

Раз таблица, два таблица

Раз таблица, два таблица.

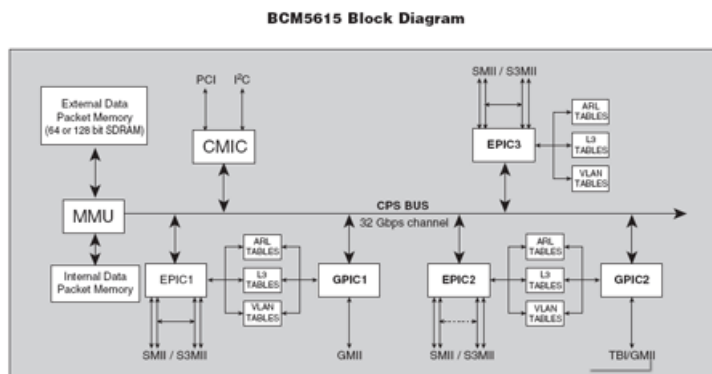
Продолжим тему, начатую в [#278](#) с ее исторического начала. Хотя если смотреть со стороны логики работы – это самая середина процесса, его основа и опора.

В незапамятной дали последних лет прошлого столетия 3com, Cisco и Nortel (Bay) начали использовать VLAN'ы. Было это еще до принятия стандарта 802.1Q, поэтому в свободные поля заголовка пакета втискивали свое – теги VLT, ISL, отличающиеся от современных лишь деталями. Внедрение 802.1Q прошло где-то на софтовом уровне, например в 3com 1100/3300 в версии софта 2.40. Или на аппаратном – Cisco перестроиться на ходу не смогла, и появилась модель-перевертыш, c2924-XL (при почти одинаковом названии есть реально два совершенно разных аппарата – один с поддержкой только ISL, другой еще и 802.1Q)

Немного позже, Nortel в BS350/450 софтовой доработкой увеличил количество поддерживаемых VLAN с 64 до 256, а в BS420 – так просто 802.1Q появился по желанию маркетологов, как только был снята с производства конкурирующая модель с 802.1Q (BS450).

В общем, нельзя сказать, что VLAN – неотъемлемое свойство ASIC.

Для понимания причин и следствий придется опять обратиться к теории. Вот логическая схема устройства серийного чипа [Broadcom 6515](#) (уровень сложности примерно как у тех, что используются в c2950T).

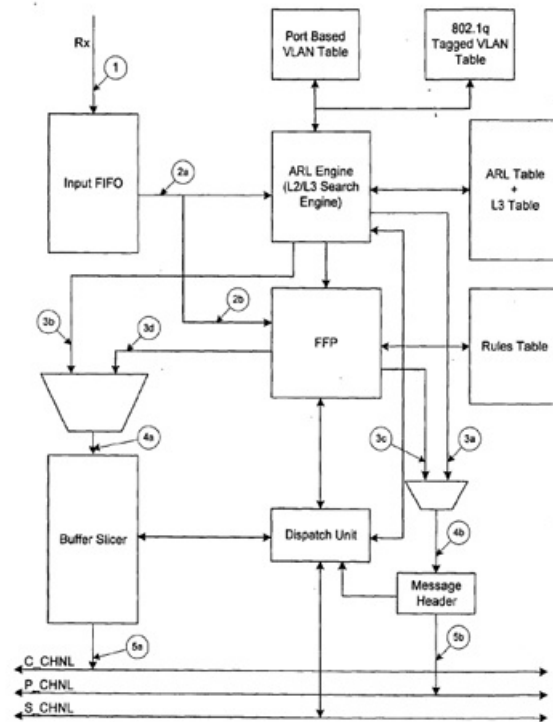


Расшифровка сокращений.

- RMII & SMII - Reduced & Serial Media Independent Interfaces
- GMII или TBIs - Gigabit Media Independent Interface or Ten Bit Interfaces
- EPIC - Ethernet Port Interface Controllers
- GPIC - Gigabit Port Interface Controllers
- MMU - Memory Management Unit. Блок скромно притулился в углу, но это в общем-то ключ к работе ASIC.
- ARL (Adress Resolve LookUp), VLAN, L3 Tables - в расшифровке не нуждаются
- CPS Bus - шина передачи данных.

Наиболее интересны три последних пункта, о которых нужно будет поговорить подробнее. Остальное, в общем, достаточно тривиально и неплохо документировано.

Вот диаграмма прохождения пакета через гребенку фильтров-таблиц:



Взята она из [этого документа](#), в котором, собственно, можно найти и подробное описание процессов. Увы, оно слишком сложно для обзора. Было бы очень интересно подробно все перевести и растолковать (прежде всего для самого себя)...

Пока лишь наиболее интересные моменты. Во-первых, надо учитывать, что таблица - реально одна на весь ASIC. Все разделение, это разные участки памяти. Обращение к ней асинхронно, инициализируется EPIC или GPIC.

Надо отметить, что память - это реально очень дорогой ресурс, подчас влияющий на стоимость свитча больше, чем ASIC (если, конечно, она не встроена в сам ASIC). Вот, например: двухпортовая память, 8Кбайт IDT7005 время выборки 10 нс. Цена в России по [efind](#) 20-30\$

При этом цена микросхемы 8-ми портового коммутатора 3\$, а вполне современные 24-х портовые Broadcom стоят порядка \$40-60. Думаю, что известные своей хитростью восточные производители коммутаторов великолепно понимают тонкость момента, и знают как и на чем можно сэкономить.

Во-вторых, таблица, например Address Resolve LookUp, хранит не сам мак, а хэш (bitmap). Хэш можно сделать только из мака (целиком), а можно - из мака + vlan id. **Если разрядность хеша была при проектировании коммутатора заложена с расчетом только на мак-адрес, то можно добавить в свитч поддержку vlan-ов, если делать хэш не из 48 бит, а, например, из 40 битов мак-адреса и 8 битов vlan id. При некотором допущении все работает.**

Например, у современных моделей (в диапазоне цен более 2-3к\$) обычно хэш размером 72 бита (2*12 два тега vlan, для поддержки QinQ + полные 48 бит мак-адреса). В более бюджетном ряду все намного хуже, вплоть до 16 бит (10 на MAC, 6 на ViD) на 8-ми портовках типа D-Link 2108.

Ну а полный L2 Flow Specifications содержит Source MAC, Destination MAC, Vlan ID, Source Port - 48 + 48 + 12 + 5 = 113 бит. IP Flow Specifications - еще 114 бит, а с подсетями - плюс 85 бит. Т.е. мягко говоря есть куда совершенствоваться в плане управления передачей данных.

Так что нужно очень осторожно относиться к чрезмерно "раскачанным" свитчам, которые очень далеко отошли от своих одноклассников в плане заявленных возможностей. Чудес не бывает - можно предполагать, что функционал будет реализован процессором, либо отобран у "соседней" функции. Причем второе более вероятно, так как менее заметно.

Следующая деликатная часть - MMU (менеджер памяти). Получает порцию пакета с данными и адрес назначения на входящем порту, и через память и шину отправляет на исходящий. А еще пакует куски данных (оказывается, это очень важный момент) и работает с приоритетами.

В принципе, все достаточно просто и на первый взгляд очевидно. Если, конечно, не углубляться в детали, от которых это понимание тает как воск на турецком солнце. Впрочем, для любителей есть [великолепнейший документ по работе MMU](#), где описано все вплоть до алгоритмов.

Кстати, выяснились интересные нюансы. Например, реальных портов у свитч-фабрики обычно

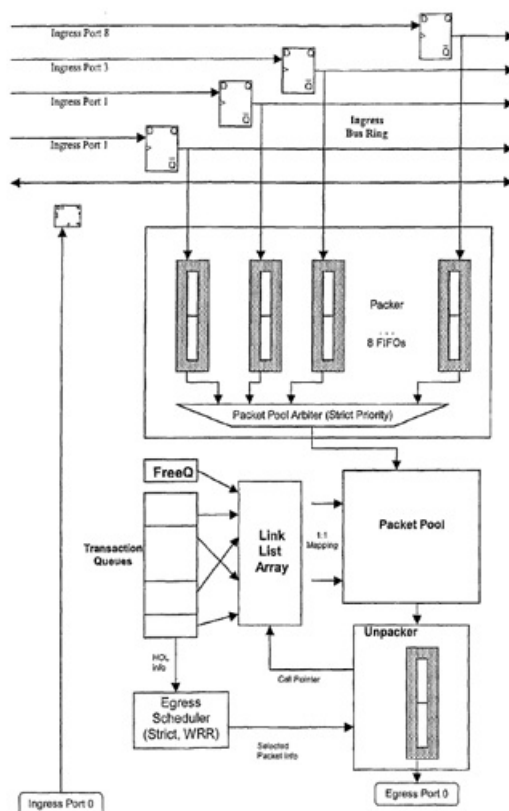
гораздо меньше, чем у ASIC (в рассмотренном примере 8 на 10G, а не 24 на 1G). Видимо, это сильно упрощает арбитраж.

А главное, нет в этом ASIC никакой коммутационной матрицы, сколько бы не вещали об этом прошлые переиздания Олиферов. Шина там, просто шина, 64-х битная круговая для данных (и если я правильно понял, есть отдельный диапазон для передачи заголовков).

Сразу становится понятно, зачем так развит в MMU механизм перепакровки частей данных (ячеек). Ведь каждый пакет не занимает шину целиком, работа идет параллельно по всем портам, и 64-х битные части по шине летают вперемешку, со всех портов, из (в) память, и на процессор.

Система сильно напоминает АТМ - даже термин "ячейки" используется самым активным образом.

Следующая проблема - QoS. Ключ к ним - опять в MMU, который работает по следующей схеме:



Интересно, что в буфера очередей попадают не пакеты целиком, а их части. Сборка идет уже после этого.

Хотя по слухам, на мыльницах порой встречается страшное. В 2-х уровневой очереди пропускается пакеты в соотношении 4:1. На каждый порт - счетчик на 4 бит, ничего сложного. В случае конфликта, пакет с низкоприоритетного порта "убивается", а счетчик на 1 инкрементируется. Если счетчик переполнился, уничтожается пакет высокого приоритета, счетчик сбрасываем. Это если вообще без всяких буферов и с минимумом программирования.

Теперь надо попробовать привязать это все к практике. Очень хочется собрать под проект приличный стенд, и скорее всего, это будет сделано в ближайшем будущем. Но для начала, нужно вспомнить уже замеченные "странности".

Вот несколько примеров, демонстрирующий свойства некоторых коммутаторов.

1. Management vlan на большинстве простых аппаратов имеет защиту от дураков (чтобы внезапно от кривых рук не терялось управление). Если прописать Management vlan к порту в tagged виде, а потом этот же порт прописать в другой (не управляемый) vlan в untagged виде, а на порту, в который включен тестируемый аппарат, оставить управляющий vlan в нетэгированном виде, то управление, как ни странно, не теряется!

Это говорит о том, что даже базовые функции ASIC можно легко "испортить" грубым "китайским" программированием. Хотя, они же могли напортачить и при разработке ASIC.

2. Знакомый провайдер умудряется принимать на управляемый коммутатор (из современных

китайских, с сертификатом CCC) поток с неуправляемой "мыльницы", в которую включено несколько обычных пользователей и фирма, выдающая пакеты с тегами (т.е. у них стоит свой специальный маршрутизатор). Этот невообразимый по стандартам винегрет приходит на один порт коммутатора, и более-менее корректно (!) на нем разбирается – т.е. Vlan опознается, на безтеговые пакеты ставится другой ViD, и все уходит в транк. Но это работает только в случае одного (!) Vlan, и не на всех версиях софта.

3. В "честных" аппаратах менять QoS не надо по умолчанию. Поток в аппарате обрабатывается в любом случае по пришедшему в заголовках его пакетов QoS. Он останется без изменения в случае транзитного QoS, или изменится на новый, согласно локальной политики QoS.

Но если аппарат в настройках по "умолчанию" (а куда без них сейчас?) "сбрасывает в нуль" поле QoS входящего пакета, и начинает руководствоваться исключительно только своей локальной политикой QoS при обработке пакетов и расстановке меток при их отправке, то... В цепочке коммутаторов достаточно одного такого устройства для спуска в унитаз общей политики QoS.

4. Пример теста (требуется заведомо "правильный" коммутатор.) В нем два Vlan 802.1q живут в одном порту каждого коммутатора. И одинаковые MACи, но с разными Vlan ID, есть на каждом коммутаторе с обеих сторон, т.е. MACи дублируются в разных Vlan.

1. Прописываем у тестируемого коммутатора vlan'ы:
 - vlan vid=3003, vlan vid=4004;
 - к vlan3003 добавляем port=1 frame=untag, port=3 frame=tag;
 - к vlan4004 добавляем port=2 frame=untag, port=3 frame=tag.
 - к vlan1(Default) добавляем port=1, 2 frame=tag, port=3 frame=untag.
2. На "правильном" (сервисном) коммутаторе прописываем:
 - vlan vid=3003, vlan vid=4004, vlan vid=2;
 - к vlan3003 добавляем port=1 frame=untag, port=3 frame=tag;
 - к vlan4004 добавляем port=2 frame=untag, port=3 frame=tag.
 - к vlan1(Default) добавляем port=1, 2 frame=tag, port=3 frame=untag;
 - к vlan2 добавляем port=4,5 frame=untag.
3. На тестируемом коммутаторе ставим короткий патчик port1 в port2. Так же замыкаем port3 тестируемого коммутатора и port3 сервисного.
4. В сервисный коммутатор в port1 и port2 включаем два компьютера и качаем фильм с одного на другой. Запоминаем время. Перетыкаем эти два компьютера в сервисный коммутатор в port4 и port5. Убираем патчик из port3 любого коммутатора. Качаем тот же фильм. Сравниваем результаты.

P.S. Прошедшие (под чутким руководством yamb'a) этот тест коммутаторы:

- Zcom - все с заявленной поддержкой 802.1Q.
- ATi - 85xx и выше - прошли. Не прошли 80xx - 83xx.
- Nortel - все с заявленной поддержкой 802.1Q.
- eXtrem - все с заявленной поддержкой 802.1Q.
- Cisco Catalyst - прошли, 3750. Не прошли 2950, 3550.
- D-Link - не прошли DES-30xx, 32xx, 33xx, 35xx.

Проверка должна показать:

1. Использование пришедшего в пакете tag'a для коммутации без всяких изменений.
2. Использование VlanID, как полноценного фактора для коммутации.
3. Отсутствие активности "навязываемых" производителями сервисов на Vlan управления (и, соответственно его защищенность). Хотя это Catalyst'a с настройками по-умолчанию не пройдет...

Есть многое на свете, друг Горацио... Остается лишь догадываться, какими еще скрытыми свойствами обладают коммутаторы - они стали более чем сложными устройствами, и закон Паркинсона оттягивается на них как Comedy Clab на поп-звездах.

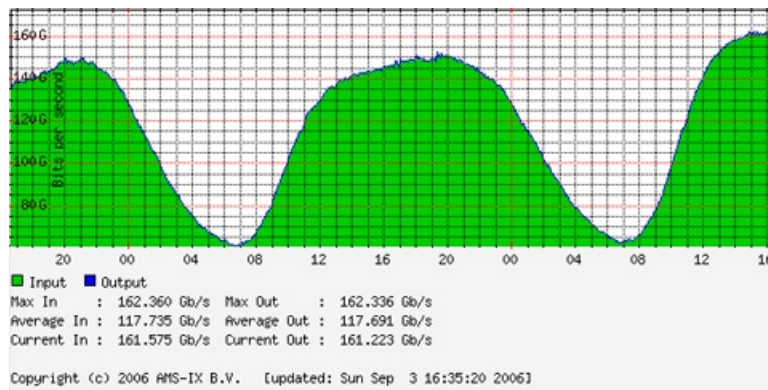
Остается только добавить, что все вышеописанное не слишком важно в сети относительно небольшого размера и простой архитектуры. Но для полносвязной MAN тонкости имеют вполне понятное прикладное значение.

Сейчас подавляющее число проблем, возникающих в больших сетях, решается на практике не путем изучения сути вопроса, а скорейшей перекомбинацией "железа", заменой, упрощением топологии. Пока это вполне эффективно, но сети растут, становятся все больше и сложнее. Растет и цена вопроса, цена ошибки. А с ней и важность хорошего и однозначного понимания процессов, происходящих внутри коммутаторов. Надеюсь, к этому сможет приблизить данная статья, или, в большей мере, форумная дискуссия по ее мотивам.

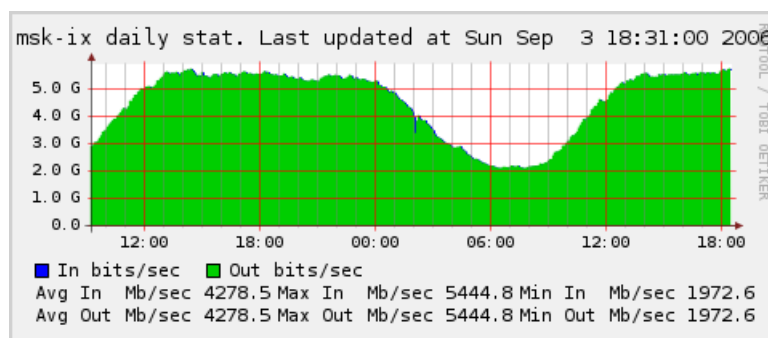
Разное

Для начала две картинки, которые лучше не показывать провайдерам со слабым сердцем. Поток в 160 гигабит в Амстердаме против 5 гигабит на М9.

AMS-IX Traffic Statistics (Daily graph) <http://www.ams-ix.net/technical/stats/>



MSK-IX. Статистика Daily graph (5 Minute Average) <http://www.msk-ix.ru/rus/tech/stat.shtm>



Несколько ссылок:

Два Новосибирских провайдера-конкурента получают общий контент-центр. "Электронный город", созданный альтернативным провайдером, открывает доступ "Сибирьтелекому". Понятно, до МРК дошло, что их АДСЛ мало интересен без дешевых локальных ресурсов. Понятно, что в Электронном городе много самостоятельных сервисов, которые нужно продавать как можно более широкому кругу пользователей.

Но зачем МРК нужна договоренность без самого нужного Webstream'у? Ведь абонентам других сетей будут доступны не все сервисы "Электронного города" - в их число не войдут IP-телевидение, файлообмен между абонентами и доступ к ftp-серверам. Это же просто праздник - возможность показать клиентам конкурента свою рекламу, привлечь внимание к недоступному, а потому сладкому плоду...

Не лучшее решение нашел и Красноярский филиал Сибирьтелекома. Там для создания "собственного" игрового портала привлекли на аутсорс совсем неизвестную фирму, и даже не расщедрились на оборудование: Все игровые и развлекательные серверы размещены на собственном оборудовании "Glanet Software"... Программное обеспечение представляет собой Unix/Linux-совместимые системы, с применением эмуляторов.

Такое впечатление, что МРК в очередной раз проиграли перспективное направление отрасли и пытаются наверстывать как умеют.

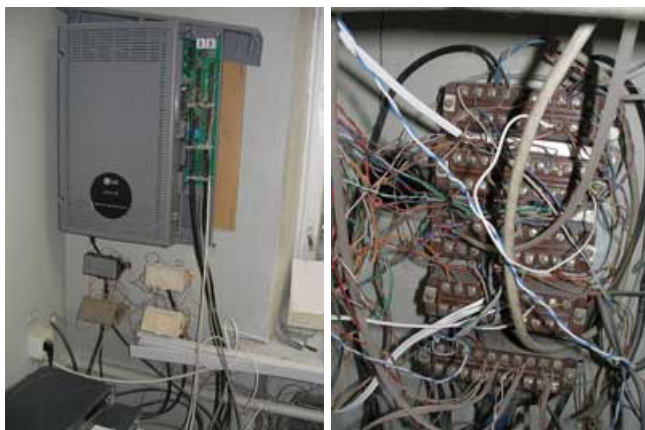
И еще несколько забавных урлов:

Биологическая атака на кабель

Закон любят нарушать не только коммерческие организации. ГУ ЦОДД ПМ завилиновали в создании радиопомех.

Пособие на тему как нельзя продвигать IPTV

Телефонные ужастики. АТС + телефонные КРТНы



Прислал **Юрок**

Обновление в разделах:

- Новые иллюстрации в [Фотографиях сетевого оборудования](#), [ATi AT-8948A](#), [Symbol AP 4131](#), [Cisco AP 1310G](#), [D-Link DSA 3100](#), [Nateks VoiceCom91-1](#).
- Новые сертификаты в [Сертификатах и лицензиях](#), [Cisco 7301 \(CCC\)](#).
- Новинки в [Коллекции инструкций к оборудованию на русском языке](#), [Практика использования тестера сети LinkRunner](#).
- В разделе [shop.nag.ru](#) продолжает пополняться **отделение патчкордов и пигтейлов**. Цены - наименьшие на Урале и в Сибири, например **шнур SC-SC от \$5,1**.

Зарисовки с натуры.

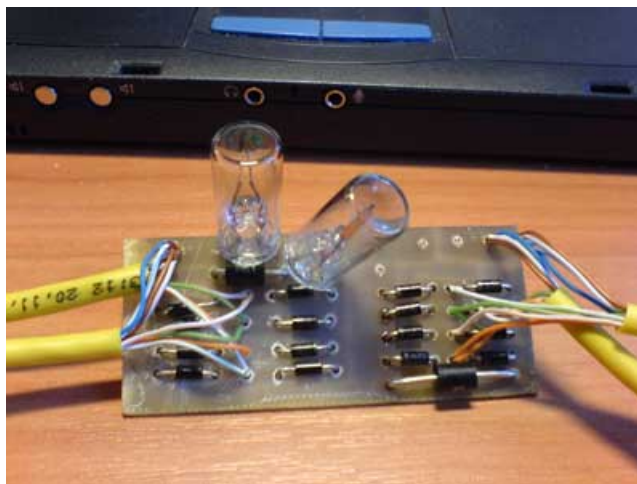
Куда деваются анлимиты. ;-) Сеть LANGNET, 7 пользователей и 1 сервер (интернет шлюз). ADSL, 512к, 2000 рублей в месяц.



- [Вид изнутри](#);
- [Открывание дверцы](#)

Прислал **Антон**, г. Лангепас ДС

Вот такую грозозащиту делают умельцы:



- [Процесс доводки;](#)
- [Крупный план;](#)
- [Вид на плату с обратной стороны.](#)

Прислал **Evgeniy**

... Вся сеть построена подвесом, потому что крайний север, вечная мерзлота, коллекторов практически нет, а те что остались со времен советской власти, демонтируют и засыпают, грунт тает.



На фото: Чердак после пожара в квартире на 9-ом этаже, город Дудинка. Сам чердак практически не пострадал, кроме кабелей. Полностью сгорел наш кабель - МКФ с вынесенным силовым элементом, 8 волокон Фуджикура, одномод.

На фотографии видно, что остались только волокна и трос. В таком состоянии связь работала 2 дня, до замены.



Кстати, сначала ЖКУ навязал платить по 3 тысячи рублей в месяц за один 5-ти этажный дом и 5 тысяч рублей за одну 9-ти этажку. Но удалось договориться по другому - у нас местной думой установлена цена аренды не жилого помещения, что-то в районе 80 рублей за 1 кв.м. В среднем у нас

получалось в районе 70 кв.сантиметром на каждом доме...

Прислал **Виктор**

Анонс

- Занимательная BRASология;
- Крыши и подвалов московских Новых Черемушек;
- Выгодно ли покупать небольшие сети? (сбор статистики, жду писем);
- Послегрозовой ремонт управляемых коммутаторов;
- Оптические магистрали Турции (фотографии);
- Традиционный пункт - ссылки на интересные места Сети. Присылайте письма - они очень нужны для обзоров. Обязательно сообщайте, нужна ли Ваша подпись, ссылка, или лучше обойтись без нее;
- В "ужастиках" - последствия пожара.

Долгострой:

- Пример настройки сервера для терминирования PPPoE (перенесен в долгострой);