MULTI-DISCIPLINARY APPROACHES TO INTELLIGENTLY SHARING LARGE-VOLUMES OF REAL-TIME SENSOR DATA DURING NATURAL DISASTERS

Stuart E Middleton[1]*, Zoheir A Sabeur[1], Peter Löwe[2], Martin Hammitzsch[2], Siamak Tavakoli[3], and Stefan Poslad[3]

*[1]University of Southampton IT Innovation Centre, UK
Email: sem@it-innovation.soton.ac.uk
[2] GFZ German Research Centre for Geosciences, Germany
Email: ploewe@gfz-potsdam.de
[3]Queen Mary University of London, UK
Email: stefan@eecs.qmul.ac.uk

ABSTRACT

We describe our knowledge-based service architecture for multi-risk environmental decision-support, capable of handling geo-distributed heterogeneous real-time data sources. Data sources include tide gauges, buoys, seismic sensors, satellites, earthquake alerts, Web 2.0 feeds to crowd source 'unconventional' measurements, and simulations of Tsunami wave propagation. Our system of systems multi-bus architecture provides a scalable and high performance messaging backbone. We are overcoming semantic interoperability between heterogeneous datasets by using a self-describing 'plug-in' data source approach. As crises develop we can agilely steer the processing server and adapt data fusion and mining algorithm configurations in real-time.

Keywords: Natural disaster, Tsunami, Semantics, Data fusion, OGC, W3C, TRIDEC

1   INTRODUCTION

This paper describes our work in the project Collaborative, Complex and Critical Decision-Support in Evolving Crises (TRIDEC) (Wächter, 2011), building on experience gained from the German Indonesian Tsunami Early Warning System (GITEWS) (Münch, 2011) and Distant Early Warning System (DEWS) (Esbri, 2010). We employ a multi-disciplinary approach to intelligently share and process large-volumes of real-time sensor data, building a knowledge-based service architecture for multi-risk environmental decision-support.

Geo-distributed heterogeneous data management is at the heart of the TRIDEC knowledge-based service architecture (Sabeur, 2011). Real-time sensor proxies publish measurement data from tide gauges, buoys, seismic sensors, and satellites, each with its own accuracy and frequency of measurement. Expert reports are published from trusted sources, such as SeisComP3 for earthquake alerts (Hanka, 2010). Real-time Web 2.0 feeds provide rapid crowd sourced multi-format measurements (text, images, video, etc.). Lastly, on-demand and high-resolution pre-computed simulations of Tsunami wave propagation are available to help (Behrens, 2010).

We describe our system of a systems approach to data management, based on a hybrid service oriented architecture (SOA) and event-driven architecture (EDA) (Haener, 2011), using a multi-bus middleware providing performance, scalability, and fault tolerance (Tao, 2012). We address semantic interoperability across sensor and information feeds through a self-described data source approach, exploiting metadata and schema from the World-Wide Web Consortium (W3C) and Open Geospatial Consortium (OGC). We provide agility of processing during crises through real-time steerable processing servers and context-aware information filtering.

2   PROBLEM STATEMENT

The TRIDEC project is looking at natural crises management for the purpose of tsunami early warning in the North-Eastern Atlantic and Mediterranean (NEAM) region. Following a crisis situation, such as a tsunamigenic earthquake, all the available sensor data will be used for decision support by a National Tsunami Warning Centre (NTWC). Figure 1 shows the effects of a Tsunami in Chile in 2010. Operators on duty need to assess quickly the relevance and the tsunamigenic properties of the earthquake event series, the likelihood of tsunami wave propagation, and the confirmation of tsunami progress. If necessary, the operator on duty will use information logistics and dissemination facilities to disseminate customized, user-tailored warnings to responsible authorities and to people under immediate risk.

It is the nature of natural crises that decision support tasks will change as the crisis evolves, requiring easy integration and re-configuration of data sources and agile processing that can be steered in accordance to current

task requirements and appropriate decision support. Multiple NTWCs will jointly share observations from multiple sensor networks in the NEAM region as well as information bulletins about imminent tsunamigenic events. The timely and reliable dissemination of bulletins and warning messages issued by governmental agencies has legal implications, adding to the challenge when setting up and testing communication channels. During a natural crisis situation, responsible agencies might also elect to forward situation information to other organisations such as Search and Rescue (SAR) or the remote sensing industry.

Figure 1. Damage by tsunami: City of Conception, Chile imaged on 10/01/2010 (left images) and 27/02/2010 (centre/right images) by the RapidEye satellite constellation. The centre image was taken eight hours after an earthquake of magnitude 8.8 had occurred and the resulting tsunami had affected the shoreline. The right map shows the areas with tsunami-related changes (red layer) on the post-tsunami image. Images courtesy of RapidEye AG, Copyright 2011, all rights reserved.

3   APPROACH

3.1   Geo-distributed heterogeneous data sources

Tsunami early warning in the NEAM region involves various warning centres operating on a local, national, and international level, linked to each other for centre-to-centre communication. On an international level information is shared across national borders following protocols negotiated among nation states in the Intergovernmental Coordination Group for the Tsunami Early Warning and Mitigation System in the NEAM (ICG/NEAMTWS) as part of the Intergovernmental Oceanographic Commission of the United Nations Educational, Scientific and Cultural Organization (IOC-UNESCO).

Due to the involvement of many nations and systems, the data being managed is very heterogeneous. In-situ sensor networks provide time series measurements from seismic sensors, tide gauges, and deep water buoys. These networks provide high quality configurable measurements but are few in number. Satellite systems provide streams of image data, and webcam footage at coastal sites provides video. The scale of measurement message throughput depends on the type and number of sensor networks connected, but as an example, a seismic sensor network of 171 sensors publishing at 50Hz will lead to about 4 GBytes of data recorded per day.

In addition to 'conventional' sensor measurement systems, there are expert reports available from simulations and alerting systems, such as tsunami simulations and the SeisComP3 earthquake alert system. In recent years vast amounts of Web 2.0 content has become available, such as Twitter messages, YouTube videos, and RSS feeds. These Web 2.0 'unconventional sensors' provide rapid in-situ crowd-sourced measurement by people actually experiencing the crisis event, e.g., using mobile devices, albeit with variable quality and having a high noise to signal ratio. In this way proven and reliable sensors are complemented by human sensors.

3.2   Scalable high performance event-driven middleware

A modern warning system following a system of systems approach has to integrate various components and sub-systems, such as different information sources, services, and simulation systems, taking into account the distributed and collaborative nature of warning systems. A system architecture implementing such a system of systems approach has to combine multiple technologies and architectural styles (Moßgraber, 2012). An important challenge is to promote a communication infrastructure that facilitates environment information services, both sensor-based and human-based, to work together managing disparate information sources providing very large volumes and dimensionality of data. Such a system needs to support: scalable, distributed messaging; asynchronous messaging; open messaging to handling changing clients, such as new and retired automated systems and human information sources becoming online or offline; flexible data filtering; and

heterogeneous access networks. In addition, the system needs to be resilient enough to handle ICT system failures, e.g., failure, degradation, and overloads, during environment events (Tao, 2012).

For the TRIDEC architecture, see Figure 2, we are looking at open source messaging platforms (Abie, 2009), which adopt a publish-subscribe model with brokers to overcome common problems, such as overload, interruptions, and the computational overhead of redundant nodes. Using a multi-bus approach supported by a hybrid MOM / SOA design, separating control channels (ESB) from content channels (multi-media) overloading a single communication channel is avoided. Tests (Sachs, 2009) of MOM's implementing a broker pattern show support for message rates of 14k/sec. The message throughput can be higher depending upon network capacity, server configuration, type of message interaction, and configuration but that of the order of 100k message rates at 100k/s (Red Hat Inc., 2008)
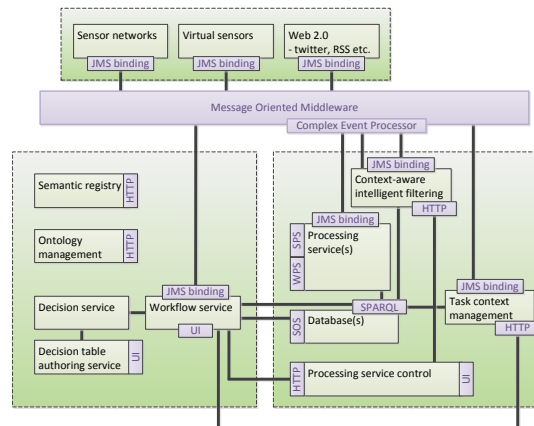
Figure 2. The TRIDEC hybrid service & event oriented multi-bus architecture

3.3   Semantic interoperability and the use of metadata driven pre-processing

We employ a self-describing 'plug-in' data source approach to manage semantic interoperability, separately publishing metadata and data. Published metadata describing the phenomena (e.g., water elevation), encoding (e.g., text encoded), and measurement device used (e.g., tide gauge type and id). This metadata is formatted according to the OGC Sensor Web Enablement (SWE) Observation & Measurement (Cox, 2011) model and used to configure 'on-demand' data parsers, which subsequently process all published data. The metadata describing the measurements (URIs, units, data access details) is uploaded to a triple store as an RDF graph. The sensor data (often large with 100,000+ measurements) and complex/binary data (e.g., images or simulations) is stored in relational databases and/or file storage, allowing high performance queries to large volumes of raw data.

Our data sources use different domain vocabularies to describe their measurements. We maintain a registry of Web Ontology Language (OWL) domain ontologies and inter-domain relationships, allowing automatic semantic mapping between identical and related measurement concepts. These semantic mappings coupled with the data source metadata allow pre-processing and dataset aggregation to be automated. Allowing 'plug and play' data sources without manual integration effort increases the scalability of our system of systems approach.

4   RESEARCH CHALLENGES AND IMPACT

One key challenge now is to use MOM technology to support complex event-driven messaging. MOMs can be designed to support message priorities, automatic client failover using configurable connection properties, queued data and metadata that is replicated across all nodes that make up a cluster, retry logic in the client code, and persistent published data queues and subscriptions (Wang, 2010). Complex event processing and event cataloguing seem inevitable (Yuan, 2009), but each event-oriented logic operation reduces the message throughput. A resilient model that is improved beyond resource hungry methods, such as broker mirroring and geo-resilience, must be implemented.

A few projects have now looked at using the OGC SWE standards for 'plug and play' sensor measurements (Middleton, 2010), exploiting OGC metadata to automate the pre-processing of data. These projects also use

catalogue services to overcome scalability issues with geo-distributed data and services. In TRIDEC we are building on these results with a semantic registry and self-described data sources via a MOM.

Coupling OGC XML and W3C RDF has been done at a small scale (Henson, 2009) or via a web portal (Janowicz, 2011), indirectly mapping RDF URLs to OGC SOS's XML queries for data. In TRIDEC we adopt a W3C linked data type approach, making our data directly accessible via a concrete URL. This allows us to add linked-data style annotations such as uncertainty information and processing provenance records.

5   CONCLUSION

Environmental information systems are becoming more and more complex as they increase in scale and scope. Data sources available to such systems of systems are geo-distributed and heterogeneous. Sensor networks such as seismic sensors, tide gauges, and satellite systems provide time series measurements, images, and video data. Expert reports from semi-automatic simulation and alerting systems are available, and recent Web 2.0 advances provide a way to crowd source 'unconventional' measurements, albeit with variable quality and a very high noise to signal ratio. Scalable middleware solutions supporting a system of systems, semantic interoperability among data sources, and agility of data processing are key challenges.

In the TRIDEC project we are adopting a multi-Bus system of system model, de-coupling geo-distributed data sources from data processing servers, and facilitating a scalable high performance messaging backbone. We are overcoming semantic interoperability within our heterogeneous multi-domain datasets by using a self-describing 'plug-in' data source approach, exploiting OGC and W3C standards for our information models, and using domain ontology mappings to automate pre-processing and aggregation of data from different domains. Lastly, we are adding agility to our processing servers by orchestrating processing workflows and deploying steerable processing server 'farms' that can adapt the data fusion and mining algorithm configuration as crises develop in real-time.

6   ACKNOWLEDGEMENTS

7   REFERENCES

Abie, H., Savola, R.M., & Dattani, I. (2009) Robust, Secure, Self-Adaptive and Resilient Messaging Middleware for Business Critical Systems. *IEEE Computation World*, pp 153-160.

Behrens, J., Androsov, A., Babeyko, A.Y., Harig, S., Klaschka, F., & Mentrup, L. (2010) A new multi-sensor approach to simulation assisted tsunami early warning. *Nat. Hazards Earth Syst. Sci. 10*, pp 1085–1100.

Cox, S. (2011) Observations and Measurements - XML Implementation, OGC ref 10-025r1.

Esbri, M.Á., Esteban, J.F., Hammitzsch, M., Lendholt, M., & Mutafungwa, E. (2010) DEWS: Distant Early Warning System - Innovative system for the early warning of tsunamis and other hazards. *In Jornadas Ibéricas de Infraestructuras de Datos Espaciales (JIIDE)*, Lisbon, Portugal, 27-29 October 2010.

Haener, R., Waechter, J., Hammitzsch, M., Lentholt, M., Sabeur, Z., & Poslad, S. (2011) Event-Driven Service Oriented Architecture Foundations for Industrial and Natural Crises Management Scenarios. *Geophysical Research Abstracts 13*.

Hanka, W., Saul, J., Weber, B., Becker, J., Harjadi, P., Fauzi, & GITEWS Seismology Group (2010) Real-time earthquake monitoring for tsunami warning in the Indian Ocean and beyond. *Nat. Hazards Earth Syst. Sci. 10*.

Henson, C.A., Pschorr, J.K., Sheth, A.P., & Thirunarayan, K. (2009) SemSOS: Semantic Sensor Observation Service. *International Symposium on Collaborative Technologies and Systems*, CTS '09, Baltimore, MD.

Janowicz, K., Broring, A., Stasch, C., Schade, S., Everding, T., & Llaves, A. (2011) A RESTful proxy and data model for linked sensor data. *International Journal of Digital Earth.*

Middleton, S.E. (2010) SANY fusion and modelling architecture. Open Geospatial Consortium (OGC) Discussion Paper, OGC ref 10-001.

Moßgraber, J., Middleton, S., Hammitzsch, M., & Poslad, S. (2012) A Distributed Architecture for Tsunami Early Warning and Collaborative Decision-support in Crises. *Geophysical Research Abstracts 14*.

Münch, U., Rudloff, A., & Lauterjung, J. (2011) The GITEWS Project – results, summary and outlook. *Nat. Hazards Earth Syst. Sci. 11*, pp 765–769.

Poslad, S. (2009) *Ubiquitous Computing: Smart Devices, Environments and Interactions*, ISBN: 978-0-470-03560-3, Wiley.

Red Hat Inc. (2008) Red Hat Enterprise MRG – MRG Messaging: Throughput, Red Hat Reference Architecture Series.

Sabeur, Z., Wächter, J., Küppers, A., & Watson, K. (2011) Large Environmental Sensor Observation Data and Knowledge Base for Collaborative Decision-support in Crises. *Geophysical Research Abstracts 13*.

Sachs, K., Kounev, S., Bacon, J., & Buchmann, A. (2009) Performance evaluation of message-oriented middleware using the SPECjms2007 benchmark. *Performance Evaluation 66* (8) pp 410-434.

Tao, R., Poslad, S., Moßgraber, J., Middleton, S., & Hammitzsch M. (2012) Scalable and Resilient Middleware to Handle Information Exchange during Environment Crisis. *Geophysical Research Abstracts 14*.

Wächter, J., Fleischer, J., Häner, R., Küppers, A.N., Lendholt, M., & Hammitzsch, M. (2011) Development of Tsunami Early Warning Systems and Future Challenges. *Geophysical Research Abstracts 13*.

Wang, J., Jiang, P., Bigham, J., Chew, B., VinalMurciano, B., Novkovic, M., & Dattani, I. (2010) Adding resilience to message oriented middleware. *2nd Int. Workshop on Software Engineering for Resilient Systems.*

Yuan, S. & Lu, M. (2009) An value-centric event driven model and architecture: A case study of adaptive complement of SOA for distributed care service delivery. *Expert Systems with Applications 36* (2).

(Article history: Available online 24 February 2013)