Hanson, R., Stutz, J., and Cheeseman, P. “Bayesian Classification Theory”, Technical Report FIA-90-12-7-01, NASA Ames Research Center, AI Branch, 1991.

[15] Quinlan, J.R. “C4.5: Programs for Machine Learning”, Morgan Kaufmann Publishers, 1993.