
Xgrid Guide



Preliminary March 17, 2004



Apple Computer, Inc.
© 2004 Apple Computer, Inc.
All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, mechanical, electronic, photocopying, recording, or otherwise, without prior written permission of Apple Computer, Inc., with the following exceptions: Any person is hereby authorized to store documentation on a single computer for personal use only and to print copies of documentation for personal use provided that the documentation contains Apple's copyright notice.

The Apple logo is a trademark of Apple Computer, Inc.

Use of the "keyboard" Apple logo (Option-Shift-K) for commercial purposes without the prior written consent of Apple may constitute trademark infringement and unfair competition in violation of federal and state laws.

No licenses, express or implied, are granted with respect to any of the technology described in this document. Apple retains all intellectual property rights associated with the technology described in this document. This document is intended to assist application developers to develop applications only for Apple-labeled or Apple-licensed computers.

Every effort has been made to ensure that the information in this document is accurate. Apple is not responsible for typographical errors.

Apple Computer, Inc.
1 Infinite Loop
Cupertino, CA 95014
408-996-1010

Apple, the Apple logo, Mac, and Mac OS are trademarks of Apple Computer, Inc., registered in the United States and other countries.

Finder and Rendezvous are trademarks of Apple Computer, Inc.

Simultaneously published in the United States and Canada.

**Even though Apple has reviewed this manual,
APPLE MAKES NO WARRANTY OR
REPRESENTATION, EITHER EXPRESS OR**

**IMPLIED, WITH RESPECT TO THIS MANUAL,
ITS QUALITY, ACCURACY,
MERCHANTABILITY, OR FITNESS FOR A
PARTICULAR PURPOSE. AS A RESULT, THIS
MANUAL IS SOLD "AS IS," AND YOU, THE
PURCHASER, ARE ASSUMING THE ENTIRE
RISK AS TO ITS QUALITY AND ACCURACY.**

**IN NO EVENT WILL APPLE BE LIABLE FOR
DIRECT, INDIRECT, SPECIAL, INCIDENTAL,
OR CONSEQUENTIAL DAMAGES
RESULTING FROM ANY DEFECT OR
INACCURACY IN THIS MANUAL, even if
advised of the possibility of such damages.**

**THE WARRANTY AND REMEDIES SET
FORTH ABOVE ARE EXCLUSIVE AND IN
LIEU OF ALL OTHERS, ORAL OR WRITTEN,
EXPRESS OR IMPLIED. No Apple dealer, agent,
or employee is authorized to make any
modification, extension, or addition to this
warranty.**

**Some states do not allow the exclusion or
limitation of implied warranties or liability for
incidental or consequential damages, so the
above limitation or exclusion may not apply to
you. This warranty gives you specific legal
rights, and you may also have other rights which
vary from state to state.**

Contents

Chapter 1 [Xgrid Guide](#) 7

Requirements	7
Overview	7
How Xgrid Is Used	9
Xgrid Architecture	9
Components	10
Graphical User Interface	11
Plug-Ins	11
Networking	12
Input/Output	12
Security	13
Operating Modes	13
Local Mode	13
Cluster Mode	15
Installing Xgrid Remotely	17
Requirements	17
Installing	17
Upgrading To a New Version	19
Setting Passwords	19
Xgrid Agent Preference Values	19
Xgrid Controller Preference Values	21
Uninstalling Xgrid	21
For More Information	22

Appendix A [Installing and Using Xgrid BLAST](#) 23

Installing the blastall Executable on Agent Nodes	23
Installing Databases On Agent Nodes	24
Installing a Single Database	24
Using Fragmented Databases for Maximum Performance	25
Using Xgrid BLAST	26
Connecting To an Xgrid Cluster	26
Starting a New BLAST Job	26
Viewing the Status of your jobs	26
Viewing the Results of Completed Jobs	27
Viewing Job Information	27
Cloning a Job	27

C O N T E N T S

- Canceling a Job 28
- Deleting a Job 28
- Selecting Multiple Jobs 28
- Searching For Specific Jobs 28
- Receiving Notifications 28
- Viewing Total CPU Power 29
- Viewing Agent Node Status 29
- Customizing the Job List Toolbar 29
- Quitting Xgrid BLAST and Reconnecting 29
- Disconnecting From an Xgrid Cluster 29

Tables

Chapter 1

Xgrid Guide 7

Table 1-1 20

Table 1-2 21

T A B L E S

Xgrid Guide

Welcome to Xgrid, a new suite of Mac OS X applications designed to bring Apple's legendary ease of use to parallel and distributed high-performance computing, providing the easiest way to run compute-intensive applications. Xgrid enables administrators to group locally networked computers or **nodes** into **clusters** or **grids** and allows users on the network to remotely submit long-running computations as **jobs** to the clusters.

Xgrid then creates multiple tasks for each job and distributes those tasks among multiple nodes, which can be either multipurpose desktops or dedicated cluster nodes. Harnessing desktop systems into a computational grid is known as **desktop recovery**, in that it allows researchers to use what would otherwise be idle time on these processors to run batch and workload processing. Xgrid's support for both desktop recovery and dedicated nodes enables you to use all your available resources for a problem, with a minimum of time and money.

This document provides a high-level overview of Xgrid and its use as of Technology Preview 2.

Note: See the Release Notes document for the most recent information on this release of Xgrid.

Requirements

Xgrid requires Mac OS X 10.2.8 or later (including 10.3), with a minimum of 128 MB RAM, but 256 MB is recommended.

Overview

In the most basic sense, a grid is a group of computers working together on a single problem. Although there is no canonical definition of a grid, the systems in a grid are most often loosely coupled, geographically dispersed and, to some extent, heterogeneous. One well-known example of a successful grid computing project is the University of California at Berkeley's [Seti@Home project](#), which allows computer users around the globe to donate processing time for analyzing radio telescope data on behalf of the project team's search for extraterrestrial intelligence. This donated processing time would otherwise go unused during much of a typical day.

Xgrid has no real limitations on the amount of computational power it can support. The performance of the grid is dependent on the systems participating, the software running, and the network, among other factors. When set up optimally, and for easily parallelizable, loosely coupled algorithm implementations, a number M of 1 GHz computers running Xgrid can work together for an effective clock speed of nearly M GHz.

There are many proprietary projects that allow you to participate in a large grid project. Often these projects, as in cases such as Seti@Home and [FightAids@home](#), are tied to a specific scientific purpose. They often have easy-to-install software packages that enable any user to participate in that particular project, and they often take the form of a screen saver or background process.

But you don't need to think in terms of thousands or millions of rarely-used computers to see the significance of a grid. For example, in a public computer lab in a university, or a company's office building full of desktop computers, the computers are used by students or employees for less time during the work day than the time they sit idle at night and on weekends.

Other grid projects such as the [Globus Alliance](#) are designed for large-scale computational grids with flexible resource management tools and more intelligent grid deployment methods. Instead of developing neatly packaged applications for a specific grid, such projects provide comprehensive frameworks for grid deployment.

Xgrid allows users to easily install a client to participate in a computational grid of their choice, as in the Seti@Home model, while still providing the flexibility of a more generic framework to grid developers in deploying their grid applications, as in the Globus Alliance model. While some compromises are made when compared to either model, Xgrid provides the primary benefits of both:

- Ease of configuration. Easier to set up than other enterprise clustering solutions.
- Ease of use.
 - Does not require clients to use the command line (but that capability is there for those who prefer it).
 - Hides complex architecture for software and data distribution, job execution, and result aggregation.
 - Uses Rendezvous for automatic discovery of available Xgrid controllers
- Security. Supports the UNIX security model.
- Extendibility. Provides a plug-in architecture that allows developers to extend the functionality of Xgrid clusters.
- Broad applicability. Aims to support all grid computational needs, not just those of bioinformatics, fluid dynamics, or another specific science.

How Xgrid Is Used

Ad Hoc Participation

In an ad hoc grid, users are permitted to add any compute resources to the grid they choose, from anywhere, even if those resources are not controlled by the organization that benefits from the grid. With Xgrid installed on their systems, they can join a grid and participate at will, simply by using the computer. Particularly with laptops, which physically connect and disconnect to and from networks often, the grid can change in size and nature very rapidly.

Imagine an office building filled with Mac OS X computers. During the day the office is busy with people using their computers to create documents, perform research, design new products, and so on. When the work day ends, the computers would be left idle until the employees returned the next work day.

But with Xgrid installed on all of the computers, as the workday comes to an end, the Xgrid screen saver appears on the screens of the computers of the people who have left for home. Each screen saver displays a tachometer indicating the total current computational power of the grid. As more and more computers' screen savers are activated, the needle on the tachometer climbs higher and higher.

As enough resources become available to perform the required calculations, those computers start working on a large engineering simulation that was submitted to the grid earlier in the day. When each computer finishes its portion of the job, it returns the results and waits for the next task to be assigned. If an employee returns to work before her computer has finished the task assigned to it, then as soon as she moves the mouse or presses a key, the screen saver disappears, and the task is suspended to wait for its next opportunity to run. If there are no available computers left because all of the office workers have returned to work, the jobs submitted to the grid remain suspended until the required resources become available again.

Dedicated Grid

Some organizations require the frequent and rapid computation of large calculations, often with large data sets. Computational biology, image rendering, and engineering simulation are areas where such large scale computation is often needed. In order to augment its compute power, an organization might purchase servers or other systems whose sole purpose is to provide a computational resource to desktop users.

For such organizations, Xgrid can quickly be installed and configured as a **dedicated grid**, such that it is always running. With a set of dedicated compute elements, users can submit jobs to the grid for immediate processing, rather than waiting for those large jobs to complete only during idle time.

Xgrid Architecture

This section provides an introduction to the key architectural features of Xgrid.

Components

An Xgrid cluster comprises three main software components:

- An **agent** runs one task at a time per CPU, in either dedicated mode or screensaver mode.
- A **controller** queues tasks, distributes those tasks to agents, and handles failover.
- A **client** submits jobs to the Xgrid controller in the form of multiple tasks.

A user interacts with the grid via the client. The client uses Rendezvous or an IP address/hostname to find a controller and BEEP to submit a **job**—a collection of execution instructions that may include data and executables. The controller accepts the job and its associated files, and then connects to the agents. Agents accept the jobs, perform the calculations, and return the results to the controller, which aggregates them and returns them to the appropriate client.

In principle, all three components can run on the same computer, but it is often more efficient to have a dedicated controller node.

Client

An Xgrid user submits a job to the cluster server via the Xgrid client application, using either the command-line (`xgrid`) or graphical (Xgrid.app) user interface. The user defines the parameters for the job to be executed in the Xgrid client, and these are sent to the controller. When the job is complete, the client is notified and can retrieve the results from the controller.

Any system can be an Xgrid client provided it has the Xgrid application installed and has a network connection to the Xgrid controller system. In general, the client can connect to only a single controller at a time.

Controller

The Xgrid controller service (GridServer process) manages the communications and the resources of the clusters. The GridServer process accepts network connections from clients and agents. It receives job submissions from the clients, breaks the jobs up into tasks, dispatches tasks to the agents and provides feedback to the clients.

Although there may be multiple controllers on a single physical subnet, there can be only one controller per logical grid. There is no theoretical limit to the number of agents per controller, but there may be a practical limit based on network bandwidth and the extent to which a given problem can be parallelized. Currently, a controller can handle no more than 10,000 connections.

Agent

The Xgrid agents handle running the computational tasks that comprise a job. When an agent (GridAgent process) starts running at startup it registers with a controller, which sends instructions and data to the agent when appropriate. An agent can be connected to only one controller at a time. Once the instructions from the controller are received, the agent then executes the appropriate code and sends the results back to the controller.

The agent systems can either be dedicated to the cluster (in **dedicated mode**, as described above) or they can be desktop systems that only become available when the system is idle. By default, the agent is configured initially to accept tasks only when the system has had no user input for 15 minutes, has the Xgrid screen-saver running, or the screen has been put to sleep.

Each task contained within a job can have a different set of arguments, so that each task run on each agent can have a different set of arguments. For example, you might wish to distribute to the grid a set of tasks with a range of numeric arguments:

Host 1: ./a.out -begin 1 -end 10

Host 2: ./a.out -begin 11 -end 20

Host n: ./a.out -begin i -end j

Xgrid can execute anything that can be executed from a shell, including scripts, applications, and commands. Output written to `stdout` or `stderr` is captured and returned to the controller and, hence, the client, along with any changed files in the working directory.

Graphical User Interface

There are four user interface components associated with Xgrid:

- /Applications/Xgrid.app, a GUI client for creating plug-ins and submitting jobs
- /usr/bin/xgrid, a command-line client for submitting and monitoring jobs
- The Xgrid Pane in System Preferences for configuring an agent (through /Library/Preferences/com.apple.xgrid.agent.plist) or controller (through /Library/Preferences/com.apple.xgrid.controller.plist)
- The Xgrid screensaver module in System Preferences, for setting idle time

Note: There is currently also a custom GUI client for XgridBLAST.app. At this time, there is no command-line client available for XgridBLAST.

Plug-Ins

The specific functionality of the Xgrid client application is provided by plug-ins, which have two roles:

- allowing the user to specify parameters, submit jobs, and retrieve results
- encapsulating resources to be distributed to the agents with the jobs

There are several types of plug-ins currently available with Xgrid:

- built-in plug-ins for a specific task (for example, Mandelbrot)
- generic plug-ins that can run multiple types of jobs (for example, Shell and Xfeed)
- the Custom plug-in, which can create an interface for an arbitrary CLI application

In the Developer folder of the Xgrid install folder, there is also:

- a source code example of the Shell plug-in (including ShellJobViewController)
- an Xgrid Shell Plug-in template

When a custom plug-in is created by the client, it actually includes a tarfile with the specified executable. The plug-in is sent to the controller and distributed to each agent each time it is given a new task. This means that you never have to install any software on the agent nodes or set up a distributed file system.

Networking

All network communication is handled via the [BEEP protocol](#). BEEP is an IETF proposed standard (RFC 3080) for asynchronous, multiplexed, peer-to-peer communication (as opposed to the client-server HTTP). The Xgrid controller supports three custom BEEP profiles. The profiles that describe the protocols used to communicate between the agent, controller, and client are based on the exchange of XML-encoded property lists, and thus are fully standards-based and platform-neutral.

In Xgrid, the controller is the only component that listens (and thus requires a privileged port; currently it uses port 4111, though this is likely to change in the future). This means the controller is also the only one that advertises on Rendezvous (as `beep.tcp`, due to the underlying protocol).

A client looks on the local subnet for a controller, but can also be assigned a specific controller elsewhere on the Internet. An agent can either look on the subnet to find a controller, in which case it will bind to the first one that has the correct password, or be assigned a specific service name or hostname/IP address. Note that if you want to use Xgrid through a firewall, you need to open port 4111 and specify an explicit controller. After connecting, Xgrid will open other ports for traffic as necessary, but those will be non-privileged.

Input/Output

The Xgrid client allows you to specify command-line arguments for the various tasks. If you specify a File (or File List) argument, Xgrid copies the relevant files into the working directory on each node. The standard output (`stdout`) and standard error (`stderr`) are written to a single file, which—along with any generated data files in the working directory—is copied by the controller back to the client's output directory.

WARNING

Files with the same name will be overwritten.

Note that the relevant plug-in and data are copied to each agent for each new task, even if it is part of the same job. This can greatly increase network traffic, so large binaries or datasets should instead be pre-installed on the agent nodes, and invoked by a shell script or custom-written plug-in.

Security

Xgrid by default requires a password for connections. A client must supply a password to connect to the controller, and the controller must supply a password to connect to each agent. The protocols are currently implemented to first attempt connecting without authentication then, if that fails, to try authenticating. Xgrid currently uses a two-way-random mutual authentication protocol that includes the use of MD5 hashes. Currently Xgrid implements data encryption only for passwords, but there are BEEP profiles for TLS (SSL) and SASL (authentication) which may be adopted in the future.

The Xgrid daemons are designed to run as the unprivileged user 'nobody', which means they can only write into /tmp. They make use of helper utilities to perform the few privileged operations required. The usual procedure is for the daemon to launch as root, then give up privileges and invoke the helper as necessary. This is designed to simplify auditing and minimize the potential for attacks, although since the agents never listen on a port (they simply maintain a connection to the controller) risk of attack is low.

Operating Modes

Xgrid can be run in two distinct modes. One mode is the normal **cluster mode**, as described in the previous section. The second mode is designed to allow a demonstration of the Xgrid technology on a single computer. This is referred to as **local mode**, because all of the components of Xgrid are running on the same system. At any time it is possible to run the Xgrid client in local mode independent of whether the system is part of a cluster.

Local Mode

This mode is intended for use as a demonstration of the interface and capabilities of Xgrid. A single Mac OS X version 10.2.8 or later system is required to run Xgrid in local mode. This system will act as client, controller, and agent.

We recommend that first time users of Xgrid install Xgrid on a single system and run the demonstration jobs in local mode. This will allow you to see how the software works without having to dedicate several systems to the evaluation.

Installing Xgrid for Local Mode Use

To install Xgrid, ensure that you have an administrator account name and password for the local machine and simply perform the following steps:

1. Mount the Xgrid installer disc or disc image. Note that installing Xgrid from network-mounted volumes is not supported.
2. Double-click the Xgrid installer. When requested, enter the administrator account name and password.
3. Follow the on-screen installer instructions. Xgrid can be installed only on the root volume of the system. All other volumes will be marked as unavailable.

The Xgrid application is installed in your Applications directory, and the documentation is installed in /Library/Xgrid/Documentation.

Running Demonstrations on a Single Computer

When the Xgrid installation is completed, the agent and controller can be configured and turned on using System Preferences, and you are ready to run Xgrid in local mode.

Local mode is suitable for testing and running demonstrations using only the local computer. To use Xgrid with multiple computers, see ["Cluster Mode" \(page 15\)](#).

Starting Local Mode

To start local mode:

1. Open the Xgrid application by double-clicking its icon in the Finder.
2. In the login panel that appears, click the Start Local Service button.

Running the local mode demonstrations

The following steps describe how to run the local mode demonstrations.

1. Double-click the Xgrid client application icon in the Applications folder. Once the Xgrid client is running it will search for Xgrid controllers on the local network using Rendezvous. At the same time it displays the Xgrid Login window. The client status is displayed in the bottom left corner of the window.

Even while the Xgrid client is searching for network based grid services it is possible to click on the "Start Local Service" button. This starts controller and agent processes on the local machine.

If you have not already set up an Xgrid cluster on your local network, the Xgrid client will not find any grid services. If no grid services have been discovered after five seconds it will display an informational sheet. This is to be expected. Dismiss the sheet by clicking the "OK" button.

2. Click the Start Local Service button in the Xgrid Login window.

Once the controller and agent processes have been successfully launched, the Xgrid client will display a list of available clusters. In the local mode the only cluster that will be listed is `localhost`. Selecting `localhost` will then display a list of job types that the cluster can perform. There are five types of job available in this version of Xgrid.

3. Select the Mandelbrot job type and click the "OK" button.

A new job window opens, and a tachometer will appear. The cluster immediately starts calculating the Mandelbrot set. As each part of the calculation is finished, the client application draws the corresponding part of the image to the screen. You can stop the continuous, automatic calculation of the image by clicking on the Stop button in the Automatic tour box.

You can have more than one job window open at once. Choose New from the File menu to open a new cluster and job type browser. A new job window will be opened once you select a cluster and job type and click on the OK button.

You can get more information about a job type by choosing Show job help from the Help menu or by clicking the Show help button while a job window is open. A new window will open displaying help about the job.

A tachometer showing the CPU power of the grid can be displayed by selecting Toggle tachometer from the Window menu or by typing the Command-1 key combination. In local mode, the tachometer will show a value equal to the speed of the local computer, but only when Xgrid is working on a job.

Cluster Mode

A minimum of two Mac OS X version 10.2.8 or later systems are required to run Xgrid in cluster mode. With only two computers, one will act as the client system, and the other will act as both controller and agent. Any number of additional computers can also be agents.

Once you have a good understanding of how Xgrid works and what can be achieved with it, try creating a two-system cluster. It is relatively trivial to expand a two-system cluster by adding additional agents. The details for installing and configuring a cluster of two or more systems are explained in the following section. For more information about installing Xgrid on remote computers using ssh and the command line, see ["Installing Xgrid Remotely" \(page 17\)](#).

Installing and Configuring for Cluster Mode

A system that has Xgrid installed for the local mode demonstration can now become an Xgrid client system. Additionally, to install and configure the necessary controller and agent on a single separate system use the following instructions.

The agent and controller can be configured and turned on using System Preferences.

1. Install the Xgrid package on the second system in the same manner as described above for local mode. This installs all of the necessary system components.
2. Use the System Preferences panel to configure the controller service and agent service. At the very least, you must turn off the password requirements and start the controller and agent.
3. To use Xgrid securely, you must first configure three passwords:
 - ❑ Using the Agent Security preference pane tab, secure the agent by configuring the it to require a password from the controller it connects to. This way you can restrict who is allowed to use your computer's computational resources.
 - ❑ Using the Controller preference pane tab, configure the controller so that it knows the password to provide to the agent. Set the same password as you set for the agent.
 - ❑ Using the Controller Security preference pane tab, configure the controller to require a password from each client that connects to it. This way you can restrict who is allowed to use the cluster's computational resources.

4. You may at this time also choose to configure the agent to always accept tasks—that is, not only when the hosting computer is deemed idle. To do so, choose the Agent tab and click the appropriate button.

Advanced User Tips

By default, agents search for a bind to a controller on the local network. If you are using remote agents, or have more than one controller on this subnet, you can instead choose to discover the controller using a specific hostname/IP address in the Agent preference pane.

The agent is set by default to accept new tasks from the server only if the computer has been idle for 15 minutes or the Xgrid screen saver is running. If you have dedicated nodes, you will want to choose to always accept tasks.

Running in Cluster Mode

Having configured Xgrid, you can use the Agent and Controller preference pane tabs to activate and run the agent and controller services. Once all of the systems are configured and you have started the services, the agents and server will automatically find each other by using Rendezvous. Rendezvous allows agents and clients to find a controller on the same subnet without any user intervention. The cluster is now ready to start accepting jobs. To test the cluster,

1. Start the Xgrid application on the client system.

The login window appears. After a short period of time, the client should detect the controller and automatically fill in the Service field.

2. If you configured the controller to require a password, enter the password in the appropriate field.
3. Click Connect.

A New Job window appears, showing a single cluster named Rendezvous (as opposed to localhost, as in local mode).

4. Double-click the Rendezvous cluster name.

A new Cluster Nodes window appears, displaying the names and status of the agents in the local area network currently or formerly bound to the controller service to which the client is currently connected.

Agent status is depicted visually in the matrix below the table. The possible states are "Offline", "Unavailable", "Available", and "Working". Agents that are hosted by computers which are either off or sleeping will be shown "Offline". Agents that are hosted by computers that are not idle (meaning they are in use by a human at the console) are shown as "Unavailable". Otherwise, agents are shown as "Available", unless they are currently performing a task for the controller, in which case the agents are shown as "Working".

5. Return to the New Job window behind the cluster nodes window, select the Mandelbrot job type, and click the "OK" button.

The Mandelbrot interface appears, and the job starts immediately.

If the job does not start and the status bar at the top of the window says that the job is pending, it is because the computer you installed the agent on is not currently idle. The Xgrid agent will only accept tasks from the Xgrid controller when the computer hosting the agent is idle. Until the first task of a job is started, the job is considered pending.

To make the job start, you must allow the computer hosting the agent to become idle, or you may activate the screen saver. The host computer is considered idle if the mouse and keyboard have not been touched for more than 15 minutes.

You can add additional nodes to your cluster by installing Xgrid on more computers and configuring and starting the agent service.

Xgrid Screen Saver

The Xgrid screen saver displays the tachometer, which depicts a value representing the CPU power of the grid to which the agent is attached.

To enable the Xgrid screen saver, go to System Preferences, click on the "Screen Effects" or "Desktop and Screen Saver" icon, and then select "Xgrid" from the list of screen savers. You may also need to make sure that screen savers are set to activate before your monitor is set to sleep.

If the Xgrid agent is running on the same computer as the Xgrid screen saver, the screen saver notifies the agent when it starts running. While the screen saver is running, the agent considers the hosting computer idle and accepts tasks from the controller. However, as soon as the screen saver wakes up or you move the mouse, the agent is notified and it suspends any running tasks (unless Choose to always accept tasks has been set in the Preferences pane). The tasks will remain suspended until the computer becomes idle again.

Installing Xgrid Remotely

This section describes how to set up an Xgrid on a remote computer using the Terminal.

Requirements

For each remote computer you want to install Xgrid on, you will need:

- An administrator account on that remote computer.
- The internet address (name or IP address) of the computer you'll be installing Xgrid on.
- Remote Login enabled (in the Sharing preferences on the remote computer).

Installing

To install Xgrid, perform the following steps:

1. Open a Terminal window.

2. Copy the Xgrid package to the remote computer.

Assuming the file Xgrid.pkg is at /Volumes/Xgrid1.0/Xgrid.pkg, enter the following command into the Terminal, substituting correct values for 'user' and 'remote-host.com':

```
scp -r /Volumes/Xgrid1.0/Xgrid.pkg user@remote-host.com:
```

(Note: Don't leave out the colon at the end of the line; it's required.)

3. Enter the user's password if prompted to do so by ssh.

4. Log in to the remote computer:

```
ssh user@remote-host.com
```

5. Install the Xgrid package on the remote computer:

```
sudo installer -pkg $HOME/Xgrid.pkg -target /
```

(WARNING: The installer command requires that you specify the pkg using a full pathname. If you use a relative path to specify the Xgrid.pkg file, the installer will appear to succeed but Xgrid will not be installed.)

6. Start the computer as an Xgrid Agent or Xgrid Server

- ❑ To use the remote computer as an Xgrid agent, that is, to have it perform jobs as part of the cluster, enter both the following commands while logged into the remote computer:

```
sudo /Library/Xgrid/Scripts/agent_on
sudo /Library/Xgrid/Scripts/agent_start
```

- ❑ To use the remote computer as an Xgrid server, which controls how all the computers in an Xgrid cluster communicate with each other, enter the following commands while logged into the remote computer:

```
sudo /Library/Xgrid/Scripts/server_on
sudo /Library/Xgrid/Scripts/server_start
```

(Note: Each Xgrid cluster only needs one machine set up as an Xgrid server. If you've already set up one machine as a server, there's probably no need to set up any more as servers. A computer set up as an Xgrid server may or may not be set up as an Xgrid agent as well. For more information about the server/ agent distinction, see the document ReadmeFirst.pdf which came with this Xgrid distribution.)

7. Make the computer join an Xgrid cluster

If the remote computer is set up as an agent and is on the same local subnet as the computer which is set up as a server, the machine on which you've just installed Xgrid should discover the Xgrid cluster and join it automatically.

If the remote computer is not on the same subnet as the server (for example, if you're adding a computer in a different building to your Xgrid cluster), you'll have to tell the agent the exact address of the Xgrid server computer by modifying the Xgrid preferences .plist file.

You must also set a password as described in ["Setting Passwords" \(page 19\)](#)

Upgrading To a New Version

If you are installing a newer version of Xgrid over an older, already running version, there's a specific order you need to follow to shut everything down. If not done in this order, communication may break down between agents and servers, and some settings may be corrupted.

First stop each agent. Next stop the controller. Install over the old controller, and start the controller. Finally install over each old agent and start agent.

(WARNING: Upgrading overwrites your old preferences.)

Setting Passwords

By default, Xgrid agents and controllers require passwords. If passwords are required, but have not been set, the agents and controller will refuse to allow connections. The following files contain the necessary password information:

```
/etc/xgrid/GridAgent/controller-password
/etc/xgrid/GridServer/agent-password
/etc/xgrid/GridServer/client-password
```

To create these files, you must use the preference pane at least once. Once they have been created with the proper password information, copy the files to the remote computers. Be sure to restore the ownership and permissions to root.wheel and 0600.

Xgrid Agent Preference Values

You may want to configure the remote computers on which you've installed the Xgrid agents more specifically than just turning them on or off. To do this, you'll need to change values in the file:

```
/Library/Preferences/com.apple.xgrid.agent.plist
```

This file is managed by the Xgrid Preferences pane, but you can also edit it manually using Property List Editor.app (if Developer Tools are installed) or any XML or text editor. The keys and value types in the file `com.apple.xgrid.agent.plist` are:

Table 1-1

Key	Value Type
AgentName	String
BindToFirstAvailable	Boolean (TRUE or FALSE)
OnlyWhenIdle	Boolean (TRUE or FALSE)

Key	Value Type
ServiceName	String
SuspendWhenNotIdle	Boolean (TRUE or FALSE)
UsesRendezvous	Boolean (TRUE or FALSE)
RequireControllerPassword	Boolean (TRUE or FALSE)
MaximumTaskCount	Integer (1 or 2)

These values are affected by the settings in the Xgrid preference pane (found in System Preferences). If you want to change an agent's settings remotely, though, you'll need to change these preferences by hand. Here are some useful settings. You can combine them; the settings below are not necessarily mutually exclusive:

- *Server on different subnet than agent:* Rendezvous won't be able to help the agent find the Xgrid server, so the address will need to be specified explicitly.

```
<key>UsesRendezvous</key>
<false/>
<key>BindToFirstAvailable</key>
<false/>
<key>ServiceName</key>
<string>xgrid-server.com</string>
```

- *Dedicated cluster:* The agent will take new Xgrid jobs regardless of what other activity is going on on the system.

```
<key>OnlyWhenIdle</key>
<false/>
```

- *Jobs likely to be interrupted often:* If you're likely to run jobs that could last several hours or days and may be interrupted by someone using the agent, then it's better to have any running jobs suspend themselves rather than simply quit, which is the default behavior. This prevents the agent from throwing away hours of processor time every time it's interrupted.

```
<key>SuspendWhenNotIdle</key>
<true/>
```

- *Multiple servers on the same subnet:* This setting would allow you to have several separate Xgrid clusters running on the same Rendezvous-enabled subnet. The service name of a controller is the same as the computer name set in the Sharing preference on the computer hosting the controller.

```
<key>UsesRendezvous</key>
<true/>
<key>BindToFirstAvailable</key>
<false/>
<key>ServiceName</key>
<string>An Xgrid service</string>
```

- *Agent computers without unique Rendezvous names:* If you use Agent computers that all have the same computer name, then by default only one of them will be able to successfully connect to a given controller at a time. You can override the default behavior of using the computer name for the agent name and specify a unique name for each agent by using the AgentName key:

```
<key>AgentName</key>
<value>unique-name</value>
```

Xgrid Controller Preference Values

You may want to configure the remote computer on which you've installed the Xgrid controller more specifically than just turning it on or off. To do this, you'll need to change values in the file:

`/Library/Preferences/com.apple.xgrid.controller.plist`

This file is managed by the Xgrid Preferences pane, but you can also edit it manually using `Property List Editor.app` (if Apple Developer Tools are installed) or any XML or text editor.

The keys and value types in the file `com.apple.xgrid.controller.plist` are:

Table 1-2

Key	Value	Type
UsesRendezvous		Boolean (TRUE or FALSE)
RequireClientPassword		Boolean (TRUE or FALSE)

Some of these values are affected by the settings in the Xgrid preference pane (found in System Preferences). If you want to change a controller's settings remotely, though, you'll need to change these preferences by hand. The following settings may be useful:

Running an Unadvertised Controller

You may choose to run your controller without advertising its presence using Rendezvous. This would be useful if you wanted to run a cluster without getting connections from the agents on the local subnet set to bind to the first available controller.

```
<key>UsesRendezvous</key>
<false/>
```

Uninstalling Xgrid

To uninstall Xgrid, you must be logged in as an administrator. From the Finder, double-click the `XgridUninstall.command`

script included on the disk image with the installer. From the command-line, type `/Library/Xgrid/Scripts/XgridUninstall.command` (or the full pathname of the file if it has been moved). The script immediately starts running, opening a new Terminal window if necessary. Enter your password when prompted.

For More Information

If you encounter any problems using Xgrid, please report them at <http://bugreport.apple.com>.

Apple websites of interest include

www.apple.com/acg/xgrid—Xgrid product page

www.apple.com/acg—Advanced Computation group

www.apple.com/xserve/cluster/resources.html—Third-party cluster resources

lists.apple.com/listinfo/xgrid-users—Clusters mailing list

lists.apple.com/listinfo/scitech—Scitech users mailing list

Clustering community sites of interest:

Note: Apple cannot guarantee the accuracy of information on external sites.

<http://www.macos.utah.edu:16080/xgrid/>

[http://unu.novajo.ca/simple/archives/000024.html/](http://unu.novajo.ca/simple/archives/000024.html)

<http://www.newsforge.com/hardware/04/01/12/1628240.shtml>

<http://www.computerworld.com/softwaretopics/os/macos/story/0,10801,88989,00.html?f=x10>

Installing and Using Xgrid BLAST

Xgrid BLAST is a popular open source biotechnology application used by life science researchers to find matches in DNA and protein sequences. This document describes how to install BLAST databases on Xgrid agent computers and how to operate the Xgrid BLAST application.

Installing the blastall Executable on Agent Nodes

Before a user of Xgrid BLAST can run query sequences against DNA or protein databases, you will need to install the `blastall` executable on each agent computer.

Note: The `blastall` executable must be installed on every computer in the cluster with which you wish to use Xgrid BLAST. Otherwise, your jobs may fail unexpectedly.

To install `blastall` onto an agent computer, perform the following steps using the Terminal:

1. Verify that Xgrid is installed on the agent computer.
2. If the agent node is a remote computer, log in to the remote computer:
`ssh user@remote-host.com`
3. Create the directory `/Library/Xgrid/blast` if it does not exist:
`sudo mkdir /Library/Xgrid/blast`
4. If you do not already have a copy of your preferred version of `blastall`, you can obtain a copy of the A/G BLAST `blastall` executable:
`curl ftp://ftp.apple.com/developer/Tool_Chest/AGBLAST/blastall.gz - gunzip > blastall`
5. Copy the `blastall` executable from the current working directory to the directory `/Library/Xgrid/blast`:
`sudo cp blastall /Library/Xgrid/blast/`
6. Make sure the `blastall` file has its executable permissions set:
`sudo chmod a+r /Library/Xgrid/blast/blastall`

7. If you wish to perform protein searches, you will need to obtain the standard BLOSUM62 scoring matrix and place it in the directory /Library/Xgrid/blast:
`sudo curl http://www.ncbi.nlm.nih.gov/Class/FieldGuide/BLOSUM62.txt > /Library/Xgrid/blast/BLOSUM62`

Installing Databases On Agent Nodes

Before a user of Xgrid BLAST can run query sequences against DNA or protein databases, you will need to install these databases on the Xgrid agent computers. It is not necessary that all agents contain the same set of installed databases; Xgrid will automatically delegate queries against a particular database only to those agents that have the database installed. A user of Xgrid BLAST will be presented with a list of databases installed on at least one agent in the cluster.

Installing a Single Database

The following instructions describe how to install a BLAST database on a remote agent node so that the agent can use the database. For this example, a database named *human_genomic*, pre-formatted using the program `formatdb`, is assumed.

1. Make sure Xgrid has been installed on the agent computer
2. Log in to the remote computer:
`ssh user@remote-host.com`
3. Create the directory /Library/BLAST if it does not exist:
`sudo mkdir /Library/BLAST`
4. Create a directory in /Library/BLAST/ for the new database files:
`sudo mkdir /Library/BLAST/human_genomic`
5. Copy the database files *human_genomic.nin*, *human_genomic.nsq*, etc. in the current working directory on the remote computer into the newly-created directory:
`sudo cp human_genomic.* /Library/BLAST/human_genomic/`
6. Inform Xgrid about your new database directory:
`/usr/libexec/xgrid/DatasetTool -source /Library/BLAST/human_genomic -userInfo '{databaseName="human_genomic";}'`
7. Repeat steps 4-6 for any additional databases you wish to install at this time.
8. Restart the Xgrid agent process:
`sudo /Library/Xgrid/Scripts/agent_stop`
`sudo /Library/Xgrid/Scripts/agent_start`

Note: If a new database is installed on an agent computer, a user of Xgrid BLAST must quit and reopen the application to cause the new database to appear in the list of available databases.

Using Fragmented Databases for Maximum Performance

For maximum performance, it is often desirable for long-running searches to be executed on multiple agents simultaneously. In order to allow Xgrid to utilize multiple agents simultaneously for one search, the database to be searched must first be formatted into multiple fragments. Such formatting can be accomplished by using NCBI's `formatdb` tool with the `-v` option.

Before you can create a fragmented database, you need to obtain NCBI's `formatdb` utility and an unformatted BLAST database. The `formatdb` utility can be obtained by visiting <http://www.ncbi.nlm.nih.gov/blast> and downloading the Mac OS X package of BLAST.

You may already have an unformatted FASTA database. In the following example, we will download the very large, compressed FASTA `human_genomic` database from NCBI, uncompress it, and format it into multiple fragments. In this case we have placed the `formatdb` tool in the current working directory:

```
curl ftp://ftp.ncbi.nih.gov/blast/db/FASTA/human_genomic.gz | gunzip |  
./formatdb -i stdin -p F -o T -v 50 -n human_genomic
```

Note: Note: The `-p F` flag informs `formatdb` that your database is DNA instead of protein, and the `-v 50` attempts to limit the maximum size of the outputted fragments, thus creating multiple fragments. For more information about the `formatdb` command please see the NCBI documentation included with BLAST.

When determining the number of fragments your database has been split up into, you will need to look at the files created by the `formatdb` command in this step; for example, if the above command creates 25 fragments, with file prefixes ranging from `human_genomic.00` up to `human_genomic.24`, then you would enter 25 as the "numberOfFragments" value as shown in the following example.

Simply creating fragmented formatted database files is not enough to get Xgrid to use them. You must first inform Xgrid that your newly-installed database is formatted into multiple fragments. To do so, add a "numberOfFragments" key to the "userInfo" dictionary in the arguments passed to the `DatasetTool`. For example, supposing your "human_genomic" database had been formatted with the `-v` option and the resulting formatted files were split up into 25 fragments, you would run the command:

```
/usr/libexec/xgrid/DatasetTool -source /Library/BLAST/human_genomic  
-userInfo '{databaseName="human_genomic_fragmented";numberOfFragments=25;}'
```

Note: If you install a database with multiple fragments, you must make sure that every agent containing that database has the same number of fragments.

Note: Xgrid BLAST automatically merges the fragmented results into a single results file in a process transparent to the user. However, due to rounding errors, reported e-values may be slightly different from those obtained from an unfragmented database. Additionally, the use of fragmented databases may cause additional matches with lower cutoff scores to be reported.

Using Xgrid BLAST

Xgrid BLAST is a Mac OS X application that allows users to perform BLAST queries on Xgrid clusters. Once you have installed at least one database as described above, you are ready to begin using Xgrid BLAST.

Connecting To an Xgrid Cluster

Xgrid BLAST is installed in your Applications folder. When you launch Xgrid BLAST, you will be presented with a login window. In the “Service” field, select the desired Xgrid cluster, or type it in, and press the “Connect” button. You will then be presented with the “Xgrid BLAST Jobs” window.

Starting a New BLAST Job

To start a job:

1. Click the New Job toolbar icon or select New Job from the Window menu. A New Job window opens.
2. In the Job name field, enter any name you like that will help you identify that job.
3. Select the desired BLAST search type in the Search type popup menu, and select a database in the Database popup menu. Note that the databases that appear in this menu will be those that were installed on at least one agent node connected to your cluster, as described in [“Installing Databases On Agent Nodes” \(page 24\)](#).
4. In the Query sequence text field, you may either type in or paste your query sequence, select it by clicking the Browse... button, or drag-and-drop a text file from the Finder into the field.
5. Once you are ready to submit your job, press the Submit Job button. You should now see your job appear in the Xgrid BLAST Jobs window.

Viewing the Status of your jobs

The Xgrid BLAST Jobs window contains a table listing all the jobs you have submitted. You can sort the table by any column by clicking the column name at the top of that column.

The Status column informs you of the status of each of your jobs, as follows:

Submitting

The job has been started, as explained in [“Starting a New BLAST Job” \(page 26\)](#).

Submitted

The Xgrid cluster has received the job request.

Running

One or more agents are available to work on your job.

Completed

The job has finished normally.

Failed

The job was unable to complete due to an error.

While a job is running, you can see its CPU power—that is, the combined CPU speeds of all agents currently working on that job.

You can also have other information displayed in the table; choosing View options... from the Edit menu opens a preference tab allowing you to select which columns are displayed in the table.

Viewing the Results of Completed Jobs

When a job's status is Completed, you can view its results. Select the job in the table and then click the Display results icon in the toolbar, select Display Results from the Job menu, or double-click the job in the table. A window containing the results of your job will open; if the results have not yet been downloaded by your computer, you see a Loading... message until the results have loaded. You can save or print your results from this window.

By default, as soon as a job is completed, Xgrid BLAST automatically begins downloading its results, so they are available for viewing as soon as possible. However, if you have many results or a slow network connection, it may be desirable for Xgrid BLAST to load job results only when you specifically request them. To change this behavior, select Preferences... from the Xgrid BLAST menu, and click on the Results tab in the Preferences window. If the Automatically load results when available checkbox is selected, results will be downloaded immediately; if not, they will be downloaded when you view the results for a job.

Xgrid BLAST also allows you to have the results for completed jobs automatically saved into a directory of your choosing. To accomplish this, select the Automatically save results to directory checkbox and choose the directory into which they will be saved.

Viewing Job Information

If you want to see the parameters you specified for a particular job, select it in the table, and either press the Show info toolbar icon or select Show info from the Job menu. An info window will open containing the job's name, database, sequence, and query type as specified when you submitted the job.

Cloning a Job

You may want to restart a job you have submitted, or to enter a job similar to one you have previously submitted. To open a New Job window that is already populated with the parameters and query sequence of a job in the Xgrid BLAST Jobs table, simply select the job in the table and either click the Clone Job toolbar icon or select Clone Job from the Job menu. Xgrid BLAST will

then fill out the New Job window with the parameters of the selected job, and load the job's query sequence from the Xgrid server. Note to avoid confusion, you may wish to give the new job a different name from that of the original job.

Canceling a Job

If a job's status is Submitting, Submitted, or Running, you can cancel the job before it finishes. Select the job in the table and either click the Cancel job toolbar icon or select Cancel job from the Job menu. The status column of the table will now indicate that the job is canceled and is no longer active.

Deleting a Job

Every job you submit to Xgrid, including ones you have canceled, remains in the job list table until you choose to remove it. When you are sure you no longer want the job, select it in the job list, and either click the Delete Job icon in the toolbar or select Delete Job from the Job menu. Regardless of whether the job you deleted was waiting to run, running, or finished, the job will be removed from Xgrid and from the table, and you will no longer be able to see it or its results.

Selecting Multiple Jobs

You can perform an action on multiple jobs at the same time; for example, you may want to delete five jobs at once. Select multiple jobs by holding down the Command key and clicking additional jobs in the table. If multiple jobs are selected, any action you perform, such as displaying results or deleting, is performed on all selected jobs.

Searching For Specific Jobs

If you have many jobs in the Xgrid BLAST Jobs window and would like to narrow down those jobs that appear, you can enter some text into the Search jobs search field in the window's toolbar. Xgrid BLAST will then search all of your jobs for that text, and display only those jobs that match the text. To choose the parameters based on which jobs will match your text, choose Preferences... from the Xgrid BLAST menu and select the Search tab. Check those parameters you wish Xgrid BLAST to match when you perform a search. For example, if only the Database checkbox is selected, then typing *chromo* in the Search jobs search field will cause all jobs searching against the *Chromosome* database to be displayed, but none of those searching against the *human_genomic* database. If more than one checkbox is selected, Xgrid BLAST will display jobs that match at least one of the selected fields.

Receiving Notifications

If you wish to be alerted when a job's status changes, select Preferences... from the Xgrid BLAST menu and select the Notifications tab. You can choose to have Xgrid BLAST's dock icon bounce if the application is in the background and one of the selected events occurs.

Viewing Total CPU Power

To see the total CPU power of all agents currently executing your searches, either click the Tachometer icon in the toolbar or select Tachometer from the Window menu. A tachometer-like display will indicate the sum of the CPU speeds of all agents currently working on your jobs.

Viewing Agent Node Status

To see the agent nodes that are working for your Xgrid cluster, as well as their current status, either click the Node list toolbar icon or select Node list from the Window menu. A drawer will be displayed indicating the number of nodes in the cluster, the number of working nodes, the number of available nodes, and the status of each node.

Customizing the Job List Toolbar

To choose which toolbar items are displayed in your toolbar, select Customize toolbar... from the Window menu and select those items that you wish to appear in your toolbar.

Quitting Xgrid BLAST and Reconnecting

If you have jobs in your job list when you quit Xgrid BLAST, all of your jobs will still show up the next time you connect to the same server. Jobs that hadn't started or jobs that were running will be run until they are complete, and jobs that are already finished will be available when you log in again. That is, quitting Xgrid BLAST or disconnecting from a server and then reconnecting to the server has no effect on any of your jobs.

Disconnecting From an Xgrid Cluster

If you wish to disconnect from an Xgrid cluster and connect to a different one, select Disconnect from the File menu. You will then be presented with the Xgrid login window, where you can then connect to another Xgrid cluster.

A P P E N D I X A

Installing and Using Xgrid BLAST