Final Report
Secure Heterogeneous Information Presentation
– SHIP –

Marc Bezem
The Programming Technology Group (PTG)
Department of Computer Science, University of Bergen

March 2012

1 Summary

The SHIP project addresses the technological challenge of presenting heterogeneous content on heterogeneous platforms in a secure way. The complete project description can be found on the project homepage `http://www.ii.uib.no/ship/`. In addition one can find there:

- A list with the main participants
- Complete progress reports 2007-2011
- Links to related activities and software produced in SHIP

This report is based on a selection of the above information, but will present the results from a different perspective, making use of the fact that the form of this part of the report is not bound to the strict limitations inherent to filling out webforms. In particular, we will not duplicate the long lists of publications and other quantitative information that can be found in the sections "Resultatindikatorer" and "Publiseringsinformasjon" of the yearly progress reports, including the current report on the web.

In summary, we have been very happy with the SHIP project. First of all, SHIP has made it possible to do research for which we otherwise wouldn't have had the resources. Secondly, SHIP has strengthened our national and international collaborations. Finally, it has given a number of our students (M.Sc. and Ph.D.) a good start of their career as an academic, both at universities, at research institutions, in companies and in the public service. In what follows we present and discuss the results in the light of the goals as formulated in the original project proposal.

2 Goals and Results

2.1 Goals

The overall goals of the SHIP project are as follows:

G1. Advancing the state-of-the-art of network technology
by developing a flexible framework for

- handling data and
- organising software

for heterogeneous content on heterogeneous platforms.

Important results in relation to this goal are: the Dynamic Presentation Generator (DPG), the SHIP Validator, and the secureXdata Protocol. All these advances of the state-of-the-art are described in more detail in Section 3 Prototypes.

G2. Strengthening the national competence base around these issues

- Increasing the knowledge base of the researchers involved in the project
- Educating new experts (M.Sc. and Ph.D.) in the field

This goal has been met to a large extent. All researchers involved in SHIP have increased their knowledge to the extent that their results have been published on an international level. During the project period, 10 M.Sc. students have graduated on SHIP-related theses. In 2011 and early 2012, the following M.Sc. students have graduated (thesis title in parentheses): Morten Høiland (Datainnsamling med XForms i Dynamic Presentation Generator)*; Øystein Lund Rolland* (Design og støtte for XForms i Dynamic Presentation Generator)*; Kelly Alexander Teigland Whiteley* (Resource management for plugins in the Dynamic Presentation Generator)*; Aleksander Vatle Waage* (Støtte for Geodata i Dynamic Presentation Generator)*; Remi Valvik* (Security API for Java ME: secureXdata)*.*

On the Ph.D. level, Dag Hovland graduated on 16 December 2010 on a Ph.D.-thesis entitled Feasible Algorithms for Semantics — Employing Automata and Inference Systems*. The other Ph.D. student in SHIP, Paul-Simon Svanberg, still works on his manuscript of currently more than 100 pages. We expect him to deliver in 2012.*

G3. Transfer of the existing and acquired knowledge to industry

- Close cooperation with industrial partners
- Development of prototype solutions based on the acquired insights

The overall majority of our M.Sc. students, including the five mentioned above, get almost immediately a job in the IT sector. In this way the transfer of knowledge to *industry is clearly warranted. On the Ph.D. and Postdoc level the situation is more diverse. Dag Hovland is currently a Postdoc at the UiO. Postdoctor Håvard Raddum has a permanent position as an expert in IT-security in the Fylkeskommune. Paul-Simon Svanberg has become a librarian at UiB. Postdoctor Federico Mancini got a permanent position at Forsvarets Forskningsinstitutt (FFI). All this underlines the importance of research projects such as SHIP for recruitment to key functions in the Norwegian society.*

On the point of cooperation with industrial partners, CellVision AS and Intelinet AS, the following should be remarked. These are small companies, both employing M.Sc. graduates coming from the PTG group (Haakon Straume, graduated in 2005, and Lars Søraas, graduated in 1994). It turned out to be difficult to maintain a stable collaboration with these companies over a long period. Small companies are required to shift their focus rapidly in response to developments in the market. In the economical turbulence in 2008, they lost their interest in SHIP. In a short period, there has been a collaboration with NoWires AS, a small company started by Klingsheim and Nettland, two Ph.D.'s graduated at our department. There are some joint publications and a contribution to the SHIP seminar, 3 June 2009. However, also this company didn't survive the economic downturn.

On the other hand, the cooperation with Høgskolen i Bergen has flourished and still continues to do so. A Ph.D. student at HiB, Adrian Rutle, graduated on a topic closely related to SHIP, which has led to several joint publications. Also, new collaborations have been initiated during SHIP. Here we can mention the collaboration with the Centre for International Health (SIH). This collaboration in the field of mHealth has led to the third prototype described below, a Mobile Data Collection System based on openXdata. A new Ph.D. student, Samson Gejibo, was first sponsored by SIH and has now a Ph.D. grant from our department. This is a clear spin-off of SHIP and will ensure a continuation for at least the next four years. Finally, former SHIP postdocs Federico Mancini (currently FFI) and Håvard Raddum (currently Fylkeskommune) have been appointed as associate professor (20%) at our department, measures to consolidate the competence built-up in SHIP.

Three prototypes have been developed. They will be described extensively below.

G4. Disseminating and exchanging the emerging knowledge

- Publications in both (inter)national conferences and journals
- Continuing and strengthening the existing cooperation with the researchers abroad

These goals have been completely met. For the publications we refer to the lists elsewhere in this report. The many co-authors from various different institutions and countries are a clear sign of good cooperation.

2.2 Results

The following results were planned:

R1. Qualified competence in the form of Ph.D. and M.Sc. degrees in connection with the project. *See under goal* **G2** *above.*

R2. A framework for handling data and organizing software for secure information presentation (goal **G1**, **G2**).

The Dynamic Presentation Generator is such a framework. The other prototypes and many of the other results concern tools and methods supporting various aspects of secure information presentation.

R3. A prototype in the field of distant learning (goal **G3**, **G2**, **G1**).

DPG as explained in Section 3.1, substantiates this result.

R4. Workshops with industrial partners included (goal **G3**).

Apart from the SHIP-seminar (3 June 2009) we organized two major international conferences, TYPES and CSL'11 in Bergen (8–15 September 2011). There were in total 133 participants from over 20 countries. The proceedings of CSL'11 are publised in the series LIPIcs, Leibniz International Proceedings in Informatics (`http://www.dagstuhl.de/dagpub/978-3-939897-32-3`). A selection of papers will be published as a special issue of the journal Logical Methods in Computer Science.

R5. Publications on the framework and on the new theoretical insights.

See under goal **G4**.

2.3 Forthcoming

- Dag Hovland, *The Inclusion Problem for Regular Expressions.* To appear in Journal of Computer and System Sciences. Published electronically: `http://dx.doi.org/10.1016/j.jcss.2011.12.003`

- Timos Antonopoulos, Dag Hovland, Wim Martens and Frank Neven, *Deciding Twig-Definability of Node Selecting Tree Automata.* To appear in Proceedings International Conference on Database Theory, March 26–29, 2012.

- F. Mancini, S. H. Gejibo, K. A. Mughal, J. I. Klungsyr, R. B. Valvik, *On the Security of Mobile Data Collection Systems in Low-Budget Settings*, submitted.

3 SHIP prototypes (science outreach version)

This section describes the prototypes developed during the SHIP project. The descriptions are intendedly non-technical, or at least not-too-technical, so that they can be appreciated by a wider public. This section also serves as the last point in the web version of the final report, under the title "Særskilt rapportering".

3.1 Dynamic Presentation Generator

By the word 'presentation' in the title we mean the general process of information disclosure, exchange and use. Secure presentation puts serious and contradicting demands on the information providers. Users want easily accessible information structured for their purposes and tailored for presentation in any available format. The provider must ensure that the information content always makes sense, no matter on what platform or in what (supported) form it is presented, and that information is only disclosed to authorized users.

E-learning (or computer-based learning) is seen by many as the holy grail to provide education to the masses, moving education out from the traditional classroom. The scope of educational material on the internet ranges from learning a new language, repairing your car, tuning an instrument, to studying for an academic degree. The advantages are clear: no travelling, no fixed schedule, no dependence on the availability of teachers, no limit on the number of students etcetera. Few people know, however, how such educational material is produced, and that this requires enormous resources. To produce educational material for e-learning in an effective way it is beneficial to distinguish between *form* and *content*. As an example, imagine two courses to repair a punctured tyre in two different languages. The form of these courses could be completely identical, the content only differs in the language. More complicated is the example of two programming courses in two different programming languages, say, Java and Python. Again the content would be rather different, but the form, the way the content is presented, would be strikingly similar. This similarity can be defined by a *presentation pattern*. With its obvious security challenges (f.e. exams), e-learning provides a challenging application field for SHIP.

The Dynamic Presentation Generator (DPG) is the prototype system behind the e-learning courses INF100F and INF101F in programming and the course MAR252 (all at UiB). Various components of the system have been developed since 1998. It has been tested extensively (each course runs every year). The innovation of DPG is that it lets you define a wide range of presentation patterns tailored for e-learning, which can then be filled with educational material for the course in question. In this way the course-designer can focus on content, and courses can be defined in an effective and uniform way. Moreover, patterns can be created to design web presentations for football clubs, nursery schools, online newspapers–virtually presenting all kinds of information on the web.

The current version of DPG 2.2 (2012) now has the capabilities to present audio, video and geodata content, among other types of media. Instead of just pushing content at the end-user, the system can now be used to generate web pages for interaction with the user. In particular, the system supports user interaction via XForms, so that forms can used to collect data from the user. Of course the data collected can in turn be processed and visualized using a presentation pattern designed for DPG. The current effort is to couple DPG with the user management in the course portal (MiSide) of the University of Bergen, making the system even more flexible for creating content for both off-campus and on-campus courses.

3.2 SHIP Validator

One important security issue is the source of many web application vulnerabilities: the lack of proper input validation. Input validation can be described as the process of checking that input to a web application is of the right type and format, before processing this input. An example of what can go wrong without input validation is SQL-scripting: Not knowing the secret password, one could enter as password "foo OR TRUE". An (insecure) system would verify "password == foo OR TRUE" which evaluates to "TRUE" even if the password "foo" is false! This should be prevented by proper input validation, which should reject the input "foo OR TRUE".

Since the task of creating input validation tests can be very repetitive and tedious, we created a framework for input validation where the tests are independent of the underlying application, and can therefore easily be customized, reused and combined in various ways to create new specific tests. The framework is developed for the Java platform, as this is one of the leading technologies used for developing web applications.

The project was initially inspired by the input validation framework by Heimdall, where the main goal is to provide a clear separation between validation and application logic. This separation was achieved by using an XML configuration file defining which tests were to be run on which object properties. The first step of our project consisted in checking whether the need for an XML external file could be eliminated by using Java-annotations to associate tests and object properties, instead. After a new input validation framework based on annotations was succesfully implemented, the focus shifted to investigate how far annotations can be pushed for validation purposes, while keeping their use as intuitive and simple as possible.

Main features that characterize the framework are:

- Easy integration in any existing Java projects
- High reusability of existing validation tests
- Possibility of creating new custom annotations with little effort
- Composed annotations: boolean combination of existing annotations to create new tests without the need of writing new code
- Cross annotations: supporting tests on multiple object properties having inter-dependent validation constraints, instead of only on single ones

The SHIP-validator is one of the main practical results of the SHIP project. For more informations, see `http://www.owasp.org/index.php/Category:OWASP_Content_Validation_using_Java_Annotations_Project`

3.3 The secureXdata Protocol

Mobile devices are having a profound impact on how services can be delivered and how information can be shared. New mobile based systems are being developed to allow communities living in remote and inaccessible areas to connect to the health care providers and get basic health services and information. The information collected in remote communities can then be relayed to local health care centers and from there to the decision makers who are thus empowered to make timely decisions. An additional challenge is that it is mandatory for any system in this setting to be robust and low-cost.

Many of the existing solutions do not systematically address very important security issues which are critical when dealing with such sensitive and private information. Moreover, many existing solutions require costly technology. Our research has focused on providing a framework that can be used to integrate security in a seamless way into the existing mHealth systems, and that provides a cost-effective way for ensuring data confidentiality, both when the data is stored on the mobile device and when it is transmitted to the server.

Unfortunately, well-known and recommended security solutions had to be ruled out because of specific requirements imposed by mobile platforms (for example, whether HTTPS is available or not, and available libraries) and because of challenges imposed by the working environment in which the system is supposed to function (eg. mobile network coverage).

We have therefore designed a protocol that guarantees, besides secure storage and communication, also data integrity, off-line and on-line authentication, account and data-recovery mechanisms, multi-user management and flexible secure configuration. Our protocol has been also implemented as an API that can easily be integrated with mobile collection systems running on the Java ME platform. The data is encapsulated both for storage and transmission, hence leaving the current implementation almost untouched.

A prototype of our secure solution has been integrated with openXdata, a Mobile Data Collection System that is primarily designed for data collection using low-end Java-enabled phones in low-budget settings. See `www.openxdata.org` for more information.