

A Study of the Reliability of Web Services using Client Sides Errors¹⁾

SangBock Lee²⁾ · MaSuk Kim³⁾

Abstract

Modeling the reliability of distributed systems requires a good understanding the reliability of the components. For thousands of web users, competitiveness in web services means a successful presence on the web. Failure rates for the presence of a web site are considered on client sides errors using RFC2068. Data were collected from some host via the internet.

Key words : reliability, web services, client sides errors, RFC2068

1. 서 론

대다수 웹 사용자들은 웹사이트 접속오류의 경험을 가지고 있다. 웹서비스 오류들 중에 서버 혹은 클라이언트 입장에서 정확히 원인을 알 수 있는 오류는 그리 많지 않다. 1990년 이후 하이퍼텍스트 전송 규약 (Hypertext Transfer Protocol, HTTP)에서 정의한 오류들을 기록한 서버 로그파일 등을 통해서만 다양한 오류 원인들을 알 수 있다. 웹 전송표준으로 현재 RFC 2068 (HTTP 1.1) 까지 나와 있으며 전 세계적으로 이 표준을 따르고 있다.

최근 플랫폼 무제약, 컴포넌트 기반을 특징으로 하는 웹서비스 개념의 출현은 웹사이트 신뢰성 문제에 대하여 다양한 연구를 필요로 하고 있다. SOAP/XML과 HTTP, TCP/IP의 분석경향은 웹서비스 로그 데이터를 분석한 단순통계치로 한 사이

1) This research was supported by the Catholic University of Daegu Research Grants in 2002

2) Professor, Faculty of Information and Environmental Sciences, Catholic University of Daegu

3) Lecturer, Institute of Continued Education, Catholic University of Daegu

트의 신뢰성을 판단하고 있다. 그러나 로그 데이터 그 자체만으로 분석할 수 있는 정보는 그다지 많지 않다. 고객 데이터와 통합한 자료분석의 경우는 고객의 프라이버시 문제나 정보 유출로 인한 문제가 발생할 수 있다(이동하, 2002). 또한 로그 파일은 불완전한 데이터이다. Long와 2인(1991)의 정의를 보면 접속실패는 분산 환경(distributed environment)에서 정의된다. 이 정의는 클라이언트가 호스트에 접속하지 못하는 것을 의미하며 호스트에 의한 하드웨어, 소프트웨어 오류, 정전과 계획된 다운타임(down time)을 포함한다. 또한 오프 사이트(off-site) 통신 실패, 물리적인 통신 연결 문제점, 정확한 경로(routing) 정보의 일시적인 부재로부터 기인한다라고 정의했다. 즉, 이들은 웹 접속 실패를 분산 환경적인 측면을 기초로 정의하였다. Kallepalli와 Tian(2001)은 웹 접속 실패는 사용자가 요청한 정보나 문서를 정확하게 전달하지 못함을 의미한다라고 정의하였다. 그리고 웹 계층(web-layer)에 따라 웹 접속 실패를 브라우저 오류(client 층), 호스트나 네트워크 오류(middleware and web server 층), 소스나 내용상의 오류(web server 층)으로 분류하였다. 또한 이 논문에서는 접속 신뢰성을 추정하기 위하여 이용한 오류는 소스나 내용상의 오류와 관련하여 생각했다. 안옥부의 2인(2002)은 웹 로그 액세스 자료에 대하여 RFC2068의 상태코드(status code)를 기본으로 하여 클라이언트 측의 접속오류 분류 및 정의를 클라이언트 입장에서 제안하였다.

본 연구에서는 SOAP/HTTP 기반의 클라이언트 측 웹서비스 오류 로그파일을 이용하여 웹서비스의 신뢰성을 구하고, 그 사례적용을 예시하였다.

2. 클라이언트 측 웹서비스 신뢰성

웹서비스의 상태코드는 요구 메시지를 이해하고 이에 따라 서비스를 제공하려는 것에 대한 결과로서 세 자리의 정수 코드이다. 상태코드의 첫 자리 숫자는 응답 클래스를 규정하며 마지막 두 자리 숫자는 아무런 구분 역할을 가지고 있지 않다. RFC 2068의 상태코드를 기준으로 코드 100에서 599까지 100단위로 5개의 그룹으로 묶여진다.

클라이언트 오류는 상태코드에서 클라이언트 오류는 400번 대의 코드이다. 클라이언트 측에서의 소프트웨어의 오류, 즉 인터넷 사용과 관련된 오류가 아닌 프로그램의 실행이나 사용과 관련된 오류와 웹브라우저의 오류로 대체로 환경설정의 오류에서 비롯된다. 즉 클라이언트 오류는 서버의 책임과는 무관하게 클라이언트와 관련된 오류이다. 서버 오류는 서버의 내부 문제로 인한 오류, 서버의 과부하로 인한 서버다운, 서버의 소스나 콘텐츠로 인한 오류이다. 상태코드에서 숫자 "5"로 시작하며 응답 상태코드는 서버가 에러를 발생시켰으며 요구를 처리할 능력이 없음을 인정한 경우를 표시한다.

안옥부의 2인(2002)에 따르면, 클라이언트 측 접속오류를 다음과 같이 정의했다. RFC 2068에서 설명한 접속 실패의 정의를 기초로 서버입장에서는 성공일지라도 클라이언트 입장에서는 요청한 정보나 문서를 정확하게 전달하지 받지 못한 경우에는 클라이언트 측 접속실패 오류로 분류하였다. 이 오류의 상태코드는 201-204, 301-304 이다. 이러한 오류 분류개념은 웹서비스 구성요소 가운데 XML문서 유통을 위한 SOAP

프로토콜과 웹접속규약 HTTP, 데이터통신규약 TCP/IP 요소의 서비스 오류로 다시 세분 정의할 수 있다.

웹서비스 신뢰성은 서비스 품질의 한 요소이다. 사용기간에 따라 기록된 서비스 오류(고장, failure)는 웹서비스의 신뢰성 척도를 나타낸다. 신뢰성은 서비스 요청자와 제공자에 의해 보내고 받은 메시지 전달의 정확도를 나타낼 수도 있다. 이런 신뢰성을 추정하기 위하여 로그 파일의 많은 기록 속성들 가운데 어떤 속성을 기준으로 추정하는지가 고려되어야 한다.

Kallepalli와 Tian(2001)는 전체 히트율과 오류율을 가지고 신뢰성을 추정하였다. 하나의 웹 페이지를 구성하는 요소는 웹 파일(html, asp, php 등), 이미지 파일(gif, jpg 등), Include 파일(js 등) 등의 파일들도 각각의 히트 수로 계산된다. 로그 파일에 저장된 모든 행이 히트의 단위가 된다. 그러므로 히트 수에 근거한 트래픽 데이터는 사용자 측면에서 웹서비스 신뢰성의 의미가 별로 없으며, 또한 신뢰성 추정량으로 문제가 있다. 그러나 페이지 뷰는 앞의 히트와 달리 사용자가 요청한 웹 페이지만을 의미한다. 보통 html, asp, php 파일 등이 이에 속한다.

보다 나은 웹서비스의 신뢰성 추정을 위하여 히트 대신 페이지 뷰와 클라이언트 오류, 서버 오류, SOAP/HTTP기반 클라이언트 측 오류를 이용하여 Nelson(1978) 신뢰성 추정량 \hat{R} (reliability)를 제안한다. Nelson 추정량 \hat{R} 은 다음과 같다.

$$\hat{R} = 1 - \frac{f}{N_{pv}} = \frac{N_{pv} - f}{N_{pv}}$$

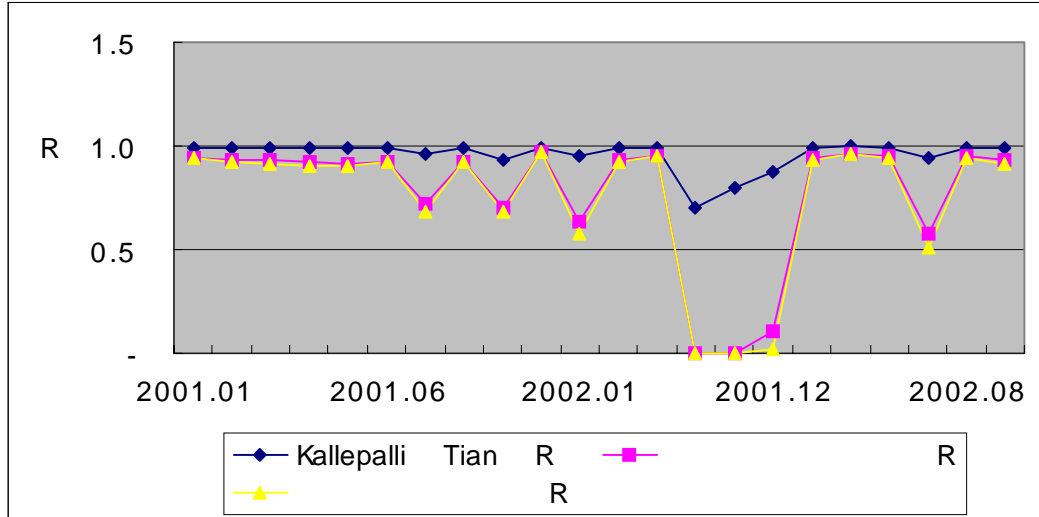
, $N_{pv} \geq f$, f : SOAP/HTTP 클라이언트측 오류, N_{pv} : 페이지 뷰.

3. 적용사례

앞 절에서 논의된 SOAP, HTTP&TCP/IP 표준에 따른 클라이언트측 오류자료는 공개된 로그 파일(URL: <http://www.logasp.com>)을 사용하였다. 이 자료에 대하여 3 종류의 웹서비스 측면의 신뢰성 추정량을 계산해 본다. 사례적용 데이터는 2000년 12월~2002년 9월까지 자료이다.

첫 번째 Kallepalli와 Tian(2001)에 의한 추정치 $\hat{R} = 0.969$, 두 번째 상태코드에 의한 오류를 이용하여 구한 추정치 $\hat{R} = 0.774$, 세 번째 본 연구에서 제시한 오류를 이용하여 구한 추정치 $\hat{R} = 0.740$ 이다.

<그림 1>은 위의 세 가지 웹서비스 신뢰성 R의 추정치를 그래프로 나타내었다. Kallepalli와 Tian (2001)에 의한 추정치 \hat{R} 은 히트 수를 이용하여 추정한 것으로 페이지 뷰를 이용한 다른 추정치보다 신뢰성이 더 높게 나온 것처럼 보인다.



<그림 1> 웹서비스 신뢰성 R의 추정치

4. 결 론

e-business 상의 웹서비스 사용증가와 더불어 많은 웹서비스 오류들이 발생한다. 본 연구에서는 하이퍼텍스트 전송규약 RFC2068의 상태코드를 기준으로 클라이언트 측 오류 분류에 따라 웹서비스 신뢰성 R을 구하였다. 그러나 현재 대부분의 웹 마이닝(web mining)은 미리 보고해 놓은 고정된 웹 로그 분석을 이용하는 경우가 많다. 또한 분석에서 사용하는 웹서비스 로그 파일은 앞에서 논의한 이유들과 더불어, 데이터 자체가 근본적으로 불완전함으로 정보통계의 신뢰성과 타당성 측면에서 문제가 발생한다.

그러므로 무엇보다 사용자들의 로그 데이터를 효과적으로 분석한 훌륭한 e-마케팅 자료로 활용하기 위해서는 사용 목적에 맞는 웹서비스 구조 데이터 수집과 UDDI, WSDL, SOAP, XML 그리고 HTTP&TCP/IP로 구성되어 있는 웹서비스 오류 분류 방법에 대한 다양한 연구가 필요하다. 나아가 사용목적별 웹 호스트에 대한 자료수집을 통하여 웹서비스 분포에 대한 규명도 뒤따라야 할 것이다.

참 고 문 헌

1. 김태근(2002). 웹서비스 콘텐츠 관리방안에 있어서 XML 기술접목 , *COMDEX Korea 2002 Conference*, 27-1- 27-21.
2. 안옥부, 이상복, 김말숙(2002). 클라이언트 입장에서 본 웹 접속오류, *Proceeding of Joint Conference of The Korean Data Analysis Society*

- and Korean Data and Information Science Society*, 157-166.
3. 이동하(2002). 웹로그를 이용한 웹사이트 방문자 분석 및 활용, *COMDEX Korea 2002 Conference*, 29-1- 29-15.
 4. Kallepalli, C. and Tian, J.(2001). Measuring and Modeling Usage and Reliability for Statistical Web Testing , *IEEE Transaction on software engineering*, 27-11.
 5. Long, D.D.E. , Carroll, J.L. and Park, C.J.(1991). A Study of the Reliability of Internet Sites, *Proceedings of the 10th IEEE Symposium on Reliable Distributed Systems*.
 6. Nelson, E.(1978). Estimating Software Reliability from Test Data , *Microelectronics and Reliability*, 17-1, 67-73.
 7. <http://www.logasp.com>
 8. http://www.nic.or.kr/index_kr.html
 9. <http://www.w3.org/TR/WD-logfile-960221.html>
 10. <http://www.w3.org/Protocols/rfc2616/rfc2616.html>

[2003년 2월 접수, 2003년 4월 채택]