Project no.	034084
Project acronym:	SELFMAN
Project title:	*Self Management for Large-Scale Distributed Systems based on Structured Overlay Networks and Components*

European Sixth Framework Programme Priority 2, Information Society Technologies

Final Activity Report - Year Four (M40)

Due date of deliverable:	Nov. 15, 2009
Actual submission date:	Nov. 15, 2009
Start date of project:	June 1, 2006
Duration:	40 months
Dissemination level:	CO

Contents

1 Project execution

The ultimate goal of the SELFMAN project was to make distributed systems self managing. The systems will reconfigure themselves to handle changes in their environment or requirements without direct human intervention but according to high-level management policies that can be set by humans. Making distributed systems self managing is crucial for building large systems, for which the task of management is overwhelming. "Abnormal" events such as software updates, faults, threats, and performance hotspots become normal and even frequent occurrences. We focus on four axes of self management, namely self configuration, self healing, self tuning, and self protection.

A major innovation of SELFMAN is to combine the complementary strengths of structured overlay networks and advanced component models. Structured overlay networks originated with peer-to-peer file sharing applications, but have matured to provide strong guarantees and efficient communication and storage operations. This makes them suitable as general platforms for hosting services instead of just for file sharing. To achieve this, the advanced component models leverages the overlay network: it make the overlay network modular and it adds the hooks needed for self management. The hooks provide introspection, reflections, and dynamic reconfiguration abilities. The overlay network can host an application that manages itself.

SELFMAN has reconstructed the structured overlay network as part of a self-managing component architecture and used it to build high-level self-managing services. We have built powerful services including replication and transactional storage and we have used them to build scalable Internet applications.

In the following sections we give the two major breakthroughs of SELFMAN as well as the most important software results and scientific results, and the media coverage that SELFMAN has enjoyed.

1.1 Breakthrough results

SELFMAN has developed two breakthrough results as well as numerous scientific advances at all levels of self management for distributed systems. The breakthrough results are in distributed transactional storage and media distribution:

- **PeerTV**. PeerTV is a state-of-the-art application developed as a product by partner Peerialism that distributes video over Internet with live streaming and progressive download. It uses advanced technology based on peer-to-peer structured overlay networks, optimization

algorithms, and component models. It uses the Chandelier dynamic peer-to-peer optimization algorithm and was developed with an integrated simulation tool MyP2PWorld, which radically sped up development and shorten time to market. It uses new techniques for firewall hole punching and NAT traversal of video streams. PeerTV has comparable QoS to leading distribution providers but at much lower costs. PeerTV was developed for and is being used by the Swedish company MPS Broadband AB.

- **Scalaris**. Scalaris is an open-source library that provides a scalable global storage service. Scalaris is built on top of a structured peer-to-peer network and implements a key/value store that supports transactions. It survives node crashes and network problems using replication and a sophisticated consensus algorithm. It scales to hundreds of nodes and provides strong data consistency in the face of concurrent operations, node failures, and network problems. It does load balancing with an algorithm that uses estimates of global knowledge to reduce the storage load deviation between nodes. Scalaris was used to implement a version of Wikipedia that won first prize in the IEEE International Scalable Computing Challenge 2008. It is competitive in performance to the actual Wikipedia backend (14,000 read+write transactions / second on a synthetic benchmark) but is more robust and scalable.

1.2 Major software results

SELFMAN has produced many other major and minor software results. We have selected the most important software:

- **Beernet**. Beernet is a modification of Scalaris that is based on a "relaxed ring" overlay network, which is easier to manage. It relaxes the connectivity condition of the ring underlying the structured overlay network. It requires only that a node be in the same ring as its successor (instead of both its successor and predecessor). Beernet also modifies the transaction algorithm to request locks quickly and to notify all nodes of modified state. We have used Beernet to implement the collaborative drawing application DeTransDraw on gPhones: HTC Magic mobile phones running the Android operating system. DeTransDraw uses transactions to maintain global coherence of the drawing as well as to overcome network latency (edits are immediately shown locally while the transaction algorithm attempts to obtain a global commit).

- **Kompics**. Kompics is an advanced component model for building reconfigurable distributed systems from event-driven components. Kompics systems can be uniformly evaluated in large-scale reproducible simulation and distributed deployment, using both the same system code and the same experiment scenarios. Kompics components are concurrent and readily exploit multi-core architectures. They are decoupled by publish/subscribe ports and channels and are compositional. They can form dynamically reconfigurable architectures and fault supervision hierarchies.

- **FructOz/LactOz/WorkflOz**. This is a complete self-configuration framework consisting of three libraries: FructOz for dynamic deployment and configuration, LactOz for dynamic navigation and monitoring, and WorkflOz for dynamic distributed workflows. FructOz implements the Fractal component model on top of Oz. LactOz provides navigation and monitoring abilities on top of FructOz structures. WorkflOz allows the construction of workflows as FructOz structures, which allows LactOz to monitor workflow execution.

- **Mozart Programming System 1.4.0**. The Mozart system is a software development platform based on the Oz multiparadigm programming language. It supports highly dynamic applications and fine-grain concurrency. Mozart 1.4.0 provides an advanced transparent distribution subsystem that allows to develop distributed applications very similar to centralized applications. All language entities (objects, components, dataflow variables, threads, ports, etc.) have distribution protocols that let them work in a distributed setting. The Mozart 1.4.0 fault model allows fault-tolerance abstractions to be built within the language, thus keeping transparency. Mozart 1.4.0 was used to build the Beernet library.

1.3 Major scientific results

SELFMAN has produced many scientific results. We have selected the three most important results:

- **Atomic transactions on a peer-to-peer network**. We have developed a transaction manager that works for data stored on structured peer-to-peer networks. The heart of the transaction manager is the algorithm that implements the atomic commit. It is an optimized version of Lamport's Paxos uniform consensus algorithm that needs only four communication steps instead of six in the common case when there are

no failures. The improvement was possible by exploiting information from the data replication in the structured peer-to-peer network. The transaction manager works correctly under realistic Internet conditions, where at any time nodes can crash or the network can have problems. The Scalaris library uses the atomic commit algorithm to implement strong data consistency and atomic transactions.

- **Merge algorithm for network partitioning**. We have solved the network partitioning problem for structured overlay networks. If the network is partitioned, the overlay splits into several smaller overlays, which each continue to provide its service as best it can with the nodes it contains. If the partition goes away (the network is repaired), then the merge algorithm will automatically combine the smaller overlays back into a single large overlay, thus restoring the complete service. This behavior can be seen a reversible phase transition, in analogy with thermodynamics. It can be used to build extremely robust Internet applications that survive network partitioning.

- **Methodology for building self-managing applications**. The SELFMAN project has pushed the state of the art in software development techniques for self-managing applications. Out of the practical experience of SELFMAN, we have derived a development methodology for large-scale distributed systems based on the concept of *weakly interacting feedback structures*. A feedback structure is a hierarchy of interacting feedback loops that together maintain one global system property. This gives a much more natural and powerful way of designing large systems than the traditional layered approach. We have applied this methodology to decentralized distributed systems and in particular to the Beernet and Scalaris systems. Considered in this way, Scalaris consists of six weakly interacting feedback structures. We are continuing to develop and extend this methodology, especially for the development of extremely robust applications, by generalizing the idea of reversible phase transitions.

1.4 Media coverage

SELFMAN has received significant media coverage:

- **Peerialism acquisition**. The Swedish company GGF (Global Gaming Factory X) attempted in Aug. 2009 to acquire The Pirate Bay (the eighth most popular website on the Internet) together with SELFMAN partner Peerialism. The PeerTV product as provided by Peerialism

would have become the media-streaming technology underlying a legal version of The Pirate Bay. The acquisition fell through because of financial difficulties at GGF, but the resulting publicity has propelled Peerialism to world recognition.

- **Articles in the press**. SELFMAN has been the subject of numerous articles on the Internet since Oct. 2009, in online technology newspapers, blogs, and the paper newspaper Automatisering Gids (in Dutch). Articles have appeared in ICT Results, ACM Technews, ScienceDaily, MegaPlatinum, VWN News, NUZE.ME. Newstin, Fast Company, ReadWriteWeb, PhysOrg.com, Technology.am, AlphaGalileo, and The Web Scene. Foreign language articles have appeared in French, Russian, and Portuguese. The technology blog RobotArmageddon has compared SELFMAN to Skynet (which takes over the world in the Terminator films) because of its self-defending and self-healing characteristics.

2 Dissemination and use

2.1 Exploitable knowledge and its use

Exploitable knowledge	Exploitable product(s)	Appl. sector	Date for comm. use	Patents or IPR	Owner
Transaction algorithm	Scalaris	Internet	2009	license	ZIB
Transaction algorithm	Beernet	Internet	2010	license	UCL
Media streaming	PeerTV	Internet	2009	trade secret	PEER
Simulation	MyP2PWorld	software		license	PEER
Self protection	Wikimedia Credibility Extension	Internet	2009	license	NUS
Simulation	SWN simulator	software	2010	license	NUS
Simulation	SicSim	software	2009	license	KTH
Component model	Kompics	software	2010	license	KTH
Deployment & configuration library	FructOz	software	2010	license	INRIA
Navigation & monitoring library	LactOz	software	2010	license	INRIA
Workflow library	WorkflOz	software	2010	license	INRIA

2.2 Dissemination of knowledge

Date	Type	Type of audience	Countries addressed	Size of audience	Partner
Jun. 2006	Project website	researchers	world		UCL
Nov. 2006	0.5-page ad Parliament Magazine	deciders	European Parliament	thousands	UCL
Aug. 2007	Almende Summer School	researchers	Europe	appx. 50	UCL
May 2008	IEEE Competition first prize	researchers	world		ZIB
Jun. 2008	3-page article eStrategies Projects	deciders	Europe	39,000	UCL
Oct. 2008	Workshop SASO 2008	researchers	world	appx. 50	UCL
Nov. 2008	2-page article Research Review	deciders	Europe	thousands	UCL
Nov. 2008	3-page article TheServerSide	IT industry	world		ZIB
Dec. 2008	Web banner PSCA Int.	deciders	Europe		UCL
Aug. 2009	Press on GGF acquisition	IT industry	world		PEER
Sep. 2009	Workshop SASO 2009	researchers	world	appx. 50	INRIA
Sep. 2009	Web banner PSCA Int.	deciders	Europe		UCL
Oct. 2009	2 articles ICT Results	researchers	world		UCL
Oct. 2009	1 article ACM Technews	researchers	world		UCL
Oct. 2009	>15 blogs	IT industry	world		UCL
Oct. 2009	1-page article Automatisering Gids (Dutch)	IT industry	Netherlands	thousands	UCL
Jan. 2010	1-page article PS Review	deciders	Europe	thousands	UCL
Jan. 2010	1-page editorial PS Review	deciders	Europe	thousands	UCL