Формальное описание аппаратного ядра L4

аппаратный планировщик

Документ описывает расширение системы команд микропроцессоров для реализации аппаратной поддержки микроядра L4 ревизии X2 и совместимых спецификаций. Ревизия документа 20130119

19.1.2013

Оглавление

Введение 2
Тезисы 2
Регистры планировщика 3
Буфер сообщения 3
Блоки регистров задачи (TCB) 4
Пул сообщений и очередь сообщений 5
Синхронизация и обмен сообщениями 6
Формальное описание планировщика задач 7
Процедура распределения времени и алгоритм планирования 9
Межпроцессорное взаимодействие: 10
Замечания 11
Задачи, которые не решены в этой работе 11

Введение

Работа опирается на исследования доктора философии профессора Jochen Liedtke. В качестве справочного руководства использовалось руководство экспериментального микроядра L4 версии X2. Цель работы – описать подсистему аппаратного планирования с возможностью передачи сообщений.

Тезисы

1. Задача это последовательность команд, обрабатываемая исполнительным устройством.
2. Каждая задача имеет собственный "Блок регистров задачи" (TCB) и однозначно определяется и описывается им. Переключение задач осуществляется посредством физической коммутации блока регистров задачи (TCB) с декодером команд (исполнительным устройством).
3. Блок регистров задачи состоит из четырех секций:
   - регистры общего назначения;
   - регистры сообщений;
   - регистры планировщика;
   - регистры управления адресным пространством.
4. Блок регистров задачи и коммутируемый буфер сообщений вместе образуют регистровое пространство задачи.
5. Процесс – это одна или несколько задач, разделяющие одну и ту же таблицу страниц (адресное пространство). В системах без MMU возможно существование единственного процесса – все задачи исполняются в едином адресном пространстве. Любые две задачи, исполняющиеся в различных адресных пространствах, по отношению друг к другу являются процессами.
6. Поток – это задача, исполняющаяся в адресном пространстве какого-либо процесса. Любые две задачи, исполняющиеся в едином адресном пространстве, по отношению друг к другу являются потоками.
7. Планировщик – функциональный блок микропроцессора, расширяющий систему команд и предоставляющий возможность обмена синхронными сообщениями между задачами.
8. Любая операция, будь то обмен сообщениями, либо прерывание (от устройства или таймера) или аппаратное исключение (page fault, protection fault, bus error и т.д.), организуется через вход в планировщик с установленными регистрами сообщений.
9. Задача имеет доступ только к регистрам своего TCB. Планировщик имеет доступ ко всем TCB.
10. Доступ EA означает, что регистры с таким типом доступа могут быть записаны планировщиком.
11. Системные вызовы спецификации L4 ревизии X2, которые не определены данным документом, могут быть реализованы в виде сообщений планировщику.

Регистры планировщика

Номер	Назначение регистра	
0	Идентификатор TCB активной задачи	RUN
1	Идентификатор TCB первой задачи в очереди таймера сообщений	MTQH
2	Определяет количество задач в очереди таймера сообщений	MTQC
3	Идентификатор первого свободного буфера сообщений	FMBH
4	Определяет количество свободных буферов сообщений	FMBC
5	Определяет количество задач, в ожидании буфера сообщения	LMBC

Любая задача имеет права на чтение регистра «Идентификатор TCB активной задачи». Доступ на запись «Идентификатор TCB активной задачи» имеет только планировщик. Запись в регистр осуществляется в момент переключения задачи.

Буфер сообщения

Буфер сообщения это блок регистров, обеспечивающих коммуникацию между задачами. Буфер сообщения состоит из 64 регистров, которые могут быть отображены в регистровое пространство любого процесса. Буфер сообщения может принадлежать только одной задаче или не принадлежать никому. Буфера, не назначенные ни одной задаче, помещаются в очередь свободных буферов сообщений. При передаче сообщения буфер сообщения удаляется из регистрового пространства задачи источника и коммутируется к регистровому пространству задачи приёмника. Метод коммутации определяется разработчиком устройства.

Первый регистр буфера однозначно описывает тип и формат сообщения. Подробности формата буфера описаны в «L4 X2 Reference manual».

Номер	Назначение регистра	Доступ
0	Тэг сообщения (имя и формат)	R/W/EA
1-62	Тело сообщения	R/W/EA
63	Поддержка цепи свободных буферов	EA

Последний регистр буфера сообщения является внутренним регистром. Он используется для организации очереди свободных буферов.

Блоки регистров задачи (TCB)

Блок регистров задачи (TCB) однозначо описывает задачу. Выполнение задачи возможно только для TCB скоммутированного с исполнительным устройвом. Выполненение задачи происходит путём выборки команд исполнительныи устройством по адресу, на который указывает регистр Instuction Pointer. Понятие «идентификатор» однозначно описывает TCB адресуемого объекта.

Номер	Назначение регистра	Доступ	Alias	Назначение
0	Регистр общего назначения 0	R/W	RAX	Регистры общего назначения
1	Регистр общего назначения 1	R/W	RBX	
2	Регистр общего назначения 2	R/W	RCX	
3	Регистр общего назначения 3	R/W	RDX	
4	Регистр общего назначения 4	R/W	RSI	
5	Регистр общего назначения 5	R/W	RDI	
6	Регистр общего назначения 6	R/W		
7	Регистр общего назначения 7	R/W		
8	Регистр флагов	R/W/EA		
9	Указатель фрейма вызова или указатель стека	R/W/EA	RSP/RBP	
10	Счётчик команд (instruction pointer)	R/W/EA	IP	
11	Идентификатор задачи источника сообщения	R/W/EA	FROM	Регистры сообщений
12	Статус операции и код ошибки	R/W/EA	STATUS	
13	Идентификатор задачи получателя сообщения	R/W	TO	
14	Максимальное время обмена сообщениями	R/W	TIME	
15	Индекс буфера сообщений	RO/EA		
16	Собственный идентификатор задачи	RO/EA	MYSELF	Регистры планировщика
17	Идентификатор задачи исключений страниц	RO/EA	PAGER	
18	Идентификатор задачи планировщика	RO/EA	SHED	
19	Приоритет задачи	RO/EA	PRIO	
20	Свободный квант времени	RO/EA		
21	Общий квант времени	RO/EA		
22	Идентификатор соседней задачи	RO/EA	NEXT	
23	Идентификатор дочерней задачи	RO/EA	CHILD	
24	Идентификатор адресного пространства	RO/EA	SPACE	Регистры MMU
25	Зарезервировано для MMU и расширений			
26	Зарезервировано для MMU и расширений			
27	Зарезервировано для MMU и расширений			
28	Значение не определено			Резерв для расширений
29	Значение не определено			
30	Значение не определено			
31	Значение не определено			

Пул сообщений и очередь сообщений

Массив, состоящий из последовательно идущих буферов сообщений, называется пул сообщений. Элементы массива являются так же и элементами односвязного списка. Два глобальных регистра - Free Message Buffer Head (**FMBH**) и Free Message Buffer Counter (**FMBC**)определяют первый элемент списка и количество элементов. В момент инициализации планировщика, каждый элемент массива помещается в этот список. Таким образом организуется очередь свободных буферов сообщений.

Для работы с очередью сообщений в систему команд вводятся две новые инструкции, назовём их условно **IPC_Lock** и **IPC_Unlock**:

- **IPC_Lock** – инструкция захвата буфера сообщений.

  1. При нулевом значении глобального регистра **FMBC**:
     - устанавливается статус «Ожидание свободного буфера» в регистре статуса **TCB** активной задачи
     - увеличивается на единицу глобальный регистр **LMBC**
     - переходит фазу планирования.

  2. При значении глобального регистра **FMBC** отличного от нуля:
     - вычитается единица из глобального регистра **FMBC**
     - отображается буфер сообщений, указываемый глобальным регистром **FMBH**, в регистровое пространство активной задачи
     - помещает в глобальный регистр **FMBH** значение регистра **63** отображённого буфера сообщений
     - продолжает выполнение со следующей за **IPC_Lock** команды.

- **IPC_Unlock** – инструкция освобождения буфера сообщений

  - снимает отображение буфера сообщений из регистрового пространства активной задачи
  - помещает в регистр **63** буфера сообщений значение глобального регистра **FMBH**
  - в регистр **FMBH** помещает значение регистра индекса буфера сообщений TCB активной задачи
  - сбрасывает значение регистра индекса буфера сообщений в TCB активной задачи
  - если значение глобального регистра **LMBC** не равно нулю, то переход в фазу планирования
  - иначе продолжается выполнение со следующей за **IPC_Unlock** команды.

Синхронизация и обмен сообщениями

Обмен сообщениями и синхронизация задач осуществляется при помощи команды **IPC_Exchange.** Команда **IPC_Exchange** может включать фазу передачи сообщения, фазу приёма сообщения или обе фазы.

1. Команда является точкой планирования и может вызвать переключение задач, блокировав задачу, исполняющую команду **IPC_Exchange.**
2. В случае, если команда сообщения имеет фазу передачи, то перед вызовом команды **IPC_Exchange** задаче предварительно должен быть назначен буфер сообщения при помощи команды **IPC_Lock.** Между инструкциями **IPC_Lock** и **IPC_Exchange** задача устанавливает регистры буфера сообщений. **IPC_Exchange** всегда освобождает буфер сообщений, назначенный предшествующей инструкцией **IPC_Lock**, после завершения фазы передачи.
3. В случае, если команда содержит фазу приёма, то в регистровое пространство задачи получателя сообщения отображается буфер сообщений передающей задачи. При этом теряется отображение буфера сообщений из регистрового пространства передающей задачи.
4. Для освобождения принятого сообщения используется команда **IPC_Unlock.**
5. Команда может быть заблокирована в фазе приёма или в фазе передачи на время, определённое регистром «Максимальное время обмена сообщениями» секции «Регистры сообщений» в блоке регистров задачи. Время приёма и время передачи упакованы в битовых полях регистра «Максимальное время обмена сообщениями» и описывают логарифмическое представление временного интервала, включая два граничных значения – без ожидания и ожидание до прихода события.

Данная спецификация не описывает метод доступа к регистрам сообщения – для доступа к регистрам сообщения может быть добавленная дополнительная инструкция к системе команд процессора или же регистры сообщения могут рассматриваться как расширение регистров общего назначения. Выбор способа доступа к регистрам сообщения назначения определяется разработчиком.

Выполнение инструкции **IPC_Exchange** осуществляет вход в планировщик. Данное описание более широко трактует понятие планировщик, нежели общепринятое значение.

Формальное описание планировщика задач

1. Проверяем регистр «Идентификатор получателя сообщения» в ТСВ активной задачи. Если пустое значение, то сообщение не имеет фазы передачи, в этом случае переходим к проверке фазы приёма - шаг №7.

2. Устанавливаем регистр «Статус операции и код ошибки» в состояние "Ожидание передачи".

3. На основании регистра «Идентификатор получателя сообщения» находим блок регистров задачи-получателя сообщения (ТСВ получателя).

4. Проверяем состояние получателя сообщения в регистре «Статус операции и код ошибки» в ТСВ задачи-получателя сообщения.

5. Если задача получатель находится в состоянии "Ожидание приёма", копируем тело сообщения из ТСВ активной задачи в ТСВ получателя, способом, заданным полями в регистре «Описатель сообщения». ТСВ процесса получателя сообщения:

   5.1. Удаляем ТСВ получателя из очереди таймера, если время ожидания сообщения в регистре «Максимальное время ожидания сообщения» ТСВ получателя не равно нулю или бесконечности.
   5.2. Устанавливаем статус завершения операции в регистре «Статус операции и код ошибки».

6. В случае ошибки при передаче сообщения - переходим в процедуру планирования – шаг №11.

7. Проверяем регистр «Идентификатор источника сообщений» в ТСВ активной задачи. Если пустое значение, то фаза приёма отсутствует, в этом случае переходим в фазу распределения времени - шаг №11.

8. Устанавливаем регистр «Статус операции и код ошибки» в состояние "Ожидание приёма".

9. На основании регистра «Идентификатор источника сообщения» находим блок регистров задачи-источника сообщения (ТСВ источника). В случае если регистр «Идентификатор источника сообщения» в ТСВ активной задачи содержит значение "Любой источник", на основании приоритета находим ТСВ задачи, находящейся в фазе "Ожидание передачи" и регистр «Идентификатор получателя сообщения» которой указывает на активную задачу.

10. Копируем тело сообщения из ТСВ задачи-отправителя в ТСВ активной задачи, способом, заданным полями в регистре «Описатель сообщения» задачи-отправителя. Устанавливаем статус завершения операции в регистре «Статус операции и код ошибки» в ТСВ задачи отправителя сообщения и в ТСВ активной задачи. Удаляем ТСВ задачи отправителя из очереди таймера.

11. Процедура предварительного планирования. Анализируется регистр «Максимальное время обмена сообщениями». Если активная задача находится в фазе передачи и максимальное время передачи равно нулю, меняем статус "Ожидание передачи" на статус "Ошибка сообщения" и устанавливаем код ошибки "Превышено время ожидания передачи".

12. Если задача находится в фазе приёма и максимальное время ожидания приёма равно нулю, меняем статус "Ожидание приёма" на статус "Ошибка приёма" и устанавливаем код ошибки "Превышено время ожидания приёма".

13. Если время ожидания активной фазы (приёма или передачи) не равно бесконечности, ставим задачу в очередь таймера, используя период времени, упакованный в регистре «Максимальное время обмена сообщениями». Для организации очереди используем один таймер и три регистра в ТСВ задачи – временной интервал относительно предыдущего события, идентификаторы ТСВ предыдущей и последующей задач в очереди. Двухсвязный список и относительное время позволит не перестраивать очередь при срабатывании таймера или завершении события.

14. Собственно планирование. На основании приоритета задачи, квантов времени и регистра «Идентификатор задачи планировщика» выбираем ТСВ задачи, чей регистр «Статус операции и код ошибки» находится в любом состоянии, отличном от состояний "Ожидание передачи" и "Ожидание приёма". Коммутируется выбранный ТСВ с декодером команд процессора и продолжаем выполнение выбранной задачи. В случае, если все задачи находятся в режиме ожидания, переходим в режим низкого потребления энергии.

15. В случае срабатывания таймера до прихода сообщения, выбираем первый ТСВ из очереди таймера и меняем статус "Ожидание передачи" на "Ошибка передачи" или статус "Ожидание приёма" на "Ошибка приёма". Переходим в процедуру планирования – шаг 14.

Процедура распределения времени и алгоритм планирования

Основной операцией синхронизации и переключения контекста является обмен сообщениями. Универсальный алгоритм планирования опирается на древовидную структуру приоритетов задач. Распределение времени осуществляется обходом дерева, узлами которого являются задачи.

- Для описания древовидной структуры используется регистр «Идентификатор задачи планировщика», указывающий на TCB задачи родительского уровня.
- Для поддержки древовидной структуры используются два регистра в TCB секции «Регистры планировщика» - идентификатор соседней задачи и идентификатор дочерней задачи. Эти поля заполняются планировщиком во время инициализации задачи, на основании регистра «Идентификатор задачи планировщика».

В случае, если задача в узле дерева находится в состоянии ожидания обмена сообщениями, её квант времени отдаётся задачам, находящихся в ветвях узла ожидающей задачи. Корневым элементом дерева является планировщик, который изначально владеет 100% своего времени и разделяет его между задачами, родительским элементом которых он является.

Время процессора распределяется между задачами, связанными в список регистром «Идентификатор соседней задачи». Если задача блокирована в фазе приёма или передачи, то её квант времени разделяется между задачами, на которые указывает регистр «Идентификатор дочерней задачи». Далее время распределяется по списку соседних задач, образуя полный обход дерева задач.

Время задач, находящихся в соседних ветвях одного узла, планируются на основе регистров «Общий квант времени» и «Свободный квант времени». С помощью комбинаций приоритета и положения задачи в дереве планирования, строится необходимая конфигурация системы с заданной латентностью. На этом базисе строится любая система приоритетов и алгоритмов планирования.

Межпроцессорное взаимодействие:

Для обмена сообщения между процессорными модулями вводится арбитражное устройство локальной шины, производящее передачу буфера сообщений. Арбитражное устройство захватывает локальную шину. После успешного захвата шины семантика шины адреса описывается следующим форматом:

Биты	Назначение поля
0-5	Индекс передаваемого регистра буфера сообщений
6-31	Идентификатор устройства локальной шины

За каждый системный такт от владельца локальной шины к устройству, адресуемому битами 6-31, по шине данных передается один регистр буфера сообщений. Необходимое количество тактов определяется первым регистром буфера сообщений. Смотрите описание тэга сообщений в справочном руководстве L4 версии X2. После передачи буфера сообщений арбитражное устройство снимает блокировку локальной шины.

Регионы памяти и отображение адресных пространств, если они присутствуют в сообщении, определяются содержимым буфера сообщений.

Замечания

1. Однозадачную операционную систему можно реализовать при помощи двух блоков регистров задач (TCB), назначив один блок регистров ядру, второй блок регистров пользовательской задаче
2. Данное описание не рассматривает процесс копирования тела сообщения между TCB. помимо передачи буфера сообщения, могут передаваться регионы памяти и осуществляться отображение виртуальных страниц от задачи источника к задаче приёмнику. Эти вопросы подробно освещены в «L4 X2 Reference manual».
3. Предложенный алгоритм передачи сообщений можно оптимизировать для случая, когда регистр «Идентификатор источника сообщения» в TCB активной задачи содержит значение "Любой источник". Для этого переходим на шаг 11 «Предварительное планирование» и во время планирования для каждой задачи, находящейся в состоянии "Ожидание передачи", находим TCB принимающего процесса, затем проверяем его состояние – если принимающий процесс в состоянии "Ожидание приёма" и его регистр «Идентификатор источника сообщения» содержит значение "Любой источник", то производим передачу сообщения.

Вопросы, которые не рассмотрены в этом документе

1. Расширенный MMU для работы с универсальными виртуальными страницами (flex pages).
2. Вытеснение TCB из внутренней памяти процессора во внешнюю оперативную память и загрузка TCB актуальных задач.