

УДК 004.05

С.С. МАМУТОВ, В.С. ХАРЧЕНКО, А.В. ГОРБЕНКО, МОХАМЕД САИД ГАЗАЛ

Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Украина

АНАЛИЗ ПРОИЗВОДИТЕЛЬНОСТИ КОММУНИКАЦИОННЫХ ПРОТОКОЛОВ SOAP И TCP

В статье проведен анализ факторов, влияющих на производительность распределенных вычислительных систем и коммуникационных протоколов TCP и SOAP.

коммуникационные протоколы, TCP, SOAP, производительность

Введение

Постановка задачи. В настоящее время web сервисы получили широкое распространение как технология организации распределенного взаимодействия.

Web сервисы, с их акцентом на открытые стандарты и гибкость использования, могут иметь преимущества по сравнению с существующими методами интеграции информационных систем.

Технологии Web сервисов основаны на XML, который обеспечивает независимость данных от языка программирования и связующего программного обеспечения. XML является гибким средством для того, чтобы приспособиться к любому типу данных или структуре или даже при необходимости создать новые типы и структуры. XML лежит также в основе транспортного протокола web сервисов.

Для использования web-сервисов в b2b приложениях они должны удовлетворять определенным техническим требованиям, включая производительность, безопасность и отказоустойчивость [1].

Для web-сервисов реального времени производительность является наиболее существенным показателем.

Однако, недостатки протоколов SOAP (Simple Access Protocol) и TCP [2] приводят к существенному увеличению времени отклика по сравнению с другими транспортными протоколами технологий распределенного взаимодействия.

Поэтому существует **актуальная задача** исследования производительности протокола SOAP и выявления факторов, влияющих на производительность взаимодействия.

Целью данной статьи является экспериментальная оценка производительности протокола SOAP, выявление факторов, влияющих на производительность, выявления зависимости времени отклика от степени загруженности сети и сервера.

Факторы, влияющие на производительность в распределенной системе

Одной из базовых идей, в значительной степени стимулировавшей развитие интернета, является децентрализованность его структуры. Интернет – это конгломерат сетей, которые существенно различаются как по принципам предоставления услуг, так и по пропускной способности каналов, их загрузке и т.п. Попытаемся выявить наиболее существенные факторы, влияющие на производительность в распределенной системе:

1. Пропускная способность – объем данных, которые сеть может передать в единицу времени от одного подсоединенного к ней устройства к другому [3].
2. Задержка – скорость прохождения единичного пакета через канал передачи данных от одного устройства к другому.

3. Время отклика (Round Trip Time – RTT) – интервал времени от момента отправки пакета до момента получения подтверждения на него.

4. Загруженность сети (канала связи) – насыщенность канала связи передаваемыми пакетами. Передачу пакета осуществляют протоколы канального уровня, которые имеют определенные механизмы управления доступом к среде (CSMA/CD в Ethernet или маркер в Token Ring). Высокая загруженность сети приводит к увеличению временного интервала между отправкой пакетов, что приводит к снижению производительности.

5. Загруженность сервера – количество запросов, обрабатываемых сервером за единицу времени. Чем больше количество запросов в единицу времени, тем больше времени необходимо серверу на обработку каждого запроса. Зависит от производительности процессора, быстродействия ОЗУ и винчестера.

6. Доступность (готовность) – зависит от надежности линий связи, а также от характеристик самого сервера. При готовности меньше 1 возникает необходимость выполнения повторных запросов

Для более глубокого понимания причин низкой производительности протокола SOAP, рассмотрим взаимодействие клиента с web сервисом на уровне стека протоколов TCP/IP (рис. 1). Условно взаимодействие клиента с web сервисом можно разделить на 4 этапа: разрешение DNS имени сервера, установление TCP соединения, обмен SOAP сообщениями и закрытие TCP соединения. Разрешение DNS имени выполняется каждый раз при выполнении единичного запроса и только один раз при выполнении серии запросов. Это связано с кэшированием.

На 2 этапе устанавливается TCP соединение. Большое время выполнения этого этапа может свидетельствовать о неисправной работе сети.

На 3 этапе выполняется непосредственно запрос к web сервису по протоколу SOAP. По этому време-

ни можно судить о скорости работы сервиса. Завершается взаимодействие закрытием TCP соединения. На рис. 2 представлено процентное отношение времени, затрачиваемого на каждом из этапов. Таким образом, взаимодействие клиента с сервером WS представляет собой обмен TCP пакетами. Протокол SOAP как и протокол TCP имеют массу недостатков [2], которые нужно учитывать при проведении эксперимента.

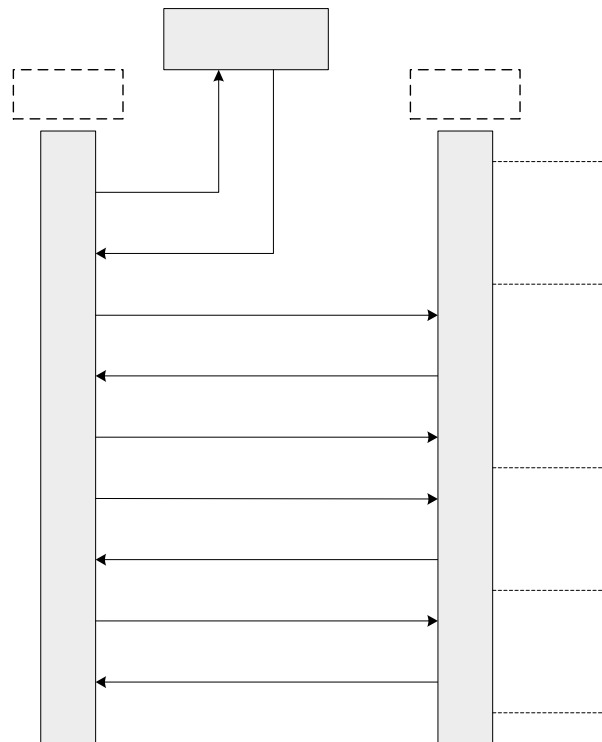


Рис. 1. Взаимодействие клиента с web службой на уровне TCP/IP

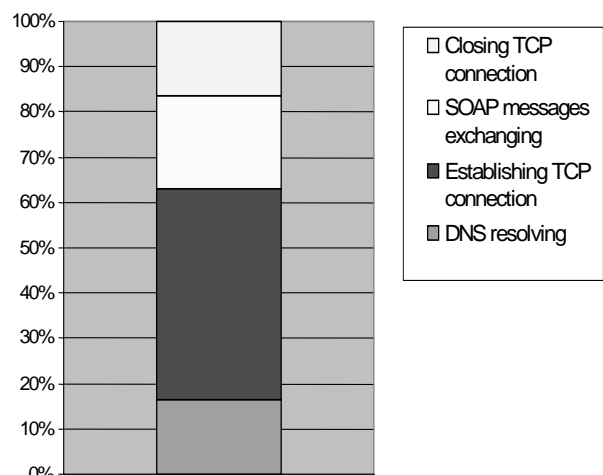


Рис. 2. Распределение временных затрат по этапам TCP-соединения

Анализ производительности протоколов SOAP и TCP

Протокол SOAP в качестве транспортного протокола использует протокол TCP стека протоколов TCP/IP. Поэтому производительность протокола SOAP в основном характеризуется величиной задержки отклика RTT (Round Trip Time).

RTT – один из важных параметров производительности TCP обмена.

Именно этот показатель используется для установления размера окна протокола TCP, время таймаута подтверждения получения пакета и др.

Для анализа зависимости времени отклика от параметров загрузки компьютерной сети и сервера был проведен следующий эксперимент.

Условия проведения эксперимента:

Сервер: Сервер приложений: Tomcat 5.5.9

Операционная система: Linux

Процессор: Celeron 2.8

ОЗУ: 1 Gb

Клиент: Операционная система: Windows

Процессор: Athlon 2.0

ОЗУ: 1 Gb

Компьютерная сеть: Internet

Цель эксперимента – выяснить влияние загрузки сети и сервера на время отклика при взаимодействии клиента с web сервисом.

Порядок проведения эксперимента и результаты. В течение определенного периода времени проводился мониторинг участка сети для выявления RTT при нулевой загрузке.

В дальнейшем, средствами операционной системы, велся учет этого показателя при проведении эксперимента. Загруженность сети постепенно увеличивалась генерацией TCP трафика.

На каждом шаге выполнялось измерение времени отклика.

Уже при 35% загрузки сети, время отклика увеличилось в 2 раза, а пиковая загрузка сети привела к его увеличению более чем в 10 раз (рис. 3).

Для выявления зависимости времени отклика от загрузки сервера на сервер создавалась нагрузка генерацией запросов. При этом число запросов увеличивалось от 1 до 300.

Конечно, время отклика при этом существенно зависит от производительности сервера и типа сервера приложений.

При заданном окружении была получена зависимость, изображенная на рис. 4.

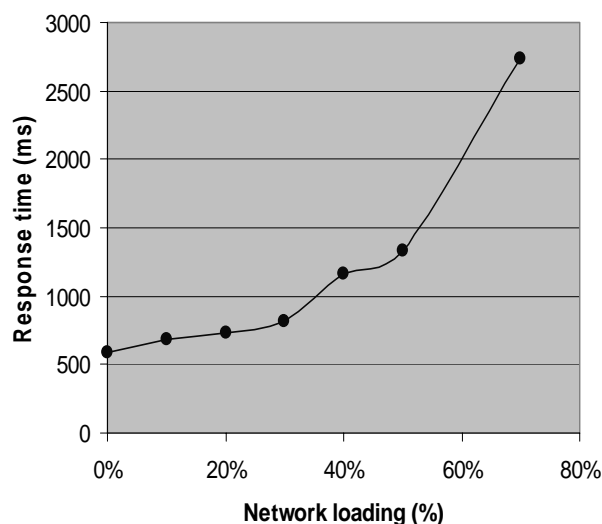


Рис. 3. Зависимость времени отклика от загрузки сети

Анализ результатов эксперимента. Как показали результаты эксперимента, величину времени отклика можно аналитически описать следующим выражением:

$$T = F(RTT) \cdot K_{NL} + t_S \cdot K_{SL} \cdot A, \quad (1)$$

где K_{NL} – коэффициент загрузки сети;

K_{NS} – коэффициент загрузки сервера;

t_S – время работы сервиса;

A – готовность сервера.

Кроме того, анализ показал, что существует линейная зависимость RTT от размера посылаемого пакета в отсутствие загрузки сети (рис. 5).

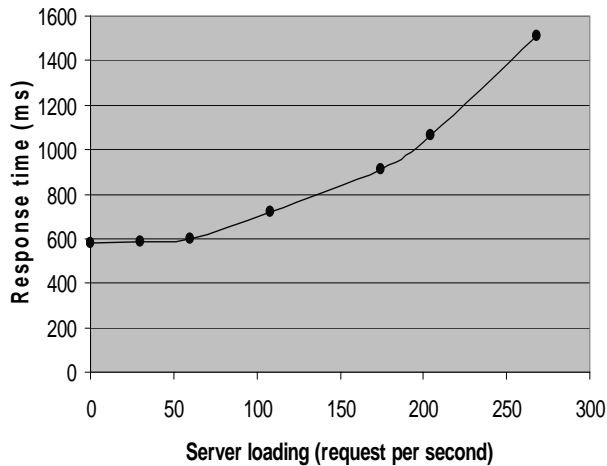


Рис. 4. Зависимость времени отклика от загрузки сервера

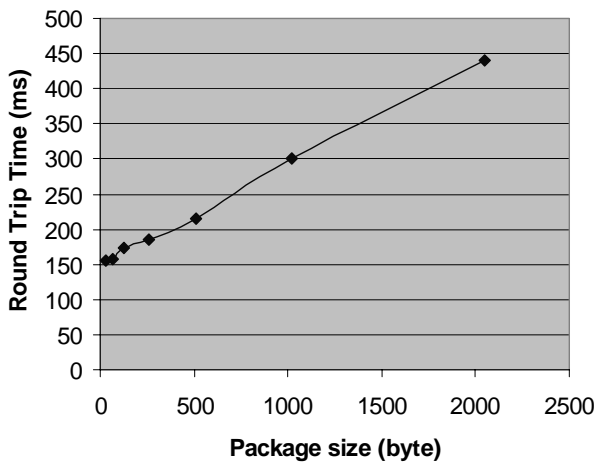


Рис. 5. Зависимость времени отклика от размера пакета

Выводы

Web сервисы, с их универсальностью и гибкостью, являются одной из основных технологий интеграции информационных систем.

Однако, избыточность служебной информации, вызванная использованием XML, и применение TCP в качестве транспортного протокола в значительной мере сказываются на производительности данной технологии.

Такое положение еще больше усугубляется повышенной загруженностью сети, связанной с трансляцией широковещательного и мультимедийной информации, VoIP и др.

Как следует из результатов эксперимента, даже небольшая загруженность сети может привести к значительному увеличению времени отклика, что категорически недопустимо в приложениях реального времени.

В связи с этим актуальными становятся задачи, связанные с модернизацией существующих или разработкой транспортных протоколов, минимизирующих накладные расходы при передаче XML-трафика Web-сервисов с учетом особенностей RPC-взаимодействия.

Литература

1. RFC-793. [Электрон. ресурс]. – Режим доступа: <http://www.ietf.org/rfc/rfc0793.txt>.
2. Kohlhoff C., Steele R. Evaluating SOAP for high performance business applications: Real-time trading systems // Proc. of WWW2003, Budapest, Hungary. – 2003. – P. 262-270.
3. Mitchell B. Bandwidth and Latency [Электрон. ресурс]. – Режим доступа: <http://compnetworking.about.com/od/speedtweaks/l/aa021902b.htm>.
4. Lowe, Hall W. Hypermedia and The Web – An engineering Approach”. – NY.: J. Wiley and sons, 1999. – 537 p.

Поступила в редакцию 6.04.2007

Рецензент: д-р техн. наук, проф. А.А. Зеленский, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.