



Project no. 001935  
Project acronym: EVERGROW  
Project title: *ever-growing global, scale-free networks,  
their provisioning, repair and unique functions*

## INTEGRATED PROJECT

### IST PRIORITY

#### Periodic Activity Report

Period covered: January 1, 2007—December 31, 2007  
Date of preparation: February 10, 2008

Start date of project: January 1, 2004  
Duration: 48 months  
Project coordinator name: Scott Kirkpatrick and Seif Haridi  
Project coordinator organisation name: SICS  
Revision: 1

# Contents

<b>1</b>	<b>Project Objectives and Achievements</b>	<b>2</b>
<b>2</b>	<b>Workpackage Progress</b>	<b>6</b>
2.1	SP1: Measurement and modelling . . . . .	6
2.1.1	Topology measurement . . . . .	6
2.1.2	Traffic measurement . . . . .	7
2.1.3	Integration of Topology and Traffic information into P2P design . . . . .	10
2.2	SP2: Virtual Network Observatory . . . . .	12
2.2.1	Virtual Network Observatory . . . . .	12
2.2.2	Data repository . . . . .	13
2.2.3	Data mining . . . . .	13
2.3	SP3: Self-healing Systems . . . . .	15
2.3.1	Algorithms for overlay networks . . . . .	15
2.3.2	Services and overlay systems . . . . .	17
2.3.3	Storage and content distribution . . . . .	20
2.3.4	Analysis of overlay systems . . . . .	21
2.4	SP4: Message Passing . . . . .	22
2.4.1	Belief and survey . . . . .	22
<b>3</b>	<b>Consortium Management</b>	<b>23</b>
3.1	Consortium management task and their achievement . . . . .	23
3.1.1	Interacting with the Commission on behalf of the project . . . . .	23
3.1.2	Preparing and submitting annual project budgets to the Commission . . . . .	23
3.1.3	Deciding reconfiguration of the consortium . . . . .	24
3.1.4	Handling financial and administrative management of the consortium . . . . .	24
3.1.5	The Administrative Team at SICS . . . . .	24
3.1.6	The Board . . . . .	24
3.1.7	List of decisions by the EVERGROW General Assembly . . . . .	24
3.1.8	List of decisions by the EVERGROW Board . . . . .	24
3.1.9	List of decisions by the EVERGROW Executive . . . . .	25
3.2	Project timetable and status . . . . .	25
3.2.1	Full IP meetings . . . . .	26
3.2.2	Subproject 1 meetings . . . . .	26
3.2.3	Subproject 2 meetings . . . . .	26
3.2.4	Subproject 3 meetings . . . . .	26
3.2.5	Subproject 4 meetings . . . . .	26
3.3	Co-operation with other projects . . . . .	26
<b>4</b>	<b>Other Issues</b>	<b>28</b>

## **Publishable Executive Summary**

Material for public dissemination will be supplied separately.

# 1 Project Objectives and Achievements

The vision of EVERGROW, turned into operational details, is that

1. *We will measure the Internet at its physical levels with unprecedented precision.*

This is the focus of SP1. The management challenge during this plan period will be making the resulting data available to the many groups within EVERGROW and the related IPs in the complexity initiative that can benefit. This is being done through publication of standard data sets on the relevant websites, and through meetings held between research groups and between subprojects to determine what data sets are needed and how to characterize their accuracy. Work coordinated between the two SP1 measurement groups is producing new measurement methods combining the unique distribution of the software active measurement with the unique high temporal precision of the hardware based methods.

2. *We will capture these measurements in a tool which generates models of the Internet for use in simulation, with parameters which let us extrapolate into the future by estimating the changes occurring in those parameters.*

WP2c has overall leadership in this effort. They have a milestone and a deliverable for data to be used in WP1c, and will draw on SP1's results and work in SP3 and SP4 to achieve this in a way that advances the interests of both P2P modelling and exploits the best applications of message passing. WP1c is taking the measurement data and making it directly relevant to Peer-to-peer applications, a step towards a more fully autonomous Internet.

3. *We will use these models of the present and future Internet to simulate architectural strategies for peer to peer services more accurately than is common today, using our clusters to facilitate this.*

In year 2 of EVERGROW, WP2a completed the integrated compute platform, EVERLAB, which permits interprocess communication between agents or applications running in the different clusters, over the open Internet. This is now in use as a network simulation platform permitting study of distributed coding of information. Operating in a slightly different configuration, our clusters support fast network modelling in which peer to peer applications interact directly with network topology and delay information to increase their performance.

4. *We will draw on theoretical developments in distributed computing from the survey propagation community to make these architectural ideas and P2P algorithms run faster and more optimally.*

This is the responsibility of SP3 and SP4, being managed by SP3. Applications to fast, approximate network topology have already been obtained.

In the first annual review of EVERGROW, the project performance was rated satisfactory, all deliverables were accepted, and the research was judged to be proceeding at good speed, of high quality, and in places unique and world-leading. However, we were criticised for insufficient overall integration between the the different subprojects and with the complex systems community more broadly. In response to the organizational issues that were raised we made modifications to the implementation plan, as articulated in the discussion above of making the EVERGROW vision an operational plan. These changes were accepted by the review panel and the EC. Further changes, including an explicit

integration workpackage, were incorporated in the second year's plan and were very favorably received in the second year review. That review again concluded that the research presented was at the highest level, and in some cases, world-leading.

There were two recommended actions that emerged from the review held in Zurich last March: Increase utilization of the clusters, perhaps by encouraging outside use and formulate plans to ensure value of the virtual observatory's data beyond the conclusion of EVERGROW. The full 2006 review report is available to this year's reviewers on our <http://www.evergrow.org> website.

We believe the first was less of a problem than it seemed at the time of the review. We will show usage data from our monitors that confirm reasonably full and level utilization throughout the concluding year of the project. In order to increase utilization, we first arranged small modifications of the software complement provided on the systems seeing high local usage in the modelling efforts. The answer to the second is principally our role within ONELAB2 and projects which hopefully will follow that. We are following a path of regular storage upgrades and compute capability migration, using whatever funds become available through our various means of support.

Budgeting: At the second year review, we presented a two-year plan taking EVERGROW to its conclusion. In the third year review, half-way through that plan, we needed to make only minor modifications. Our review at the end of the third year was extremely positive. Our budgets for the final two years are in effect as stated at last year's review and approved by the EVERGROW General Assembly in early 2006.

One unstated requirement on EVERGROW as both a complex systems project and a participant in the networking community has been to make its insights available in planning for FP7. EVERGROW partners and individual researchers were active during 2005 and 2006 in the ONCE-CS and Beyond the Horizon efforts, sponsored by FET, to identify the elements of complex systems science and large scale network-based activities that will help to address the challenges of the coming decade. We were also visible among the lecturers and tutorial speakers of the complex systems conferences, such as ECSS2005, 2006 and 2007. Several EVERGROW members founded the ARCADIA action to discuss coordinating future internet research across the European, US and Far Eastern geographies. In addition, we have participated in the EIFFEL task force and the FIRE effort to define more specifically what key topics should be focused upon in FP7 for future Internet research. EVERGROW's tendency has been to argue for greater attention to be paid to the edge of the Internet, to how applications will require more flexible infrastructures and thus drive virtualization. Our EVERLAB (a private PlanetLab) has provided a concrete example of the problems that must be solved in federating advanced testbeds and ultimately next generation resources, as well as offering some new tools to make this possible. We have worked closely with the US PlanetLab and European ONELAB people in this. As a result of this activity, three parts of the EVERGROW project, DIMES, ETOMIC and EVERLAB will continue uninterrupted into 2008-2010 as key elements of the recently funded ONELAB2 activity, in which a number of advanced European testbeds are federated, made more widely available, and used as a platform for pursuing potentially mold-breaking research in networking for the coming decade, under FIRE sponsorship.

In the year 2005, we made item 1 a reality, with Over 8000 DIMES agents registered and nearly one billion measurements of the Internet's topology (its logical structure as a graph of connections between points where computing and data reside). In 2006, this growth continued, reaching over 12000 DIMES agents registered, over two billion measurements completed, and the beginnings of a data archive from which extrapolations into the future will be possible. In 2007, the number of registered agents has passed 18000,

and the total number of measurements archived now exceeds 3.5 billion. We have begun to automatically identify "point-of-presence" or POP structure within the largest subnetworks that comprise the Internet, allowing us a granularity of description that exceeds the traditional "Autonomous System" or AS-level, provides geographic locality in the description, yet is more manageable and useful than attempting to work with a map of the globe's IP-addresses. Further analysis has given insight into the question of accuracy and reliability of predictions that can be made with this growing body of data.

Throughout 2006 and 2007, regular measurements using the high precision ETOMIC systems have been accumulated over the central portion of the European data network, and are accessible from a central database with published schema. The DIMES and ETOMIC tools are now becoming integrated into a single complex, with packettrains (an ETOMIC procedure intended to explore bandwidth and capacity) now deployed as a measurement tool on every DIMES agent. This to date is research data, from which the first papers are steadily emerging in the hands of the DIMES and ETOMIC experimenters. Our challenge of the final two years will be to make this data broadly usable and accessible to researchers in both the networking and complex systems fields. The deliverables D.1a.4 and D.1b.4 summarize our status in this as the EVERGROW project concludes.

Last year we initiated new activity across the EVERGROW SPs to make this data useable and accessible to programs as well as to researchers, specifically to future peer-to-peer (P2P) applications and services. We have also completed several first-pass analyses of the topology data from DIMES and the delay/bandwidth data from ETOMIC, drawing on partners in SP2 and SP4 for analytical tools, insights into the graph theory involved, and new algorithms for fitting and interpreting the data. So item 2 in the above vision will be the centerpiece of our activities in the years 3 and 4 of EVERGROW. The deliverables in this activity occur at Month 48. Last year we presented a white paper which was a result of the activity in this WP. This year we present the modelling system, GODS, which was developed and describe some of the first results obtained with this tool, using data obtained by the EVERGROW measurement tools.

Item 3, experimentation with novel P2P architectures for content distribution, search and other distributed applications (our steps toward the ultimate versions of Google and Akamai) has continued, using the PlanetLab platform developed at Princeton University. As a foundation for scaling up these experiments and extrapolating them to the Internet of future decades, we have established a private PlanetLab, called EVERLAB. This will answer to the needs of both high performance "embarrassingly parallel" simulation computations and network measurements of distributed computation between tens of thousands of interacting agents in a way that existing commercial GRID developments still cannot. Deliverable 2a.3 describes the EVERLAB status. As the platform on which we can deploy our Internet predictive (or at least extrapolative) generated models, it is playing a central role in the EVERGROW work of the final two years.

About one-third of the EVERGROW partners have been also making significant scientific contributions to the science of distributed computing, its algorithms, analysis and applications. This work is presented at leading scientific conferences, and has in some cases flowed back into the technologies which we employ in the tools that EVERGROW has realized. As the project wraps up, we have results to report in four workpackages, led by researchers in four institutions. Each of the workpackages has involved collaborative work across two to four institutions, work which probably would not have been possible without the stimulus of the EVERGROW consortium and its common goals. For the final review, we will present work on basic algorithms for resilient distributed processes, in which a stochastic overlay takes on most of the desirable properties of structured overlays without incurring the costs of the rigid structure. We will also describe a range of

novel, efficient distributed information-rich services, which take advantage of some of the theoretical advances made in the course of the EVERGROW project. The specific service of content distribution and distributed storage received special focus in EVERGROW. This year we report advances such as network coding, which increase the performance of these processes. Finally, analytic procedures developed within EVERGROW allow exact solution of many dynamical properties of distributed overlays, while incorporating accurate information such as that available from our Internet topography.

We are making progress on Item 4 along several fronts. The message-passing or "belief propagation" approach to distributed computation, when successful, replaces an exact centrally computed optimization or control problem with compute cost  $N^3$  or even  $exp(N)$  with an approximate, but often very accurate distributed algorithms that require roughly linear in  $N$  time, or constant computing effort per participating node. Through a collaboration between SP3 and SP4, we have succeeded in demonstrating belief propagation algorithms that simplify and accelerate content distribution and other classic tasks needed in providing P2P services. A number of internal workshops have helped to accelerate and spread this approach within the EVERGROW population. We have also participated in two very visible external schools on message passing, its theoretical underpinnings and its applications.

Similarly, in SP4 we have developed distributed algorithms that solve the hard under-determined problem of assigning delays measured, e.g. by ETOMIC) from points at the outside of a network to the internal links of that network in linear  $N$  time, in contrast to the best current approaches which require of order  $N^3$ . Steady progress in the understanding of finite block-size stochastic codes, in SP4, has lead to efficient codes which can be optimized with fast algorithms (replacing methods which were suitable only for research on this subject). These form a building block for the error-correction codes that are needed for robust distributed archives and content distribution schemes. Because the manager of the WP4b effort has taken up a professorship in the US, we have decided to focus all the theory efforts in WP4a for the last year of the project. Collecting the scientific insights obtained on message-passing and complex systems optimization during the period of the EVERGROW project, we find that a new and more complete synthesis of the results to be expected for a wide class of complex optimization, logistics and planning problems is now available and can be used in both exact and heuristic algorithms of great practical value.

Finally, the research groups making up EVERGROW were invited to participate because they are among the world's best in the relevant fields. As part of their normal activities, they continue to publish many papers in quality journals and have been invited to speak to top-ranking conferences.

## 2 Workpackage Progress

### 2.1 SP1: Measurement and modelling

#### 2.1.1 Topology measurement

**Objectives** The objective of this WP was to create a large-scale distributed measurements infrastructure to measure and track the evolution of the Internet from hundreds of different view-points. The data collection aim was to be processed and archived in order to provide a better understanding of the Internet topology and evolution. This was facilitated through the DIMES project at TAU.

**Progress** In the last year we made great strides both in analyzing DIMES data and in the engineering of DIMES as a system. The later is important to maintain DIMES liveness in the future. DIMES will continue to be funded by the EU under FP7 by playing important role in two FP7 projects, MOMENT and OneLabII.

DIMES data analysis this year took several forms. Our most celebrated result was the analysis of the Internet AS graph using  $k$ -shell decomposition which was published in *Proceedings of the National Academy of Sciences USA (PNAS)* [2]. The analysis has several interesting findings among them we suggest a new way to define the Internet core, which includes close to 60 ASes. The new core is larger than the one, which is in use now and is generated by looking for the max clique, and is not as North American centric as the max clique. In a separated published paper [6, 7] we define a new measure to node importance in complex networks.

An important step in tracking the Internet evolution in time the ability to generate periodic PoP level map of the Internet since the AS level is too coarse while the router level is too fine for this purpose. A paper which describe such a method was submitted recently and the method was tested on the 100 top degree ASes. This year we will continue working in this direction to place the PoPs geographically, and to post periodic PoP maps on our topology web page.

During 2007, we invested considerable efforts to better engineering of the DIMES infrastructure. This include revision of large portions of our operating code, which made it stabler and easier to maintain. The revision include the agent, the planner (which needs more work), the default experiment generator, and the agent-center communication scheme. The agent was also enhanced to include a revised and improved packettrain module which is now is initial use in scientific experiments both by TAU and CB.

Another important engineering investment was the database. Overall we collected more than 3.5 Billion measurement and our old design pushed our MySQL server to its performance limit and filled up our storage. To tackle this we split our main raw traceroute tables to yearly chunks, but keeping a unified view of the table such that queries can be sent seamlessly to the union of the yearly tables. We also upgraded the hardware with an additional RAID (brute storage of 10.5T), and a more powerful blade server.

Finally, we devoted time this year to analyze DIMES as a measurement tool. We looked at system aspects such as measurement stability, rouge agents identification, and filtering techniques; and at the scientific aspects of DIMES, namely the contribution of distributing the measurement effort to the topology and its characteristics. A paper that dwells into these question is now in an advanced stage of work.

#### Deviations from programme and corrective actions

## Deliverables

Del. no	Deliverable name	WP no	Date due	Actual date	Lead
D.1a.4	Model of Internet evolution	1a	48	48	TAU

## Milestones

Mil. no	Milestone name	WP no	Date due	Actual date	Lead
M.1a-3.4	maps with additional measures; unidirectional measurements. Detailed router level map.	1a	42	48	TAU

### 2.1.2 Traffic measurement

**Objectives** To develop and finalize a complex systems based model of the Internet traffic. To use our measurement archives and databases to validate a statistically accurate model which can be used in simulations of peer to peer systems. To finish ETOMIC visualization tools. Open and public ETOMIC measurement archive. Development of permanent interoperability between ETOMIC and DIMES. ETOMIC tomography map. Periodic ETOMIC tomography in the European academic network. To contribute to the joint large scale DIMES - ETOMIC tomography and bandwidth measurements. Develop a short timescale bandwidth monitoring capability in the BART framework. Implementation of BART in the ETOMIC infrastructure. Investigation of implementability of BART in DIMES.

**Progress** The above goals were set for a two years period (2006 and 07). In the previous reporting period we completed the tasks set in the Milestones

- Implementation of BART tool in etomic,
- ETOMIC-DIMES permanent interoperability,
- Final etomic visualization tools,
- European periodic traffic tomography map.

The groups collaborating in this workpackage concentrated on the remaining tasks and finalized also some activities which started in 2006. The three main remaining goals for 2007 were to reach the Milestones and Deliverable

- Final form and documentation of BART,
- Advanced final ETOMIC measurement archives,
- Models of Internet delays.

The role of partners in these activities were:

**CB** concentrated mainly on the finalization of its part of the ETOMIC measurement archive: Worked out new data schemes for network tomography, network topology and available bandwidth measurements (BART). Extracted delay distributions of network tomography experiments. Then investigated the statistical properties of the data and provided it for the purposes of WP 1c. It continued to work on network tomography data collection, data archivation, data visualization and maintenance of the physical infrastructure of ETOMIC.

**UPN** concentrated mainly on the finalization of its part of the ETOMIC measurement archive: Periodically collected and archived one way delay data, have been working on

alias resolution in order to be able to identify all different IP addresses belonging to the same router to reduce complexity of graphs showing routers and not IP addresses. It also developed new capabilities to collect and provide information about the measurement node states (GPS signal availability, connection speed changes at ETOMIC nodes and general statistics of node utilization). This information is now archived too and helps the interpretation of measurement results. It investigated the statistical properties of one way delay at different probing rates. It continued to work on the maintenance of the ETOMIC Central Management System.

**EAB** mainly on the bandwidth estimation in real time (BART): Finalized and documented the tool, which is reported in the Deliverable 1b.4 in detail. Continued its collaboration with CB on the BART version integrated into ETOMIC in the previous reporting period.

In 2007 the architecture of the ETOMIC measurement database has undergone significant developments. First of all, the concept of modeling network data has been considerably refined to be able to go beyond the functionality of a simple network tomography database. We now describe measurements on a far more general level, so as to be able to store practically any kind of ETOMIC measurements in the database. Another notable improvement is that we extended the database schema in order to store evaluation results too, along with the raw measurement data. This feature enables us to efficiently compare, aggregate and cross-match different measurement results. Furthermore, we implemented important server side functions and stored procedures to simplify the usability of the data stored in the database.

An important step was taken forward to make the virtual network measurement observatory capabilities easier to handle and make use of. In parallel with the process of the database schema redeclaration, an application level interface was defined that well maps to the new database tables, structures and stored procedures. This new interface enables the implementation of connection based applications. After a proper authentication via the interface a database connection is built up to the ETOMIC repository and rights allocated to the application, so data can be easily retrieved and stored in the database via simple function calls.

The application interface is maintained in python scripting language and in C++ code. The python interface is tested on 32 bit, while the C++ interface is tested on both 32 bit and 64 bit architecture machines running debian linux operating system. Dependency packages are pyodbc, libodbc++. Connection to the MsSQL server is supplied via tdsodbc driver.

Along with the interface, example scripts and program codes are at hand of program developers. Based on the new interface the network tomography algorithm is reimplemented. An application is developed to keep track of possible tomography tree variations stored in the repository.

In the past decade it became obvious that Internet has several interesting fractal and scaling properties, such as the self-similar nature of traffic and the power law scaling of the network topology.

In order to handle large scale network tomography in practice we faced performance problems. The computation of delay distributions requires a summation over all possible combination of internal delays in the network rising a non-polynomial computational complexity. The simplest algorithm would not scale well with the size of the network. By modifying the classical message-passing algorithm the computations of network tomography can be accelerated substantially, introducing a new highly reliable estimator the complexity of distribution computation becomes polynomial. The linear scaling of the runtime of the new algorithm with the network size is shown numerically. The link

delay distribution inference is calculated independently for each sender host, hence the topology and inner link delay distributions can be evaluated in parallel, which also makes calculation faster.

The accelerated inference algorithm and the high temporal resolution let us infer internal delay distributions with higher resolution than earlier. The richness of detail makes it possible to analyze the shape of these distributions further and to find interesting new scaling properties in our delay data. We highlight that our distributions show signs of self-similar traffic on the links and establish a new scaling law for the distribution of the average link delays.

Evaluated network tomography data have been made public and have been shared with the WP1c team coordinated by SICS working on the ModelNet implementation. This data includes individual queuing delay distributions that belong to various network segments. These distributions are well approximated by the class of Weibull functions. The ModelNet uses these parameters to describe and fine tune the model parameters of background traffic built in the emulator.

UPN worked over 2007 on the ETOMIC Central Management System (CMS WP1B, <http://measurement.etomic.org>) that provides an user interface to the ETOMIC nodes that allow very high precision active monitoring. We have been working in the improvement of general functionalities. Now it is possible to manage monitoring nodes with some component out of order, for example, if GPS signal is not available or with bad quality, or if there is very low connection speed with the node. The researcher is informed about the limitations of each node and he can decide if the node is enough for his planned experiment. A new section has been implemented with general statistics of system utilization: number of logins over time, number of experiments executed, data size transferred, experiments per agent and user, etc. Some of these statistics are private because they provide specific information per user.

The periodic measurements section have been improved with new measurements based on paris-traceroute tool. Default traceroute is affected by load balancing and alternative routes to the same destination, getting some times not enough data to provide certain information of path routes. With paris-traceroute, all packets in the traceroute are handled as the same flow and, as demonstrated in some studies, they are not affected by load balancing and we can get all the hops in the path to some destination. Different variations of paris-traceroute (ICMP, UDP and TCP based) are offered now.

Related to these traceroute measurements we have been working on alias resolution in order to be able to identify all different IP addresses belonging to the same router. This is useful in topology creation and representation so we can reduce complexity of graphs showing routers and not IP addresses. Another research line using ETOMIC data is related with one-way delay and its relation with router forwarding capacities.

Finally, UPN have been providing continuous maintenance of the service and user support. User support and help desk is very important and time-consuming, overall for new non-expert users. Several problems are found with these system that needs 24x7 availability, specially related to nodes distributed around Europe: hardware failure of any kind (several nodes are running for 4 years), GPS performance problems, power outages, connectivity problems and so on.

The EAB group has been focusing on work in the field of available bandwidth estimation. The method BART, Bandwidth Available in Real-Time, which was first described in 2004, has been further developed and documented in final form for the purposes of the present project. The method uses active probing with packet pairs sent over an IP network path, and filter-based analysis of the probe-packet time stamps in order to obtain estimates of the time-varying available bandwidth over the path. BART allows for

tracking the path available bandwidth at arbitrary (configurable) time scale.

During 2007, effort has been devoted to studying different ways of doing change detection. We compared results obtained using the more elaborate method Generalized Likelihood Ratio (GLR) to those obtained using the CUSUM method. Using GLR it is also possible to not only detect whether there has been a sudden state change, but also the magnitude and time of the change. The trade-off is increased computational complexity, which also grows in time. We have published one conference paper and submitted one journal paper during 2007. Also, EAB has focused on internal work to exploit BART. We have filed three more patents during 2007. During the course of EVERGROW, EAB has filed five patent applications related to BART. The company is planning to use this bandwidth estimation technology in various products and systems. Further, EAB is in various stages of negotiation with external companies who wish to license this technology for use in their products and systems.

**Deviations from programme and corrective actions** None.

### Deliverables

D.1b.4	Models of Internet delays	WP 1b	M48	M48	CB
--------	---------------------------	-------	-----	-----	----

### Milestones

M.1b-4.1	Final form and documentation of BART	WP 1b	M48	M48	EAB
M.1b-4.2	Advanced final ETOMIC measurement archives	WP 1b	M48	M48	CB

### 2.1.3 Integration of Topology and Traffic information into P2P design

**Objectives** One of the goals of the Evergrow project in its final year is to bring together the results obtained in WP1, concerned with observing the Internet and measuring its properties, and WP3, concerned with developing scalable algorithms for Internet-based systems, by creating an evaluation test-bed for large-scale dynamic distributed systems. The test-bed allows participants to evaluate their systems developed as part of WP3 in emulated realistic settings available from WP1.

**Progress** We have decided to use ModelNet, the large-scale network emulator most widely used by the peer-to-peer research community, as a basic emulation mechanism for our evaluation infrastructure. To achieve an emulation infrastructure useful for research in peer-to-peer systems, we complement the network emulation with the ability to emulate the lifetime dynamism (churn) of the peers participating in the system under evaluation, and the ability to emulate the partitioning and merging of the emulated network.

Because ModelNet was not directly usable on Evergrow machines, KTH(P14) and SICS(CO) undertook an effort to make it available on the Evergrow IBM blades infrastructure. This effort meant porting ModelNet to the FreeBSD 6.2 (the latest stable version at that time) operating system kernel. This port was disseminated to the ModelNet authors and will be made available for all users of ModelNet. ModelNet has been enhanced to support the emulation of the partitioning and merging of the emulated virtual network, and support for network partitioning emulation has been implemented in the GODS tool which provides an infrastructure for managing the running peers of the system under evaluation, and provides emulation of churn.

CB(P5) worked on extracting subsets of relevant data for peer-to-peer system experiments, from the ETOMIC database. Distributions of individual queuing delay distributions that belong to various network segments and network tomography data have been

produced from measurements between ETOMIC boxes. CB(P5) and KTH(P14) have collaborated on designing an enhancement to ModelNet that allows for emulating background traffic on the links of the emulated network, according to background traffic models extracted by CB(P5). The background traffic emulation imposes very little overhead on the emulator. The implementation of this design is the subject of ongoing work.

KTH(P14) collaborated with TAU(P19) on defining ways of extracting from the DIMES dataset models of the Internet that are relevant for peer-to-peer systems evaluation. TAU (P19) extracted Internet topologies and delay maps from the DIMES AS-level data and has produced peer-to-peer delay maps from traceroute measurements between DIMES agents running on hosts at the edge of the Internet. These delay maps have been incorporated into ModelNet network models and were used in experiments on the evaluation infrastructure.

**Deviations from programme and corrective actions** None.

### **Deliverables**

Del. no	Deliverable name	WP no	Date due	Actual date	Lead contractor
D.1c2	Network emulation tool for distributed systems	WP1c	M48	M48	KTH, SICS

### **Milestones**

Mil. no	Milestone name	WP no	Date due	Actual date	Lead contractor
M.1c-4.1	Extraction of key statistical data from the measurement databases	WP1c	M45	M45	TAU, CB
M.1c-4.2	Building the emulation model from the extracted data	WP1c	M48	M48	TAU, CB

## 2.2 SP2: Virtual Network Observatory

### 2.2.1 Virtual Network Observatory

**Objectives** Develop the EVERGROW computing facility into a shared resource of CPUs, communications links and storage. The facility will support measurement services, data repositories and modeling tools. Researchers will have access to a real-life widely distributed computation and networking platform for use in experiments that take advantage of the information and models flowing from EVERGROW's measurements, as well as a platform for large scale theoretical calculations.

**Progress** The EverGrow infrastructure has been an integral part of the EverGrow project, providing computation and experimental resources to our project and its partners. During the 2007 calendar year, the system was stable and saw considerable usage. As in the previous year, the hardware infrastructure was split between local usage and EverLab usage.

Local usage refers to systems operated and used by the researchers in the institution where the hardware was physically deployed. For example, the Rome cluster was integrated into a larger cluster operated by the INFM and used for EverGrow research in complex systems. Most of the Collegium Budapest Egyesulet resources were used by the Etoomic project and most the Tel Aviv University blades were used by the DIMES project. The hardware and software for these systems seems to have been very stable in 2007. The resources were well utilized by EverGrow partners and sub-projects.

The EverLab system is a private PlanetLab deployed and managed by The Hebrew University on the remaining blades of the EverGrow clusters. EverLab provides a real-world environment for testing distributed algorithms and applications. As reported last year, the EverStats system monitors the EverLab clusters and reports on usage by node and by researcher. EverLab was used by many of the EverGrow partners during 2007. Ours reports indicate that UPSXI, UCL, HUJI, AST, SICS, EPFL, TUC and CLAES utilized the system for their research in 2007.

With the end of the EverGrow project, our focus is now on retaining as much of the EverGrow infrastructure as possible for use in the OneLab2 project. Some researchers have already indicated that they will not be participating in OneLab2 and hence we have removed support and monitoring for those clusters.

Our efforts supporting EverLab were documented in a USENIX paper *Everlab - A production platform for research in network experimentation and computation*, *USENIX LISA 2007* included in the 2a.3 deliverable for 2007. Among the lessons learned was the need to educate the user base to the capabilities of Everlab and on how to use it. These issues will be critical to the success of similar testbeds and to OneLab2 in particular.

**Deviations from programme and corrective actions** None

#### Deliverables

Del. no	Deliverable name	WP no	Date due	Actual date	Lead contractor
D.2a.3	Virtual Network Observatory —experience and conclusions	WP2a	48	48	HUJI

#### Milestones

Mil. no	Milestone name	WP no	Date due	Actual date	Lead contractor
---------	----------------	-------	----------	-------------	-----------------

### 2.2.2 Data repository

**Objectives** Define and implement data formats and standards of access to permit wide use of the SP1 measurements across EVERGROW and by scientific users around the world. Establish the validity of this data for use in modelling realistic distributed applications and in extrapolating to the future growth of the Internet.

**Progress** As the EVERGROW project concludes, Both DIMES (WP1a) and ETOMIC (WP1b) are in full production mode, making regular measurements and accumulating them into public databases for subsequent analysis. Both also have interfaces by which researchers can define and conduct their own measurements. Analyses of historical data are also supported in a parallel fashion.

The experiment design and management interfaces have been described in past deliverables D.1a.3. The filtering and processing pipeline that DIMES employs, and the experience that has led to this set of tools, is described in D.2c.3. Further analysis of the danger of systematic bias in this data is given in D.2c.4.

The analytical tools for identifying mid-level structure (the "point of presence" or POP level) have been extensively tested during 2007 and are ready for release along with data aggregated in this fashion in early 2008. They will be extended in the coming year, as continued support for DIMES through the ONELAB2 project becomes available (or, more probably, just before this migration occurs.) This is described in Deliverable D.1a.4.

Permanent archiving of this data, as it accumulates, is a serious challenge, but one which we believe we will meet for the coming two years. A final investment of EVERGROW funds has added 7 TB of archival storage to our server cluster.

**Deviations from programme and corrective actions** none

#### Deliverables

Del. no	Deliverable name	WP	Due	Actual	Lead
D.2c.2	Review of DIMES network data	2c	M48	M48	HUJI
D.2c.3	Analysis of sources of bias in network data	2c	M48	M48	TAU

#### Milestones

Mil. no	Milestone name	WP	Due	Actual	Lead
M.2c.2.4	Extraction of network models based on DIMES and ETOMIC data	2c	M48	M48	HUJI

### 2.2.3 Data mining

**Objectives** Develop a tool to explore the spatial and temporal patterns in the measured network topology data.

#### Progress

- A finalized our AS data exploration tool (M-PASM). The tool allows for the analysis of AS data from a number of sources. Currently we have support for CAIDA and DIMES data and are working on further plugins
- We have fixed a number of bugs and the tool now is in a very stable state.

- We have launched a web site downloading the tool and learning about its features. The website is at <http://evergrow.claes.sci.eg> and is mirrored at <http://www.natural-computation.com/mpasm>

**Deviations from programme and corrective actions** none

**Deliverables**

Del. no	Deliverable name	WP no	Date due	Actual date	Lead contractor
D.2d.4	Fully featured tool	WP2D	M48	M48	CLAES

**Milestones**

Mil. no	Milestone name	WP no	Date due	Actual date	Lead contractor
M.2d-2.3	Fully Featured tool	WP2D	M48	M48	CLAES

## 2.3 SP3: Self-healing Systems

### 2.3.1 Algorithms for overlay networks

**Objectives** The main objective of this WP is to work on algorithms for maintaining overlay network structures, improving routing and fault tolerance. This work addresses both the development of novel, efficient algorithms for specific tasks as well as foundational questions on feasibility and applicability of concepts from the complex systems domain. More particularly during M37-m48, the following objectives have been set:

- Investigate and develop overlay network design approaches with low maintenance cost that deal efficiently with the non-transitivity issue.
- Investigate efficient and practical ways to approximate Internet latencies using trees and embedding techniques.
- Develop schemes for handling network partitioning and merging in overlay networks.
- Develop a self-managing structured overlay network that provides lookup consistency in high churn environments.

**Progress** During the forth (last) year of the Evergrow project, the partners involved in Work Package WP3k have worked on several aspects of maintaining overlay networks, providing efficient routing within them and performing task-specific algorithms. More particular:

- EPFL have designed an overlay called Fuzzynet, which does not rely on the ring invariant, yet have all the functionalities of structured overlays. Fuzzynet takes the idea of lazy overlay maintenance further by dropping any explicit connectivity and data maintenance requirement, relying merely on the actions performed when new peers join the network. The suggested relaxed structure of Fuzzynet has the following differences compared to tightly structured DHTs: i) No explicit ring maintenance. ii) Peers are not deterministically responsible for a particular key section but probabilistically. iii) Data keys are disseminated and replicated in the vicinity of the targeted key. Fuzzynet peers develop their neighbors according to a policy for optimal routing at the joining phase and this effort helps older peers update their stale connections. As it is shown later, this suffices assuming stable churn rate. In contrast to the related literature which tries to improve the ring maintenance mechanisms, we take a complementary approach, where we want to ensure functional correctness (of querying and new data insertions) even in case the ring invariant is violated. Whenever the ring invariant is met, our approach has no message overheads compared to the traditional approaches given similar data replication factor (which is anyway needed for fault tolerance). Therefore, the mechanism can be integrated to work seamlessly in a ring-based P2P network, while avoiding the non-transitivity problems and obviating the need for any aggressive and expensive ring self-stabilization. We show that with sufficient amount of neighbors ( $O(\log N)$ , comparable to traditional structured overlays), even under high churn, data can be retrieved in Fuzzynet w.h.p. We validate our novel design principles by simulations as well as PlanetLab experiments and compare it with ring based overlays.
- EPFL has developed a new routing algorithm for DHTs to deal with skewed peer distributions, while utilizing order preserving hash functions. The concept of "hop count" have been introduced to approximate the distance in DHTs and derive a

new efficient routing algorithm, which works under skewed peer distributions. An important advantage of this algorithm is that it is very flexible with respect to the number of routing entries each peer maintains. Experimental and simulation results show the promising properties of the proposed system.

- HUJI have developed an algorithmic technique with formal guarantees for finding faithful and low dimensional representations of data lying in high dimensional space. Our main theorem states that every finite metric space  $X$  embeds into Euclidean space with dimension  $O(\dim(X)/\epsilon)$  and distortion  $O(\log^{1+\epsilon} n)$ , where  $\dim(X)$  is the *doubling* dimension of the space  $X$ . Moreover, we show that  $X$  can be embedded into dimension  $\tilde{O}(\dim(X))$  with *constant* average distortion and  $\ell_q$ -distortion for any  $q < \infty$ . Our technique also provides a dimension-distortion tradeoff and an extension of Assouad's theorem, providing distance oracles that improve known construction when  $\dim(X) = o(\log |X|)$ .
- HUJI studied an intuitive and practical approach for predicting network latency, namely *embedding into a tree metric*. A metric  $V$  is a *tree metric* if there exists a tree with non-negative weights such that  $V \subseteq T$  and  $d_V(u, v) = d_T(u, v)$  for all  $u, v \in V$ . Note that the embedded tree might contain additional Steiner nodes not in  $V$ . A tree embedding is more intuitive because even though the Internet is not exactly a tree, it has an inherent hierarchy in the relationships between end hosts and Internet Service Providers (ISPs) at different tiers (Tier 1, Tier 2, and Tier 3). It is more practical because trees have been proposed as a basic primitive in distributed systems for multi-cast communication, locality-aware clustering, and data aggregation. Moreover, trees provide the same ability as coordinate-based systems for instantaneous distance estimation through short and efficient distance labels.
- HUJI provided the first sparse covers and probabilistic partitions for graphs excluding a fixed minor that have *strong* diameter bounds; i.e. each set of the cover/partition has a small diameter as an induced sub-graph. Using these results we provide improved distributed name-independent routing schemes. Specifically, given a graph excluding a minor on  $r$  vertices and a parameter  $\rho > 0$  we obtain the following results: (1) a polynomial algorithm that constructs a set of clusters such that each cluster has a strong-diameter of  $O(r^2\rho)$  and each vertex belongs to  $2^{O(r)}r!$  clusters; (2) a name-independent routing scheme with a stretch of  $O(r^2)$  and tables of size  $2^{O(r)}r!\log^4 n$  bits; (3) a randomized algorithm that partitions the graph such that each cluster has strong-diameter  $O(r6^r\rho)$  and the probability an edge  $(u, v)$  is cut is  $O(r d(u, v)/\rho)$ .
- EPFL looked at the problem of merging two separate overlay networks. In particular, EPFL elaborated how two networks using the same protocols can be merged, looking specifically into two different overlay design principles: (i) maintaining the ring invariant and (ii) structural replications. Depending of the peculiarities of a specific overlay - ring based, or structural replication, the mechanisms which are necessary to execute the merger process and indicate the minimal effort it will require in order to merge two networks have been identified. Two case studies for two different structured overlays - ring based overlay (like Chord) and P-Grid (which utilizes structural replication) have been performed, in order to identify the properties of an overlay network which would facilitate or hinder successful merger of distinct overlay networks.

- KTH/SICS developed an algorithm for merging multiple similar ring-based overlays when the underlying network merges. The algorithm has been designed such that it is resilient to churn while the merger takes place and is flexible as the trade-off between message complexity and time complexity can be adjusted by a parameter.
- UCL have developed the Relaxed-Ring architecture, which provides self-organization and self-healing properties. The relaxed-ring is the new topology of the P2PS system. The newly designed algorithms are using the feedback loop model, which is taken from Control Theory. The ring maintenance protocols is described as a set of feedback loops. One of them concerned with the join of peers, and the other one in charge of the failure recovery.
- KTH/SICS studied key-based consistency and availability in structured overlay networks (SONs). They focused on the study of frequency of occurrence of lookup inconsistencies due to imperfect failure detectors and churn. The derived solution shows how the effect of lookup inconsistencies can be reduced by using node responsibilities, which depicts a trade-off between consistency and availability of keys. Further, since many distributed applications require quorum techniques at their core, the work also focused on analyzing the probability that majority-based quorum techniques will function correctly in a SON with inconsistent lookups. It was derived that the probability of majority-based algorithms to function correctly despite lookup inconsistencies is very high.

**Deviations from programme and corrective actions** No deviations, so no corrective actions were necessary.

#### **Deliverables**

Del. no	Deliverable name	WP no	Date due	Actual date	Lead contractor
D.3k.3	Final Report on Overlay	WP3k	12/2007	12/2007	EPFL
D.3k.3	Network Algorithms				

#### **Milestones**

Mil. no	Milestone name	WP no	Date due	Actual date	Lead contractor
M.3k-3.1	Maintenance of small-world structured overlays under multi-dimensional constraints.	WP3k	12/2007	12/2007	EPFL
M.3k-3.2	Scalable routing schemes and latency approximation	WP3k	12/2007	12/2007	HUJI
M.3k-3.3	Schemes for handling network partitions and mergers in overlay networks	WP3k	04/2007	04/2007	KTH/SICS
M.3k-3.4	Algorithms for self management of large-scale systems deployed over the Evergrow cluster infrastructure	WP3k	12/2007	12/2007	UCL

### **2.3.2 Services and overlay systems**

**Objectives** The objective of this WP is to provide higher-level services for large scale networks that in turn allow the development of applications for highly heterogeneous,

dynamic and decentralized environments.

**Progress** In the final year of Evergrow, Work package WP3I completed its work on the design, prototype implementation and evaluation of the following kinds of services for large-scale networks:

- Large-scale, distributed, information retrieval and filtering.
- Large-scale, distributed inference by belief propagation.
- Peer-to-peer network monitoring and visualization.
- Asynchronous failure handling for concurrent components.

The research we carried out this year is reported in Deliverable D3I.3 and can be summarized as follows:

- In collaboration with the DELIS partner Max-Planck Institute for Informatics, we extended our work on the information filtering service MAPS we first presented in Deliverable D3I.2.

Most approaches to information filtering taken so far have the underlying hypothesis of potentially delivering notifications from every information producer to subscribers. This exact information filtering model creates an efficiency and scalability bottleneck, and might not even be desirable in certain applications. In Deliverable D3I.2, we put forward MAPS, a novel approach to support approximate information filtering in a peer-to-peer environment. In MAPS, a user subscribes to and monitors only carefully selected data sources, and receives notifications about interesting events from these sources only. This way scalability is enhanced by trading recall for lower message traffic.

Last year, in Deliverable D3I.2, we presented MAPS and compared it in a high-level way with DHTrie, an exact information filtering service developed previously by us in WP3I. This year, in Deliverable D3I.3, we carried out a detailed experimental evaluation of MAPS which shows the efficiency and scalability of our approach in many diverse scenarios.

In summary, WP3I has implemented, evaluated and compared three information retrieval and filtering systems during the four years of Evergrow: MAPS, DHTrie and LibraRing.

- We studied efficient distributed algorithms for belief propagation on Bayesian networks stored on a DHT. We go beyond the work we reported in Deliverable D3I.2 last year, which involved simulations on Matlab, and implement our algorithms on the DHT P-Grid. Our work shows that it is possible to solve a Bayesian network in a distributed fashion in a reasonable time. We implemented a novel algorithm to decrease the number of messages that must be sent over the network to solve the Bayesian network. Our algorithm balances the load between the hosts in the system, has a satisfactory convergence time and works for various topologies of Bayesian networks.
- We implemented PEPINO, a graphical PEer-to-Peer network INspectOr for peer-to-peer network monitoring and visualization. PEPINO is running on top of P2PS, a structured overlay network providing a DHT based on the Chord ring. The goal of PEPINO is to monitor the network by detecting failures, and by observing the

messages sent between peers. A dynamic and self-organizing view of the network is presented to the user, who can interact with it to inject failures or send messages in order to study the network protocols. Since many systems implement a DHT in different ways - in particular by choosing the finger table with a different strategy - PEPINO also helps to study three different strategies, in particular finger tables following the strategy of DKS, Tango, and Chord.

- We implemented PEPINO, a graphical PEer-to-Peer network INspectOr for peer-to-peer network monitoring and visualization. PEPINO is running on top of P2PS, a structured overlay network providing a DHT based on the Chord ring. The goal of PEPINO is to monitor the network by detecting failures, and by observing the messages sent between peers. A dynamic and self-organizing view of the network is presented to the user, who can interact with it to inject failures or send messages in order to study the network protocols. Since many systems implement DHT in different ways - in particular by choosing the finger table with a different strategy - PEPINO also helps to study three different strategies, in particular finger tables following the strategy of DKS, Tango, and Chord.
- In collaboration with the SELFMAN project, we made progress on designing and building a service architecture over structured overlay networks. We designed and implemented an asynchronous failure handling model for a network-transparent distributed platform. This is reported in Raphael Collet's Ph.D. thesis, *The Limits of Network Transparency in a Distributed Programming Language*, Dec. 2007. This is the foundation of a completely asynchronous programming model based on concurrent components. The programming model will be based on a formal model for concurrent components and services, called Oz/K, which is a process calculus based on INRIA's Kell calculus and Fractal model. This work will continue in the SELFMAN project as a continuation of EVERGROW.

**Deviations from programme and corrective actions** UCL did not complete the service architecture because our work has mainly concentrated on WP3k, where we designed and implemented the relaxed ring, which maintains connectivity and the finger tables of a structured overlay network in a completely asynchronous manner (no locking whatsoever). This is due to budget limitations in 2007, which did not let us allocate sufficient man-power to WP3l. Nevertheless, we did collaborate with the SELFMAN project to make progress on the service architecture of WP3l.

### Deliverables

Del. no	Deliverable name	WP no	Date due	Actual date	Lead
D3l.3	Final Report on Services on Overlay Networks	WP3l	December 31, 2007	December 31, 2007	TUC

### Milestones

Mil. no	Milestone name	WP no	Date due	Actual date	Lead
M3l-3.1	Belief propagation as a generic service in decentralized overlay networks	3l	M48	M48	EPFL
M3l-3.2	Prototype implementation of information retrieval and filtering services on top of overlay networks	3l	M48	M48	TUC

### 2.3.3 Storage and content distribution

**Objectives** The main objectives of this work were to continue the work related to storage and content distribution. We have pursued three main goals:

1. The Julia content distribution network – adding network coding support.
2. Data aggregation in sensor networks using clustering.
3. Distributed rating of users in file sharing networks for reducing spam and improving efficiency of the download.

Detailed information is found on the next section.

**Progress** This year, we have continued the work on the Julia content distribution network [1], reported in previous years. We have added a support for network coding [4], [3], [5]: an efficient mechanism for allowing higher flexibility in selection of file chunks to be transferred. Currently, network coding is not implemented in most of the file sharing networks – probably because it complicates the protocols. We propose three simple heuristics that are simple to implement and show using simulations the significantly improve the network utilization and shorten download times.

Another path pursued this year is in adapting some of the algorithms to new types of networks. Sensor network is an emerging research topic, where content dissemination in sensor network has many challenges besides of the fact that the nodes are distributed: sensors have limited energy, and weak computation power. The communication is done using wireless communications, unlike Peer-to-Peer networks which has usually fixed wired connections. We have decided to tackle these problems by creating an algorithm for selecting cluster heads. This allows us to aggregate information from the network using hierarchy, thus significantly saving in energy and reducing communication. Proposed algorithm was implemented using a TinyOS simulator and compared to state-of-the-art algorithm. Following EVERGROW spirit, we have decided to implement message-passing algorithms from the complex systems domain to be used for the task of clustering.

Interesting work was done in the context of Social Networks. One of the fundamental problems in file sharing network is that usually there is no authenticity of files, many times resulting in the download of bogus information and wasting of time and bandwidth. We have decided to look at this problem at the aspect of social networks are gaining increased popularity. In this work, which is complementary to content distribution, we propose to rank trust in neighboring nodes. By assigning different trust levels to different neighbors in our social networks, we are able to filter out information. This mechanism can be used in file sharing network to evaluate the quality of information received from neighboring nodes. That way future file sharing network will be able to filter out connections which provides information of low quality. We show using simulations on real topologies, that our method is resilient to malicious users that are trying to affect the rating computed. This work again utilizes message-passing algorithms from the complex system domain.

### Deviations from programme and corrective actions

#### Deliverables

Del. no	Deliverable name	WP no	Date due	Actual date	Lead
d3.m3	Final Report on Content Distribution and Storage	3m	Month 48	Month 48	P11

### Milestones

Mil. no	Milestone name	WP no	Date due	Actual date	Lead
M.3m-2.4	Simulation of extended support in network Coding	3m	Month 48	Month 48	P11
M.3m-2.5	Simulation of data aggregation in sensor networks using TinyOS	3m	Month 48	Month 48	P11

### 2.3.4 Analysis of overlay systems

**Objectives** The main objective of WP3n during the entire period of Evergrow was to develop analytical tools to predict the performance of overlays, over and above those that exist. We have done that in developing the master equation framework for understanding the functioning of structured overlay networks. We have shown that this framework can be used to understand very precisely, the effects of any design choice made in the construction of an overlay, as well as the effects of dynamic membership and overlay maintenance algorithms.

In 2007, our main objective was to extend the master equation formalism to understand the effect that delays caused by the underlying network may have on the functioning of an overlay.

**Progress** We have successfully concluded our analysis of delays in the context of the master equation formalism. We find from the analysis, that there is a region in the parameter space (parameters are, the number of nodes, rates of joins, failures or maintenance operations, and the average delay per link in the network), where the network though still connected does not function as a small world graph anymore, due to the 'long' links of a node always being dead. This is a phase transition occurring in the context of peer-to-peer networks. Beyond this region of instability, the network functions properly. We also analyse various 'proximity-aware' schemes usually adopted by peer-to-peer networks. These are proximity neighbor selection (PNS) and proximity route selection (PRS). In the former case, each node only keeps the 'fast' links as its contacts. In the latter case, each node keeps any contact it finds, but only routes through the fast links. We compare the performance of the network for these different schemes for different parameter values. To our knowledge this is the first time such a detailed analysis has been carried out on proximity-related issues.

**Deviations from programme and corrective actions** There have been no deviations from the programme.

### Deliverables

Del. no	Deliverable name	WP	Due	Actual	Lead
D.3n-3.1	Final report on Analytical Results for Overlays	3n	M48	M48	SICS

### Milestones

Mil. no	Milestone name	WP	Due	Actual	Lead
M.3n-3.1	Understanding the effect of delays on the functioning of Overlays	3n	M48	M48	SICS

## 2.4 SP4: Message Passing

### 2.4.1 Belief and survey

**Objectives** Our principal aim is to use message passing algorithms (MPA) as a general paradigm for solving hard optimization and inference problems. More generally we would like to use MPA to design and control complex systems.

On the one hand we need to complete the studies of prototypical constraint satisfaction problems, defined on random structures, and solving them through MPA.

On the other hand we want to extend the use of MPA to problems defined on non-random structures, that is to networks arising in realistic applications.

**Progress** We have completed the study of the space of solutions in constraint satisfaction problems (CSP) defined on random structures, identifying all relevant thresholds and phases. This is clearly a major achievement of the present project!

With this picture in mind (a picture much richer than what was believed before), we have started the study of possible connections between the properties of a cluster of solutions and the algorithmic complexity for finding those solutions. We already presented some conjectures, mainly based on the presence of frozen variables, which agree with present numerical investigations, and deserve more attention in future studies.

We have been able to analyze a simple search algorithm for random k-SAT, which fixes variables in a random order on the basis of the information obtained from a message passing procedure (namely belief propagation). This is the first non-trivial ‘decimation’ algorithm solved analytically and the comparison with numerical experiments supports the validity of our analysis.

Regarding the use of message passing algorithms for solving inference and optimization problems defined on non-random and/or dense networks, we have achieved several interesting results. Among these it is worth mentioning the MPA for solving the Sudoku game, which, despite its simplicity, has a dense and non-random inference network.

A more scientific application of MPA to densely connected networks comes from the problem of inference in multiple access systems (Code Division Multiple Access, CDMA) and that of learning in the Ising perceptron. In these cases we have considered more complex correlations among solutions, and modified MPA accordingly.

Last, but not least, we have studied the problem of identifying defective items by ‘group testing’, which is a common problem in many disciplines. We have resolved the long-lasting issue of optimal pool design when the probability of defect is small, and we have developed MPA to identify the defective items from a bunch of tests.

### Deviations from programme and corrective actions

#### Deliverables

No.	Deliverable name	WP no	Date due	Actual date	Lead
4a.3	Final report on message passing	4a	31/12/07	31/12/07	INFM

#### Milestones

Mil. no	Milestone name	WP no	Date due	Actual date	Lead contractor
---------	----------------	-------	----------	-------------	-----------------

## 3 Consortium Management

The Consortium in EVERGROW is fairly large for the funds available in the complex systems Proactive Initiative. An effort was therefore made in the proposal to find a management structure which would be efficient, accountable, and which would provide clear scientific and administrative leadership. In short, the structure chosen was a project run by the Executive and the Board, where administrative issues and day-to-day operations were mainly to be handled by the Executive, assisted by the management and staff of the Coordinator. In addition, Subproject managers, also members of the Board, were given large leeway to structure work in their Subprojects as they found most appropriate. Partners have a representation on a project General Assembly. The detailed structure is described in Annex I to the Contract, and, in more elaborate detail, in the Consortium Agreement, both documents available at <http://bscw.sics.se/bscw/> (internal project web site).

### 3.1 Consortium management task and their achievement

#### 3.1.1 Interacting with the Commission on behalf of the project

A number of EVERGROW members (especially from WP1a., WP1b and WP2a, and WP3) have participated in the FIRE initiative series of workshops and are now members of the ONELAB2 program, which is funded under call 2 of FP7. In the course of these discussions, Prof. Kirkpatrick has had several meetings with Dr. Sestini and other members of the directorate sponsoring FIRE. This directorate also sponsors the SAC projects, in which EVERGROW is now included, so this consultation has been essentially continuous, extending from the existing project to the foreseen interactions required of the new one.

The administrative coordinator has been in frequent contact with representatives of the commission.

#### 3.1.2 Preparing and submitting annual project budgets to the Commission

The budget accepted in 2006 covered the final two years of the project. The 2007 budget is formed from that budget together with the redistribution of the unclaimed budget of Sheer, who has previously left the project. No other changes were made in that budget for the final year of the project. Progress in the project and objectives for the next planning period were discussed in depth at the General Assembly meeting, in Rome, Italy, December 2006 and updated at our all-project meeting in Torino, December 2007.

Budgeting for the final two years of the project was finalised after the review in March 2006. Taking into consideration the result of the review and the unspent funds as well as unallocated funds remaining available to us. Budgets were prepared by the Executive and negotiated with the SP managers, who have confirmed their allocations with the members of each SP. The consortium elected to have it cover M31–M48 rather than M31–M42, thus running to the end of the project. Thus the budget we submit is accepted by voting of the General Assembly (in April 2006), as provided for in the Consortium Agreement.

### **3.1.3 Deciding reconfiguration of the consortium**

The consortium also in November 2006 requested that the contract be amended by inclusion of special clause 39, which eliminates the need for audit certificates for partners claiming less than 150 000 Euro (except at the end of the project). The amendment request was accepted by the commission.

In consequence, several partners omitted audit certificates for 2006 and will this year submit audit certificates covering 2006 and 2007.

### **3.1.4 Handling financial and administrative management of the consortium**

Partners are required to report financial information bi-annually. Payment of the second tranche of Commission funding is by the Consortium Agreement contingent on the Coordinator's approval of these reports.

### **3.1.5 The Administrative Team at SICS**

Dr. Karl-Filip Faxén serves as EVERGROW Head of Administration since January 2006. He is further supported by the administrative staff at SICS, including Ms. Charlotta Jörsäter, Ms. Eva Gudmundsson and Mr. Bengt Wahlström (financial issues) as well as Mr. Janusz Launberg (Senior Advisor, especially on legal and contractual issues).

### **3.1.6 The Board**

Members of the EVERGROW Board elected by the General Assembly on January 31, 2004 were Prof. Scott Kirkpatrick (Executive), Prof Erik Aurell (Executive), Prof. Karl Aberer (EPFL), Dr. Svante Ekelin (Ericsson Research), Prof. Yuval Shavitt (Tel Aviv University), Prof. Enzo Marinari (Rome, Subproject 4 manager), Prof. Giorgio Parisi (Rome), Prof. Gabor Vattay (Budapest, Subproject 1 manager), Prof. Peter Van Roy (Louvain, Subproject 3 manager) and Prof. David Saad (Aston). Of these, Prof. Van Roy resigned from the Board after resigning as Subproject 3 manager on June 30, 2004. In his stead, Prof. Seif Haridi, SICS, was appointed SP3 manager on July 1, 2004. Prof. Aurell resigned as coordinator of EVERGROW in May, 2005, because of personal and scientific time constraints, and has been replaced by Prof. Haridi. Prof. Danny Dolev (HUJI) replaces Prof. Haridi as the manager of SP3.

The EVERGROW General Assembly elected Prof. Haridi, Dr. Krishnamurthy, Dr. Gambäck and Prof. Danny Dolev to the EVERGROW Board. Prof. Dolev gave valuable advice to the Executive over the year, and participated by invitation at the Second Board meeting (Jerusalem, November 5-7, 2004). This resolution was carried by the General Assembly on January 31, 2005, and was effective from that date.

### **3.1.7 List of decisions by the EVERGROW General Assembly**

None during 2007.

### **3.1.8 List of decisions by the EVERGROW Board**

The executive and the board have consulted informally on numerous occasions using email and telephone but no formal decisions have been taken in 2007.

### 3.1.9 List of decisions by the EVERGROW Executive

**November 2006** Decision taken in Stockholm/Jerusalem.

Resolved, To request that special clause 39 be added to Article 9 of the Contract, obviating the need for audit certificates in some cases.

Dissemination status: Consortium

### 3.2 Project timetable and status

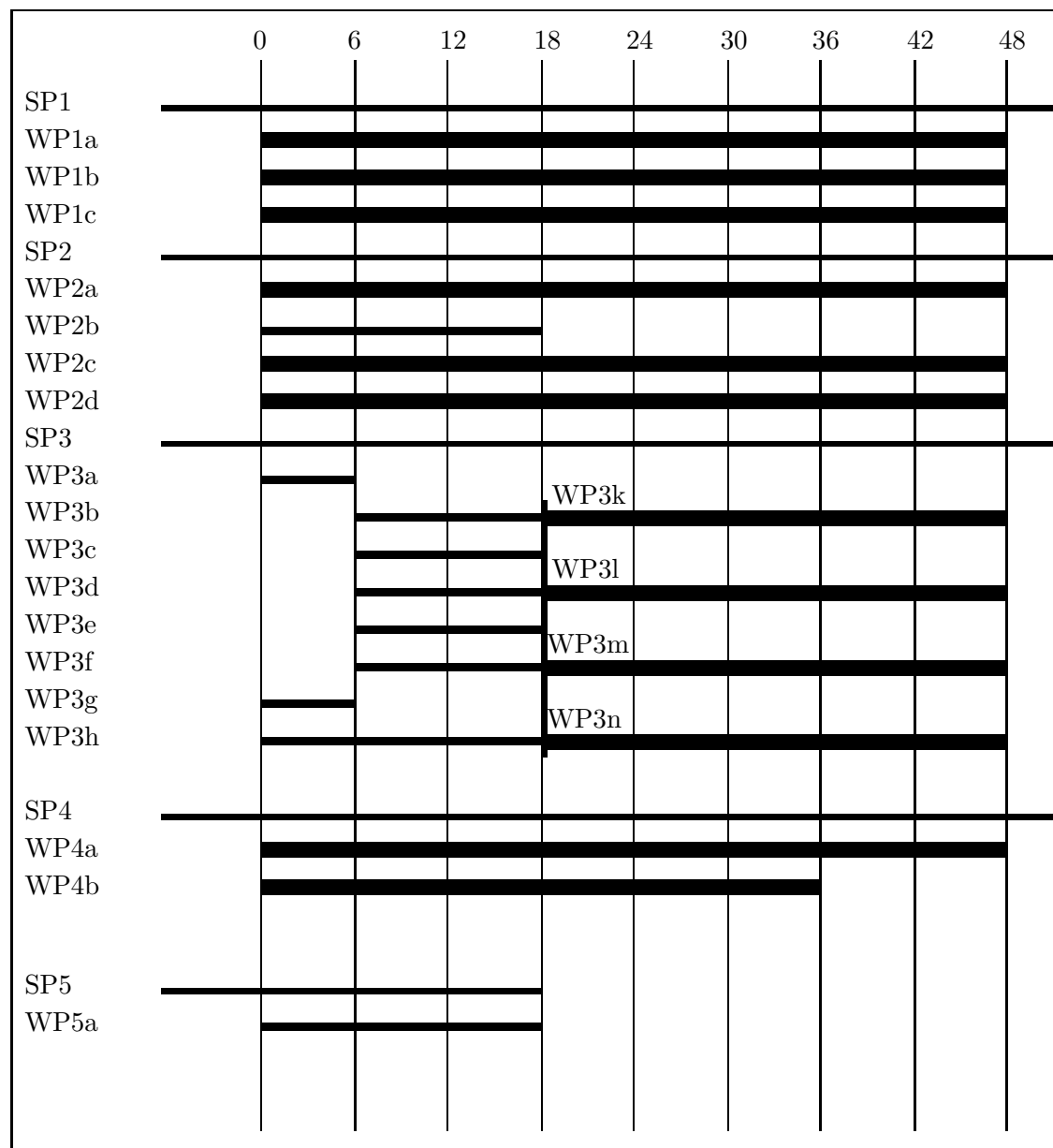


Figure 1: **EVERGROW IP level GANTT Chart:** Thin Workpackage lines represent Workpackages that have ended. Uninterrupted (thick) Workpackage lines indicate activities foreseen to the end of the project (month 48). SP3 has regrouped in four main activities starting month 18. Workpackage WP5a was exploratory. Its activities either terminated or continue as WP3n. WP4b terminates at M36.

The project barchart is given in Fig. 1. The main change implemented when moving from year 1 to year 2 is that SP3 has been reorganized along the same lines as the other Subprojects, i.e. in a smaller number of Workpackages. Also, WP4b terminates after the third year.

### **3.2.1 Full IP meetings**

- The Fourth EVERGROW full project workshop was held in Torino, IT, Dec 13, 2007. This was a one-day meeting with roughly 20 participants, reviewing all parts of the project but focusing on the cross-project interactions that dominated the final year. These are centered in WP1c. The group also planned presentations for the final review.

### **3.2.2 Subproject 1 meetings**

All the partners in SP1 and SP3 participated in a series of cross project meetings, coordinated through a special GODS mailing list set up for this purpose, in order to execute our joint deliverable. The group met in Stockholm in spring of 2007, in Athens in September 2007 and in Ireland (as an informal satellite to a P2P software research conference) in November 2007.

### **3.2.3 Subproject 2 meetings**

WP2a (EVERLAB) met with users and talked to potential collaborators from other projects at ROADS2007 and the PlanetLab Europe conference which was a satellite to ROADS (July 2007, Warsaw). The work was also presented at USENIX's LISA conference in Houston, TX in Nov 2007.

The CLAES group was unable to obtain permission to visit the DIMES group in Tel Aviv, but has participated in EVERGROW meetings in Stockholm and elsewhere in Europe. Their day to day contact is managed over Skype and proceeds through collaboration with Prof. Kirkpatrick and WP2c.

### **3.2.4 Subproject 3 meetings**

All the partners in SP1 and SP3 participated in a series of cross project meetings, coordinated through a special GODS mailing list set up for this purpose, in order to execute our joint deliverable. The group met in Stockholm in spring of 2007, in Athens in September 2007 and in Ireland (as an informal satellite to a P2P software research conference) in November 2007.

### **3.2.5 Subproject 4 meetings**

SP4 has strong pre-existing links between the collaborating institutions, as most of its members participate in the SPHINX network of excellence, and some have been members of complexity initiatives like EXYSTENCE and ONCE-CS.

## **3.3 Co-operation with other projects**

There have been collaborations between our TUC team and the DELIS (FP6) group at the Max-Planck Institut fuer Informatik in Saarbruecken, as both are working on advanced distributed services. The DIMES team has interacted with the UPMC group responsible for the novel "Paris traceroute" algorithms which decrease the possibility of seeing false

links. These ideas have been incorporated in the version of the DIMES agent that was released in mid 2007. The ARCADIA series of meetings provided an occasion to have the technical discussions that led to this transfer. Both DIMES and ETOMIC have joined up with FP7 partners to participate in ongoing efforts. They have participated in the drafting of the MOMENT and ONELAB2 proposals. EVERGROW partners from SP1, 2, and 3 participated in extensive discussions during 2007 through the ONCE-CS and FIRE coordinating umbrellas, leading to white papers that have informed FP7 calls.

## 4 Other Issues

N/A

## References

- [1] D. Bickson and D. Malkhi. The julia content distribution network. In *In the 2nd Usenix of Real World Distributed Systems (WORLDS 05')*.
- [2] Shai Carmi, Shlomo Havlin, Scott Kirkpatrick, Yuval Shavitt, and Eran Shir. A model of Internet topology using  $k$ -shell decomposition. *Proceedings of the National Academy of Sciences USA (PNAS)*, 104(27):11150–11154, July 3 2007.
- [3] C. Gkantsidis, J. Miller, and P. Rodriguez. Anatomy of a p2p content distribution system with network coding. In *IPTPS'06 Santa Barbara, CA, February 2006*.
- [4] C. Gkantsidis and P. Rodriguez. Network coding for large scale content distribution. In *proc. of INFOCOM 2005*, 2005.
- [5] Petar Maymounkov and David Mazires. Rateless codes and big downloads. In *IPTPS 03*, February 2003.
- [6] Yuval Shavitt and Yaron Singer. Beyond centrality - classifying topological significance using backup efficiency and alternative paths. In *Networking 2007*, Atlanta, GA, USA, May 2007.
- [7] Yuval Shavitt and Yaron Singer. Beyond centrality - classifying topological significance using backup efficiency and alternative paths. *New Journal of Physics*, 9, August 2007.