Self-Managing Sensor-based Middleware for Performance Monitoring and Data Integration in Grids [*]

Hong-Linh Truong
Institute for Software Science,
University of Vienna
Nordbergstr. 15/C/3, A-1090 Vienna, Austria
truong@par.univie.ac.at

Thomas Fahringer
Institute for Computer Science,
University of Innsbruck
Technikerstr. 13, A-6020 Innsbruck, Austria
Thomas.Fahringer@uibk.ac.at

Abstract

This paper describes a sensor-based middleware for performance monitoring and data integration in the Grid that is capable of self-management. The middleware unifies both system and application monitoring in a single system, storing various types of monitoring and performance data in decentralized storages, and providing a uniform interface to access that data. We have developed event-driven and demand-driven sensors to support rule-based monitoring and data integration. Grid service-based operations and TCP-based data delivery are exploited to balance tradeoffs between interoperability, flexibility and performance. Peer-to-peer features have been integrated into the middleware, enabling self-managing capabilities, supporting group-based and automatic data discovery, data query and subscription of performance and monitoring data.

1. Introduction

Grid monitoring is a crucial task as it provides performance and monitoring data for several functions such as performance analysis and tuning, performance prediction, fault detection, and scheduling. Grid monitoring middleware has to support monitoring of disparate different resources and applications, seamlessly integrating performance and monitoring data from many sources. More importantly, it should provide the same mechanism for data requesters to efficiently access various types of data.

To monitor various resources in Grids, a large number of monitoring sensors needs to be developed and deployed in different domains. In our view, such sensors are very similar to those in sensor networks [2, 33] in which the sensor follows resource constraints such as communication (networks connecting sensors usually vary, having latency with high variance, sensors must use limited bandwidth), computation (sensors have to use limited computing power and memory sizes, otherwise the monitoring may change the state of the monitored resources). These constraints limit data processing capability of a sensor thus normally the sensor sends collected data to a sink node which stores the data. In many cases the sink node also controls or requests data from sensors; we call such sink node a sensor manager. In Grids, it is impossible for all sensors to communicate with a central sensor manager. Also because resources on which sensors execute and resources sensors monitor may join and leave, the structure of sensor networks is frequently changed. Therefore, sensors and sensor managers must operate in self-managed, decentralized manner.

Most existing Grid monitoring tools have monitoring sensors operating in distributed manner and the network connecting sensors to sensor managers exploits the various types of communication such as shared memory [6], TCP [35], UDP [8], multicast [25]. However, these tools do not focus on the interoperability among sensor networks and the self-organization within them, and support limited types of sensors. Mostly they support event-driven sensors (e.g. in [6, 25, 35]). Sensor managers are configured into tree of point-to-point connections (e.g. in [6, 25]) or directory services, supporting discovery of data and sensor managers, do not interact with each other (e.g. in [35]). Thus they do not cope well with sensor networks topology which is changed frequently.

Lack of interoperability among sensor networks and lack of self-organization within them have hindered distributed data discovery, data query and subscription (DQS) in Grid monitoring tools, not to mention fault-tolerance. Currently data discovery and DQS are mostly based on hierarchical or centralized models, as studied in [18]. But such models do not work well with more dynamic, large-scale distributed environments, in which useful informa-

[*]This research is supported by the Austrian Science Fund as part of the Aurora Project under contract SFBF1104. A short version of this paper is published in the Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2005), IEEE Computer Society Press, April 4-8, 2005 Denver, Colorado, USA.

tion services are not known in advance. As suggested, e.g. in [17, 29], and demonstrated, e.g. in [20, 30, 21], the super-peer model and service group, powered by peer-to-peer (P2P) computing [26], have advantages in solving the above-mentioned issues, but have not been exploited in Grid monitoring middleware. Moreover, most Grid monitoring tools are not capable of self-configuration and -reconfiguration under varying conditions which occur frequently in the Grid. Autonomic computing [23] which aims to cope with the unpredictable conditions of systems should be exploited.

Integrating performance and monitoring data in Grids is crucial because it is likely that no single tool will be deployed to provide performance data for all Grid sites and we need to utilize and analyze monitoring data across multiple Grids at the same time. Each Grid system is equipped with different computing capabilities, platforms, libraries that require various performance monitoring and measurement tools. But Grid users should not be forced, when possible, to access monitoring data in Grid systems by using different mechanisms. Instead, Grid users should be able to utilize that diverse monitoring data by using the same mechanism.

Seamlessly integration and highly interoperability require well-defined interfaces, rich expressive customized data representations, and more power to process and store data. However, involving more function and processing results in slower performance. Therefore, we need to balance tradeoffs between interoperability and performance. Using Grid/Web service-based operations and XML data supports highly interoperability among different tools, easily customizing collected data, however, the performance considerably suffers when data is delivered via Web service operations with SOAP [14]. On the other hand, (parallel) TCP-based data streams can be utilized to achieve higher performance in delivering data in Grids [3]. Current Grid monitoring tools exploit either purely Grid service-based operations or TCP-based data streams.

In this paper, we describe our first step in exploiting, developing and incorporating self-managing and P2P features into a sensor-based middleware for Grid monitoring and performance data integration within the SCALEA-G system [32]. We exploit the sensor networks model to provide and integrate various types of performance and monitoring data in a unified system. Regardless of the monitoring data of Grid infrastructures or applications, data is treated and accessed in the same way. Both event-driven and demand-driven sensors are developed to serve different purposes. Sensors also use a rich set of rules and inference engine to control the monitoring. In order to balance tradeoffs between interoperability and performance, we employ both Grid service-based operations and TCP-based stream data delivery. Self-managing and P2P features are exploited to increase the scalability and fault-tolerance, and to foster distributed, group-based and automatic DQS of performance and monitoring data.

The rest of this paper is organized as follows: Section 2 outlines our sensor-based architecture. Section 3 discusses our sensor model for performance monitoring and data integration. We then describe self-managing capabilities in Section 4. We discuss DQS in Section 5. Hybrid communication based on service-based operations and TCP-based streams are presented in Section 6. Section 7 illustrates experiments and examples. We present some related work in Section 8 before presenting our conclusions in Section 9.

2. Sensor-based Middleware Overview

Figure 1 depicts the architecture of sensor-based middleware implemented in SCALEA-G with the main Grid services named Directory Service, Sensor Manager Service, Client Service. These services, based on OGSA [15] and organized into service groups, support managing, storing and providing various types of performance and monitoring data measured and gathered by an extensive set of distributed sensors. They are capable of self-management and can collaborate to serve the requests from clients.

Directory Service (DS) stores information (e.g. schema, availability) about performance and monitoring data, SM and other services of the middleware. **Sensor Manager Service (SM)** manages sensors and data collected and gathered by sensors, and provides these data to consumers via DQS operations. An SM can interact with several sensor instances executed on distributed machines; sensor instances will send their collected data to SMs. SM uses XML containers to store performance and monitoring data. **Client Service (CS)** provides interfaces for administrating activities of SMs, querying data registered in DS, subscribing and/or querying data stored in SMs, etc. Other services (e.g. scheduling service) access performance data by exploiting facilities provided by CS.

SM and DS are organized into two types of groups (communities): *SM Group* and *DS group*. Within a Virtual Organization (VO) [16] there could be several SM groups. A DS group is deployed for multiple VOs; each VO provides a number of DSs which form the DS group. DSs register their information with a set of Registry Services. By using CS, the client of the monitoring middleware, exploring the monitoring service through existing Registry Services, can find DSs, SMs and then access performance and monitoring data. In our framework, we reuse existing implementation of Registry Service. However, DS and SM are specially designed for performance and monitoring purpose.

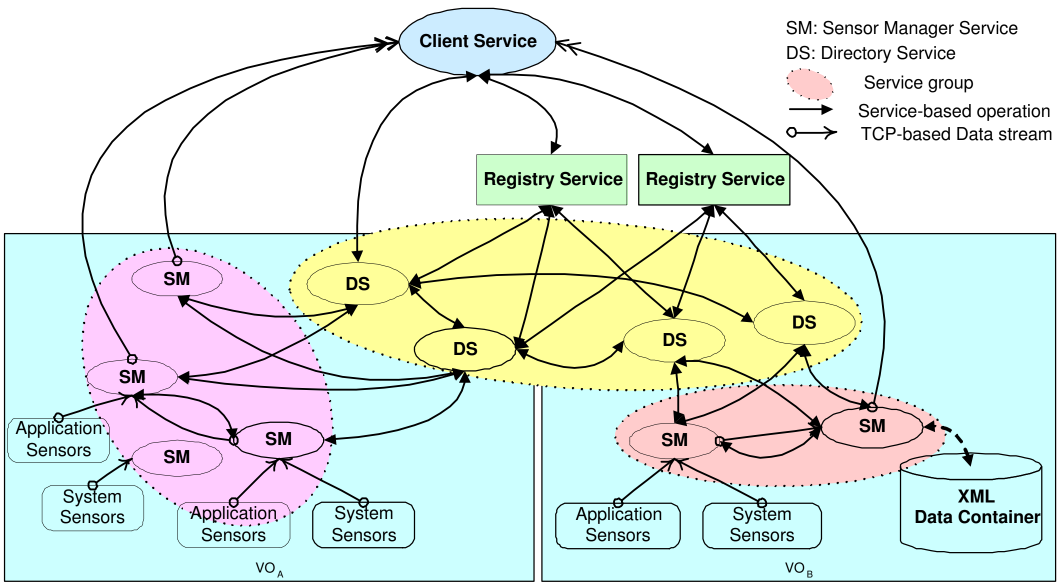


Figure 1. High-level view of self-managing sensor-based middleware.

3. Sensor-based Model for Performance Monitoring and Data Integration

3.1. Monitoring Sensor Conceptualization

Figure 2 presents the underlying concept of our monitoring sensors. Sensors are used to capture performance data, to monitor resources including computational and network resources, and Grid applications. Every sensor monitors one or more *resources* (e.g. machine, network, Grid applications) and provides *measurement data* of the monitored resources; each resource is determined by a unique resource identifier (`ResourceID`) and measurement data is described in XML. Each sensor presents a *sensor profile* which describes the sensor, e.g. unique sensor identifier (`SensorID`), sensor description and lifetime, how to control the sensor (e.g. calling parameters) and information about the provided sensor data (e.g. XML schema of data). How a sensor works is described by the *sensor model*, e.g. event-driven or demand-driven, or rule-based monitoring. For measurement data, the tuple (`SensorID`, `ResourceID`) is unique that is used to determine monitoring data of a resource.

A sensor can be invoked at different times with different parameters. Each invocation of a sensor is called *sensor instance* referring to a particular instantiation of a sensor at run-time. Each sensor instance has its own lifetime.

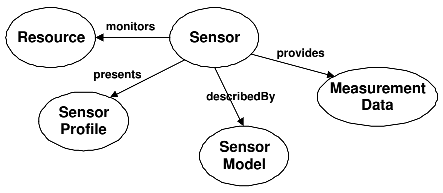


Figure 2. Conceptualization of monitoring sensors.

3.2. Event-driven and Demand-driven sensors

Sensors in most existing Grid monitoring tools are based on event-driven model; a sensor measures and collects data based on events, mostly time-based event. Event-driven sensors collect the data and store the collected data when an event happens at a time, without consideration at that time the data is needed. Demand-driven sensors collect and provide data only when receiving requests. Demand-driven sensors are particularly useful for integrating data provided by other sources. To realize the importance of both types of sensors, our middleware supports both event-driven and demand-driven sensors.

3.3. System Sensors and Application Sensors

We distinguish two kinds of sensors: *system sensors* and *application sensors*. System sensors are used to monitor and measure the performance of Grid computational services (e.g. computational hosts) and network services (e.g. network connections). Application sensors embedded in Grid applications are used to measure execution behavior of code regions and to monitor pre-defined events in these applications. The two kinds of sensors are treated basically the same. They, however, differ in control and security model. This distinction allows us to simplify the management of two different kinds of sensors.

3.4. Processing and Storing Customizable Data

In order to allow sensors to freely customize their provided data, SM must receive, process and store multiple data types with unknown (XML) structures into its data buffer. To do so, we design a three-phase protocol for exchanging data between sensors and SMs.

In this protocol, the interactions between sensors and SMs involve the exchange of three XML messages: `sensorinit` message in the initialization phase, `sensordataentry` message in the measurement phase, and `sensorfinal` message in the final phase. Measurement data is encapsulated into `sensordataentry` message. The measurement data is enclosed by <`![CDATA[ ... ]]`> tag. Thus, sensors can customize the structure of their collected data. Before it stops sending collected data, the sensor instance sends a `sensorfinal` XML message to notify the SM. The three XML messages that always contain `SensorID` and `ResourceID` with other information (e.g XML schema of data which sensor produces, lifetime and description information about the sensor) are self-explained. Based on (`SensorID, ResourceID`), SM can setup appropriate buffers, storing data into the buffers. Therefore, multiple types of monitoring data can be delivered via a transient connection, not just through a persistent one.

Sensors can send XML schemas to SM even though SM does not need the schemas in processing the received data. SM will publish these schemas to DS so that other services consuming collected data can get schemas in order to make use of the data. Instead of sending XML schema to SM, the schemas can be stored at SM. Our approach allows any sensor that implements the above-mentioned protocol to send the data to SM while SM is not necessarily aware of the structure of the measurement data.

3.5. Rule-based Monitoring

Different with event-driven sensors in existing Grid monitoring tools, our event-driven sensors support rule-based monitoring. Instead of sending monitoring data it collects, the sensor uses rules to analyze monitoring data, and reacts with appropriate functions.

As different resources have different characteristics, we need to setup different rule sets for monitoring different resources. For example, with two different instances of a sensor used to monitor the bandwidth of two network paths which have different characteristics, e.g. one with maximum 100 MBytes/s, the other with 10 MBytes/s, the rules used to detect whether the bandwidth is low or high must be different even though the way to monitor these paths is the same. Therefore, a rule set is normally associated with each sensor instance. However, multiple sensor instances can share the same rule set, e.g. when their monitored resources have the same characteristics.

We use ABLE Rule Language (ARL)[10], which supports if-then-else rules, when-do pattern match rules, etc., to define rule sets for sensors. ABLE toolkit [9] provides a wide range of inference engines to process the ARL rulesets, e.g. boolean forward/backward chaining, fuzzy forward chaining, pattern match engine. For example, to define a fuzzy variable for monitoring bandwidth of a network path in the Austrian Grid [4], we used Iperf [22] to test the bandwidth, and obtained the maximum observed bandwidth which never exceeds 5 MBytes/s. We divided the bandwidth into 5 states by using fuzzy logic as shown in Figure 3. Based on this fuzzy variable, we define a rule set, presented in Figure 4. With this rule set, depending on the status of bandwidth of the network path, e.g. *very low, low* or *very high*, the sensor will react with appropriate functions, e.g. to send events to SM.

In the case where rules are not specified when creating a sensor instance, the instance will work as in the normal model (e.g. sending monitoring data when the event happens). Rule-based monitoring approach brings many advantages as it allows to control the monitoring actions. In addition, we can implement autonomic features that consider changing systems as an effect of the monitoring behavior. However, there is no common rule set for all resources even those monitored by a single sensor. Rules have to be built for each resources based on best practices.

3.6. Performance Data Integration by Using Demand-driven Sensor

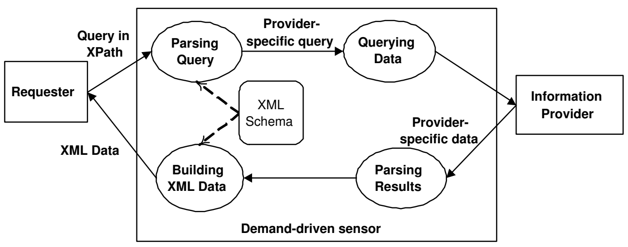


Figure 5. Using a demand-driven sensor to integrate performance data.

Besides using demand-driven sensors to monitor resources, we also exploit them for data integration. Figure 5 presents the model of using demand-driven sen-

```
Fuzzy bandwidth= new Fuzzy( 0,5) {
      Shoulder VERYLOW = new Shoulder(0, 1, ARL.Left);
      Triangle LOW = new Triangle(1, 1.5, 2);
      Trapezoid MEDIUM = new Trapezoid(2,2.2,2.8 ,3);
      Triangle HIGH = new Triangle(3, 3.5, 4);
      Shoulder VERYHIGH = new Shoulder(4, 5, ARL.Right);
    };
```

Figure 3. A fuzzy variable describing status of the bandwidth of a network path.

```
S1: bandwidth = getBandwidth();
R_VERYLOW: if (bandwidth is VERYLOW) {
             doReactionWhenBandwidthVeryLow();
           }
R_LOW:     if (bandwidth is LOW) {
             doReactionWhenBandwidthLow();
           }
R_VERYHIGH: if (bandwidth is VERYHIGH) {
             doReactionWhenBandwidthVeryHigh();
            }
R_OTHER:    doNormalReaction();
```

Figure 4. Example of rule set for bandwidth of a network path.

sor for integrating performance and monitoring data from other providers, e.g. MDS (Monitoring and Directory System) [12], NWS (Network Weather Service) [35], Ganglia [25]. To access different providers, we develop different demand-driven sensors taking the role of data mediators. As shown in Figure 5, when the sensor receives an XPath-based request from a requester, based on XML schema, it parses the request, extracting information of the request such as tag names with their associated attributes. The sensor then constructs a provider-specific request, calling the information provider with that request, and obtaining the result in provider-specific format. The sensor then parses this result and builds a new result described in XML. The XML-based result will be sent back to the requester. With this approach, other services use the same mechanism to access data in other providers as in our service.

When a demand-driven sensor is activated, the sensor returns information about resources whose monitoring data it can collect to the SM which in turn publishes the information to DSs. With that information, consumers can create requests for monitoring data. For instance, consider the `path.predict.bandwidth.capacity.TCP` sensor which obtains predicted network bandwidth from NWS. The startup information for an instance of this sensor is configured in Figure 6.

When this sensor starts it queries the NWS NameServer and returns network paths whose predicted bandwidth values it can collect to the SM. A consumer specifies an XPath-based request as follows

```
/sensordata[@SensorID="path.predict.band
width.capacity.TCP"][source="blindis.dps.
uibk.ac.at:8060"][destination="olperer.dps.
uibk.ac.at:8060"]
```

in order to obtain the predicted bandwidth of network path (`blindis.dps.uibk.ac.at:8060`, `olperer.dps.uibk.ac.at:8060`). This request is then translated into an `nws_extract` command as follows

```
nws_extract -f,time,mae_forecast -N
olperer.dps.uibk.ac.at:8090  bandwidthTcp
blindis.dps.uibk.ac.at:8060 olperer.dps.
uibk.ac.at:8060
```

which is used to obtain predicted bandwidth values from NWS. The output of `nws_extract` is then translated into XML that is sent back to the requester. The requester only knows the XML schema of requested data in order to specify the request. The rest, where the requested data locates and how to get the requested data, are done by the middleware. Our integration approach aims at providing a uniform, flexible interface for accessing data collected by lower-level, domain-specific information providers.

```
<startupsensor name="path.predict.bandwidth.capacity.TCP">
 <startupparam name="NWSNameServer"
                     value="olperer.dps.uibk.ac.at:8090"/>
</startupsensor>
```

Figure 6. Startup information for `path.predict.bandwidth.capacity.TCP` sensor.

4. Self-Managing Services

4.1. Service Group

Each SM or DS group has a set of operations associated with the group. These operations address (i) how the requests for performance data are handled, and (ii) how the requested data are delivered. The real number of members of a group is dependent on the actual deployment which can dynamically change. In the whole system, there could exist many different groups.

The group operations associated with SM group are group-based DQS. One member in the group can act as a mediator for other members. Given a DQS request, an SM can provide requested data even though its storage does not contain the requested data by collaborating with other SMs in the same SM group. For a DS group, the group-based operation supports the discovery of data providers. Given a request to find the provider of a needed data type, DSs in a DS group can cooperate to determine the data provider.

4.2. Data Dissemination and Maintenance

Instances of sensors are executed on monitored nodes and send data collected to SM which in turn stores the data into its data container. SM automatically publishes characteristics of received data to a set of DSs, not to a single DS. Each SM keeps its group name and a list of DSs to which it publishes data. The list of DSs can be dynamically changed over the time. Each SM keeps a list of Registries where it can search information about DSs. When an SM is created, the SM gets a maximum number p of DSs it should register with and a pre-defined updated interval t seconds. In the cycle of t, SM looks up Registries to get n DSs. The SM then selects `min(n,p)` DSs from n ones. A DS is chosen based on the following procedures. Firstly DS is selected based on domain-name approximation, and then based on the latency of `ping` operation, and finally based on random selection. SM then disseminates information about data it stores to its selected DSs.

A DS can publish information about itself to multiple Registries. In the current implementation, one DS belongs to a DS group. A DS keeps a list of Registries with which it registers its information. A DS maintains a list of DSs in its groups; these DSs are its edge peers. Repeatedly with predefined t_r seconds, the DS searches Registries for up-to-date information about its edge peers. DS also performs `ping` test to its edge peers to check whether its edge peers are alive. To make sure that it provides the updated information, the DS checks its data in the database periodically based on a pre-defined t_d seconds. During the checking procedure, DS invokes `ping` operation of its registered SMs. If a `ping` to an SM failed, DS assumes the SM is out of service and then DS removes all information associated with that SM. In the cycle of t_u seconds, DS publishes its information to Registries. Before DS finishes its execution, it unregisters its information from the Registries.

As common in service-oriented computing, the availability of Registries, DSs and SMs is the key issue to fault-tolerance of the middleware. Instead of storing data into a centralized SM, collected data are stored over a set of distributed SMs, thus guaranteeing that a failure of one or many SMs does not bring the whole service down. Differing from existing monitoring tools where SM mostly publishes data to a single DS, our SM publishes its information to multiple DSs. Thus, not only data is widely disseminated, highly available but also it makes sure that if a DS is failed to serve requests from clients, still other DSs can do.

4.3. Discovery of Data Providers

Any client that wants to subscribe or query performance data of a resource has to locate a corresponding SM which provides the data. The discovery of data providers is based on requests containing tuples of (`SensorID`, `ResourceID`). A tuple (`SensorID`, `ResourceID`) is unique that determines monitoring data of a resource. Each DS provides a set of operations for other services to retrieve and search its registered data. Based on a tuple (`SensorID`, `ResourceID`), a client can call operations of a DS in order to discover data providers registered with that DS. (`SensorID`, `ResourceID`) can also be specified in the data content filters of DQS requests.

Data discovery can also be done automatically by CS thus a client does not need to interact with DSs. CS parses client requests to get detailed elements such as `SensorID`, `ResourceID`. A list of DSs will be obtained from given Registries. The request is then sent to DSs which in turn cooperate to locate SMs by using group-based operations. When a DS cannot locate the provider of the requested data, it forwards the request to all its edge peers, otherwise it just sends back the results. These edge peers conduct the search and return the result to the peer who calls them. The DS then sends back the result to the requester. A parameter is used to control the request forwarding policy.

5. Data Query and Subscription

5.1. Query and Subscription Operations

SM provides a set of service operations for other services to subscribe and query data available in the SM. Below, we just outline main operations.

- *subscribeData(consumer_dr, SensorID, ResourceID, content_filter, ResultID, duration, relay)*: `consumer_dr` specifies the *Data Receiver* of the consumer; the Data Receiver indicates the endpoint receiving the resulting data. A consumer performs a subscription of monitoring data of a resource determined by `ResourceID`; the data is collected by a sensor determined by `SensorID`. Parameter `content_filter` specifies the content filter applied to the requested data. `SensorID` and `ResourceID` are optional as they can be specified in the content filter. Subscription time is specified by a `duration`. `ResultID` specifies the identifier of the resulting data. If the subscription operation is successful, a `subscription_id` will be returned to the caller whereas resulting data is delivered to consumer via TCP-based streams. `relay` specifies whether SM should relay the subscription to other SMs when it cannot serve.

- *unsubscribe(subscription_id)*: this operation is used to terminate an existing subscription.

- *renew(subscription_id, duration)*: renew or extend an existing subscription (determined by `subscription_id`) to new `duration`.

- *queryData(consumer_dr, SensorID, ResourceID, content_filter, ResultID, relay)*: similar to *subscribeData* operation but for querying data.

Content filters are described in XPath that can be easily written based on XML schemas of data provided by sensors.

By using operations of SM, CS supports both *one-to-one* and *one-to-many* DQS requests. In one-to-one mode, a subscription or query request is used to obtain performance data provided by a single SM whereas in one-to-many mode a client subscribes or queries data from many SMs by using a single subscription or query request.

5.2. Automatic Query and Subscription

Clients can also perform DQS automatically without knowing where the requested data is located. The content filter, specified in DQS requests that clients pass to CS, can contain characteristics of data such as `SensorID`, `ResourceID`. Given an XPath-based content filter, we can obtain XML tags, attributes (e.g. `SensorID`, `ResourceID`), etc., of the filter by using an XPath parser. For example, by processing the following content filter

```
/sensordata[@SensorID='host.mem.used']
[ResourceID='bridge.vcpc.univie.ac.at']
/...
```

, we obtain (`host.mem.used`, `bridge.vcpc.univie.ac.at`) as the value of (`SensorID`, `ResourceID`).

When receiving a request from clients, CS processes the content filter and obtains `(SensorID, ResourceID)` information. It then searches DSs in order to find SMs that provide the requested data; the search is mentioned in Section 4.3. CS then sends requests to SMs which provides the requested data. As a result, the client does not necessary know where the monitoring data is stored. If DQS requests contain information about sensors and monitoring resources, the middleware can automatically handle DQS requests.

The middleware provides simplified APIs for clients to conduct automatic querying and subscribing performance and monitoring data. The APIs hide all the lower-level details of the middleware. For example, Figure 7 presents a simple code which is used to query available monitoring data of CPU usage of the machine `schareck.dps.uibk.ac.at`. The `ConsumerService` class, part of CS, is responsible for processing DQS tasks. The client knows a Registry Service. It indicates the service handle of Registry Service (variable `handle`), specifies the content filter (variable `content_filter`), and calls the CS. CS returns a `DataSensorReader` from which the resulting data is retrieved. The client can call APIs (blocking and non-blocking) or set up a call-back on `DataSensorReader` to obtain the data.

5.3. Group-based Data Query and Subscription

An SM can act as a mediator for other services to access data provided by other SMs in its group. When a client sends a DQS request to an SM, if the SM does not provide the requested data, SM will search its registered DSs to find SMs that can serve the request. If the search is successful, the SM acts as a super-peer between the data requester and the SM provider by forwarding the request to the SM

```
ConsumerService cs = null ;
cs = new ConsumerService();
cs.activateUpDataService();
String handle="http://bridge.vcpc.univie.ac.at:8765/ogsa/services/
                              samples/registry/VORegistryService";
String content_filter="/sensordata[@SensorID=\"host.cpu.used\"]
                        [@ResourceID=\"schareck.dps.uibk.ac.at\"]";
SensorDataReader out =
       cs.distributedQueryDataWithRegistry(handle, content_filter);
```

Figure 7. Example of querying monitoring data by using information from Registry Service.

provider. The provider first tries to communicate with the requester. If successful, the provider sends requested data to requester, otherwise it sends data back to the caller SM. If an SM receives request from another SM, it will not propagate the request when it cannot serve the request.

In this model, an SM can take the role of the super-peer in either/both forwarding requests or/and delivering data. Any SM may become a super peer at the runtime.

5.4. Notification

In all mentioned DQS, the client conducts DQS based on available information about monitoring data published in DS. However, there are many cases in which the client wishes to subscribe for a notification of interesting data which is not available at the time of the subscription. For example, the client may inform the monitoring system that it wishes to receive execution status of activities of a workflow application being executed by the workflow enactment before it submits the workflow application. We call this type of data subscription *notification subscription*.

DS and SM provide two service operations named `subscribeNotification`, `unsubscribeNotification` for subscribing and unsubscribing notification data. DS and SM use a table to keep existing subscriptions of notifications. The client can subscribe the notification on a specific SM or on the whole monitoring system. If the client wishes to receive notification message from a specific SM, the client can register with the SM by calling `subscribeNotification` operation of that SM. In this case, the client will not receive notification data collected by other SMs even though that data satisfies the client's request.

In our framework, SM gathers and stores performance data collected from sensors. However, there is no mechanism to determine SMs which are capable of distributing a specific notification data because SMs and sensors can enter and exit the monitoring system arbitrarily. The client may only know of a few services to which it contacts, e.g. a DS or an SM, but it wishes to receive a notification without knowing the service which is capable of providing this notification. We support this type of notification subscription by implementing a global notification mechanism. By using the CS, the client registers with a set of DSs $\{DS_1, DS_2, \cdots, DS_n\}$ that it knows, and indicates information about interesting data which it wishes to be notified. Each DS_i updates the table containing subscriptions of notifications and then calls `subscribeNotification` operation of registered $\{SM_{i1}, SM_{i2}, \cdots, SM_{im}\}$ in its directory with that indicated information. Similarly, when a new SM registers with a DS, the DS calls that operation of the SM with existing subscriptions in its table. When receiving a `subscribeNotification` call, the SM updates a table containing tuples of `(ResultID, Subscription)`. Whenever SM receives data satisfying notification constraint, SM delivers the data to CS. If SM cannot deliver a notification to a client, the SM will remove the subscription of that notification from the table. To unsubscribe a notification, CS sends unsubscription requests to DSs which in turn pass these requests to SMs.

As each SM registers its information with multiple DSs, an SM can receive duplicate subscriptions for a notification. Currently, SM accepts the duplication because ensuring SM not to receive duplicate subscriptions of a notification might be at a higher cost than calling `subscribeNotification` which just updates the table containing subscriptions of notifications.

6. Service-based Operations and TCP-based Data Delivery

Each SM can be viewed as a peer in a P2P network. It, however, also is a Grid service. In most P2P systems, a peer processes request and delivers data via TCP/UDP channels. Our peer node is unique as we try to integrate both concepts, P2P model and Grid service, into a single peer. A peer provides Grid service operations for other peers and high-level clients to access and control its service. But, peers use TCP-based streams to deliver monitoring data to each other, thus relay functions can easily be implemented to support data delivery among peers. That also offers higher performance for data delivery.

Figure 8 depicts how requests for data and requested data are handled. CS or SM requests data through *Grid service-based invocations* whereas requested data is delivered via *TCP-based streams*. *Data Sender, Data Receiver* and *Data Relay* of SM and CS are responsible for sending, receiving, and relaying performance data, respectively. Each Data Receiver or Data Relay is associated with an *endpoint* describing the network transport. An endpoint is described by a unique XML message containing network transport information such as host name, port, etc. An SM has only one connection to a consumer for delivering all kinds of subscribed data. The connection is created at the first subscription and will be freed after pre-defined t_δ seconds since the last subscription finishes. For delivering resulting data of queries, an on-demand connection will be created and freed when the delivery finishes. A request for data always specifies a unique `ResultID` which is associated with requested data that satisfies the subscribed/queried constraints. SM uses `ResultID` to route requested data to the destination while CS uses `ResultID` to aggregate results of the same request delivered from multiple SMs.

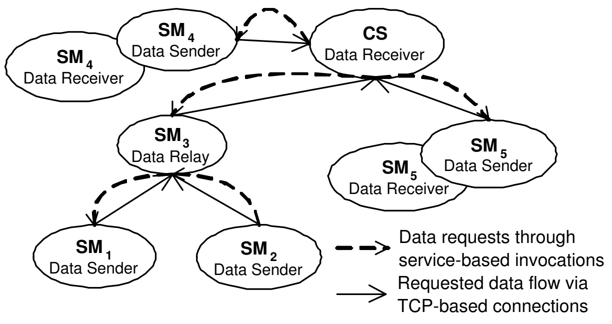


Figure 8. Service-based invocations and TCP-based data streams.

In our middleware, monitoring data is described in XML. While using XML to describe performance and monitoring data provides a widely accessible interface and simplifies the interoperability among services, XML data grows in size. In subscription mode, the size of monitoring data delivered to the consumer each time is small and almost unchanged for a given subscription. However, in data query mode, the size of monitoring data can be large, depending on the query, for example, the consumer may want to retrieve resource monitoring data for the last 10 hours. To reduce the size of monitoring data transfered, we compress monitoring data before sending it over the network.

Data compression may increase the throughput and transfer rate. In Table 1, we monitored the data size and transfer time of CPU usage data from an SM in UIBK domain to a client in PAR domain (see Section 7 for more detail about the experimental test bed); transfer time is the average value of observed values at different times. In our measurement, compression ratio, compression and decompression time are all dependent on the data size and the type of monitoring data. When the data size is small, compressing data will not reduce transfer time because of the small compression ratio and the impact of compression and decompression time on the overall transfer time. However, when the data size is large, compressing data reduces the data size substantially which improves both data transfer time and throughput significantly. For most types of monitoring data supported, when the size of data to be transfered is less than 512 bytes, the compression does not achieve a better transfer time and throughput because the compression ratio is close to 1. Moreover, there is an extra overhead due to the compression and decompression of data. Therefore, we develop a simple self-adaptive mechanism for deciding whether the resulting data should be compressed before sending to the requester that is based on the size of delivered data. If the data size is larger than s_δ, the data will be compressed, otherwise data is transfered as normal. Currently, s_δ is set to 512 bytes.

Data size (bytes)	T_f **(ms)**	T_{fcd} **(ms)**	r	$T_c + T_d$ **(ms)**
588	110	109	1.863	2
1560	119.33	115.24	4.537	2.11
3019	120.45	114.85	8.137	2.19
3991	120.68	114.68	10.207	2.44
5449	129.2	116.15	13.257	2.53
6421	130.45	116.05	14.967	2.73
7879	131.63	117.85	17.316	2.76
8851	136.28	117.48	18.712	3.08
10309	141.39	117.68	20.742	3.09
11281	143.25	117.93	21.949	3.22
12739	147.83	117.5	23.548	3.42
13711	149.75	117.25	24.355	3.67

Table 1. Example of transfer time without compression (T_f), transfer time of compressed data (T_{fcd}), compression ratio (r), compression and decompression time ($T_c + T_d$) for CPU usage data. Time is measured with Java `System.currentTimeMillis()` call. Compression and decompression are implemented based on `java.util.zip` package.

7. Experiments

Our middleware is implemented based on GT 3.2 [19], Java Cogkit [24], with various other libraries. We have deployed our sensor-based monitoring infrastructure on three domains: VCPC (University of Vienna), UIBK (University of Innsbruck) and GUP (Linz University) in the Austrian Grid [4]. Figure 9 presents our experimental test-bed. We set up three SM groups named SM-VCPC, SM-UIBK and SM-GUP in VCPC, UIBK, GUP, respectively. We establish a DS group that includes one DS in VCPC (DS-VCPC) and one in UIBK (DS-UIBK). Each DS stores data in a PostgreSQL database server which can execute in the same domain (e.g. in case of DS-UIBK) or different one (e.g. DS-VCPC). SM stores collected data into XML containers implemented atop Berkeley DB XML [1]. There are two Registries in VCPC and UIBK. A client is deployed in PAR domain (in University of Vienna). All DSs and Registries can be accessed by all SMs and clients, but only SMs executed on `bridge/VCPC`, `olperer/UIBK`, `iris/GUP` can directly deliver data to the client executed on `kim/PAR` due to the firewall. Most machines in our test-bed are non dedicated.

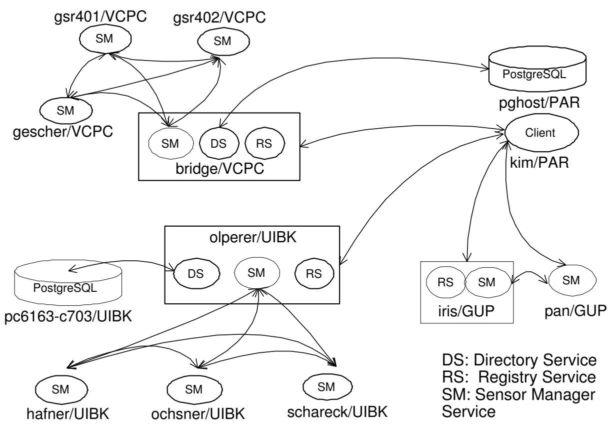


Figure 9. Experimental test-bed.

7.1. Performance Analysis of Data Discovery

To evaluate the performance of the discovery of data providers within the test-bed, we setup two modes. In *one-to-one mode*, a client sends requests directly to DS which in turns finds data providers of the requested data. If the DS cannot locate the data provider, it will not send the request to other DSs in the same group. In *group mode*, if the DS cannot answer the request, it sends the request to its edge peers in its group asking for the location of data providers.

At a random time, for each mode we conducted a series of 20 tests with the interval 60 seconds between two consecutive tests. Each test repeats 10 times and we selected the best timing value as result of the test. In both modes, a client in PAR domain sends requests to DSs at the same time. We also measure the latency of *ping operation* from clients to DSs. The time is measured by using Java `System.currentTimeMillis()` call.

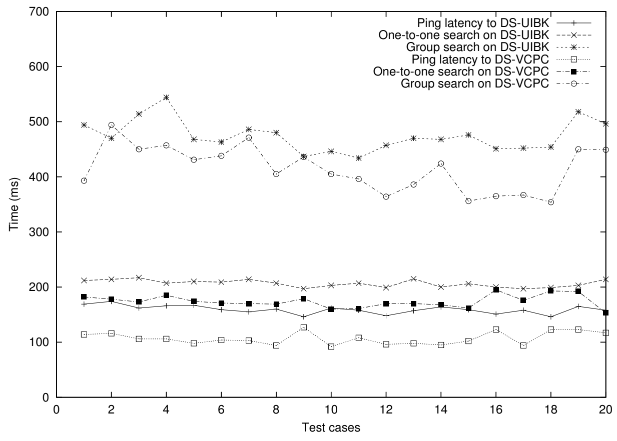


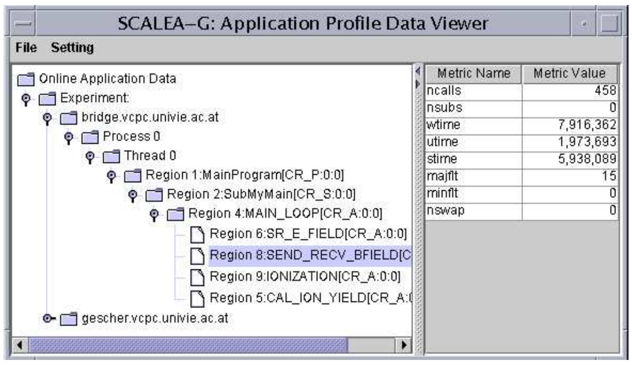
Figure 10. Ping latency and searching time.

Figure 10 presents searching time in one-to-one and group mode, and latency of `ping` operation. Overall for both DSs, the ping latency is larger than a half of the time spent on the search for data providers in one-to-one mode, suggesting that the time DS spends in searching its database is small when compared with ping latency. A considerable portion of time spent in the discovery process is service-operation latency from client to service. In our implementation, for group mode, DS creates a new thread which calls an edge peer when the DS cannot locate the data provider. Searching time in group mode nearly doubles that in one-to-one mode partially because at the same time a DS has to fulfill a request from an edge DS and to forward a request to its edge DS. (In other experiments, not shown in this paper, when a DS just forwarded requests to its edge DS, the searching time was substantially reduced.) The latency from client domain (PAR) to DS-UIBK is higher than that to DS-VCPC, also DS-VCPC is executed on an SMP machine where DS-UIBK is executed on a single CPU machine. Therefore, conducting group-based and one-to-one discovery through DS-VCPC is considerably faster than via DS-UIBK due to the differences of network latency and computation power. Also search times in group mode imply high variance partially because the search now involves different DSs running on wide area networks.

7.2. Monitoring and Data Integration Example

We have implemented a GUI client which accesses a variety of types of monitoring data provided by the middleware and conducts the analysis of that data. We present some examples of using that GUI client to monitor and analysis the monitoring data.

Figure 11 presents an analysis of profiling data collected by application sensors. Incremental profiling data of Grid applications is sent online to SMs. The application profile analyzer then conducts DQS on profiling data, analyzing and visualizing the result to the user. The left window of Figure 11 shows code regions associated with their processing units (compute node, process, thread). For each code region, profiling metrics are incrementally displayed in the right window.

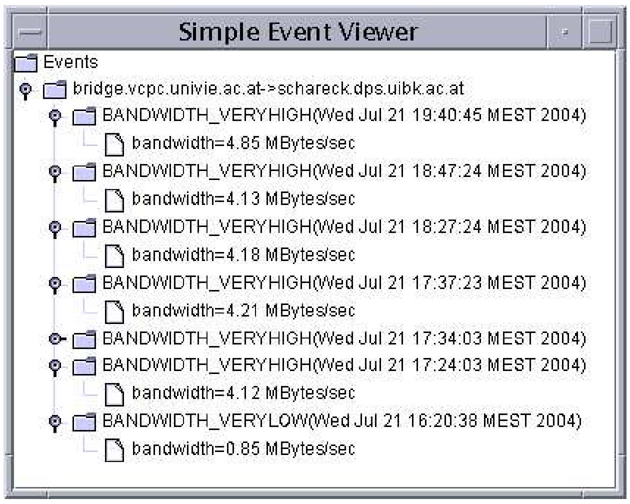

SCALEA–G: Application Profile Data Viewer

File Setting

```
Online Application Data
  Experiment:
    bridge.vcpc.univie.ac.at
      Process 0
        Thread 0
          Region 1:MainProgram[CR_P:0:0]
            Region 2:SubMyMain[CR_S:0:0]
              Region 4:MAIN_LOOP[CR_A:0:0]
                Region 6:SR_E_FIELD[CR_A:0:0]
                Region 8:SEND_RECV_BFIELD[C
                Region 9:IONIZATION[CR_A:0:0]
                Region 5:CAL_ION_YIELD[CR_A:(
    gescher.vcpc.univie.ac.at
```

Metric Name	Metric Value
ncalls	458
nsubs	0
wtime	7,916,362
utime	1,973,693
stime	5,938,089
majflt	15
minflt	0
nswap	0

Figure 11. Analysis of profiling data provided by application sensors.


Simple Event Viewer

```
Events
  bridge.vcpc.univie.ac.at->schareck.dps.uibk.ac.at
    BANDWIDTH_VERYHIGH(Wed Jul 21 19:40:45 MEST 2004)
      bandwidth=4.85 MBytes/sec
    BANDWIDTH_VERYHIGH(Wed Jul 21 18:47:24 MEST 2004)
      bandwidth=4.13 MBytes/sec
    BANDWIDTH_VERYHIGH(Wed Jul 21 18:27:24 MEST 2004)
      bandwidth=4.18 MBytes/sec
    BANDWIDTH_VERYHIGH(Wed Jul 21 17:37:23 MEST 2004)
      bandwidth=4.21 MBytes/sec
    BANDWIDTH_VERYHIGH(Wed Jul 21 17:34:03 MEST 2004)
    BANDWIDTH_VERYHIGH(Wed Jul 21 17:24:03 MEST 2004)
      bandwidth=4.12 MBytes/sec
    BANDWIDTH_VERYLOW(Wed Jul 21 16:20:38 MEST 2004)
      bandwidth=0.85 MBytes/sec
```

Figure 12. Events generated from a rule-based sensor monitoring network bandwidth.

Figure 12 presents an example of bandwidth monitoring of a network path from VCPC to UIBK. We setup a simple rule set based on fuzzy variable for the bandwidth, as presented in Section 3.5. Only when the bandwidth of the network path is *very low, low* and *very high*, the sensor sends events to SM. Events are subscribed and visualized by a simple generic event viewer as shown in Figure 12.

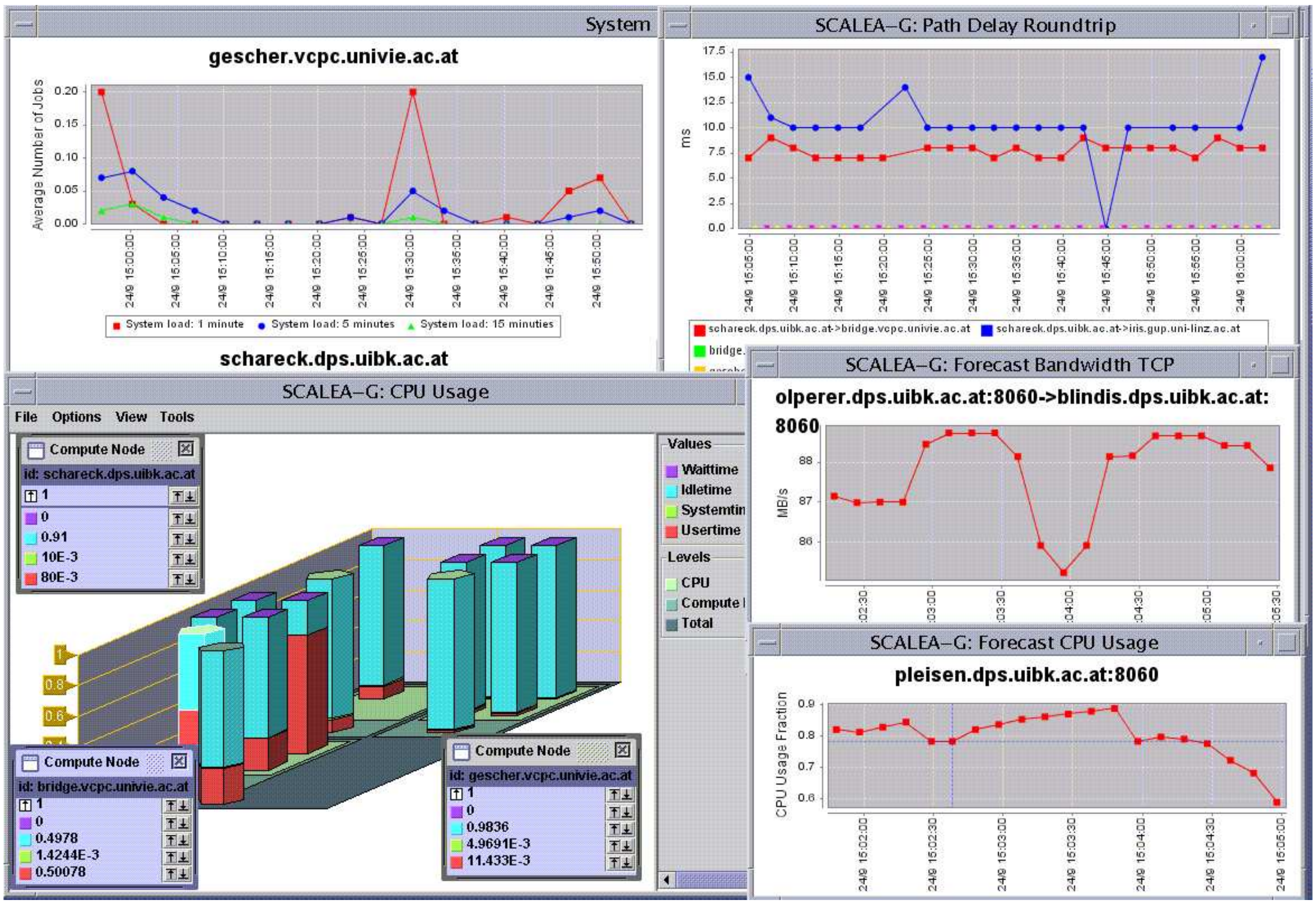
Figure 13 presents few snapshots of monitoring system load, CPU usage and network delay roundtrip (monitoring data provided by event-driven sensors), and of forecasting CPU usage and TCP bandwidth (forecasted data provided by demand-driven sensors). For example, data about CPU usage (waiting time, idle time, system time, user time) are measured per CPU. CPU monitoring data is periodically collected and the change of CPU usage can be observed on the fly through data subscription (see window *CPU Usage*). The machine `gescher` almost is idle whereas `bridge` is not fully utilized as only one of its four CPUs has high user time. Demand-driven sensors are used to obtain forecasted CPU usage and TCP bandwidth from NWS.

8. Related Work

Recently, work that exploits P2P for Grid computing has shown many advantages, e.g. in [20, 30, 21]. Our work complements the existing work as we try to integrate P2P features into a Grid monitoring middleware.

Over the past few years, many Grid monitoring and performance tools have been developed as cataloged in [18]. Several existing tools are available for monitoring Grid computing resources and networks, e.g. MDS (a.k.a GRIS) [12], NWS [35], GridRM [5], Gangila [25], GDMonitoring [11]. However, few tools have been developed for monitoring and analyzing performance of Grid applications. For example, GRM [27] is a semi-on-line monitor that supports tracing applications running in a distributed heterogeneous system. OCM-G [7] is an infrastructure for monitoring interactive Grid applications.

None of aforementioned systems, except MDS, is OGSA-based service. They support the monitoring of either infrastructures or applications while we unify both in a single system. Most existing tools either employ low level communication based on TCP streams or high level communication based on Web/Grid service. Differing from them, our middleware exploits both Grid service invocations and TCP data streams. These tools, although conduct distributed monitoring, mostly support data discovery and DQS based on hierarchical and centralized models, and event-driven sensors without rule-based monitoring. We use decentralized data storages, and support event- and demand-driven sensors, and rule-based monitoring. By featuring P2P computing, we support group-based operations for data discovery and DQS, and enhance the availability, reliability and self-managing capability of the tool.



System
gescher.vcpc.univie.ac.at
Average Number of Jobs
System load: 1 minute
System load: 5 minutes
System load: 15 minuties
schareck.dps.uibk.ac.at
SCALEA–G: Path Delay Roundtrip
ms
schareck.dps.uibk.ac.at->bridge.vcpc.univie.ac.at
schareck.dps.uibk.ac.at->iris.gup.uni-linz.ac.at
SCALEA–G: CPU Usage
File Options View Tools
Compute Node
id: schareck.dps.uibk.ac.at
id: bridge.vcpc.univie.ac.at
id: gescher.vcpc.univie.ac.at
Values
Waittime
Idletime
Systemtim
Usertime
Levels
CPU
Compute
Total
SCALEA–G: Forecast Bandwidth TCP
olperer.dps.uibk.ac.at:8060->blindis.dps.uibk.ac.at:8060
MB/s
SCALEA–G: Forecast CPU Usage
pleisen.dps.uibk.ac.at:8060
CPU Usage Fraction

Figure 13. Snapshots of online monitoring system load, CPU usage and networks.

The use of actuators to enable and configure resource management, e.g. in [28], is one aspect of using monitoring data for self-configuring. Based on rule sets, our sensors can self-manage its actions in processing monitoring data of monitored resources but we do not provide actuators/effectors that control monitored resources yet. Our rule-based monitoring also differs from the use of fuzzy-based monitoring to check performance contracts [34]. In our framework, the rule set which is built based on best practices can be embedded into the sensor for different control purposes, not only for checking performance contracts.

9. Conclusion and Future Work

In this paper we have presented a self-managing sensor-based middleware for performance monitoring and performance data integration in the Grid. Due to the diversity and dynamics of the Grid, monitoring middleware needs to unify and provide different types of monitoring sensors such as system and application sensors, event- and demand-driven sensor, to integrate various types of data from many sources. Exploiting both Grid service-based operation and TCP-based data stream can help balancing tradeoffs among interoperability, manageability and performance. Middleware must store collected data on distributed sites, providing the same mechanism for accessing that distributed data. By incorporating P2P and autonomic technologies, Grid monitoring middleware is capable of self-organization, supporting group-based data discovery and DQS. As a result, it helps increasing availability and reliability of the middleware as well as dealing with the dynamics of large distributed environments. This paper contributes on the design and implementation of a Grid monitoring middleware that exploits the above-mentioned points.

We are currently improving our prototype and investigating to port our framework to WSRF [13]. Although P2P and autonomic features give many promising solutions to solve challenges in Grid monitoring, to incorporate these features into an OGSA-based middleware is not a simple task. Still our work is just at an early stage of developing and exploiting these features as a part of a middleware for Grid monitoring and performance data integration. To continue our effort on utilizing sensor networks, P2P and autonomic computing features, the set of sensors will be extended, together with effectors, to support self-healing. We plan to provide adaptive capabilities for SM and DS so that they can self-adjust their functions under the computing capabilities of the hosting environment. Besides, we also work on supporting monitoring based on ontological performance data [31].

References

[1] Sleepcat Berkeley DB, http://www.sleepycat.com.

[2] Ian F. Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. A Survey on Sensor Networks. *IEEE Communications Magazine*, 40(8):102–114, Aug 2002.

[3] Bill Allcock, Joe Bester, John Bresnahan, Ann L. Chervenak, Ian Foster, Carl Kesselman, Sam Meder, Veronika Nefedova, Darcy Quesnel, and Steven Tuecke. Data management and transfer in high-performance computational grid environments. *Parallel Comput.*, 28(5):749–771, 2002.

[4] AustrianGrid. http://www.austriangrid.at/.

[5] Mark Baker and Garry Smith. GridRM: An Extensible Resource Management System. In *Proceedings of IEEE International Conference on Cluster Computing (CLUSTER'03)*, pages 207–215, Hong Kong, December 01-04 2003. IEEE Computer Society Press.

[6] Z. Balaton and G. Gombas. Resource and Job Monitoring in the Grid. In *Proceedings. Euro-Par 2003 Parallel Processings*, Klagenfurt, Austria, 2003.

[7] Bartosz Balis, Marian Bubak, Włodzimierz Funika, Tomasz Szepieniec, and Roland Wismüller. An infrastructure for Grid application monitoring. *LNCS*, 2474:41–49, 2002.

[8] Olof Barring. Towards automation of computing fabrics using tools from the fabric management workpackage of the eu datagrid project. *ECONF*, C0303241:MODT004, 2003.

[9] J. P. Bigus, D. A. Schlosnagle, J. R. Pilgrim, W. N. Mills III, and Y. Diao. ABLE: A toolkit for building multiagent autonomic systems. *IBM Systems Journal*, 41(3):350–371, 2002.

[10] IBM T.J. Watson Research Center. ABLE Rule Language: User's Guide and Reference. Version 2.1.0.

[11] P. Cicotti, Michela Taufer, and Andrew A. Chien. DGMonitor: A Performance Monitoring Tool for Sandbox-Based Desktop Grid Platforms. In *Third International Workshop on Performance Modeling, Evaluation, and Optimization of Parallel and Distributed Systems (PMEO), CD-ROM / Abstracts Proceedings, 26-30 April 2004,Santa Fe, New Mexico, USA*. IEEE Computer Society, 2004.

[12] K. Czajkowski, S. Fitzgerald, I. Foster, and C. Kesselman. Grid Information Services for Distributed Resource Sharing. In *Proceedings of the 10th IEEE International Symposium on High-Performance Distributed Computing (HPDC-10)*. IEEE Press, August 2001.

[13] Ian Foster el al. Modeling Stateful Resources with Web Services. Specification, Globus Alliance, Argonne National Laboratory, IBM, USC ISI, Hewle tt-Packard, January 2004.

[14] Robert Elfwing, Ulf Paulsson, and Lars Lundberg. Performance of SOAP in Web Service Environment Compared to CORBA. In *Ninth Asia-Pacific Software Engineering Conference (APSEC'02)*, pages 84–, Gold Coast, Australia, December 04 - 06. IEEE Computer Society.

[15] I. Foster, C. Kesselman, J. Nick, and S. Tuecke. Grid Services for Distributed System Integration. *IEEE Computer*, pages 37–46, June 2002.

[16] I. Foster, C. Kesselman, and S. Tuecke. The anatomy of the grid: Enabling scalable virtual organizations. *International Journal of Supercomputer Applications*, 15(3), 2001.

[17] Ian Foster and Adriana Iamnitchi. On Death, Taxes, and the Convergence of Peer-to-Peer and Grid Computing. In *2nd International Workshop on Peer-to-Peer Systems (IPTPS'03)*, Berkeley, CA, February 2003.

[18] M. Gerndt, R. Wismueller, Z. Balaton, G. Gombas, P. Kacsuk, Z. Nemeth, N. Podhorszki, H.-L. Truong, T. Fahringer, M. Bubak, E. Laure, and T. Margalef. *Performance Tools for the Grid: State of the Art and Future*, volume 30 of *Research Report Series, Lehrstuhl fuer Rechnertechnik und Rechnerorganisation (LRR-TUM) Technische Universitaet Muenchen*. Shaker Verlag, 2004. ISBN 3-8322-2413-0.

[19] Globus Project. http://www.globus.org.

[20] Wolfgang Hoschek. *Grid Computing: Making the Global Infrastructure a Reality*, chapter Peer-to-Peer Grid Databases for Web Service Discovery, pages 491–539. Wiley Press, 2003.

[21] Junseok Hwang and Praveen Aravamudham. Middleware Services for P2P Computing in Wireless Grid Networks. *IEEE Internet Computing*, 8(4), 2004.

[22] Iperf. http://dast.nlanr.net/projects/iperf/.

[23] Jeffrey O. Kephart and David M. Chess. Cover feature: The vision of autonomic computing. *Computer*, 36(1):41–50, January 2003.

[24] G. Laszewski, I. Foster, J. Gawor, and P. Lane. A java commodity grid kit. *Concurrency and Computation: Practice and Experience*, 13(643-662), 2001.

[25] Matthew L. Massie, Brent N. Chun, and David E. Culler. The Ganglia Distributed Monitoring System: Design, Implementation, and Experience. *Parallel Computing*, May 2004.

[26] D. Milojicic, V. Kalogeraki, R. Lukose, K. Nagaraja, J. Pruyne, B. Richard, S. Rollins, and Z. Xu. Peer-To-Peer Computing. Technical Report HPL-2002-57, HP Labs, March 2002.

[27] N. Podhorszki and P. Kacsuk. Design and implementation of a distributed monitor for semi-on-line monitoring of visualmp applications. In *DAPSYS'2000 Distributed and Parallel System, From Instruction Parallelism to Cluster Computing*, pages 23–32, Balatonfured,Hungary, 2000.

[28] Daniel A. Reed, Huseyin Simitci, and Y L. Ribler. Autopilot performance-directed adaptive control system. January 10 1998.

[29] Domenico Talia and Paolo Trunfio. Toward a synergy between p2p and grids. *IEEE Internet Computing*, 7(04):96–95, 2003.

[30] Domenico Talia and Paolo Trunfio. Web Services for Peer-to-Peer Resource Discovery on the Grid. In *DELOS Workshop: Digital Library Architectures*, S. Margherita di Pula, Cagliari, Italy, 24-25, June 2004. Edizioni Libreria Progetto, Padova.

[31] Hong-Linh Truong and Thomas Fahringer. Performance Analysis, Data Sharing and Tools Integration in Grids: New Approach based on Ontology. In *Proceedings of International Conference on Computational Science (ICCS 2004)*, LNCS 3038, pages 424 – 431, Krakow, Poland, Jun 7-9 2004. Springer-Verlag.

[32] Hong-Linh Truong and Thomas Fahringer. SCALEA-G: a Unified Monitoring and Performance Analysis System for the Grid. In *Proceedings of 2nd European Across Grids Conference (AxGrids 2004)*, LNCS 3165, pages 202–211, Nicosia, Cyprus, Jan 28-30 2004. Springer-Verlag.

[33] M. Tubaishat and S. Madria. Sensor networks: an overview. *IEEE Potentials*, 22(2):20–23, April-May 2003.

[34] Fredrik Vraalsen, Ruth A. Aydt, Celso L. Mendes, and Daniel A. Reed. Performance contracts: Predicting and monitoring grid application behavior. In *Proceedings of GRID 2001*, volume LNCS 2242, pages 154–165, Denver, Colorado, Nov 2001. Springer-Verlag.

[35] R. Wolski, N. Spring, and J. Hayes. The Network Weather Service: A Distributed Resource Performance Forecasting Service for Metacomputing. *Future Generation Computing Systems*, 15:757–768, 1999.