СИСТЕМА УПРАВЛЕНИЯ ПРОХОЖДЕНИЕМ ЗАДАЧ И ПЛАНИРОВЩИК MAUI ДЛЯ МВС-100К

А.В. Баранов, Д.М. Голинка

Многие отечественные многопроцессорные системы типа МВС-1000, в том числе и самая мощная российская супер-ЭВМ МВС-100К, работают под управлением Системы управления прохождением задач (СУПЗ). Важнейшей компонентой СУПЗ является сервер очередей, отвечающий за распределение параллельных задач по вычислительным ресурсам и за ведение очереди задач.

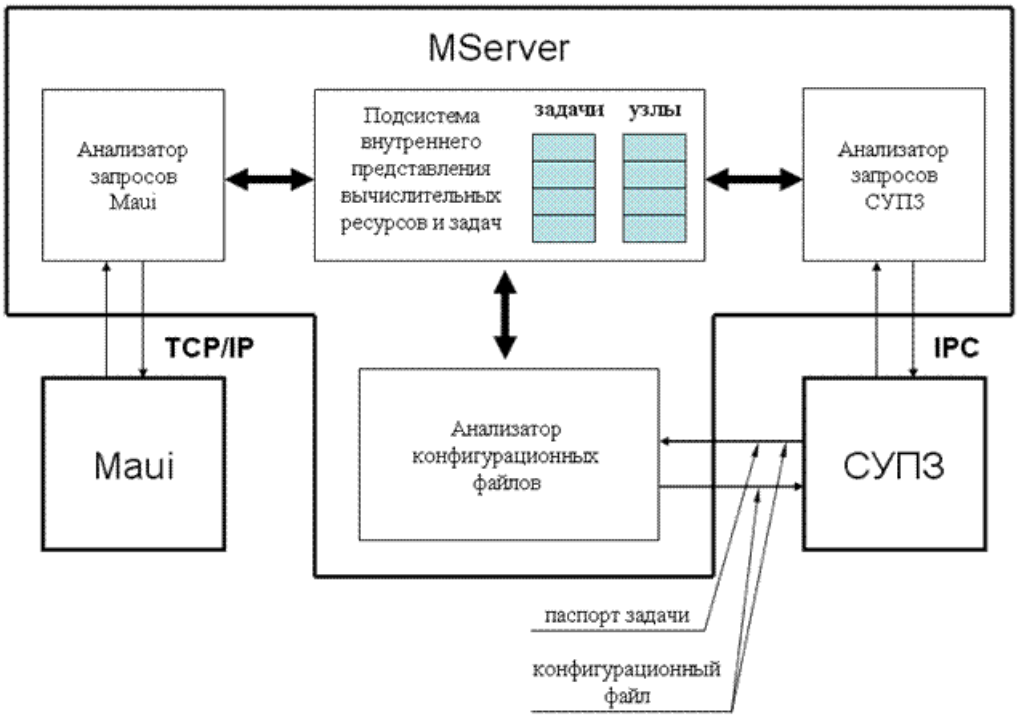
В настоящее время разработчики сервера очередей не имеют возможности осуществлять его техническую поддержку и модернизацию. В то же время такая модернизация является жизненно необходимой – существующая версия СУПЗ с трудом обслуживает текущую конфигурацию МВС-100К, демонстрируя нестабильную работу и низкую производительность. При этом крайне нежелательна замена всей СУПЗ на альтернативную систему пакетной обработки (СПО) (например, OpenPBS/Torque), т.к. в этом случае придется изменять годами сложившийся порядок работы пользователей и администраторов.

Одним из возможных выходов из ситуации авторам представляется использование планировщика Maui [2] в составе СУПЗ вместо существующего сервера очередей. Maui разработан и поддерживается фирмой Maui High Performance Computing Center и обладает рядом преимуществ, важнейшими из которых являются поддержка гетерогенных вычислительных систем и широкие возможности администрирования. Для ряда СПО, в том числе для OpenPBS, у Maui реализованы интерфейсы взаимодействия. Интеграция Maui в состав СУПЗ позволит получить новое качество системы без изменения схемы работы пользователей и порядка администрирования.

Как отмечалось в публикации авторов [1], исследование исходных текстов и документации показало, что OpenPBS и Maui очень тесно интегрированы друг с другом. Между СПО и планировщиком невозможно провести четкий интерфейсный разрез, который позволил бы заменить OpenPBS на СУПЗ МВС-1000. Поэтому для объединения Maui и СУПЗ был использован универсальный интерфейс Wiki [3].

Хотя существует документация, содержащая описание интерфейса Wiki, авторам неизвестны примеры реально работающих систем с использованием этого интерфейса. На практике оказалось, что интерфейс Wiki имеет ряд недокументированных особенностей, критичных для организации взаимодействия компонент СУПЗ и Maui. Недокументированные особенности протокола Wiki были выявлены авторами, в результате чего удалось «заставить» Maui взаимодействовать с СУПЗ.

Для интеграции планировщика Maui в состав СУПЗ МВС-1000 был разработан специальный сервер очередей, получивший название Mserver. Mserver с одной стороны реализует интерфейс Wiki для планировщика Maui, с другой стороны – интерфейсы сервера очередей СУПЗ, что позволяет без модификации использовать существующие компоненты СУПЗ с сохранением порядка их взаимодействия.



Интерфейс Wiki предполагает клиент-серверную организацию взаимодействия, в роли сервера (Wiki-сервер) выступает сервер очередей Mserver, в роли клиента(Wiki-клиент) – планировщик Maui.

В процессе функционирования планировщик Maui устанавливает TCP-соединение с сервером и отправляет ему команду. Сервер очередей обрабатывает команду, посылает ответ клиенту и закрывает соединение. Команды планировщика Maui делятся на категории:

- опрос состояния (запрос текущего состояния очереди задач и вычислительных узлов);
- управление задачами (запрос на постановку и снятие задачи с выполнения).

Структура сервера очередей Mserver показана на рисунке. Центром сервера являет подсистема внутреннего представления вычислительных ресурсов и задач – она отвечает за учет состояния задач и вычислительных ресурсов. За взаимодействие с планировщиком Maui и с СУПЗ отвечают соответствующие анализаторы запросов. Анализатор конфигурационных файлов поддерживает существующие конфигурационные файлы и формат паспортов задач СУПЗ, обеспечивая тем самым полную совместимость с существующей версией.

Кроме этого, для Mserver авторами был разработан программный интерфейс, позволяющий третьей стороне разрабатывать или подключать к СУПЗ альтернативные планировщики.

Для проведения опытной эксплуатации макета СУПЗ с интегрированным планировщиком Maui, Межведомственным суперкомпьютерным центром РАН был выделен стенд, состоящий из 21 восьмипроцессорного вычислительного узла из состава МВС-100К. Опытная эксплуатация подтвердила работоспособность макета, ближайшей перспективой работы является внедрение разработанной системы на МВС-100К.

Главными результатами работы являются:

- удалось создать полнофункциональную СПО, реализующую Wiki-интерфейс с планировщиком Maui;
- для СУПЗ МВС-1000 создан программный интерфейс, позволяющий подключать планировщики третьей стороны.

ЛИТЕРАТУРА:

1. А.В. Баранов, Д.М. Голинка. Исследование возможности использования планировщика Maui в составе СУПЗ-МВС-1000 // Материалы Всероссийской научной конференции "Научный сервис в сети ИНТЕРНЕТ", Новороссийск, 22-27 сентября 2003 г., Изд-во Московского Университета, с.223.
2. Maui cluster scheduler // http://www.clusterresources.com/pages/products/maui-cluster-scheduler.php
3. Wiki Interface Specification // http://www.clusterresources.com/products/maui/docs/wiki/wikiinterface.shtml