

# Архитектура IP-маршрутизаторов серии Cisco 12000 – коммутация пакетов

## Содержание

[Введение](#)

[Предварительные условия](#)

[Требования](#)

[Используемые компоненты](#)

[Условные обозначения](#)

[Общие сведения](#)

[Коммутация пакетов: Обзор](#)

[Коммутация пакетов: Платы линии обработчика 0 и обработчика 1](#)

[Коммутация пакетов: Линейные платы Engine 2](#)

[Коммутация пакетов: Коммутация ячеек через коммутирующую матрицу](#)

[Коммутация пакетов: Передача пакетов](#)

[Сводка потока пакетов](#)

[Дополнительные сведения](#)

## **Введение**

Этот документ исследует самые важные элементы архитектуры Интернет-маршрутизатора Cisco 12000 серии - коммутируемые пакеты. Коммутация пакетов радикально отличается от любых архитектур общей памяти или архитектур Cisco на основе шины. При помощи перекрестной матрицы Cisco 12000 предоставляет очень большие количества пропускной способности и масштабируемости. Кроме того, 12000 использует виртуальные выходные очереди для устранения блокирования заголовка линии внутри коммутационной матрицы.

## **Предварительные условия**

### **Требования**

Для этого документа отсутствуют особые требования.

### **Используемые компоненты**

Сведения, содержащиеся в данном документе, относятся к следующему оборудованию:

- IP-маршрутизатор Cisco серии 12000

Сведения, представленные в этом документе, были получены от устройств, работающих в специальной лабораторной среде. Все устройства, описанные в этом документе, были запущены с чистой (стандартной) конфигурацией. В рабочей сети необходимо изучить

потенциальное воздействие всех команд до их использования.

## Условные обозначения

[Дополнительные сведения об условных обозначениях см. в документе Технические рекомендации Cisco. Условные обозначения.](#)

## Общие сведения

(Решения о переключениях на коммутаторе Cisco 12000 принимаются линейными картами (LC)). Предназначенная для некоторых LC специализированная интегральная схема фактически выполняет коммутацию пакетов. Distributed Cisco Express Forwarding (dCEF) является единственным доступным методом коммутации.

**Комментарий:** Механизмы 0, 1, и 2 не являются последними механизмами, разработанными Cisco. Это также линейные платы Engine 3, 4 и 4+ и последующие выпуски. Линейные карты устройства 3 могут выполнять краевые функции, используя пропускную способность линии. Чем выше находится механизм уровня 3, тем больше пакетов коммутирует оборудование. Можно найти некоторые полезные сведения о других линейных картах доступными для маршрутизатора Cisco серии 12000 и механизма, на котором они базируются в [Интернет-маршрутизаторе Cisco 12000 серии: Вопросы и ответы.](#)

## Коммутация пакетов: Обзор

Переадресацию пакетов всегда выполняет входная линейная карта (LC). Выходной LC только выполняет исходящее Качество обслуживания (QoS), которое зависит от очереди (например, Взвешенное произвольное раннее обнаружение (WRED) или Согласованная скорость доступа (CAR)). Большинство пакетов коммутировано LC с помощью Distributed Cisco Express Forwarding (dCEF). Только управляющие пакеты (например, обновления маршрутизации) отправляются на гигабитный процессор маршрута (GRP) для обработки. Путь перенаправления пакетов зависит от типов коммутирующих устройств, используемых на LC.

Это - то, что происходит, когда входит пакет:

1. Пакет приходит в модуль интерфейса физического уровня (PLIM). Здесь происходят разные вещи: Приемопередатчик преобразует оптические сигналы в электрические (у большинства линейных плат CSR есть оптоволоконные соединители) Формирование кадров L2 удалено (SANE, Асинхронный режим передачи (ATM), Ethernet, High-Level Data Link Control (HDLC)/Point-to-Point Протокол - PPP) Ячейки ATM повторно собраны Пакеты, которые не проходят проверку циклическим избыточным кодом (CRC), сбрасываются
2. Поскольку пакет получен и обработан, это - Непосредственный доступ к памяти, к которому Обращаются в маленькое (приблизительно 2 x буфера максимального размера передаваемого блока данных (MTU)) память, названная "First In, First Out (FIFO) пакетная память". Объем памяти зависит от типа линейной платы (от 128 Кбайт до 1 Мбайта).
3. Как только пакет находится полностью в FIFO - памяти, специализированная интегральная схема (ASIC) на PLIM связывается с Buffer Management ASIC (BMA) и

просит буфер вставлять пакет. BMA получает сведения о размере пакета и размещает его в соответствующем буфере. Если BMA не может получить буфер правильного размера, пакет отбрасывается и счетчик игнорирований увеличивается для входящего интерфейса. В отличие от других платформ механизм восстановления в данном случае не применим. Пока это продолжается, PLIM может получить другой пакет в пакетной памяти FIFO, вот почему она имеет размер  $2 \times \text{MTU}$ .

4. Если в правой очереди имеется доступный свободный буфер, BMA сохраняет пакет в списке свободной очереди соответствующего размера. Этот буфер размещается в прямой очереди, проверенной Salsa ASIC или R5K CPU. CPU R5K определяет место назначения пакета, предварительно обращаясь к локальной таблице dCEF в DRAM, а затем перемещает буфер из прямой очереди в очередь ToFabric в соответствии со слотом назначения. Если назначение не находится в таблице CEF, пакет отброшен. Если пакет является управляющим пакетом (например, обновления маршрута), он ставится в очередь очереди GRP и будет обработан GRP. Существует 17 очередей ToFab (16 одноадресных и 1 многоадресная). Существует одна очередь toFab на линейную карту (это включает RP). Эти очереди известны как "виртуальные очереди вывода (VOQ)" и важны так, чтобы не происходило блокирование заголовка очереди.
5. ToFab BMA разделяет пакет на 44-байтовые части, являющиеся полезной нагрузкой для того, что в итоге будет известно под именем "ячейки Cisco". Эти ячейки образуют 8-байтовый заголовок и 4-байтовый заголовок буфера посредством frFab BMA (общий размер данных = 56 байт) и затем ставятся в соответствующую очередь ToFab (в этой точке счетчик #Qelem пула, откуда поступил буфер, уменьшается на единицу, а счетчик очереди ToFab увеличивается на единицу). Блок принятия решений зависит от типа модулей коммутации: На Engine 2 + карты, специальный ASIC используется для улучшения способа, которым коммутированы пакеты. Нормальные пакеты (IP/метка, нет опций, проверочная сумма) обрабатываются напрямую ASIC-микросхемой коммутации пакетов (PSA), затем проходят мимо связи "сырая очередь/CPU/Salsa" и попадают прямо в очередь toFab. Только первые 64 байта пакета передаются через коммутацию пакетов ASIC. Если пакет не может быть маршрутизован PSA, то он встает в очередь RawQ на обработку CPU, принадлежащим LC, как было описано ранее. На данном этапе решение о коммутации принято, и пакет помещен в правильную исходящую очередь ToFab.
6. ToFab BMA передает ячейки пакета по прямому доступу к памяти (DMA) в небольшие буферы FIFO на матричном интерфейсе ASIC (FIA). Имеется 17 буферов ФИФО (по одному на каждую очередь ToFab). Когда FIA получает ячейку от BMA toFab, он добавляет 8-байтовый CRC (общий размер ячейки - 64 байта; 44 байтовых полезных нагрузки, 8 заголовков байтной ячейки, 4-байтовый заголовок буфера). FIA имеет интегральные схемы интерфейса последовательной линии (SLI), выполняющие кодирование 8B/10B для ячеек (аналогично интерфейсу FDDI 4B/5B) и подготавливающие их к передаче через коммутационную матрицу. Создается впечатление большого объема служебной информации (44 байта данных преобразуются в 80 байт, используя матрицу!!), но это не проблема, так как емкость оптоволоконной была настроена соответственно.
7. Теперь, когда FIA готов передать, FIA запрашивает доступ к матрице от в настоящее время планировщик активной карты и часы (CSC). Плата CSC работает по сложному алгоритму равнодоступности. Идея состоит в том, что никакому LC не позволяют монополизировать пропускную способность исходящего канала никакой другой карты. Обратите внимание на то, что, даже если LC хочет передать данные из одного из его

собственных портов, это все еще должно перейти матрицу. Это важно, так как, если это не происходит, то один из портов на облаке LC монополизирует всю пропускную способность облака. Эта также сделало бы план коммутации более сложным. FIA отправляет ячейки через коммутационную матрицу на их исходящий LC (указанный данными в заголовке ячейки Cisco, помещенными туда коммутационным процессором). Алгоритм равнодоступности также разработан для оптимального сопоставления; если плате 1 нужно переслать данные на плату 2 и, одновременно с этим, плате 3 нужно передать данные на плату 4, процессы запускаются параллельно. Это - большая разница между коммутационной матрицей и архитектурой шины. Думайте о нем как аналогичном Коммутатору Ethernet по сравнению с концентратором; на коммутаторе, если порт А хочет передать к порту В, и порт С хочет говорить с портом D, те два потока происходят друг независимо от друга. На концентраторе существуют полудуплексные проблемы, такие как коллизии и откат и повторяют алгоритмы.

8. Ячейки Cisco, которые выходят из матрицы, проходят Обработку SLI для удаления кодирования 8B/10B. Если бы там какие-либо ошибки здесь, они появились бы в выходных данных команды show controller fia как "четность ячейки". Посмотрите, [Как Считать Выходные данные Команды show controller fia](#) для дополнительных сведений.
9. Эти ячейки Cisco являются DMA'd в FIFOs на FIA frFab, и затем в буфер на BMA frFab. BMA frFab является тем, который фактически делает повторную сборку ячеек в пакет. Как BMA frFab знает, какой буфер вставить ячейки, прежде чем это повторно соберет их? Это еще одно решение, созданное модулем коммутатора входящей линейной карты; так как все очереди на всей коробке являются одинаковым размером и в том же заказе, устройство переключения просто имеет LC Tx, помещает пакет в ту же очередь номера, от которой это ввело маршрутизатор. Очереди frFab BMA SDRAM можно просматривать с помощью команды show controller frfab queue на LC. Посмотрите, [Как Считать Выходные данные Команд show controller frfab | tofab queue на Интернет-маршрутизаторе Cisco 12000 серии](#) для подробных данных. Это по существу то же самое понятие, что выходные данные toFab BMA. Пакеты поступают и помещаются в пакеты, исключенные из соответствующих свободных очередей. Эти пакеты размещены в очередь от матрицы, ставившую в очередь на любом очередь интерфейса (существует одна очередь на физический порт), или rawQ для обработки выходных данных. Не много происходит в rawQ: многоадресная репликация для отдельных портов, измененный кольцевой алгоритм MDRR (Modified Deficit Round Robin), аналогичный распределенной очереди с весами (DWFQ), а также выходной CAR. Если очередь передачи загружена, пакет отбрасывается и счетчик исходящих отбрасываний увеличивается.
10. BMA frFab ждет, пока часть TX PLIM не готова передать пакет. BMA frFab делает перезапись реального MAC (основанный, помните на информации, содержащейся в заголовке ячейки Cisco), и DMA пакет к маленькому (снова, 2xMTU) буфер в схеме PLIM. PLIM по возможности осуществляет инкапсуляцию SAR ATM и SONET и передает пакет.
11. Трафик ATM собран повторно (при помощи SAR), сегментирован (при помощи tofab BMA), собран повторно (при помощи fromfab BMA) и снова сегментирован (при помощи fromfab SAR). Это происходит очень быстро.

Это время жизни пакета, от начала до конца. Если вы хотите знать то, чему GSR чувствует себя подобно в конце дня, прочитайте эту всю газету 500,000 раз!

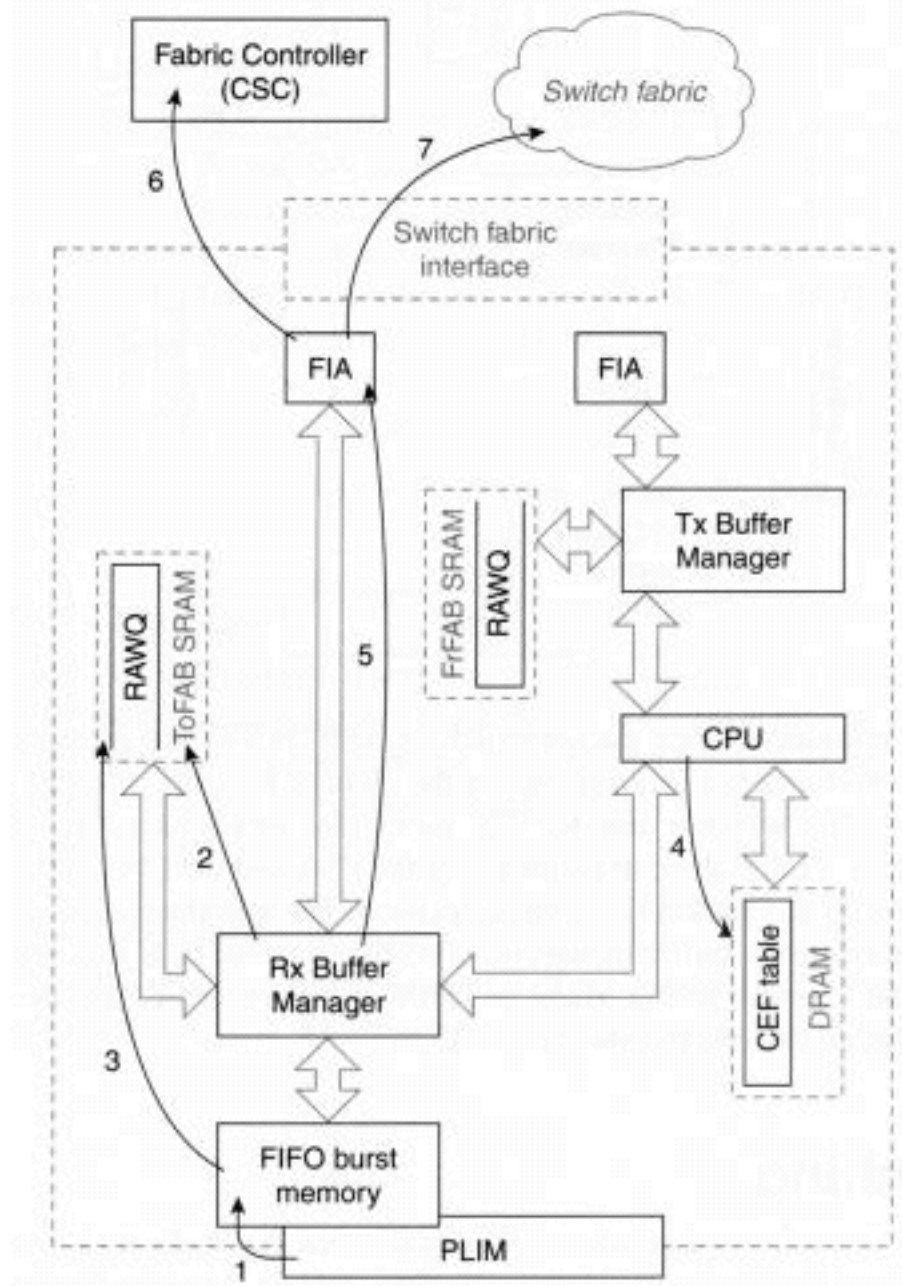
Путь коммутации пакета на GSR зависит от типа механизма переадресации на LC. Теперь выполним все этапы процедуры для устройств Engine 0, Engine 1 и двух LC.

## Коммутация пакетов: Платы линии обработчика 0 и обработчика 1

Представленные ниже разделы основаны на книге Внутренней архитектуры программного обеспечения Cisco IOS, печатных изданий Cisco.

[Рисунок 1](#) ниже иллюстрирует другие шаги во время коммутации пакетов для Механизма 0 или Механизма 1 LC.

Рисунок 1: Путь коммутации обработчика 0 и обработчика 1



Коммутируемый путь для Механизма 0 и Механизма, 1 LC является по существу тем же, невзирая на то, что Механизм 1 LC имеет расширенное устройство переключения и буферного менеджера для повышения производительности. Коммутируемый путь имеет

следующий вид:

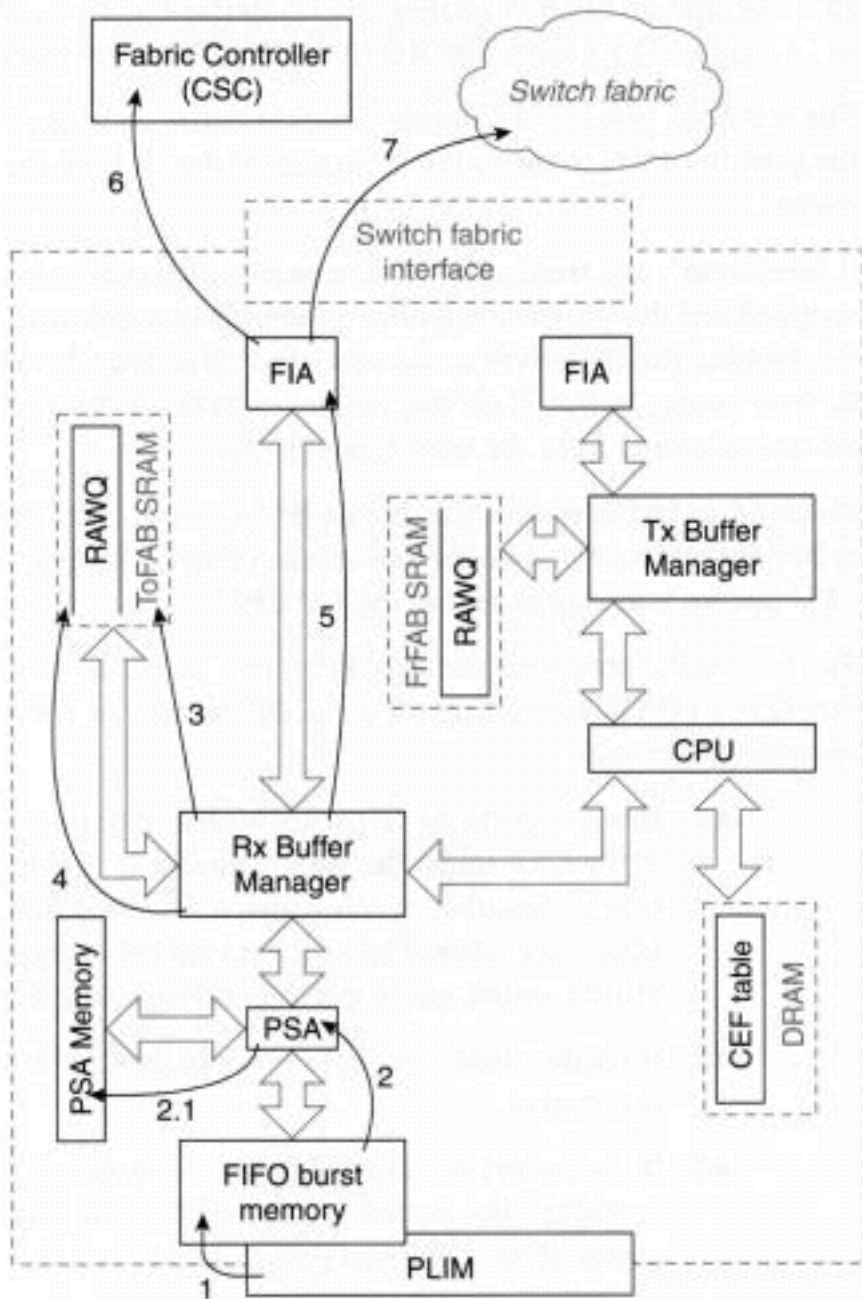
- **Шаг 1.** Интерфейсный процессор (PLIM) обнаруживает пакет на сетевом накопителе и начинает копировать его в память FIFO, которая называется пакетной памятью (burst memory) в LC. Сумма пакетной памяти, которую имеет каждый интерфейс, зависит от типа LC; обычно LC имеют от 128 Кб до 1 Мб пакетной памяти.
- **Шаг 2** - интерфейсный процессор запрашивает буфер пакетов от получить ВМА; пул, от которого запрашивают буфер, зависит от длины пакета. При отсутствии свободных буферов интерфейс отключается, и счетчик "пропусков" интерфейса увеличивается. Например, если 64-байтный пакет достигает интерфейса, ВМА пытается выделить 80-байтный буфер пакетов. Если в 80-байтном пуле отсутствуют свободные буферы, выделение буферов из следующего доступного пула не производится.
- **Шаг 3** – когда ВМА выделяет свободный буфер, пакет копируется в буфер и ставится в прямую очередь (RawQ) для обработки CPU. Прерывание передается ЦП LC.
- **Шаг 4** - Процессы ЦПУ LC каждый пакет в RawQ, поскольку это получено (RawQ является FIFO), консультируясь с локальной таблицей Distributed Cisco Express Forwarding в DRAM для создания решения о коммутации.**4.1** Если это - пакет IP индивидуальной рассылки с допустимым адресом назначения (DA) в таблице CEF, заголовок пакета переписан с новой информацией об инкапсуляции, полученной из таблицы смежности CEF. Коммутируемый пакет ставится в виртуальную очередь вывода, соответствующую гнезду назначения.**4.2** Если адрес назначения (DA) не находится в таблице CEF, пакет отброшен.**4.3** Если пакет является управляющим (например, пакетом обновления маршрута), он будет направлен в виртуальную исходящую очередь GRP и обработан GRP.
- **Шаг 5.** Входящий ВМА фрагментирует пакет на 64-байтовые ячейки и направляет эти ячейки к FIA для передачи на исходящую LC.

В конце Шага 5 пакет, который поступил в Механизм 0/1 LC, был коммутирован и готов быть транспортированным через коммутационную матрицу как ячейки. Перейдите к Шагу 6 в [Коммутацию пакетов](#) раздела: [Коммутация ячеек через коммутирующую матрицу](#).

## [Коммутация пакетов: Линейные платы Engine 2](#)

Когда пакеты поступают в LC Engine 2, как описано в следующем списке шагов, [рисунок 2](#) ниже иллюстрирует путь перенаправления пакетов.

Рис. 2: Модуль 2, коммутируемый путь



- Шаг 1. Интерфейсный процессор (PLIM) обнаруживает пакет на сетевом накопителе и начинает копировать его в память FIFO, которая называется пакетной памятью (burst memory) в LC. Сумма пакетной памяти, которую имеет каждый интерфейс, зависит от типа LC; обычно LC имеют от 128 Кб до 1 Мб пакетной памяти.
- Шаг 2 - первые 64 байта пакета, называемые заголовком, проходят через специализированную интегральную микросхему коммутации пакетов (PSA). 2.1 PSA направляет пакет, сверяясь с локальной таблицей CEF в памяти PSA. Если пакет не может быть коммутирован PSA, перейдите к Шагу 4; иначе перейдите к шагу 3.
- Шаг 3 – диспетчер буфера получения (RBM) принимает заголовок от PSA и копирует его в свободный заголовок буфера. Если пакет больше 64 байт, хвост пакета также копируется в тот же свободный буфер памяти пакета и ставится в виртуальную выходную очередь выходного LC. Переходите к шагу 5.
- Шаг 4 - пакет поступает в этот шаг, если это не может быть коммутировано PSA. Данные пакеты помещаются в необработанную очередь (RawQ). С этого момента путь коммутации по существу совпадает с путем для Engine 1 и Engine 0 LC (действие 4 для Engine 0). Обратите внимание на то, что пакеты, которые коммутированы PSA, никогда

не размещаются в RawQ, и никакое прерывание не передается ЦП.

- **Шаг 5** - Модуль интерфейса коммутационной матрицы (FIM) ответственен за сегментацию пакетов в [ячейки Cisco](#) и передачу ячеек к Fabric Interface ASIC (FIA) для передачи к исходящему LC.

## Коммутация пакетов: Коммутация ячеек через коммутирующую матрицу

Вы поступаете на данном этапе после того, как механизм коммутации пакетов переключает пакеты. На этом этапе пакеты сегментируются в ячейки Cisco и ожидают передачи через коммутационную матрицу. Действия для данного этапа таковы:

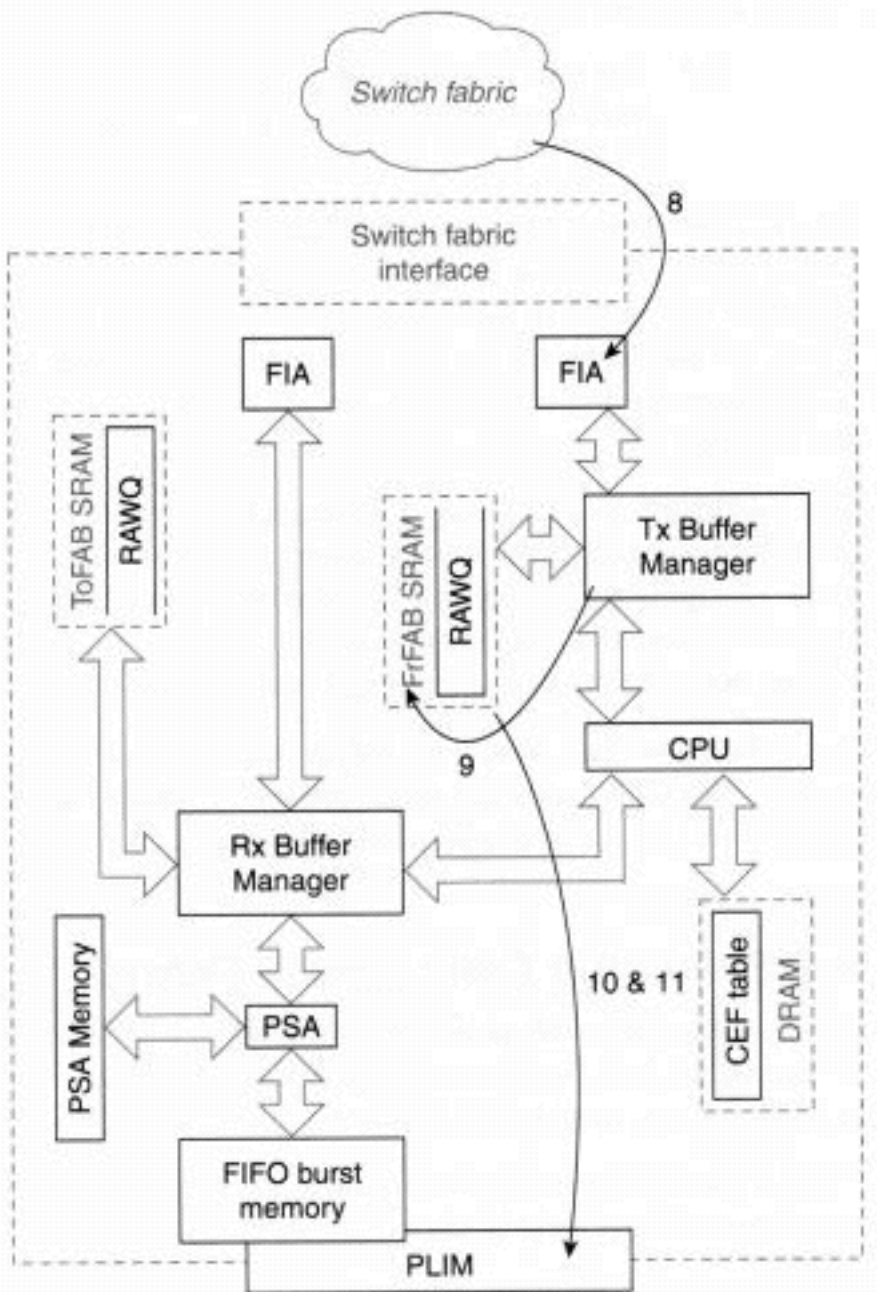
- **Шаг 6** – FIA посылает CSC запрос на подтверждение, который планирует передачу каждой ячейки по коммутационной матрице.
- **Шаг 7** - Когда планировщик предоставляет доступ к коммутационной матрице, ячейки, передан конечному слоту. Обратите внимание на то, что ячейки не могли бы быть переданы внезапно; другие ячейки в других пакетах могли бы быть чередованы.

## Коммутация пакетов: Передача пакетов

Рисунок 3 ниже показав последняя стадия коммутации пакетов. Ячейки собираются заново, а пакет передается в среду. Это имеет место на исходящей линейной плате.

Рис. 3: Коммутация пакетов Cisco 12000: Стадия передачи данных





- Шаг 8 — ячейки, перенаправляемые в коммутационной матрице, подаются на линейную плату назначения через FIA.
- Шаг 9 – передающий диспетчер буфера выделяет буфер из памяти пакетной передачи и повторно собирает пакет в этом буфере.
- Шаг 10 – когда пакет восстановлен, передающий BMA ставит его в очередь передачи конечного интерфейса на LC. Если очередь передачи интерфейса загружена (пакет нельзя поставить в очередь), пакет отбрасывается и счетчик исходящих отбрасываний увеличивается. **Примечание:** Единственный раз, когда пакеты в направлении передачи помещаются в RawQ – это случай, когда LC CPU должен произвести дополнительную обработку до передачи. Примеры описывают IP-фрагментацию, многоадресность и исходящую CAR.
- Шаг 11 – Процессор интерфейса обнаруживает пакет, ожидающий отправки, выводит буфер из очереди передачи, копирует его во внутреннюю FIFO-память и передает пакет в среду.

## Сводка потока пакетов

IP пакеты, которые проходят 12000, обрабатываются в три этапа:

- Три раздела входной линейной платы: Входной PLIM (Блок взаимодействия Физической линии) - Оптические сети к Электрическому преобразованию, Synchronous Optical Network (SONET) (синхронная оптоволоконная сеть)/синхронная Разрядная иерархия (SDH) неформирование кадров, HDLC и Обработка PPP. IP-переадресация – решение о переадресации на основе поиска FIB и постановка в одну из одноадресных или многоадресных очередей входа. Управление очередью входа и интерфейс матрицы – обработка очередей входа и выведение пакетов из очереди к матрице для увеличения использования матрицы на основе механизма произвольное раннего обнаружения (RED)/взвешенного произвольного раннего обнаружения (WRED).
- Коммутация IP-пакетов через матрицу 12000 с входной платы на выходную плату или выходные платы (для многоадресной маршрутизации).
- Выходная линейная карта в трех разделах: Выходной Интерфейс коммутационной матрицы - Повторная сборка пакетов IP, которые будут передаваться и организация очереди в выходные очереди; обработка пакетов групповой адресации. Управление очередью выхода - обработка очередей входа и выведение пакетов из очереди к выходному PLIM для увеличения использования выходного канала на основе механизма RED/WRED. Выходной PLIM - HDLC и обработка PPP, формирование кадров SONET/SDH, преобразование из электрического в оптический.

## Дополнительные сведения

- [Техническая поддержка - Cisco Systems](#)