

## Об аппаратной поддержке виртуализации системы прерываний в многоядерных микропроцессорах семейства «Эльбрус».

*Р.В. Деменко<sup>1,2</sup>, В.Б. Трофимов<sup>2</sup>*

<sup>1</sup> Московский физико-технический институт (государственный университет)

<sup>2</sup> ЗАО «МЦСТ»

Рост производительности современных компьютерных платформ и необходимость иметь гибкую инфраструктуру вычислительных средств обуславливают эволюцию технологий виртуализации. Реализация средств аппаратной поддержки виртуализации позволяет получить значительный прирост производительности виртуальных машин. В современных многоядерных микропроцессорах такая аппаратная поддержка реализуется, помимо прочего, для виртуализации операций ввода/вывода и системы прерываний [1, 2]. В данной работе предложены принципы реализации аппаратной поддержки виртуализации системы прерываний МП «Эльбрус», которые являются частью общего подхода к созданию этих средств для архитектуры «Эльбрус» [3] и подразумевают:

- виртуализацию обращения к регистрам контроллера прерываний
- обеспечение доставки прерываний нужному гостевому ядру

Обращения к регистрам контроллера прерываний

Как правило, управляющая структура виртуальной машины, включающая состояние системы прерываний (копии регистров контроллера прерываний), хранится в памяти машины. В отсутствие аппаратной поддержки каждое обращение гостевого ядра к регистрам контроллера прерываний перехватывается гипервизором и обрабатывается программно.

Такой перехват для каждого ядра можно не делать, если реализовать дублирующий набор управляющих регистров, который используется гостевым ядром, активным на данном физическом ядре в текущий момент. При постановке гостевого ядра гипервизор заполняет эти регистры значениями из управляющей структуры. При снятии гостевого ядра, значения регистров откачиваются в управляющую структуру.

Доставка прерываний нужному гостевому ядру

Для поддержки многопроцессорной конфигурации контроллер прерываний в МП «Эльбрус» имеет распределенную структуру. Элементы системы прерываний взаимодействуют через аппаратно формируемые сообщения (сообщения о прерываниях, служебные сообщения). Введение гостевого набора регистров приводит к появлению в аппаратуре гостевых сообщений о прерываниях, которые необходимо изолировать как от сообщений хоста, так и от сообщений других гостей. С этой целью сообщения сопровождаются номером гостевой машины, поскольку три параметра <номер гостевой машины, номер гостевого ядра, тип (и вектор) прерывания> позволяют однозначно идентифицировать сообщение.

Чтобы доставить прерывания нужному гостевому ядру, пара <номер гостевой машины, номер гостевого ядра> преобразуется в физический номер ядра (если данное гостевое ядро активно) с использованием аппаратной таблицы соответствия, управляемой гипервизором. Ее состояние должно быть одинаковым во всех процессорах многопроцессорной системы, это свойство гарантируется аппаратной схемой синхронизации, реализованной через служебные сообщения.

В случае, если гостевое ядро неактивно, это прерывание фиксируется в управляющей структуре соответствующей гостевой машины, аппаратно генерируемой атомарной DMA-операцией записи.

При заполнении гостевых регистров во время процедур снятия/постановки гостевого ядра, необходимо гарантировать, что уже сформированные сообщения о прерываниях не будет потеряны. Такая потеря может возникнуть вследствие гонок между сообщениями о прерываниях, служебными сообщениями и DMA-операциями, связанными с доставкой прерываний отложенному гостевому ядру. Для исключения ситуаций гонок аппаратура предоставляет системному ПО возможность точно определить момент завершения процесса синхронизации таблиц соответствия и завершения всех соответствующих DMA-операций.

Для прерываний от внешних устройств, отданных под непосредственный контроль гостевой ОС («проброс» устройства в виртуальную машину), стоит проблема определения номера гостевой машины. В самом распространенном случае внешнее прерывание приходит в виде Message Signaled Interrupt (MSI), представляющем собой с точки зрения процессора особый DMA-

запрос. Содержащийся в нем виртуальный адрес проходит процедуру трансляции в физический адрес. Применительно к MSI соответствующие поля в таблице трансляции можно использовать для хранения параметров прерывания внешнего устройства, отданного гостевой ОС, в нашем случае — номера гостевой машины.

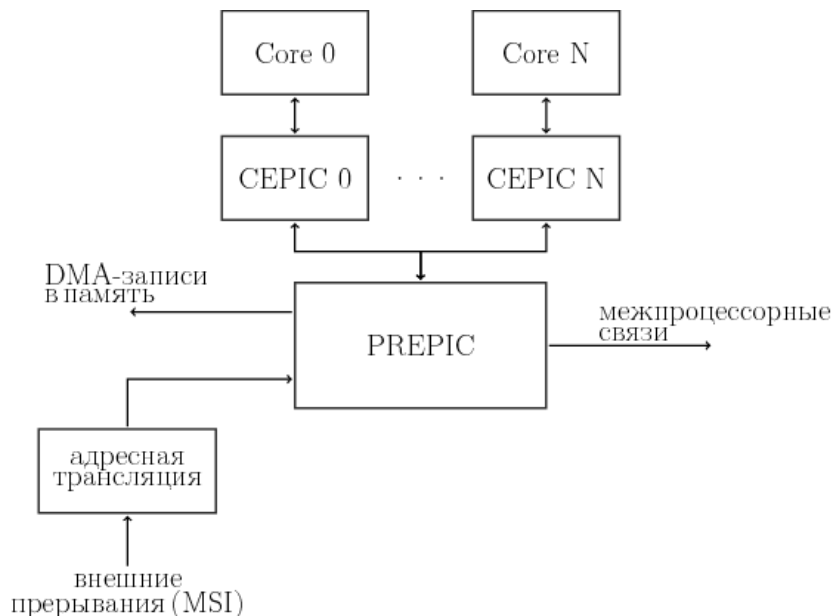


Рис.1 Упрощенная схема контроллера прерываний (Elbrus Programmable Interrupt Controller, EPIC)  
Условные обозначения:

Core 0, Core N — ядра микропроцессора

CEPIC 0, CEPIC N — Core EPIC, локальная для каждого ядра часть контроллера прерываний

PREPIC — Processor EPIC, общая для процессора часть контроллера прерываний, содержит таблицу соответствия физических и виртуальных ядер.

### Литература

1. Intel Virtualization Technology for Directed I/O, Architecture Specification, 2016.
2. ARM Generic Interrupt Controller Architecture Specification v2.0, 2013.
3. Знаменский Д.В. Выбор вариантов реализации средств аппаратной поддержки виртуализации архитектуры Эльбрус. «Вопросы радиоэлектроники», сер. ЭВТ. 2014. V. 3.