The DORII Project e-Infrastructure: Deployment, Applications, and Measurements

Davide Adami[1], Alexey Chepstov[2], Franco Davoli[1], Bastian Koller[2], Matteo Lanati[3], Ioannis Liabotis[4], Stefano Vignola[1], Anastasios Zafeiropoulos[4], and Sandro Zappatore[1]

[1] CNIT, University of Pisa/University of Genoa Research Units, Italy
[2] High Performance Computing Center Stuttgart (HLRS), University of Stuttgart, Germany
[3] EUCENTRE Pavia, Italy
[4] GRNET, Athens, Greece

Abstract. Remote Instrumentation Services go far beyond offering networked access to remote instrument resources. They are establishing as a way of fully integrating instruments (including laboratory equipment, large-scale experimental facilities, and sensor networks) in a Service Oriented Architecture, where users can view and operate them in the same fashion with computing and storage resources. The deployment of test beds for a large basis of scientific instrumentation and e-Science applications is mandatory to develop new functionalities to be embedded in the existing middleware to enable such integration, to test them on the field, and to promote their usage in scientific communities. The DORII (Deployment of Remote Instrumentation Infrastructure) project is a major effort in this direction. The paper presents the performance monitoring infrastructure that has been built in DORII and the results concerning a selected application in seismic engineering.

Keywords: Remote Instrumentation Services, SOA, e-Science.

1 Introduction

Almost all scientific areas and a good deal of technological developments use specialized instrumentation – laboratory equipment, measurement devices, large- and small-scale experimental facilities, sensor networks for data acquisition – other than computational and storage resources. The complex of activities that allow automated data processing and analysis, by exploiting distributed computational service like those offered by Grid architectures and cloud computing, and that can be referred to as *e-Science* (or, with a more precise term, Service-Oriented Science [1]) would greatly benefit from the full integration of such experimental instrumentation with the computational infrastructure into one powerful pool of resources that can be searched, selected, composed and configured, accessed, and controlled by their users.

The extension of the e-Infrastructure with this complex of activities, which can be termed *Remote Instrumentation Services (RIS)* [2], is not straightforward, and it requires addressing a number of issues in middleware and network architectural design, middleware development, and instrumentation and measurement related aspects. Recently, a number of European research projects, among others, have been

dedicated to it, and a community of researchers in the field has been forming and actively investigating aspects in this field (see, e.g., [3], [4]).

At the same time, efforts have been dedicated to the deployment of the infrastructure and of test beds that allow user communities to become acquainted with the related technology, to perform experiments on-line, and to be involved in the development of new applications and in the extension of existing ones. This is one of the main objectives of the DORII (Deployment Of Remote Instrumentation Infrastructure) project [5], [6], funded by the European Commission in the 7th Framework Program.

The general architecture, the deployed applications and the test bed organization of DORII were described in [6]. In the present paper, we describe the currently deployed DORII e-Infrastructure, the performance monitoring tools' customization and deployment, and present the results of a selected application in earthquake engineering. The paper is organized as follows. Sections 2 and 3 describe the e-Infrastructure and the monitoring tools, respectively. Section 4 presents the deployment of the selected application, and Section 5 reports related experimental results. Section 6 contains the conclusions.

2   Deployment of the e-Infrastructure

One of the main requirements posed by applications of many strategic areas in science and technology (as the ones specified by ESFRI - European Strategy Forum on Research Infrastructure [7]) is to design a service-oriented IT architecture which should allow users manage, maintain and exploit diverse instrumentation and acquisition devices together with heterogeneous computation and storage facilities granted by the traditional Grid, as those set up by EGEE (Enabling Grids for E-sciencE) [8], DEISA (Distributed European Infrastructure for Supercomputing Applications) [9] and many other Grid projects. Unlike the traditional Grid, the e-Infrastructure should practically enable access to remote instrumentation in high-performance computing and storage environments, and allow users and their applications to get an easy and secure access to various remote instrumentation resources, supported by high-performance Grid computation and storage facilities. The e-Infrastructure does that by providing standardized services to access integrated instrumentation resources (including expensive experimental equipment, but also smaller network-connected sensors and mobile devices), in a unified way with the traditional Grid services (as provided, e.g., by gLite [10]).

The DORII e-Infrastructure is based both on the EGEE infrastructure and its middleware of choice gLite, and on specific middleware services built within the DORII project. The interaction of the users with the instruments is effected via the Instrument Element (IE), originally conceived in the GRIDCC [11] project, and then redesigned within DORII. The IE includes specific Instrument Managers (IMs). Information about the resources and services of the infrastructure is provided by the Berkeley Database Information Index (BDII). The Workload Management System (WMS) is the service responsible for the distribution and management of tasks across Grid resources, in such a way that applications are conveniently, efficiently and

effectively executed. The LCG Computing Element (LCG-CE) is responsible for submitting jobs to the underlying local cluster of Worker Nodes (WNs). Storage Elements (SEs) are responsible for data storage and management, while the LCG File catalogue (LFC) offers a hierarchical view of files to users, with a UNIX-like client interface. From the security perspective, the Virtual Organization Management Service (VOMS) is a full-fledged Attribute Authority, whose job is to assign attributes like group membership and role ownership to members of a Virtual Organization (VO).

At the time of writing of this paper the DORII e-Infrastructure consists of 9 sites offering computational and storage resources distributed among the partners of the project. Table 1 shows the sites that support the catch all DORII VO, where most of the DORII applications have been deployed:

Table 1. vo.dorii.eu Computational and Storage Resources

Country	Partner Name	Site Name	CPU Cores	Storage (TB)
Poland	PSNC	PSNC	1068	16
Spain	CSIC	IFCA-CSIC	372	107
		IFCA-I2G*	372*	107*
Italy	ELETTRA	ELETTRA	80	0.1
Greece	GRNET	HG-01-GRNET	64	4.78
		HG-02-IASA	118	3.14
		HG-03-AUTH	120	3.13
		HG-04-CTI-CEID	114	2.87
		HG-05-FORTH	120	2.33
		HG-06-EKT	228	7.76
		Totals	2284	147.11

Three scientific communities have deployed their instruments and the corresponding applications using them in the DORII infrastructure: i) environmental observation and monitoring; ii) earthquake engineering; iii) experimental science.

The e-Infrastructure addressed by DORII encompasses the following main environments:

- *Application environment*

This layer comprises a diverse set of applications, from those providing Remote Instrumentation Services to parallel applications performing simulation and modelling on the acquired data.

- *Middleware framework*

The framework consolidates traditional middleware solutions for Grid, enhancing and adapting to the requirements of the remote instrumentation. In the context of DORII, the middleware framework is comprised of the tools that are best practices for remote instrumentation (like the Instrument Element – IE [12]) and management of Grid resources (the Virtual Control Room – VCR [13]), application development

(g- Eclipse [14]) and workflow management (VLab [15]), interactivity and visualization (GLogin and GVid [16]), as well as parallel application support (Open MPI and PACX-MPI [17]).

- *Fabric layer*

This layer, comprising network-connected Grid resources, is extended by scientific instrumentation and remotely controlled devices.

- *Networking technologies*

This layer includes the networking solutions for connecting all the upper layers of the Grid architecture. TCP/IP is the common networking technology, along with specific access network protocols.

3   Monitoring the Network and the Infrastructure

The network monitoring infrastructure deployed for the DORII project consists of the following tools:

- Smokeping, for network latency measurement;
- Pathload, for the estimation of the available bandwidth along a network path;
- SNMP-based Web applications, for monitoring network interface utilization.

3.1   Smokeping

Smokeping [18] is a software tool that can be used to measure the network latency. More specifically, a Smokeping probe sends test packets out to the network and measures the amount of time they need to travel to a target host node and back. The RRDtool [19] is used to maintain a long-term data-store with latency measurement, and the presentation of the data on the web is done by means of a CGI with some AJAX capabilities for interactive graph exploration.

In the framework of the DORII project, Smokeping is used in master/slave mode: this way, Smokeping probes (slaves) are allowed to run remotely and to perform latency measurements from multiple locations to the target hosts.

As shown in Fig. 1, Smokeping has been deployed as follows:

- the Smokeping master is located at CNIT [20]; it maintains a configuration file with a specific section for each slave, and it stores and presents all monitoring data collected by the slaves.
- Remote probes (Smokeping slaves) have been installed at DORII partners' sites EUCENTRE, ELETTRA, GRNET, CSIC-IFCA, OGS and PSNC. Based on settings contained in the configuration file retrieved from the master (e.g., measurement utility, target host address, measurement length and period, etc.), each slave performs latency measurements and sends back the results to the Master server by using the HTTP protocol. In the DORII network infrastructure, the following targets have been identified: Computing Elements (CEs), Storage Elements (SEs), Instrument Elements (IEs), remote sites' Access Gateways (AGs).
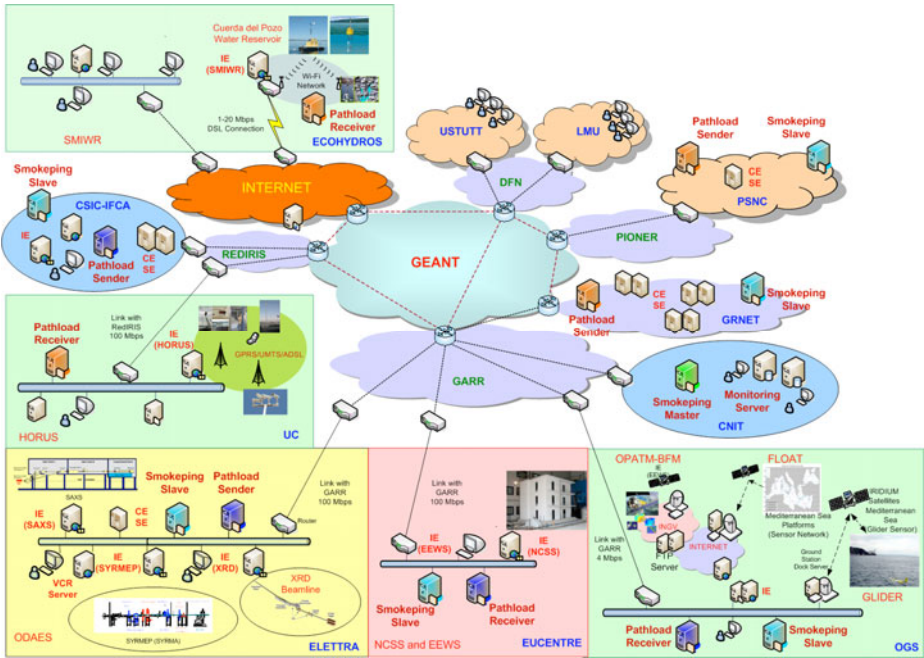
Fig. 1. Smokeping and Pathload deployment for the DORII project infrastructure

3.2 Pathload

Pathload [21] is a monitoring tool that estimates the available bandwidth of a network path. The basic idea behind Pathload is that the one-way delays of a periodic packet stream show an increasing trend when the stream rate is larger than the available bandwidth.

Pathload is based on a client-server architecture and consists of two main components:

- *pathload_snd* that listens on TCP port 55002 and acts as a traffic generator;
- *pathload_rcv* that starts a Pathload session and acts as a traffic receiver.

Pathload has been customized for the DORII project. In addition to the previous components, some scripts have been introduced to monitor the status of the sender and receiver processes and to automatically export the measurement data collected by the receiver to the management station located at CNIT via HTTP. The tool has been installed as follows (see Fig. 1):

- Pathload_sender: at each site where CEs and/or SEs of the DORII e-Infrastrcture are located (GRNET, PSNC, CSIC-IFCA);
- Pathload_receiver: at each site where IEs are deployed (EUCENTRE, OGS, ELETTRA, UC, etc.) and, therefore, DORII applications are running.

This way, the bandwidth available from the sites hosting CEs and SEs to the sites with applications (IEs and VCR) can be estimated.

3.3   SNMP-Based Network Monitoring

Various applications exist to collect and consolidate network usage information. At a basic level, such applications (also called managers) use the Simple Network Management Protocol (SNMP) to read statistics from each monitored device (router or host) where an SNMP agent is configured and running. A standard Management Information Base (MIB) collects counters of the number of datagrams and bytes sent and received on each interface of a device, and it also gives the number of packets discarded because of congestion. An SNMP application can periodically poll each device and convert the returned information into a view of usage across the whole network. SNMP can also help identify network interface failures or outage conditions. In the framework of the DORII, SNMP is required to be enabled on IEs, CEs, SEs and routers. Data are collected by an SNMP manager and interfaced with a Web server by using ad-hoc CGI programs.

3.4   Nagios

Nagios [22] is an open source monitoring system providing comprehensive and scalable monitoring of all mission-critical infrastructure components, including applications, services, operating system and system metrics or network protocols and infrastructure. Nagios is integrated in the monitoring framework of the DORII e-Infrastructure providing information on problems and failures related to the computational, storage and instrument resources of this infrastructure, and monitoring services such as the CE, SE, BDII, WMS, and IE.

4   Deployment of a Selected Application

4.1   The Application

The EEWS (Earthquake Early Warning System) aims at recording seismic data from sensors, possibly in real time, and at processing them in order to extract time history for ground velocity, ground acceleration and displacements. This is the starting point to calculate some interesting parameters widely adopted in the seismic community, such as the acceleration Fourier amplitude spectrum and the acceleration response spectrum (useful to evaluate a building's response to the force imposed by the earthquake). The rationale of this application is to provide scientists a unified environment mixing access to instruments and computational tools, speeding up the analysis carried out after a seismic event.

All the operations are performed remotely and on the grid, employing the DORII infrastructure. The VCR plays the role of the user interface: all the actions are carried out and all the resources are accessed through this web portal.

The IE is located at EUCENTRE (Pavia, Italy) and it hosts the IM devoted to access the server collecting data from sensors. This node is in Genoa, Italy, while the seismic sensors are spread over the Liguria Region. Measurements from each channel are time-stamped and saved locally on the IE in a separate file, then moved to a SE. Finally, a CE retrieves each file from the SE and performs the computation. Since the same computation is repeated for each input file, the job is parametric,

where the parameters are the file names. The JDL (Job Description Language) file characterizing the job is created using a VCR application, which is a Jython script that customizes a given template. The user is asked only to select the input folder on the right SE. The output is downloaded to the home folder on the VCR. An alternative approach is represented by the Workflow Manager, a graphical and friendly interface to specify the parameters.

4.2   The Instruments

A set of seismic sensors is connected to a central point over the UDP protocol, by means of wired or wireless links. Each device measures the ground velocity along the three Cartesian directions, so each station broadcasts at least three channels, plus state of health information. All the data are gathered by a central server, which manages replicated and out of order packets. The reconstructed stream is stored and made available as DAT service: the user can access the historical data series, specifying the starting point of information flow and the time window of interest. On the contrary, if a user or an application is focused on near-real-time access, the NAQ service is more suitable. Given that some parameters are tuned properly, it is possible to configure the service to forward the original packets from instruments to the application, minimising the delay. In fact this is the meaning of near-real-time acquisition. Seismic sensors send measurements to the central server as soon as possible, but the server has to store them, organizing in data structures called "bundles" and "packets". These operations take some time, however limited to few seconds, depending on the distance between the server and the client.

Packets are uniquely identified by a sequence number and a time stamp; they include an odd number, from 1 to 255, of 17-byte bundles. This solution allows adapting the packet size to the network. Moreover, it is worth noting that data contained in a packet are homogeneous, for example the measurements of a particular channel or the status information of a station. The first bundle in a packet always acts as header, specifying some useful details such as station's unique ID, time stamp, sampling frequency and sequence number. The following groups of bytes carry only data. State of health messages are obviously strings, while measurements are integer values, in compressed or uncompressed format.

4.3   The Instrument Manager

The information flow is provided to the IM by means of a TCP connection as a time series or a transparent serial stream. A transparent serial stream handles only uncompressed format and all packets have the same length, employing padding where necessary. On the contrary, a time series stream can deal with compressed data, too, and the packet size may vary. Both streams exist also in the buffered version, so it is possible to retrieve additional packets prior to the beginning of subscription, moving the starting point slightly to the past. Finally, one of the most important parameters is the Short Term Completion (STC) time. It represents the time interval, from 0 to 300 seconds, the server waits to fill the gaps in the stream in case of retransmitted packets. Our goal consists in reading seismic data in near-real-time, so we chose to subscribe a time series compressed stream (as it is more efficient) in its unbuffered version,

disabling STC. This means that measurements are not guaranteed to be in order because of errors and losses, so the gaps are filled using interpolation and delayed packets are discarded. The approach guarantees a continuous flow, fundamental feature for the subsequent computation, complying with the strict time constraint. All the operations are performed by a Java client library developed for the VCR architecture on the basis of the user's manual provided by the server manufacturer.

5   Experimental Results

The user task list previously described represents the starting point for the test bed set-up employed in this work. Moreover, network performance over the grid infrastructure is monitored during the entire life cycle of the application execution. In our experiments, we skipped the acquisition phase, since the server gathering sensors' data is not part of the infrastructure, unlike the IE sending the query. Moreover, the required network resources are very limited: as a matter of fact, a single station channel only needs few kbps. The size of the file containing the initial data set for the computation is about 115 MB and corresponds to measurement data coming from two channels and collected for a whole day. The archive is stored on a server at EUCEN-TRE that acts as IE and VCR; then, it is transferred to a SE at GRNET (se01.isabella.grnet.gr). The average throughput is about 7.5 Mb/s. The analysis is carried out by a parametric job, where the parameters correspond to different settings for the two seismic stations being monitored, so a single execution of the application produces two children nodes. The job is launched three times, and the target CEs are located   at   different   sites: GRNET   (ce01.athena.hellasgrid.gr),   PSNC (ce.reef.man.poznan.pl) and IFCA-CSIC (egeece01.ifca.es).

Table 2. Upload time from se01.isabella.grnet.gr to each CE and processing time

CE	Upload Start Time	Upload Time [s]	Throughput [Mb/s]	Average Latency [ms]	Computation Start Time	Processing Time [s]
ce01.athena.hellasgrid.gr	27/11/2009, 12.59	2.0	478.4	-	27/11/2009, 12.59	26
ce01.athena.hellasgrid.gr	27/11/2009, 12.59	2.1	460.2	-	27/11/2009, 13.18	65
ce.reef.man.poznan.pl	27/11/2009, 13.07	351	2.6	59.2	27/11/2009, 13.07	16
ce.reef.man.poznan.pl	27/11/2009, 13.04	241	3.9	59.2	27/11/2009, 13.04	47
egeece01.ifca.es	27/11/2009, 13.46	95	9.8	94.3	27/11/2009, 13.48	22
egeece01.ifca.es	27/11/2009, 13.46	82	11.7	94.3	27/11/2009, 13.48	45

Table 2 reports the time necessary to upload the file from the SE to each CE, along with the processing time. The last column of the table reports the latency measured by using Smokeping. If the user chooses a local CE (ce01.athena.hellasgrid.gr) the latency is in the order of a few ms, and therefore it is not reported in the table, but

also in the other two cases the latency is very low, since it is less than 100 ms.

The job output consists of files whose size is comparable with the size of the input file. Finally, the output files are retrieved, by using the VCR at EUCENTRE. Table 3 reports the time necessary to download the output files stored by each CE from EUCENTRE.

Table 3. Download Time from each CE to EUCENTRE VCR

CE	Getting Output Start Time	Download Time [s]	Throughput [Mb/s]	Available Bandwidth [Mb/s]	Average Latency [ms]
ce01.athena.hellasgrid.gr	27/11/2009, 13.18	36	26.8	92	49.2
ce01.athena.hellasgrid.gr	27/11/2009, 13.19	32	30.2	92	49.2
.2ce.reef.man.poznan.pl	27/11/2009, 13.22	33	29.3	102	29.9
ce.reef.man.poznan.pl	27/11/2009, 13.23	89	10.9	102	29.9
egeece01.ifca.es	27/11/2009, 14.16	32	30.2	96	46.1
egeece01.ifca.es	27/11/2009, 14.17	32	30.2	96	46.1

Table 4. Total Time

CE	Total Time [s]	Elaboration Time [s]	Communication Time [s]
ce01.athena.hellasgrid.gr	64	26	38
ce01.athena.hellasgrid.gr	99	65	34
ce.reef.man.poznan.pl	400	16	384
ce.reef.man.poznan.pl	377	47	330
egeece01.ifca.es	149	22	127
egeece01.ifca.es	159	45	114

The last two columns contain the available bandwidth (estimated by Pathload) and the average latency (measured by Smokeping) from each CE to EUCENTRE VCR. It is relevant to highlight that the throughput is significantly less than the available bandwidth: this means that the communication protocols are not efficient enough to utilize the available bandwidth of the communication channel.

Finally, Table 4 reports the overall time for the execution of the application. As clearly shown in the table, the amount of time necessary for exchanging the data may significantly affect the performance of the application and represents (except in the second case) the major component of the overall execution time.

6   Conclusions

The DORII project has created and is managing a Virtual Organization for the deployment of applications in a number of different scientific fields. Applications that were traditionally executed locally have been ported to a Grid environment, where they can find and ask for computational and storage resources, and carry out the data processing operations they require efficiently and timely. More importantly, DORII middleware enables scientists to expose and access their instrumental resources on the Grid, by means of universal abstractions that apply to diverse e-Science domains. At

the same time, a ubiquitous monitoring service has been deployed, allowing the continuous evaluation of the distributed system performance and the discovery of possible problems and bottlenecks.

The architecture and the main features of the deployed test bed have been presented in the paper, together with an example in performance monitoring of a selected application. Scientists in the different disciplines involved are actively participating in DORII applications' deployment and experimental activity, with the goal of evaluating the new middleware functionalities and suggesting improvements. Future work will be aimed towards the development of additional functionalities and to providing input to standardization activities in Remote Instrumentation Services.

Acknowledgments

References

1. Foster, I.: Service-oriented science. Science Mag. 308(5723), 814–817 (2005)
2. The RINGrid Consortium: Whitepaper on Remote Instrumentation. Computational Methods in Science and Technol. 15(1), 119–138 (2009)
3. Davoli, F., Meyer, N., Pugliese, R., Zappatore, S. (eds.): Grid-Enabled Remote Instrumentation. Springer, New York (2008)
4. Davoli, F., Meyer, N., Pugliese, R., Zappatore, S. (eds.): Remote Instrumentation and Virtual Laboratories. Springer, New York (to apppear, 2010)
5. DORII Project Home Page, http://www.dorii.eu
6. Adami, D., Cheptsov, A., Davoli, F., Liabotis, I., Pugliese, R., Zafeiropoulos, A.: The DORII project test bed: Distributed eScience applications at work. In: Proc. 1st Internat. Workshop on Pervasive Computing Systems and Infrastructures (PCSI), Washington, DC (April 2009)
7. ESFRI Home Page, http://cordis.europa.eu/esfri/
8. EGEE Project Home Page, http://www.eu-egee.org/
9. DEISA project Home Page, http://www.deisa.eu/
10. http://glite.web.cern.ch/glite/
11. GRIDCC Project Home Page, http://www.gridcc.org/
12. Frizziero, E., Gulmini, M., Lelli, F., Maron, G., Oh, A., Orlando, S., Petrucci, A., Squizzato, S., Traldi, S.: Instrument Element: a new Grid component that enables the control of remote instrumentation. In: Proc. 6th IEEE Internat. Symp. on Cluster Computing and the Grid Workshops (CCGRIDW 2006), Singapore (May 2006)
13. Ranon, R., De Marco, L., Senerchia, A., Gabrielli, S., Chittaro, L., Pugliese, R., Del Cano, L., Asnicar, F., Prica, M.: A web-based tool for collaborative access to scientific instruments in cyberinfrastructures. In: Davoli, F., Meyer, N., Pugliese, R., Zappatore, S. (eds.) Grid Enabled Remote Instrumentation, pp. 237–251. Springer, New York (2008)
14. Kornmayer, H., Stümpert, M., Knauer, M., Wolniewicz, P.: g-Eclipse – an integrated workbench tool for Grid application users, Grid operators and Grid application developers. In: Proc. Cracow Grid Workshop 2006, Cracow, Poland (October 2006)

15. Okoń, M., Kaliszan, D., Lawenda, M., Stokłosa, D., Rajtar, T., Meyer, N., Stroiński, M.: Virtual Laboratory as a remote and interactive access to the scientific instrumentation embedded in Grid environment. In: Proc. 2nd IEEE Internat. Conf. on e-Science and Grid Computing (e-Science 2006), Amsterdam, The Netherlands (December 2006)
16. Köckerbauer, T., Polak, M., Stütz, T., Uhl, A.: GVid - Video Coding and Encryption for Advanced Grid Visualization. In: Proc. 1st Austrian Grid Symp., Schloß Hagenberg, Austria (December 2005)
17. Keller, R., Liebing, M.: Using PACX-MPI in MetaComputing applications. In: Proc. ASIM 2005 - 18th Symp. on Simulation Technique (Symposium Simulationstechnik), Erlangen, Germany (September 2005)
18. Smokeping Home Page, `http://oss.oetiker.ch/smokeping/`
19. RRD Tool Home Page, `http://oss.oetiker.ch/rrdtool/index.en.html`
20. DORII Monitoring Platform Home Page, `http://monitor2.cnit.it`
21. Pathload Home Page,
    `http://www.cc.gatech.edu/fac/Constantinos.Dovrolis/`
    `bw-est/pathload.html`
22. NAGIOS Home Page, `http://www.nagios.org`