UvA-DARE (Digital Academic Repository)

Studies in Interactive Visualization

van Liere, R.

Link to publication

_Citation for published version (APA):_

**General rights**
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

_UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)_
Chapter 2

A Distributed Blackboard Architecture for Interactive Data Visualization

In this chapter the motivation, design and application of a distributed blackboard architecture for interactive data visualization is discussed. The main advantages of the architecture are twofold. First, it allows visualization tools to be tightly integrated with simulations. Second, it allows qualitative and quantitative analysis to be combined during the visualization process.

2.1 Introduction

The need for enhanced data modeling and data integration in visualization environments has been widely recognized and has been a topic of interest in recent workshops and conferences [1, 2]. Although visualization environments provide rich support for graphical modeling and rendering, many visualization researchers feel that integration of the visualization environment with the simulation environment remains a problem to be solved.

In this chapter a distributed blackboard architecture for interactive data visualization is discussed. The architecture has been used in various projects related to high performance computing [3] and computational steering [4].

Motivation This work was motivated by the following observations:

- State-of-the-art scientific visualization tools are not sufficiently integrated into simulation environments. In computational steering, users can investigate in-

---

1This chapter has been published as: R. van Liere, J.A. Harkes and W.C. de Leeuw, A Distributed Blackboard Architecture for Interactive Data Visualization. In R. Yagel and H. Hagen, IEEE Visualization '98, pages 235–244, 1998.
termediate results and can modify the computation before completion. Feedback during the investigation may be required, for example to steer computation around local minima. Users may interactively steer simulations through adjustment of an application's critical parameters. To support computational steering tight integration between simulation and visualization is required.

- Large data sets defy inspection by visualization alone. Analysis of simulation output often consists of a combination of visual inspection and numerical or statistical analysis. Scientific visualization environments lack the integration of general purpose analysis tools. Visualization combined with interactive analysis will allow the scientist to explore the data both visually and numerically. Combining the two allows the user to perform qualitative and quantitative analysis simultaneously. In addition, the output of the analysis tools can be combined with the visualization of the simulation data.

- Efficient support for very large data sets is limited. Very flexible data models have been defined that allow the import/export of data from/to the simulation environment [5, 6]. However, in order to effectively embed these data models in the visualization environment, two system related issues have to be addressed: scalability of data, and access to remote computations.

Many visualization researchers believe that state-of-the-art visualization environments do not adequately address these problems. For example, Foley and Ribarsky [7] point out that next-generation visualization environments require, amongst others, a means to bind data to geometry and a general analysis model.

**Blackboard Models** Blackboard models have been widely used in the AI community as a particular kind of problem solving model [8, 9]. The blackboard model allows multiple independent agents (usually called knowledge sources) to share information in a central store (called the blackboard). The model serves as a scheme for organizing reasoning steps and domain knowledge to construct a solution to a particular problem. For example, in a forward reasoning model, problem solving begins by reasoning forward from initial data towards a goal. In this case each knowledge source will contribute its specific knowledge towards the goal.

Knowledge is segmented into modules and a separate inference engine is provided for each module. Communication between modules is realized by reading and writing in the blackboard. The blackboard can be partitioned so that it contains regions with differing, but perhaps related, data structures. In this way applications can organize the solution space into one or more application dependent hierarchies.

The blackboard model does not explicitly specify a control component. It merely specifies a general problem solving behavior. The actual locus of control can be in the knowledge modules, in the blackboard itself, in a separate module, or in a combination
of these.\footnote{Initial proposals considered the blackboard only as a passive memory with external control modules that monitored the changes in the blackboard. Later – efficiency related – refinements integrated the control modules into the blackboard.}

The difficulty with this description of the blackboard model is that it only outlines the organizational principles. For those who want to build a blackboard system, the model does not specify how it is to be realized as a computational entity. However, given a problem to be solved, the blackboard model provides enough guidelines for sketching a solution.

The blackboard model can be used as a framework for implementing visualization environments. This will be substantiated in Section 2.2.6 after we discuss some details of the distributed blackboard architecture.

**Chapter Format** The format of the chapter is as follows. In the next sections some design details of the distributed blackboard architecture are presented. First we give an overview of the architecture, its ingredients, and its programming abstractions. Then, in Section 2.2.6 we discuss the merits of the architecture and how our original motivations are addressed. In Section 2.3, some related work that resembles the blackboard architecture is discussed. Finally, Section 2.4 shows how general analysis tools can be integrated into the visualization process using the blackboard architecture.

### 2.2 Distributed Blackboard Architecture

#### 2.2.1 Overview

A simplified overview of the architecture is shown in figure 2.1. The architecture pro-

![Figure 2.1: The user view of the blackboard architecture.](image)
Distributed Blackboard Architecture

vides an interface between a user and a simulation. It is centered around blackboard and satellite processes that produce and consume data. Satellites implement the simulation, analysis programs, geometry mapping and rendering algorithms. The purpose of the blackboard is twofold. First, it manages a database of variables. Satellites can create, open, close, read, and write variables. Second, it acts as an event notification manager. Satellites can subscribe to events that represent state changes in the blackboard. Whenever such an event occurs, it will publish the event to all subscribing satellites.

A large collection of general purpose satellites have been developed. For example, a data slicing satellite, a calculator, a data logger, a VTK satellite that provides all VTK functionality [10], etc. Also, a 3D interactive editor and rendering satellite that binds data to geometry has been developed [11].

2.2.2 Variables

The basic blackboard object is the variable, which encapsulates all information required to access a blackboard object. Variables are defined as a tuple consisting of four components: a name, a type descriptor, raw data and a list of attributes (see figure 2.2). Names uniquely identify the variable. The variable descriptor determines the type, size and layout of the data. The data component is the storage container for the raw data. Attributes are name value pairs that may be used to describe meta-data of the variable.

The underlying data model supports two composite data types:

- **Regular topology.**

  Data which has been generated from the following grid types; i.e., uniform, rectilinear and curvilinear grids. The regular topology data type is very similar to the data model supported by NetCDF [12]. In this case, the type descriptor contains all information concerning the shape and dimensionality of the variable.

- **Geometry lists.**
A geometry list is a list of geometric elements. Each element can be a simple geometric object, such as polygon, polyline, mesh, light, camera, etc. The functionality offered by the geometry list is very similar to that offered by the low level interface of P3D [13].

Future extensions to the data model will include explicit support for data with an irregular topology and scattered data. Note that these extensions do not affect the semantics of the variable, but only the expressiveness of the underlying data model.

Operations on variables are very similar to low level file operations: create, open, close, read, write, and getdescriptor. Simple set/get operations are available to manipulate attribute lists.

Two scatter/gather techniques are supported to optimize I/O operations on variables. First, many variables can be read/written simultaneously in one atomic I/O operation. Second, a comprehensive data mapping mechanism is provided to allow data to be sliced, sub-sampled and the like during the I/O operation. This allows satellites to use a different data layout compared to the data structure stored in the blackboard. The identity mapping allows a one-to-one copy between storage in the satellite and the blackboard.

### 2.2.3 Architecture

The distributed architecture consists of four building blocks: a global name manager (denoted as GNM), one or more local blackboards (LBB), one or more satellites, and typed streams.

- **Global Name Manager:** The GNM maintains the bookkeeping information of all LBBs, satellites and variables in the system. Only the variable names, descriptors and attributes are stored in the GNM. Variable data is not stored in the global name manager.

- **Local Blackboard:** A local blackboard resides on each host in the distributed environment. Local blackboards accept connections from satellites executing on the same host and other LBBs. Variable data is stored in the LBB, and is shared by all connecting satellites. Each LBB maintains a copy of the variable data. The LBB manages only those variables that are opened by the connected satellites. When a satellite opens a variable, the LBB consults the GNM to check if the same variable exists in other LBBs. If this is the case, the LBB will connect with these LBBs. A LBB-LBB connection is used to maintain variable consistency (variable consistency is addressed in the next paragraph).

- **Satellites:** A satellite is a process which communicates with its corresponding LBB. Satellites may create, open, close, and read/write variables, as well as subscribe to variable events.
An abstract satellite is shown in figure 2.3. Basically, it consists of an operator that transforms input data into output data. Control determines when this operation has to be carried out, or, in other words, when a satellite is triggered. Operators can also be controlled by additional parameters manipulated via user interface widgets.

![Figure 2.3: Interfaces to an abstract satellite.](image)

Data input and output is performed by read/writing variables. Input and output triggering is discussed in Section 2.2.5.

- Command, event and data streams: A connection between a satellite and the LBB consists of a command, an event and a data stream. Commands from the satellite to the LBB are sent over the command stream. The LBB sends events to the satellites via the event stream. Data streams are used to transport data between LBBs and satellites.

Figure 2.4 shows an example configuration of the distributed blackboard architecture. This configuration shows two LBBs and four satellites. Both local blackboards are always connected to the GNM. The local blackboards share a variable, hence, they are connected.

Satellites execute in parallel but LBBs are single threaded, so that only one satellite can access the LBB simultaneously. However, access to different LBBs is concurrent.

When a satellite writes a variable, the LBB will broadcast a mutate event to all connected satellites and LBBs that share the variable. When a satellite reads a variable, the LBB will first check if the data is up to date and, if so, will service the read request. If the data is not up to date, the LBB will first get the latest copy of the data from another LBB before servicing the read request. The details of this algorithm are very similar to the cache consistency algorithms found on cache based SMP machines.

### 2.2.4 Programming Abstractions

Satellite programmers can access the LBB using three layered APIs. Each layer provides a higher level of abstraction. Higher layers are easier to use, but provide less functionality than the lower layer.
1. The local blackboard API is a low level library interface which provides functionality for LBB communication, variable management and event management. This layer provides all details of the underlying LBB protocols. It requires detailed knowledge and is difficult to use due to the inherent parallelism in the system.

2. The high level data input output layer is built on top of the local blackboard API. Many cumbersome low level details are shielded from the user. In particular, the data input output layer hides the notion of events and has built-in support for structuring variables into sets, and support for handling efficient set I/O.

A design goal of the data input output layer was to keep the required changes to the simulation code minimal. As an example of the data input output layer, consider the following C program:

```c
simulation(float *s, float *t, int *size, float *x)
{
    int continue = TRUE;

    /* Open connection, connect and subscribe vars */
    dioOpen("borneo.cwi.nl");
    dioConnectFloat("s", s, READ);
    dioConnectInt("continue", \&continue, READ);
    dioConnectFloatArray("x", x, 1, size, UPDATE);
    dioConnectFloat("t", t, WRITE);

    /* simulation loop and update data */
```
while (continue) {
    t = t + 1.0;
    calculate\_values(t, s, size, x);
    dioUpdate();
}

dioClose();

The structure of this example, which is typical for continuous simulations, consists of two parts. First, variables are initialized. The required changes to existing source code are limited to opening and closing a connection with the blackboard and connecting the variables via the dioConnect routines. Second, a main loop is entered where time is incremented and new values are calculated. The required change to the source code is a single call to exchange data. The locations where to insert these calls are easy to find; typically at the outer level of the simulation program.

The first two parameters of the dioConnect routines are the name of the variable and its address. For the connection of arrays the number of dimensions and their sizes must also be specified. The last parameter is used by the dioUpdate routine to determine the direction of the data flow. In dioUpdate first the event stream from the LBB is checked if variables to be read or updated have changed. If so, these variables are read from the LBB. Next the values of all unread variables are written to the LBB. The net result of dioUpdate is that all connected variables have the same value in the simulation and the LBB. With these few calls the user can interact with parameters (s) of the simulation, to stop the simulation (continue), monitor its progress (t, x) or change state variables (x).

To deal with more hierarchical situations, variables may be grouped into sets. In the main loop the satellite can read and write specific sets, and wait until a particular set mutates.

3. A extensible and embedded scripting language built on top of the data input output. Scripting can be used for simple operations on variables, such as slicing and logging. The advantage of scripting is its ease of use in developing satellites.

2.2.5 Satellite Control and Synchronization

Satellites cooperate via the basic input/output mechanisms that are provided for variables. Writing to a variable will cause an event to be sent to all satellites subscribed to that variable. This mechanism is used to mediate the execution of satellites. The user can specify that if a particular variable – the input trigger variable – is mutated, the operator has to be evaluated. The action of operator evaluation is called triggering. Furthermore, the user can also specify an output trigger variable, which is to be written to each time the operator has been evaluated. Input and output trigger variables can
be linked together to mediate the execution of satellites. In general, linking two trigger variables defines a data dependency between these two variables. Linking a number of variables results in a directed graph, which we call the trigger graph.

A high level trigger scripting language and script interpreter satellite have been developed to simplify the definition of trigger variables. The task of the trigger script interpreter satellite is to manage the trigger graph. As an example, consider the trigger graph shown in figure 2.5: A slicing and dicing satellite operates on simulation output, which in turn is the input for the rendering satellite. The simulation may only compute the next step after the rendering satellite has drawn the previous frame. The rendering satellite depends on variables computed by the simulation and the dicer. The sample script to realize this synchronized configuration is:

\[
\text{simulation} \rightarrow \text{slicer} \\
\text{slicer} \rightarrow \text{dicer} \\
\text{simulation} \& \text{dicer} \rightarrow \text{renderer} \\
\text{renderer} \rightarrow \text{simulation}
\]

A trigger script is defined as a sequence of trigger rules. The syntax of each rule is:

\[
\text{rule} ::= \text{expr} \ '=>' \ \text{name} \\
\text{expr} ::= \text{expr} \ '\&' \ \text{expr} \mid \text{expr} \ '!' \ \text{expr} \mid \text{expr} \ ';;' \ \text{expr} \mid \text{expr} \ '()' \ \text{expr} \mid \text{name} \\
\text{name} ::= \text{satellitename}
\]

The trigger script interpreter satellite has been integrated in the data input output layer. Whenever the trigger script satellite interprets a trigger rule, it sets the attributes of the trigger input variable with a representation of this rule. The data input output layer uses this information to determine when a satellite is to be triggered. Note that attributes of the trigger input variable are set by the trigger script interpreter satellite and can be reset during the lifetime of the satellite.
2.2.6 Discussion

The distributed blackboard extends the centralized blackboard in many ways: it provides support for efficient data movement among heterogeneous hosts, scales for large data sets, and offers a richer data model. Although the architecture is distributed, it is important to note that the programming model is not affected, i.e., the user and programmer view of a blackboard is still a centralized data store.

The architecture has been implemented on a number of heterogeneous UNIX and NT machines. TCP/IP is used for communication between LBB's and, when possible, shared memory is used for satellite to LBB communication. Rendering satellites are available for many display types, ranging from low end laptop displays to sophisticated VR displays, such as the CAVE.

In retrospect, we believe that the architecture provides support to fulfill our original requirements:

- Integration.
  
The variable naming mechanism is used to bind data structures in the blackboard to data structures in the satellite. The event mechanism is used to maintain the consistency of these data structures. The net effect is that a two-way binding exists between data in the blackboard and data in the satellite.

  For example, a rendering satellite can utilize this by binding geometry to variables in the blackboard. A simulation can also bind the same variables. Hence, computational steering is supported.

- Qualitative vs. quantitative analysis.
  
  Integration of data structures is not restricted to simulation and rendering satellites alone, but can be used by any satellite. General purpose analysis tools can be packaged as satellites. Hence, the analysis of simulation output can be a combination of visual inspection and numerical/statistical analysis.

- Efficiency of data transport.
  
  Distributed blackboards maintain local copies of a data structure. A single event will be broadcasted when a satellite mutates the data structure. The data structure will be transported to another blackboard only when it is needed. This mechanism - called 'transport by demand' - saves bandwidth if data structures are written frequently but read only occasionally.

- Ease of use.
  
  Using the low level libraries require knowledge about event driven and parallel programming abstractions. However, higher level libraries shield all these details and allow a programmer to easily bind for instance Fortran data structures to variables in the blackboard. In this way existing simulation code can rapidly be integrated into the environment.
Also, programmers need not know that the blackboard is distributed. Abstractions for opening and manipulation variables are very similar to the familiar file handling abstractions.

2.3 Related work

Many research and development teams have designed and implemented interactive visualization environments. Giving an in depth analysis of other visualization environments is outside the scope of this chapter – see [14] for an elaborate annotated bibliography on various aspects of interactive data visualization, including interactive program monitoring and steering. Instead of an extensive overview of related work, we discuss work dealing with issues that relate to our blackboard architecture.

Williams, Rasure and Hanson [15] provide a framework to understand design tradeoffs when developing data flow based visualization systems. Data flow systems are attractive because of the similarities with the visualization pipeline: users can easily organize their simulation, filter, map and render modules in an intuitive way. However, data flow environments do not provide support to deal directly with the underlying data, except for importing/export data from a file or simulation. Hence, the integration with the underlying data is limited.

CAVEvis [16] is a distributed real-time visualization system for streaming large scalar and vector fields into the CAVE. The governing idea is to render the geometry as fast as possible in order to maintain the highest level of interactivity. It uses separately running modules to asynchronously generate the geometry data, such as modules to generate iso-surfaces or particle paths. Sequence names and time-stamps are used to gather data related to a frame, regardless of the order in which the data is received. Our blackboard architecture does not explicitly support any form of sequencing and control. Rather, the synchronization is used to provide similar functionality.

SuperGlue [17] is a programming environment for scientific visualization whose main goal is extensibility and ease of use. The approach used is to offer a very generic set of primitive data structures and a inter-language interface, which programmers use to integrate data into the SuperGlue system.

2.4 Radon Forecasting

The distributed blackboard architecture has been applied to an atmospheric transport application. In the hope that systematic simulation errors can be found, researchers are interested in comparing simulated concentrations with actual measurements. Simulation errors can arise from modeling errors, numerical errors, visualization errors, and input errors. Using our system we want to discover systematic errors that occur due to a combination of:
- Spatial errors: geographical locations of the simulated data differ from the measured data.
- Temporal errors: the simulated data differ from the measured data in time.
- Scaling errors: the simulated data is systematically higher or lower than the measured data; e.g., due to an inaccurate emission sources.

Various automated data analysis techniques have been developed that search for regions in the simulated data that fit the measured data. Details of these techniques, which are based on statistical comparison and fitting methods, are outside the scope of this chapter.

The goal of this particular case is the accurate forecasting of radioactive noble gas Radon ($^{222}Rn$) concentrations based on measured wind and emission fields. The simulated Radon concentrations were compared with measured concentrations on three islands in the Indian ocean.

**Visualization** Figure 2.6 gives an overview of the ongoing Radon transport simulation over the Indian ocean. Spot noise was used to display the wind fields. A rainbow color mapping was used to display the Radon concentrations and small colored circles showing the measured concentrations are drawn on the three sites where measured data is available. The three sites are located at: Amsterdam Island (77 deg 34'E, 37 deg 50'S), Crozet Island (51 deg 52'E, 46 deg 26'S), and Kerguelen (70 deg 15'E, 49 deg 21'S), all on French territory.

![Figure 2.6: Radon concentrations over the Indian ocean. Small colored circles show measured sites data. (see color plate on page 152)](image)

Figure 2.7 shows a sequence of snapshots of the automated point fitting process. Semi-transparent circles are drawn at points calculated by the data analysis techniques.
Distributed Blackboard Architecture

The opacity of the circles is mapped to the fit of the data. Transparent circles indicate points of poor fit; opaque circles indicate points of better fit. The left image of figure 2.7 depicts the points in an initial configuration around Amsterdam Island. A new configuration is derived by deleting a poor fitting point and taking for a new point at a random position close to the best fit point. In this way the process converges to a minimum, the area which best fits the data. The middle image shows the configuration after a number of steps. Finally the points may converge to a stable configuration, as indicated in the right image.

Figure 2.7: Comparative time sequence of converging point set. (see color plate on page 152)

A plotting satellite was used to show time series of a scalar value. The output of the plotting satellite is shown on the bottom of figure 2.8. The three plots show: the measured data at Amsterdam Island (top plot), the simulated data at the point of measurement (middle plot), and the best fit found by the analysis satellite (bottom plot).

The user may at any time also edit the set of points by dragging any point to a different location. This is useful if the user suspects other local minima in the data which may be missed by the analysis software.

Blackboard Figure 2.9 is a diagram of the blackboard and the satellites around it. The Radon simulation satellite creates a set of three variables containing the wind fields and the simulated scalar Radon field. After each time step this set will be dumped into the LBB. A reader satellite reads the measured site data from a data base and writes this data onto the blackboard.

The analysis satellite creates the variable containing the candidate points. It continuously dumps new candidate points into the LBB until a stop criterion is reached. In addition, the analysis satellite opens the variables created by the simulation. Whenever a new data set is produced it will read it, and if any of the candidate points have been mutated, it will read the new candidate points. The visualization satellite will read and display the data sets and candidate points.

Upon any variable mutation, the corresponding satellite will be triggered. After
Distributed Blackboard Architecture

Discussion  The governing idea of this example is to show how the blackboard model is used in a non-trivial setting. The application combines qualitative user actions (direct manipulation of the visualization) with quantitative analysis tools (computations of numerical algorithms). Several levels of information can be differentiated: on the lowest level computed data and measured data is available, analysis satellites consume this data to produce information of a higher level. The user, in turn, can interact with the simulation or analysis satellites as a reaction to this information.

The distributed blackboard architecture is a natural framework for solving such problems. Whenever information at a certain level is mutated, the appropriate satellite recalculates its output using the new information and mutates the next level of information. Due to the blackboard architecture, data can be shared among all satellites.

2.5 Conclusions

In this chapter we presented a distributed blackboard architecture for scientific visualization. The blackboard architecture allowed us to address two important issues concerning interactive visualization environments. First, tight integration between simulation and visualization. This is realized through the name concept of a variable which tightly binds data stored in the blackboard with data in the satellite. Second, to combine qualitative and quantitative data analysis. This is realized by allowing general
analysis satellites to operate in close cooperation with the visualization satellites.

The Radon application is very simple and should be seen as an elementary case study. In the future we plan to apply the distributed architecture to a 3D ozone simulation over the Netherlands. Here, the fitting criteria will be a volume, the chemical reactions involved in computing ozone are much more complicated, and the measured ozone data is less reliable. Nevertheless, there are signs that this type of visualization based analysis will provide added value to the atmospheric researcher [18].

Figure 2.9: Blackboard configuration of Radon application.
References


