UiO : Department of Informatics
University of Oslo

Improving the performance of Web Services in Disconnected, Intermittent and Limited Environments

Joakim Johanson Lindquister

June 11, 2015

Contents

1 Introduction

Military units operate under conditions where the reliability of the network connection is low. They may operate far from existing communication infrastructure and rely only on wireless communication. Such networks are often characterized by unreliable connections with low bandwidth and high error rates making communication difficult. In a military scenario it is necessary for units at all operational levels to seamlessly exchange information across different types of communication systems. The North Atlantic Treaty Organization (NATO) has identified the Service-Oriented Architecture (SOA) and Web Services as key elements for enabling interoperable military communication.

In my thesis I will in cooperation with the Norwegian Defence Research Establishment (FFI) investigate different approaches to improve the performance of Web services in military tactical networks.

2 Background and Motivation

NATO is an military alliance consisting of 28 member countries[1] and which primary goal is to protect the freedom and security of its members through political and military means. In joint operations the relatively large number of member countries can be a challenge when communicating. Difference in communication systems and equipment can attribute to make communicating difficult. In order to address this issue, NATO studies has identified the SOA concept and Web Services as key enablers[2]. Web services are widespread and used successfully in civil applications, but there exist some key differences when applied in a military setting. In a civil environment networks are stable and bandwidth is abundant, while in a military scenario this may not be the case.

Military networks are complex and consist of different operational levels. At the highest level, network conditions are stable. However at the lowest level, where small tactical units operate, the network conditions are more dynamic and challenging. Military units may move into areas where little or no communication infrastructure exists making communication difficult. In a military scenario, friendly networks may also be target of deliberate enemy interference, called *jamming*.

2.1 Service-Oriented Architecture

SOA is a design pattern for organizing and utilizing distributed capabilities that may be under the control of different ownership domains. As discussed, this is the case in NATO. Furthermore SOA provides a uniform means to offer, discover, interact with and use capabilities to produce desired effects consistent with measurable preconditions and expectations[3, chapter 1.1.1].

2.2 Mobile tactical networks

Mobile tactical networks are characterized by that the units use tactical communication equipment which includes technologies like VHF, UHF, HF, tactical broadband and satellites. Examples of such units are mobile units like vehicles, foot soldiers and field headquarters. These types of networks have low bandwidth, possibly high delay, high error rates and frequent disconnections. They are often called disadvantaged grids or Disconnected, Intermittent and Limited environments (DIL). NATO studies has identified such networks to have the following characteristics:

Disadvantaged grids are characterized by low bandwidth, variable throughput, unreliable connectivity, and energy constraints imposed by the wireless communications grid that link the nodes[4].

These constrains of mobile tactical networks are central in order to understand the problem at hand, and I will therefore explain the concepts here:

Bandwidth and throughput The terms bandwidth and throughput are used interchangeably in the networking community and refers to the data transfer rate; how fast data can be transported from one point to another in given time period. This is often expressed in bits per second.

Unreliable connectivity Units that are participating in a tactical network are highly mobile and may disconnect from a network either voluntarily or not. Unplanned loss of connectivity can be due to various reasons, such as loss of signal or equipment malfunction.

Energy constraints imposed by the wireless communication grid The battery capacity and the transmission range of the communication equipment for mobile units may be limited. Another issue is that in some cases military units are required to enter radio silence in order to avoid being detected by the enemy. During radio silence units may only receive data and not send any.

2.3 Problem Statements

In my thesis I will investigate different approaches to improving Web services in DIL. I will study all or some of the following topics:

1. How can the performance of Web Services, both Simple Object Access Protocol (SOAP) and Representational State Transfer (REST), be improved in typical tactical network environments?

2. How can a proxy implemented at both the client and server be utilized?

3. Application server parameter tuning. How does the configuration influence performance of Web Services?

4. Running Web Services over HTTP/TCP is very common in civil application, but not necessarily the best approach in DIL environments. I will therefore investigate alternative transport protocols.

5. Alternate data representations. SOAP is bound to using XML, but REST puts no such limitations on payload.

In this essay I will present each problem statement in detail and present their theoretical background. Furthermore, I will discuss what related work has previously been done in the different areas. Finally, I will present my research methodology.

3 Web Services

NATO has chosen Web services as the technology for achieving interoperability with respect to machine-to-machine information exchange in NATO. The World Wide Web Consortium has defined Web Services as[5]:

A Web service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP-messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards.

However, there also exist other types of Web services which does not follow this definition. RESTful web services let users manipulate data using a set of stateless operations. REST services have gained a lot of traction in the civil industry in the latest years and are discussed in section 3.4.

3.1 Problem statement

Most of the Web Service solutions used today are aimed for civilian use and does not necessarily perform well in military environments. In contrast to civilian networks where bandwidth are abundant, mobile tactical networks may suffer from high error rates and low bandwidth. In my master thesis I will investigate which Web services that can be utilized in such environments, and in specific SOAP Web Services versus RESTful Web Services.

3.2 Related work

In their paper "Recommendations for increased efficiency of Web services in the tactial domain," Johnsen, Bloebaum and Karud[6], looked on how server application parameter tuning affected the performance of Web Services. They performed experiments using both SOAP and RESTful Web Services and discovered that RESTful services using JSON, performed better than SOAP services using XML. Finally they concluded that RESTful services can be an interesting alternative to SOAP for applications where performance is more important than interoperability concerns. It should be noted that the experiments was not tested with Internet Protocol Security (IPsec). IPsec is commonly used in military networks and in order to perform a credible network emulation it should be considered.

3.3 SOAP

SOAP is an application layer protocol and a way of implementing Web services that performs remote procedure calls between programs in a language- and system-independent way[7, p 682]. To send and receive messages or requests, Extensible Markup Language (XML) messages are constructed and sent between an client and server. SOAP messages are usually sent over HTTP on top of the transport protocol TCP. HTTP is synchronous, which means that while a SOAP request is sent, the HTTP connection is kept open until the SOAP response is returned in the HTTP acknowledgment. If the connection times out because of delay or for any other reason, the SOAP response is not returned.

SOAP is transport protocol independent which means that different protocols other than HTTP over TCP can be used to transport SOAP messages.

3.4 RESTful Web Services

REST is a architectural style that provides engineering principles and interaction constrains chosen to retain those principles[8]. It differs in from Service-Oriented Architecture in that it places constrains on connector semantics rather than component semantics. REST is closely associated with HTTP but enforces no restriction on the type of the data. The REST approach is often selected in data-oriented applications. One potential issue against REST is that there exist no industry standard for RESTful web services.

4 Proxies

A proxy is an application which acts as an intermediary between an client and a server. The advantage of this is that the web services running at client and server side does not need to take account for an unreliable network, it is all done by the proxy transparent from the web services. This helps the need for implementation and reduce costs which is always helpful.

4.1 Related work

4.2 Plan

In my master thesis I plan to set up an architecture with a client and a server, with a proxy pair acting as an intermediary between. The web services will be ran on the same machine and communicate with the proxy over HTTP. The actual transport between the proxy pair is transparent from the web services.

4.2.1 HTTP proxy

5 Tuning application server parameters

5.1 Problem Statement

When setting up an application server, several parameters which can affect the performance of Web Services running on the application server can be configured. Wrong or bad configuration may cause inaccurate timeouts and congestion in the network. Tuning server parameters can be performed independently of the data being sent over a Web service. In my master thesis I will build up on existing related studies and explore how server parameters can be tuned in order to increase performance in mobile tactical networks.

5.2 Related work

In their study[6], which I previously discussed in section 3.2, the authors investigated how tuning server parameters affect the performs of both REST and SOAP Web services. They identified a number of HTTP and TCP tuning parameters which are discussed in section 5.3. After running their experiments they concluded that few of the parameters actually had any significant impact on the performance of the Web Service. However, they identified HTTP Chunking as having the most impact on the performance. It significantly improved the performance in different types of networks and for both SOAP and RESTful Web services.

5.3 Parameters

HTTP Timeout Controls how long a HTTP connection can be deemed as idle and kept in the "keep-alive" state. Having a to low timeout on networks with low bandwidth, can potentially flood the network with packets that have timed out. Consideration should therefor be taken when setting setting this parameter for mobile tactical networks.

HTTP Compression Enables HTTP/1.1 GZIP compression.

HTTP Chunking Allows the server to send data in dynamic chunks.

HTTP Header and Send Buffer Sizes Can vary the size of the buffers that hold the request and send data, respectively.

TCP Idle Key Timeout Sets the time before an idle TCP channel closes.

TCP Read and Write Timeouts Set the timeout for TCP read and write operations, respectively.

TCP Selector Poll Timeout Sets the time a Java new/non-blocking I/O (NIO) selector will block waiting for user requests.

TCP Buffer Size Sets the size of the buffer that holds input streams created by the network listener.

TCP Batching/TCP NO_DELAY Batches together small TCP packets into larger packets.

MTU Size The maximum transmission unit size regulates the largest data unit that can be passed onwards. In tactical military communication the MTU size can be very low(down to 128 bytes).

6 Alternate Transport protocols

In order for data to be transmitted from a machine to another, it must be transported over a network. In the OSI model, the transport layer builds on the network layer to provide data transport from a process on a source machine to a process on a destination machine with a desired level of reliability that is independent of the physical networks currently in use[7, p 495]. The transport code, runs entirely on the user's machine, and must be able to handle errors in underlying layers. The transport protocols has no control over the speed and reliability of routers or the network and must therefore among other things, have techniques for handling packet loss, sequencing and slow transmissions.

6.1 Problem statement

In disadvantaged grids, bandwidth can be extremely sparse. Bandwidth down to 40 bits/s is not uncommon, which is to slow for transport protocols such as traditional TCP[3, chapter 2.1.3]. In mobile tactical networks we need protocols that are able to cope with the characteristics of tactical networks. They will need to:

- Withstand long and variable round trip times, while at the same time having very little communication overhead.
- Store-and-forward capabilities, where intermediate nodes stores messages until they can successfully be delivered to the receiver.

The most common transport protocol for transporting data over the web are HTTP over TCP. However, in disadvantaged grids, other protocols could perform better. In my thesis I will investigate different protocols.

6.2 Related work

In an article published by FFI, Johnsen, Bloebaum et al[9], investigated what can be gained by replacing the commonly used HTTP/TCP transport for Web services with alternative transport protocols. They studied how TCP, UDP, SCTP and AMQP performs for conveying Web services traffic evaluated under military and civilian networking conditions. Their results are summarized here:

- Given the nature of UDP, it comes as no surprise that UDP struggled when the network was congested and started dropping packets.
- SCTP performed better than TCP in many cases. However, when the bandwidth got low SCTP generated more overhead than TCP.
- AMQP had an overall low success rate at low bandwidths, but the success rate improved as the bandwidth increased.

6.3 Relevant transport protocols

TCP Transmission Control Protocol(TCP) is transport layer protocol for reliable transfer of data. It is connection-oriented protocol, which means that a communication sessions must be established before any data can be transferred. TCP is a widely used and is a core protocol of the Internet protocol suite.

TCP congestion control is algorithms that aims to prevent or reduce congestion on a network. Different TCP congestion control algorithms exist e.g TCP Tahoe and Reno.

UDP User Datagram Protocol(UDP) is as TCP a transport layer protocol but it has no mechanisms for providing reliable transfer, flow control or packet ordering. It is suitable in scenarios where error checking and sequence of messages is not necessary. However, in military scenarios(and in Web services in general), correct and reliable data transmission is necessary and I will therefor not focus on UDP in my thesis.

SCTP The Stream Control Transmission Protocol(SCTP) serves a similar roles as both TCP and UDP by being message-oriented, reliable, and with in-sequence transport and congestion control. SCTP features multi-homing, which enables the respective endpoints to communicate over multiple IP addresses and network interfaces. This grants a higher chance of survivability in the presence of network failure, a feature which is interesting when considering solutions for mobile tactical networks.

AMQP The Advanced Message Queuing Protocol(AMPQ) is an application level protocol which can utilize different transport protocols. It is interesting in DIL environments because it employs a broker-based architecture with store-and-forward capabilities[9].

HTTP/2 Although not a transport level protocol, it can be interesting to explore the new standardization of HTTP. Version 2 is a major revision of the HTTP network protocol and the first new version since HTTP 1.1 was standardized in 1997[10]. The HTTP/2 specification was published in May 2015[11] and promises many improvements which would improve performance for Web Services utilizing HTTP. It is more friendly to the networks because fewer TCP connections can be used in comparison to HTTP/1.x. This means less competition with other flows and longer-lived connections, which in turn lead to better utilization of available network capacity.

However, HTTP/2 is not in use by NATO today so it is not so relevant to test at this point in time.

6.3.1 Running SOAP directly on TCP

Using TCP transport for SOAP without using HTTP, has been implemented by multiple organizations, e.g Oracle. Their application server Glassfish supports this feature.

Since there currently exist no standard for SOAP over TCP, as well as the main focus of my thesis will be to implement a HTTP proxy, this will not be prioritized.

7 Alternate data representations

Data and messages sent over Web Services can be represented in different ways. SOAP Web services is limited to using XML, which has an relatively high overhead, while RESTful web services puts no such limitations on the payload. Another approach is data compression which is discussed in a later section.

7.1 Problem Statement

In disadvantaged grids, bandwidth can be very limited. Reducing the overhead of data being sent over such networks could therefore provide a great increase in performance. It is therefore interesting to study which data types can be used in order to decrease the overhead.

7.2 Related Work

In their paper "Recommendations for increased efficiency of Web services in the tactial domain," Johnsen, Bloebaum and Karud[6], concluded that RESTful Web services using JSON performed better than the one that used XML. Furthermore, non-lossy compression has been identified as an important technique for reducing the overhead of XML[12].

7.3 Compression

Non-lossy compression allows representation of the same information using fewer bits than in the original representation. The downside with compression is the overhead it causes, mainly CPU-time which can be limited on battery-driven communication equipment.

Gzip is a popular data processing program which I will use in my simulations.

8 Research Methodology

8.1 Scope of master thesis

As we have seen, many approaches can be taken in order to improve the performance of Web Services. In this essay I have discussed a range of approaches, but in order perform an in depth study I will probably only pursue a subset of the problem areas discussed in this essay.

8.2 Research Methodology

There exist multiple research methods for the computer science discipline. In my thesis I will follow the design approach, which is based on engineering and is commonly used in software projects. It consist of the following four stages[13]:

1. Perform requirements analysis.
2. Derive a specification based on the requirements.
3. Design and implement the system.
4. Test the system.

8.3 Measuring Network Performance

It is difficult to correctly measure and identify bottlenecks in computer networks. Networks are often very large with many nodes and have a complexity which makes it difficult to find isolated issues which have impact on the over all performance. There are important considerations that must be taken into account when performing network experiments in order to get results with scientific value[7, page 584]:

- Make sure that the sample size is large enough.
- Make sure the samples are representative.
- Caching can wreak havoc with measurements.
- Be sure that nothing unexpected is going on during your tests.
- Be careful when using a course-grained clock

8.3.1 Network emulation

In order to perform networks emulation to see how the different approaches affect the performance of Web services, I will in my experiments use a network emulator called *netem*. It provides functionality for testing protocols by emulating properties of wide area networks. It can emulate variable delay, loss, duplication and reordering making it a suitable tool for simulating DIL environments.

8.4 Thesis outcome

The outcome of my master thesis will be a prototype of a HTTP-proxy pair which will act as a intermediary between Web Services operating in a DIL environment. As well as the proxy, I will give a recommendation regarding which optimization techniques can be applied in DIL environments in order to improve the efficiency of Web services.

9 Progress Plan

This section contains a tentative progress plan for my master thesis.

Milestones	**Due date**
Readings of relevant work	1. October 2015
Finished implementation of HTTP-proxy	Before Christmas 2015
Testing and evaluation of protocols	1. March 2016
Draft of master thesis finished	1. April 2016
Delivery of master thesis	2. May 2016

References

[1] NATO. *NATO - Member Countries*. http://www.nato.int/cps/en/natohq/nato_countries.htm. Accessed: 2015-05-04.

[2] P. Bartolomasi et al. *NATO network enabled capability feasibility study*. 2005.

[3] Frank T. Johnsen. "Pervasive Web Services Discovery and Invocation in Military Networks." In: *FFI-rapport 2011/00257* (2011).

[4] A. Gibb et al. "Information Management over Disadvantaged Grids." In: *Task Group IST-030/ RTG-012, RTO-TR-IST-030* (2007). Final report of the RTO Information Systems Technology Panel.

[5] Hugo Haas and Allen Brown. *Web Services Glossary*. http://www.w3.org/TR/ws-gloss/\#webservice. Accessed: 2015-05-06.

[6] Frank T. Johnsen and Trude H. Bloebaum. "Recommendations for increased efficiency of Web services in the tactial domain." 2015.

[7] David J. Wetherall Andrew S. Tanenbaum. *Computer Networks*. Fifth Edition. Pearson New International Edition.

[8] Roy T. Fielding and Richard N. Taylor. "Principled Design of the Modern Web Architecture." In: *Proceedings of the 22Nd International Conference on Software Engineering*. ACM, 2000, pp. 407–416. URL: http://doi.acm.org/10.1145/337180.337228.

[9] Frank T. Johnsen et al. "Evaluation of Transport Protocols for Web Services." In: *MCC 2013* (2013).

[10] R. Fielding et al. *RFC 2068 - Hypertext Transfer Protocol – HTTP/1.1*. http://tools.ietf.org/html/rfc2068. Accessed: 2015-05-20.

[11] M. Belshe et al. *RFC 7540 - Hypertext Transfer Protocol Version 2 (HTTP/2)*. http://tools.ietf.org/html/rfc7540. Accessed: 2015-05-20.

[12] Christopher J. Augeri et al. "An Analysis of XML Compression Effieciency." In: *Workshop on Experimental Computer Science(ExpCS)* (2007).

[13] Peter J. Denning et al. "Computing as a dicipline." In: *Communications of the ACM* (1989).