

LAN 환경에서의 트래픽 해석에 관한 연구

正會員 李 鍾 榮*, 吳 英 煥**

A Study on The Traffic Analysis in LAN Environment

Joung-Young Lee*, Young Hwan Oh** *Regular Members*

요 약

본 논문에서는 실제 구축되어 있는 Ethernet LAN상의 데이터 트래픽의 특성을 연구하였다. 망상의 패킷 도착 형태로는 포아송 분포가 아니라 weibull 분포가 적합함을 보였다. 망의 트래픽을 분석하면서 패킷간 전송특성이 'source locality'로 나타나며 망상의 트래픽중 약 92%가 화일전송임을 알 수 있었다. 본 결과는 통신망의 모델링에 유용할 것으로 기대 된다.

ABSTRACT

The characteristics of data traffic on the Ethernet LAN are investigated on the basis of measurements. The analysis on the arrival pattern of packets on the network is found not to be a Poission process but to be Weibull distributions. As analysis of network traffic, packet arrivals are found to exhibit a 'source locality'. It is observed that file transfers are responsible for about 92% of the traffic on the network. our results will be useful for modeling purposes.

I. 서 론

LAN(Local Area Network)은 오늘날 조직에서 그 역할의 중요성이 증대되어 조직의 중추신경 역할을 담당하고 있다. 일반적으로 LAN이란 제한된 지역내에서 사용한다는 측면에서 WAN(Wide Area Network)과는 분리할 수 있다.

망관리(Network Management)의 필요성은 망의 확장, 장애 원인 분리 및 신속한 복구등과 감시자료를 이용하여 향후 망의 최적설계에 이용하는데 있다. 특히 LAN의 경우에는 LAN 특성상 데이터가 모든 곳으로 전달되고(Ethernet), LAN에 연결된 모든 단말들이 대역폭을 공유하기 때문에 망이 커지면 커질수록 망관리의 중요성이 대두된다. 망 감시의 중요요소는 LAN의 효과적 관리를 위한 성능을 측정하는 것이다. 이는 단순히 이론적인 연구의 결과뿐만아니라 경험적인 측면이 가미되어야 될 것이다. 예를들어, 향후 패킷 도착에 관해 예측할 수 있다면, 망 설계자(관리

*광운대학교 신기술연구소
**광운대학교 전자공학부 교수
論文番號:96169-0605
接受日字:1996年 6月 5日

자)는 더 좋은 패킷 조절(packet handling) 전략을 설계할 수 있다. 또한 적어도 더욱 효과적인 방법으로 현재의 망 상태를 증진할 수 있을 것이다. 관련된 연구결과를 보면, Shoch와 Hupp[1]는 대역폭이 LAN의 성능 문제에 중요한 원인이 아니고, 짧은 데이터그램에 대한 최소의 latency가 매우 중요하다는 점을, Lazowska등은[2] 화일서버의 응용에서 성능을 다스리는 기본적인 요소가 화일서버의 CPU 속도임을 밝혔다. Falaki등은[3] 패킷 도착 상태가 weibull이나 hyperexponential 분포에 가깝다는 것을 밝혔다.

본 논문에서는 실제 구축되어 있는 LAN을 통해 데이터를 추출했다. 망의 트래픽을 분석하면서 패킷간 전송특성이 'source locality'로 나타 난다는 것을 알았다. 즉, 하나의 패킷이 A에서 B로(일반적인 LAN 환경에서는 client와 server간의 관계이다.) 전송된다면, 다음 패킷이 전송될 확률은 $A \Rightarrow B$ 또는 $B \Rightarrow A$ 가 매우 높다. 또한 원천 데이터를 활용하여 패킷 트래픽에 다양한 bustiness의 존재를 알아냈고, 기존 연구 결과 [3]와 시험 결과를 갖고 적당한 패킷 도착시간의 확률 분포를 찾았다.

II. 트래픽 시험

2.1 환경

실제적인 데이터를 추출하기 위한 Ethernet LAN 구성은 그림 1과 같다. LAN에 연결되어 있는 기기는 대부분 PC(486급 이상)이며 약간의 UNIX W/S(SUN 및 HP 서버 급)과 국산 주 전산기인 Ticom(DB 서버)이 있고 그 수는 약 150대이다. 또 PC급의 3대가 file/printer/image 서버(LAN Manager)로 사용하고 있다. 외부 망과의 연계는 2대의 라우터와 1대의 브릿지 및 패킷 교환기가 사용되고 있으며, LAN 상에서 주로 사용되는 DARPA 프로토콜은 Telnet, UDP, TCP, IP, ARP, FTP 등이다. 또한 이미 개발된 LAN 대응 응용업무들의 트래픽 특성을 화일전송 응용(서버의 공용 디렉토리 사용), 대화형 응용(Ticom OS), 혼합 운용(MIS/OA)으로 분리하여 볼 수 있다(표 1참조). 시험에 사용된 LAN은 B급의 IP 주소체계를 사용하며, 서버의 주소는 XX.XX.XX.10 이하(단, Ticom 주소만 159이다.)이고 클라이언트는 호스트 주소가 10 이상이다.

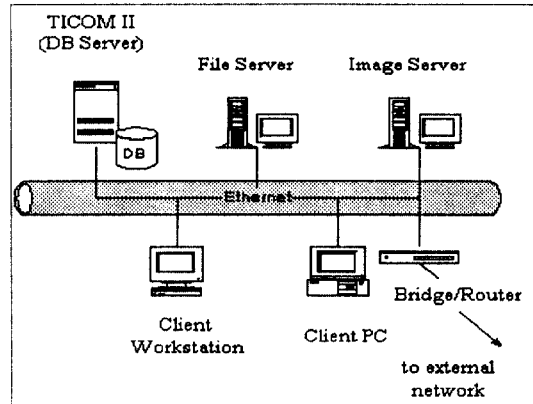


그림 1. 네트워크 구성
Fig. 1 Network ConFIGuration

2.2 원천 데이터 수집 도구

사용된 원천 데이터 수집 도구는 UB(Ungerma Bass)사의 Netscope이다.

- Netscope의 기능

- Network Activity(all packet, broadcast packet, multicast packet, packet from other networks, packet to other networks)

:Network 상에 있는 패킷의 형태를 초당 통계적으로 화면에 표시

- Throughput

:Network 상의 처리량을 30초 간격으로 화면에 표시

- Packet Display

:Network 상에 있는 패킷의 구성을 화면에 표시

- Size Distribution

:Network 상에 있는 패킷을 크기별로 10초 간격으로 화면에 표시

- 시간

일반적으로 최대 트래픽(peak time)은 출근 직후 및 퇴근 직전에 발생되기때문에 오전에 두번 원천 데이터를 수집했고, 통상의 트래픽 확인을 위해 오후에 두번 원천 데이터를 수집했다. 사용자의 망 사용이 없을시 허브(HUB), 라우터, 브리지 및 서버의 제어 패킷을 총괄적으로 파악하기 위해 한 번 더 원천 데이터를 수집했다(물론 상기 네번의 시험에서도 알수 있지만, 제어 패킷 발생 주기가 시험 기간에 발생 안

표 1. 시스템 환경

Table 1. System environment

구성	내용
데이터베이스 서버	<ul style="list-style-type: none"> • 기종: 타이컴(SSM7000) • 1 • OS: UNIX SVR 3.2 • DBMS: ORACLE 7
파일 서버	<ul style="list-style-type: none"> • 기종: IBM 486DX2/66 호환(32M RAM, 1GB HDD) • 2 • OS: OS/2, Microsoft LAN Manager(NOS) Ver 2.2
이미지 서버	<ul style="list-style-type: none"> • 기종: IBM 486DX/33 호환 (16M RAM, 200MB HDD 600MB ODD) • 1 • OS: OS/2, Microsoft LAN Manager(NOS) Ver 2.2
통신	<ul style="list-style-type: none"> • TCP/IP(Socket) • Application: FTP, TELNET
클라이언트	<ul style="list-style-type: none"> • 기종: IBM PC 호환(486, 펜티엄) • OS: MS-DOS, MS-Windows(사용자 인터페이스)
개발도구	<ul style="list-style-type: none"> • 서버: C언어, Oracle PL/SQL • 클라이언트: C언어, Visual Basic, SQL • Windows, Clipper

될 수도 있다). 다음은 시험일자를 나타낸 것이다.

- '95.3.20, 09:42-09:47, 14:49-14:54
- '95.3.21, 09:22-09:27, 14:26-14:31
- '95.3.22, 07:10-07:15

2.3 분석내역

Netscope의 상기 기능만으로 분석을 수행할 수 없다. 따라서 Netscope의 상기 기능 중 packet display를 이용하여 망상의 비트 스트림의 결과를 파일로 받는다. 그 다음 원천 데이터가 저장된 파일을 아스키 파일로 전환하여 의미있는 정보를 찾기 위해 "C"로 코딩한 프로그램(IP_Time라 한다.)을 수행한다. 또한 IP_Time을 실행한 결과 데이터를 엑셀로 읽어들이고 그 결과를 그래픽하게 처리한다(다음절에서 상세 기능 설명).

10M Ethernet에 대한 1,000m 케이블에 전송되는 전파지연은 5µs이다. 또한 1,000bit에 대한 전송시간은 약 100µs이며, 매체의 총 지연은 0.2µs이다[4].

III. 측정 내역

3.1 송수신간의 트래픽 형태

150대의 노드가 설치되어 있는 경우 약 6643개의

source-destination(S-D)쌍의 통신이 가능하다. 이것은 LAN상에 n노드가 주어지고 패킷 스트림은 n(n-1)/2 노드 쌍이 되기 때문이다. 그러나 IP_Time을 이용하여 분석해 보면, 모든 쌍이 통신을 하는 것은 아니며 트래픽 중 대부분은 어떤 특정 서버들에게 집중되어 있다. 이것은 표 2의 S-D 트래픽 리스트로부터 파악할 수 있다. 표에서 timestamp는 시스템에서 제공하는 시간이고, source와 destination gap은 각각의 전송 시간 차이이다. 또한 E_{length}는 Ethernet 프레임 크기이고 I_{length}는 IP 패킷의 크기이며, type은 사용 프로토콜을 의미한다. 그림 2에서보면, 두 대의 파일서버(47.53%, 1.28%)와 한대의 DB서버(27.50%) 및 한대의 이미지 서버(1.03%)의 각각의 트래픽 점유율은 시험시간 동안에 망의 총 트래픽중 약 77.34%를 점유한다. 그림 2에서 건수와 데이터량의 차이는 표 3에서와 같이 UDP, Link Data(Multicast)등이 평균적으로 한번에 적은양(약 100byte 이하)을 전송하는 반면에 TCP 데이터는 한번 전송량(238byte)이 상대적으로 많기 때문이다. Netscope의 원천데이터를 IP_Time을 이용하여 분석하면 다음과 같은 정보로 구성된 S-D 트래픽 형태를 알아낼 수 있다.

- name of packet source

표 2. S-D 트래픽 목록

Table 2. List of S-D traffic

timestamp	source_addr	destin_addr	source_gap	destin_gap	Elength	Ilength	type
14:55:37.578	190.0.40.55	190.0.40.159	00:00:00.166	00:00:00.166	466	452	TCP
14:55:37.591	190.0.40.159	190.0.40.55	00:00:00.149	00:00:00.149	60	40	TCP
14:55:37.740	190.0.40.159	190.0.40.55	00:00:00.036	00:00:00.036	466	452	TCP
14:55:37.744	190.0.40.55	190.0.40.159	00:00:00.036	00:00:00.036	466	452	TCP
14:55:37.776	190.0.40.159	190.0.40.55	00:00:00.168	00:00:00.168	60	40	TCP
14:55:37.780	190.0.40.55	190.0.40.159	00:00:00.171	00:00:00.171	466	452	TCP
14:55:37.944	190.0.40.159	190.0.40.55	00:00:00.028	00:00:00.028	466	452	TCP
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
14:55:39.025	190.0.40.4	190.0.40.50	00:00:00.002	00:00:00.002	60	44	TCP
14:55:39.026	190.0.40.50	190.0.40.4	00:00:00.001	00:00:00.001	126	112	TCP
14:55:39.027	190.0.40.50	190.0.40.4	00:00:00.004	00:00:00.004	190	176	TCP
14:55:39.027	190.0.40.4	190.0.40.50	00:00:00.003	00:00:00.003	60	44	TCP

- name of packet destination
- number of message across these node pair
- inter-packet arrival times distribution
- inter-protocol arrival time distribution for each protocol type(TCP, UDP, RARP, ICMP, ARP, LINK)
- total(minimum/maximum/average) packet(frame) size
- number of records
- total records for source gap
 - minimum(maximum/average) source gap
- total records for destination gap
 - minimum(maximum/average) destination gap

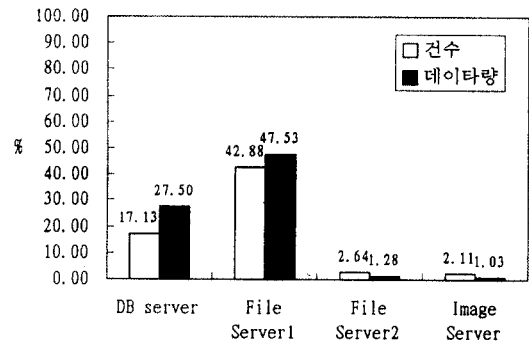


그림 2 서버별 트래픽 점유율
Fig. 2 Traffic ratio of server

상기 결과로부터 쉽게 다음과 같은 성능 변수를 추출한다.

- network load distribution
- average response time of each server or node on the network
- network utilization
- average packet size

3.2 망 트래픽 형태

망의 전반적인 상황을 알기 위해서 모든 정보가 시험 주기 동안에 포함되어 있다고 가정한다. 망에 대

한 시험 주기가 정확하게 일치하는 노드의 쌍은 아니라도 데이터는 앞節의 S-D 트래픽 형태와 같은 정보를 포함할 것이다. 분석은 이들 데이터의 집합을 통해 수행된다. 패킷간 도착간격(inter packet arrival)은 망 설계에 중요한 정보를 제공하기 때문에 분석 대상이 된다.

본 연구에서 분포의 의미있는 관찰은 5분간으로 제한된다. 그 결과 보다 작은 간격으로 해서 얻은 결과 [1, 2, 3]와 차이가 없음을 알 수 있었다. 그림 3은 패킷 도착 시간 분포와 주요 프로토콜(TCP, UDP)이 총 트래픽에 차지하는 빈도를 보여준다. 그림 3을 보면 총 트래픽 중 약 82%가 평균 패킷 도착간 간격이 25ms

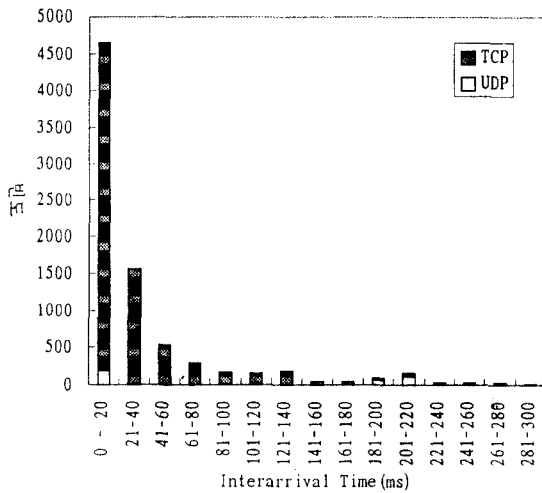


그림 3 20ms 간격에 대한 관찰된 도착간 빈도수
Fig. 3 interarrival frequencies in 20ms intervals

표 3. 통신망 프로토콜 분석

Table 3. Network protocol analysis

프로토콜 형태	트래픽 점유율(%)		평균 패킷 크기(bytes)
	건수	데이터량	
TCP	81.08	91.60	238
UDP	10.49	5.02	99
ICMP	0	0	0
Link	7.60	3.02	87
ARP	0.82	0.36	68
RARP	0	0	0

이하 임을 알 수 있다. 표 3은 프로토콜 형태별 총 트래픽 중 차지하는 비율 및 평균 패킷 크기를 보여준다. 표 3으로부터 TCP가 시험기간 동안 망의 총 부하 중 90% 이상 차지함을 알 수 있다. ICMP는 패킷의 잘못된 전달 또는 어떤 문제를 보고하기 위해 사용되고, ARP는 Ethernet 주소를 찾기 위해, RARP는 주로 터미널에서 사용되기 때문에 사용빈도가 매우 낮다.

IV. 수학적 분석

프로토콜 형태에 상관없이 총 트래픽의 분석결과 는 표 4에 보여준다. 이것은 평균, 표준편차, 도착간 시간 간격의 분산계수 및 k의 가장 낮은 세값에 대한

autocorrelation function(ACF)의 값 등이다. ACF는 식(1)과 같이 얻을 수 있다.

$$\gamma_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \text{ for } k=0, 1, 2, \dots \quad (1)$$

where \bar{Y} : Sample Mean

\bar{Y} 는 Y_1 서부터 Y_n 까지 통계적 독립으로 측정된다.

4.1 Weibull 분포

관찰된 분산계수 1.41이 포아송과정에 부적당 하다는 것을 표 4로부터 즉시 알 수 있다.

포아송 과정에서 도착간 시간 간격은 지수분포를 갖고 당연히 분산계수(coefficient of variation)는 1이다. 지수분포는 와이블 분포의 특별한 경우이다. 식 (2)와 (3)은 와이블 분포의 밀도함수와 분포함수이다.

$$f(t) = \frac{\beta t^{\beta-1}}{\alpha^\beta} e^{-(t/\alpha)^\beta} \quad (2)$$

$$F(t) = 1 - e^{-(t/\alpha)^\beta} \quad (3)$$

이것은 형태의 다양성에 대한 적합도에 사용된다. 본 연구에서 수행된 데이터 트래픽 세기(intensity)의 측정은 포아송 보다 와이블 분포에 유사함을 알 수 있다. 식(4)와 식(5)는 분포의 평균과 분산이다.

$$\mu = \frac{\alpha}{\beta} \Gamma\left(\frac{1}{\beta}\right) \quad (4)$$

$$V = \frac{\alpha^2}{\beta^2} \left[2\beta \Gamma\left(\frac{2}{\beta}\right) - \Gamma^2\left(\frac{1}{\beta}\right) \right] \quad (5)$$

여기서 $\Gamma(x)$ 는 감마 함수이다.

$$\Gamma(n) = \int_0^\infty e^{-t} t^{(n-1)} dt \quad (6)$$

형태모수(shape parameter) β 는 분포곡선의 형태를 결정하고, 그 결과 경험적 및 이론적 분포의 다양한 적합도에 사용될 수 있다. β 값의 특별한 경우는 다음과 같다.

$\beta=1$ simple 지수분포
 $\beta=2$ Rayleigh 분포
 $\beta \approx 3.44$ Weibull 분포가 정규분포에 접근할 때

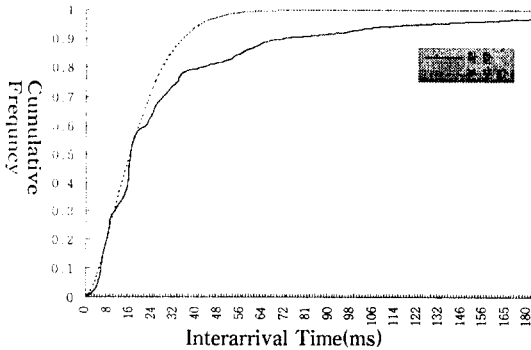


그림 4 와이불 적합도
 Fig. 4 Weibull fit

V. 결 론

150대의 노드를 갖는 Ethernet LAN 상의 트래픽 측정 결과 패킷간 전송특성이 'source locality'임을 알 수 있다. 이것은 시험 데이터의 S-D 패킷 트래픽에서 명확하게 알 수 있다. 이는 클라이언트와 서버 간 트래픽을 의미하며, 클라이언트 간의 트래픽은 존재하지 않는다. 표 3의 결과에서 보듯이 전체 트래픽에서 TCP 데이터(즉, 망의 화일전송 utilization)가 약 92% 임을 알 수 있다. 또한 TCP 데이터 중 82%가 평균 도착 간격이 25ms이하를 나타내며, 18%는 평균 도착간 간격 5.31초가 된다. 패킷 도착간 간격은 A가 송신한 후 또 다시 A가 송신할 때 까지의 간격을 의미하므로, 18%는 세션이 설정된 후 상대적으로 오랜 시간이 경과한 다음(대략 30초 이상, 제어를 위한 방송형 패킷을 제외하면 대부분 OA시스템의 계시판 기능에 사용할 경우)에 통신을 하는 경우이다.

그림 4는 패킷 도착 형태를 나타내며, 와이불 분포가 적합하다는 것을 보여준다(hyperexponential 분포도 그래프 형태가 유사하다[3]). 시각적으로도 패킷 도착이 포아송 분포가 아님을 알 수 있다.

참 고 문 헌

1. Shoch, JF and Hupp, JA, "Measured performance of an Ethernet local network", *Commun., ACM*, Vol 23 No 12, 1980
2. Lazowska, E D et al., "File access performance of diskless workstations", *ACM, Trans. Comput. Syst.*, Vol 4 No 3, 1986
3. S O Falaki and S-A Sorensen, "Traffic measurements on a local area computer network", *Computer Commun.*, Vol 15 No 3, 1992
4. Mark A. Miller, "LAN Protocol Handbook", *M&T Publishing, Inc.* 1990
5. J. F. Lawlwws, "Statistical models and methods for lifetime data". 1994
6. T. Johannsen, G. Mansfield, "A study of FTP traffic", *International journal of network management*, December 1994
7. Guy T. Almes, "Engineering Trans-oceanic 44Mbps TCP/IP networks", *Computer Networks and ISDN system* 25, 1994

이 종 영(Joung Young Lee) 정희원

1949년 2월 8일생

1973년 2월 : 광운전자공과대학 응용전자공학과(공학사)

1982년 2월 : 광운공과대학 대학원 전자통신공학과 (공학석사)

1994년 8월 : 광운대학교 대학원 전자통신공학과 박사 과정 수료

1979년 9월~1993년 6월 : 충주공업전문대학 전자통신과 교수

1993년 7월~현재 : 충주산업대학교 전자통신공학과 교수

※관심분야: 통신망 설계, 통신망 트래픽해석, SONET

오 영 환(Young Hwan Oh)

정희원

한국통신학회 논문지 제21권 제7호 참조

현재 : 광운대학교 전자공학부(전자통신) 교수