

## Mean Time Between Failures of Web Services on SOAP/HTTP<sup>1)</sup>

SangBock Lee<sup>2)</sup>

### Abstract

The reliability modeling 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. Mean time between failures of a web site on SOAP12/HTTP is considered, where SOAP12 is a protocol intended for exchanging structured information in a decentralized, distributed environment. Data were collected from [statlab.cam.ac.uk/webstat/stat.html](http://statlab.cam.ac.uk/webstat/stat.html) via the internet.

**Key words** : MTBF, web services, SOAP, HTTP

### 1. 서론

최근 SOAP/UDDI(simple object access protocol/universal description-discovery-integration) 기반의 c-commerce로 약칭되는 웹서비스 개념의 출현은 웹서비스 신뢰성 문제에 대하여 새롭고 다양한 연구를 필요로 하고 있다. 일반적으로 웹서비스 신뢰성 추정 혹은 평균고장시간의 예측을 통하여 웹 서브 또는 서비스의 신뢰성과 안정성, 또한 정기점검 시기를 알 수 있다. 과거 10여 년 간 TCP/IP(telecommuaction protocol/Internet protocol), 혹은 HTTP(Hypertext transfer protocol)기반 로그파일 데이터를 분석한 웹 서비스의 신뢰성 연구가 다양하게 진행되었다. TCP/IP 기반 연구로 Long외 2인(1991)의 정의를 보면 웹 접속실패는 분산환경(distributed environment)에서 정의된다. 이 정의는 클라이언트가 호스트에 접속하지 못하는 것을 의미하며 호스트에 의한 하드웨어, 소프트웨어 오류, 정전과 계획된 다운타임(down

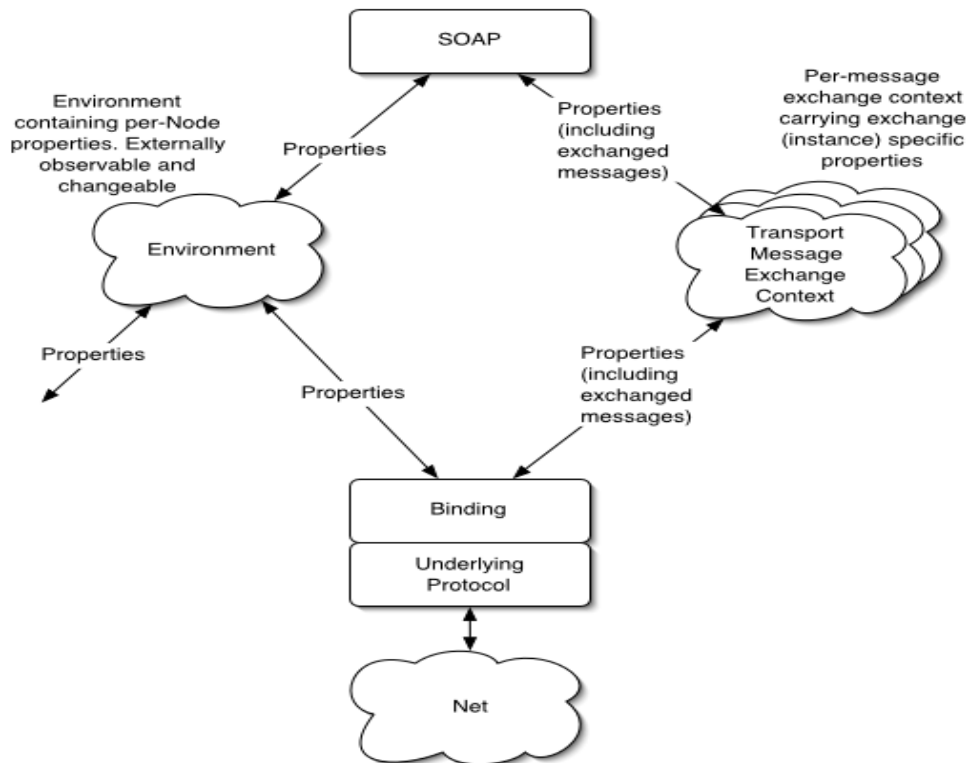
---

1) This research was supported by the Catholic University of Daegu Research Grants in 2003

2) Professor, Faculty of Information and Environmental Sciences, Catholic University of Daegu, [sangbock@cu.ac.kr](mailto:sangbock@cu.ac.kr)

time)을 의미한다. 또한 오프 사이트(off-site) 통신 실패, 물리적인 통신 연결 문제점, 정확한 경로(routing) 정보의 일시적인 부재로부터 기인한다라고 정의했다. HTTP 기반 연구로 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, 호스트나 네트워크 오류, 소스나 내용상의 오류으로 분류하였다. 또한 이 연구에서는 접속 신뢰성을 추정하기 위하여 이용한 오류는 RFC2068의 소스나 내용상의 오류와 관련하여 생각했다. 안옥부의 2인(2002)은 웹 로그 액세스 자료에 대하여 RFC2068의 상태코드(status code)를 기본으로 하여 접속오류 분류 및 정의를 클라이언트 입장을 규명하였다.

Kallepalli와 Tian(2001)은 자신들이 제안한 웹서브 오류를 사용하여 웹서브의 신뢰성과 평균고장시간을 제안하였다. Spiliopoulou와 Pohle(2001)는 데이터마이닝을 응용한 웹사이트 효율성과 개선방안을 제안하였다. 클라이언트 측 SOAP/ HTTP오류를 사용하여 이상복과 김말숙(2003)은 웹서비스의 신뢰성을 계산하였다.



[그림 1] SOAP/HTTP의 특성결합 개념도

그 밖에 스포트 벡타 기계학습(support vector machine learning)을 응용한 웹사이트 신뢰성과 효용성을 연구한 사례(전해성의 6인, 2003)도 있다.

본 연구에서는 분산환경하에서 구조화된 XML(extended markup language)내용 정보교환을 위한 최근 표준인 SOAP12/HTTP 기반의 웹서비스 로그파일을 이용하여 웹

서비스의 평균고장시간(MTBF, mean time between failures)을 구하고, 그 사례적용을 예시하였다. [그림 1]은 SOAP와 프로토콜의 특성 결합관계 개념을 설명한다(www.w3.org/tr/2003/REC-SOAP12).

## 2. SOAP/HTTP 기반 웹서비스 평균고장시간

80년대 클라이언트/서브 개념은 TCP/IP 인터넷 표준 프로토콜을 사용한 시기였으나, 90년대 HTML/XML을 거쳐 현재의 표준 웹기반은 WSDL(web services description language), UDDI, SOAP로 발전되고 있다(백종현, 2002; 김태근, 2002). 특히 웹서비스 아키텍처는 UDDI, 웹서비스 클라이언트, 웹서비스 서버로 구성된다. UDDI는 중개자역할을 웹서비스 클라이언트와 서버는 SOAP에 의한 WSDL을 매개한다. 그러므로 웹서비스의 성공은 잘 구축된 아키텍처에 있으며, 이를 위해 최근 IBM을 비롯한 앞서있는 상용 개발회사들 모두가 어느 정도 가시적 결과를 얻고 있다.

<표 1> SOAP12/HTTP 상태코드 일부분

Status Code	Reason phrase	NextState
2xx	Successful	
4xx	Client Error	
400	Bad Request	"Sending+Receiving", "Receiving" or "Fail"
401	Unauthorized	"Requesting" or "Fail"
415	Unsupported Media Type	"Fail"
5xx	Server Error	
500	Internal Server Error	"Sending+Receiving", "Receiving" or "Fail"

SOAP12/HTTP의 상태코드는 요구 메시지를 이해하고 이에 따라 서비스를 제공하는 것에 대한 결과로서 세 자리의 정수코드이다. 표준상태코드의 첫 자리 숫자는 응답 클래스를 규정하며 4는 클라이언트, 5는 서버를 의미한다. 나머지 두 자리 숫자는 RFC2068과 달리 오류의 종류를 구분한다. RFC2616의 상태코드를 기준으로 코드 100에서 599까지 100단위로 5개의 그룹으로 묶여진다. <표 1>에서 보면, RFC2616 400- 500번 대에 "Sending+Receiving" 정보교환 개념을 도입하였다. 이는 분산환경 하에서 하위 웹 프로토콜 무관(free)의 신뢰성, 보안성, 상관성, 라우팅 그리고 메시지 교환형식에 구애받지 않기 때문에 단순화와 확장성을 더욱 높였다.

안옥부의 2인(2002)은 클라이언트 측 접속오류를 RFC2068에서 규약된 접속 실패의 정의를 기초로 상태코드 201-204, 301-304번을 포함하여 분류하였다. 그러나, 하위 프로토콜 무관의 단순화와 확장성이 좋고 구조화된 정보교환 표준인 SOAP12/ HTTP에 의한 클라이언트-서브 웹서비스 신뢰성을 새롭게 도입할 수 있다.

특히, 웹서비스 신뢰성은 서비스 품질의 중요한 요소이다. 사용기간에 따라 기록된

서비스 오류(고장, failure)는 웹서비스의 신뢰성 척도를 나타낸다. Kallepalli와 Tian(2001)는 전체 히트율(수)과 RFC2068 표준으로 평균고장시간을 추정하였다. 웹 페이지 뷰(view)를 구성하는 정보내용은 웹 파일(htm, asp, php 등), 이미지 파일(gif, jpg 등), Include 파일(js 등) 등의 파일들로 구성되며, 페이지 뷰에 이들은 각각의 히트 수로 계산된다. 로그 파일에 저장된 모든 행이 히트의 단위가 된다. 그러므로 히트 수에 근거한 트래픽 데이터는 사용자 측면에서 웹서비스 평균고장시간의 의미가 없을 수 있으며, 또한 신뢰성 추정치로 약간의 문제가 있을 수 있다. 그러나 SOAP12 표준의 페이지 뷰는 사용자가 요청한 내용 혹은 정보가 있는 접속된 웹 페이지를 의미한다.

웹서브의 신뢰성이나 정기점검시기 예측을 위하여, 구조화된 정보교환 표준인 SOAP12/HTTP기반 평균고장시간 MTBF(Mean time between failures)은 사용시간  $t_i$ 가 존재할 경우의 평균 고장시간 MTBF은 일반적으로 다음과 같이 정의된다.

$$MTBF = \frac{1}{f} \sum_i t_i, \quad (2.1)$$

여기서  $t_i$ : 사용시간,  $f$ : SOAP12/HTTP 표준의 서브-클라이언트측 혹은 정보교환 오류 총합.

그러나, 사용시간  $t_i$ 가 존재하지 않을 경우의 평균 고장시간 MTBF는 식(2.2)로 정의할 수 있다.

$$MTBF = \frac{N_{pv}}{f}, \quad (2.2)$$

$N_{pv} \leq f$ ,  $N_{pv}$ : SOAP12/HTTP 표준에 의한 접속한 총 페이지뷰.

### 3. 적용사례

앞 절에 논의된 SOAP12/HTTP 표준에 따른 웹서비스 로그파일 자료는 공개된 로그 파일(URL: <http://www.statslab.cam.ac.uk/webstats/stats.html>)을 사용하였다. 이 자료에 대하여 2종류의 웹서비스의 신뢰성과 평균고장시간 추정량을 계산해 본다. <표 2>의 데이터는 1994년6월28일 20시31분~2003년4월29일 23시59분(총3197.14일)까지의 자료이다.

본 연구사례 <표 2>의 경우는 요청(requests)을 페이지뷰로 계산하였으며, 사용시간이 예시되어 있지 않으므로 평균고장시간의 의미보다는 평균고장간격의 의미로 해석될 수 있다. 첫 번째, Kallepalli와 Tian(2001)에 의하면 RFC2068 오류는 400-500번 대로 오류총합은 3,641,672, 웹서비스 접속성공은 54,103,990이고 총 페이지뷰는 57,745,662이다. 그러므로, 신뢰성 추정치  $\hat{R} = 0.937$  (54,103,990/ 57,745,662), 평균고장시간 추정치  $\widehat{MTBF} = 15.9$ (57,745,662/3,641,672)가 된다. 본 연구에서 제안한 SOAP12/HTTP 표준을 이용하여 구한  $f$ 는 400-500번 대 접속오류와 301번, 304번 정보교환 실패오류를 포함하여 11,951,871, 웹서비스 성공은 45,793,791으로 신뢰성 추정치  $\hat{R} = 0.793$ (45,793,791/57,745,662), 평균고장시간 추정치  $\widehat{MTBF} = 4.83$

(57,745,662/11,951,871)이다.

< 표 2 > 캠프릿지대학교 통계학실험실 웹서브 로그 자료

Requests	Status code	Reason phrase	Significance/Action	NextState
45,300,171	200	Successful	OK	
473,563	206		Partial content	
1,276,691	301		Document moved permanently	"Init"
20,057	302		Document found elsewhere	
7,033,508	304		Not modified since last retrieval	
	4xx	Client Error		
6,922	400		Bad request	"Sending+Receiving", "Receiving" or "Fail"
4,393	401	Unauthorized	Authentication required	"Requesting" or "Fail"
210,010	403		Access forbidden	"Fail"
3,400,464	404		Document not found	"Fail"
402	405		Method not allowed	"Fail"
7,964	408		Request timeout	"Fail"
2	412		Precondition failed	"Fail"
1	414		Requested filename too long	"Fail"
344	416		Requested range not valid	"Fail"
	5xx	Server Error		
10,377	500	Internal Server Error	Internal server error	"Sending+Receiving", "Receiving" or "Fail"
793	501		Request Type not Supported	"Fail"

최근의 SOAP 표준과 RFC2068을 직접 비교를 하는 것은 그 본질상 다소 무리가 있다. 그러나, Kallepalli와 Tian (2001)의 오류개념에 의한 추정치  $\hat{R}$  과 평균고장시간 추정치  $\widehat{MTBF}$ 이 본 연구의 추정치보다 신뢰성이 더 높게 나온 것은 구조화된 정보 교환상의 오류를 간과한 데 따른 것이다.

#### 4. 결 론

일반적으로 웹서비스의 신뢰성 혹은 평균고장시간의 예측을 통하여 웹 서버 정기 점검시기를 알 수 있다. 본 연구에서는 SOAP12/HTTP 표준에 따라 웹서비스 신뢰성 R과 평균고장시간을 구하였다. 현재 대부분의 상용 이나 연구용 웹 마이닝은 미리 보고해 놓은 고정된 TCP/IP 혹은 HTTP 기반 웹 로그 분석을 사용하는 경우가 대부분이다. 물론, 사용자가 적은 이유로서, SOAP12/HTTP 표준이 최근에야 완성된 데에 그 원인을 찾아 볼 수도 있다. 그러나, 분산환경하에서 하위 웹 프로토콜 무관 (free)의 신뢰성, 보안성, 상관성, 라우팅 그리고 메시지 교환형식에 구애받지 않아 단순화와 확장성이 높은 SOAP 표준은 앞으로 웹서비스 성장의 기본기술로 예상된다. 또한, 웹서비스 신뢰성뿐만 아니라, 구조화된 XML 정보교환이라는 측면에서 SOAP/HTTP 기반 웹서비스 로그파일 마이닝기법 분야에서 많은 발전이 있을 것으로 예상된다.

그리고, 무엇보다 사용자들의 로그 데이터를 효과적으로 분석한 훌륭한 e-마케팅 자료로 활용하기 위해서는 사용 목적에 맞는 웹서비스 구조 데이터 수집과 UDDI, WSDL, SOAP, XML을 비롯하여, 나아가 XAML(transaction markup language), WSFL(web services flow language), BTP(business transaction protocol) 등과 같은 예고된 웹서비스 표준자료에 대한 다양한 연구가 필요한 때이다.

#### 참 고 문 헌

1. 김태근(2002). 웹서비스 콘텐츠 관리방안에 있어서 XML 기술접목 , *COMDEX Korea 2002 Conference*, 27, 1-21.
2. 백종현(2002). Web Services Trands & CAse Studies, *COMDEX Korea 2002 Conference*, 28,1-19.
3. 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.
4. Jun, SH., Lim, MT., Jorn, HS., Hwang, JS., Choi, SY., Kim, JY. and Oh, KW.(2003). Web Log Analysis Using Support Vector Regression, *The Korean Communication in Statics*, 10,1,61-78.
5. Kallepalli, C. and Tian, J.(2001). Measuring and Modeling Usage and Reliability for Statistical Web Testing , *IEEE Transaction on software engineering*, 27-11.
6. Lee, S.B. and Kim, M.S.(2003). A Study of the Relibility of Web Services using Client Sides Errors, *Journal of The Korean Data Analysis Society and Korean Data and Information Science Society*, 14, 2, 23-36.
7. 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*.

8. Nelson, E.(1978).Estimating Software Reliability from Test Data, *Microelectro- nics and Reliability*, 17,1, 67-73.
9. <http://www.statslab.cam.ac.uk/webstats/stats.html>
10. <http://www.w3.org/TR/WD-logfile-960221.html>
11. <http://www.w3.org/tr/2003/REC-SOAP12>
12. <http://www.w3.org/Protocols/rfc2616/rfc2616.html>

[ 2003년 4월 접수, 2003년 7월 채택 ]