

Some Internet Architectural Guidelines and Philosophy

Некоторые архитектурные рекомендации и философия Internet

Статус документа

Этот документ содержит информацию для сообщества Internet. Документ не задаёт каких-либо стандартов Internet и может распространяться без ограничений.

Авторские права

Copyright (C) The Internet Society (2002). Все права защищены.

Аннотация

Этот документ расширяет RFC 1958, очерчивая некоторые философские принципы, которых следует придерживаться архитекторам и проектировщикам магистральных сетей Internet. Описан принцип простоты (Simplicity Principle), который утверждает, что сложность является основным препятствием эффективному расширению и рассмотрено его влияние на проблемы архитектуры и устройства, возникающие на крупных магистралях Internet.

Оглавление

1. Введение.....	1
2. Большие системы и принцип простоты.....	2
2.1. «Сквозной аргумент» и простота.....	2
2.2. Нелинейность и сложность сети.....	2
2.2.1. Принцип усиления.....	2
2.2.2. Принцип взаимосвязи.....	3
2.3. Уроки телефонных сетей.....	3
2.4. Расходы на обновление сложных систем.....	3
3. Многоуровневая структура может быть вредной.....	3
3.1. Оптимизация может быть вредной.....	4
3.2. Изобилие функций может быть вредным.....	4
3.3. Развитие транспортной эффективности IP.....	4
3.4. Уровни схождения.....	4
3.4.1. Уровни транспортных протоколов.....	5
3.5. Эффекты второго порядка.....	5
3.6. Создание экземпляра модели EOSL с IP.....	5
4. Исключение универсального межсетевого взаимодействия.....	5
4.1. Исключение межсетевого взаимодействия в плоскости управления.....	6
5. Фундаментальные различия в коммутации пакетов и каналов.....	6
5.1. Действительно ли PS эффективней, чем CS, по своей природе?.....	6
5.2. Действительно ли PS проще, чем CS?.....	6
5.2.1. Сложность программ и микрокода.....	6
5.2.2. Сложность макроопераций.....	7
5.2.3. Сложность оборудования.....	7
5.2.4. Потребляемая мощность.....	7
5.2.5. Плотность.....	7
5.2.6. Фиксированные и переменные расходы.....	7
5.2.7. QoS.....	7
5.2.8. Гибкость.....	7
5.3. Относительная сложность.....	7
5.3.1. NBNI и OPEX.....	8
6. Миф об избыточности.....	8
7. Миф о пяти девятках.....	8
8. Закон пропорциональности архитектурных компонент.....	8
8.1. Пути доставки услуг.....	9
9. Заключение.....	9
10. Вопросы безопасности.....	9
11. Благодарности.....	9
12. Литература.....	9
13. Адреса авторов.....	11
14. Полное заявление авторских прав.....	11

1. Введение

В RFC 1958 [RFC1958] описаны базовые принципы архитектуры Internet. Данный документ расширяет эту работу, излагая некоторые философские принципы, которые следует применять архитекторам и проектировщикам магистральных сетей Internet. Хотя многие из описанных здесь областей могут противоречить друг другу, в документе приведён объединяющий принцип, контролирующий сложность как механизм влияния на стоимость и надёжность.

Сложность сетей операторов может возникать по разным причинам. Однако, как отмечено в [DOYLE2002]: «Сложность большинства систем обусловлена необходимостью устойчивости к неопределённым в средах и компонентах, а не базовой функциональностью». Таким образом, основная цель этого документа состоит в повышении осведомлённости о сложностях текущей архитектуры и изучении влияния этой сложности на успешность отрасли операторов IP.

Далее в документе рассматривается принцип простоты и его влияние на проектирование очень больших систем. В оставшейся части документа описаны базовые следствия принципа простоты и способы определения на его основе подходов к архитектуре и проектированию крупномасштабных сетей.

2. Большие системы и принцип простоты

Принцип простоты, сформулированный, возможно впервые, Mike O'Dell, бывшим главным архитектором (Chief Architect) UUNET, гласит, что сложность является основным препятствием эффективному расширению и, как следствие, основным фактором роста капитальных (CAPEX) и эксплуатационных (OPEX) расходов. Для IP-сетей операторов это означает, что в архитектуре и проектировании нужно выбирать простейшие из возможных решений.

2.1. «Сквозной аргумент» и простота

«Сквозной аргумент», описанный в [SALTZER] (а также в [RFC1958]), утверждает: «Сквозному протоколу не следует полагаться на поддержку состояния (т. е. сведений о состоянии сквозных коммуникаций) внутри сети. Такие состояния следует поддерживать лишь на конечных точках, чтобы состояние разрушалось лишь самой конечной точкой». Это свойство связано также с концепцией «общей судьбы» в работе Кларка [CLARK]. Можно видеть, что «сквозной принцип» напрямую ведёт к принципу простоты, исследуя формулировку архитектуры Internet, называемую «песочными часами» (hourglass) [WILLINGER2002]. В этой модели тонкая перемычка песочных часов рассматривается как (минималистичный) уровень IP, а все сложности добавляются над этим уровнем. Вкратце можно сказать, что сложности Internet возникают на краях, а уровню IP в Internet следует оставаться как можно более простым.

Следует отметить, что End-to-End Argument не означает, что в ядре Internet не будет поддерживаться никаких состояний. Фактически в ядре поддерживается огромное число состояний (например, состояние маршрутизации). Однако важно то, что эти состояния почти полностью ортогональны состояниям конечных точек (например, хостов). Именно такая минимизация способствует простоте. В результате учёт взаимодействия состояний в ядре и конечных точках становится критически важным при анализе протоколов, таких как трансляция адресов (Network Address Translation или NAT), снижающих степень прозрачности между сетью и хостами.

2.2. Нелинейность и сложность сети

Сложные архитектуры и конструкции были (и остаются) одними из самых значительных препятствий при построении рентабельных крупномасштабных сетей IP. Рассмотрим, например, задачу создания большой пакетной сети. Опыт показал, что построение такой сети требует иных действий (и иного набора навыков), нежели создание небольших и средних сетей и поэтому имеет другие свойства. В частности, крупнейшие сети демонстрируют в теории и на практике нелинейность архитектуры, устройства и инженерных аспектов, которая не проявляется при меньших размерах. Мы называем это ADE-нелинейностью (Architecture, Design, and Engineering). Таким образом, системы, подобные Internet, можно охарактеризовать как сильно разнородные (self-dissimilar) со значительно различающимися масштабами и уровнями абстракции [CARLSON]. ADE-нелинейность связана с двумя общеизвестными принципами теории нелинейных систем [THOMPSON], описанными ниже.

2.2.1. Принцип усиления

Принцип усиления (Amplification Principle) утверждает, что в больших системах имеются нелинейности, отсутствующие в мелких и средних системах.

Следствие. Во многих крупных сетях даже небольшие «вещи» могут вызывать очень крупные события. С точки зрения теории систем даже небольшие изменения на входе процесса в большой системе могут дестабилизировать её выход.

Важным примером принципа усиления является нелинейное резонансное усиление, которое может существенно трансформировать динамические системы, такие как большие сети, в результате представляющихся незначительными флуктуаций. Эти небольшие флуктуации могут медленно накапливаться, а при возникновении синхронизации вызывать серьёзные изменения. Резонансные явления являются примерами нелинейного поведения, когда небольшие изменения могут усиливаться и оказывать влияние, многократно превышающее начальный размер флуктуаций. В природе много примеров таких резонансов, таких как разрушение моста Tacoma Narrows в результате усиления небольших порывов ветра. Другими примерами являются разрывы в поясах астероидов и кольцах Сатурна, созданные нелинейным резонансным усилением. Некоторые особенности человеческого поведения и большинство систем паломничества подвержены влиянию резонансов, связанных с динамикой Солнечной системы, таких как световой день, сидерический (27,3 суток) и синодический (29,5 суток) цикл Луны, годичный (365,25 суток) цикл Солнца.

Для сети Internet было показано, что увеличение связности ведёт к усложнению и более частому замедлению схождения маршрутизации BGP [AHUJA]. Другое наблюдение говорит, что слабая связность делает вывод системы маршрутизации более сложным, чем ввод [GRIFFIN]. Важным способом снижения усиления является ограничение влияния локальных изменений локальной областью (в отличие от систем, где локальное изменение имеет глобальный эффект). ATM является прекрасным примером усиления - потеря одной ячейки ведёт к уничтожению всего пакета (а при отсутствии таких механизмов, как раннее отбрасывание пакетов - Early Packet Discard [ROMANOV] - будет продолжаться приём заведомо повреждённого пакета).

Другой интересный пример усиления из инженерной сферы описан в [CARLSON]. В работе рассмотрен самолёт Boeing 777, работающий по принципу «полёта по проводу» (fly-by-wire) и содержащий до 150 000 подсистем и около 1000 CPU. Отмечено, что, несмотря на устойчивость Boeing 777 к масштабным возмущениям атмосферы, границам зон турбулентности и изменению загрузки, полет может быть нарушен микроскопическими изменениями в нескольких больших CPU (к счастью, это происходит редко). Этот пример показывает, что: «сложность может усилить слабые возмущения и инженеры-конструкторы должны гарантировать, что такие возмущения чрезвычайно редки» [CARLSON].

2.2.2. Принцип взаимосвязи

Принцип взаимосвязи (Coupling Principle) гласит, что по мере увеличения объектов растёт уровень взаимозависимости между компонентами.

Следствие. Чем больше событий происходит одновременно, тем больше вероятность взаимодействия между ними. Это называют также «непредвиденным взаимодействием функций» (unforeseen feature interaction) [WILLINGER2002].

Значительная часть встречающихся в больших системах нелинейностей обусловлена взаимосвязями, которые могут быть вертикальными и горизонтальными. В контексте сетей горизонтальные связи возникают внутри одного протокольного уровня, а вертикальные - между уровнями протоколов.

Взаимосвязи проявляются в большом числе природных систем, включая микронеустойчивости плазмы (гидромагнитные, например, скручивание, отражение, вспучивание, разрыв, захват) [NAVE], а также различные виды электрохимических систем (нестандартный флюоресцентный синтез нуклеотидов, проблема маркировки нуклеиновых кислот [WARD]). Наблюдались также взаимосвязи временных циклов [JACOBSON], а также разных типов биологических циклов.

Существуют хорошо известные примеры и в сетевых системах, включая синхронизацию различных контуров управления, например, синхронизация маршрутных обновлений и TCP Slow Start [FLOYD, JACOBSON]. Важным результатом этих наблюдений является то, что взаимозависимости тесно связаны с синхронизацией. Внесение случайных факторов в эти системы является одним из способов снижения взаимозависимости. Интересно, что при анализе факторов риска для телефонных сетей общего пользования (Public Switched Telephone Network или PSTN) Charles Perrow разложил проблему сложности по двум связанным осям, которые он назвал взаимодействиями (interactions) и взаимозависимостью (coupling) [PERROW]. Perrow называет взаимодействия и взаимозависимости важными факторами, определяющими надёжность сложной системы (в частности, PSTN). В этой модели взаимодействием называются зависимости между компонентами (линейные или нелинейные), а взаимозависимости относятся к гибкости системы. Системы с простыми, линейными взаимодействиями имеют компоненты, влияющие лишь на другие компоненты, расположенные ниже в функциональном потоке. Компоненты сложной системы взаимодействуют со многими другими компонентами в разных и, возможно, удалённых частях системы. Считается, что системы со слабой взаимозависимостью более гибки в плане временных ограничений, упорядоченности и допущений о среде, нежели системы с сильными взаимозависимостями. Кроме того, системы со сложными взаимодействиями и сильной взаимозависимостью могут сталкиваться с непредвиденными отказами (сложные взаимодействия допускают большее число осложнений, которые сложнее понять и предсказать). Такое поведение описано в [WILLINGER2002]. Тесная взаимозависимость также означает снижение гибкости системы в плане восстановления при отказе.

Управление сетью PSTN SS7 является интересным примером неожиданного поведения в сильно связанной сложной сети. Такие отказы, как широко известный случай 1991 г. в сети AT&T SS7, являются показательными - сбой был вызван программной ошибкой в коде аварийного восстановления коммутатора. Когда коммутатор снова включился, он (и достаточно вероятное событие синхронизации) вызвал отказы соседей. Те при восстановлении работы вызвали отказы своих соседей и т. д. [NEUMANN] (основной причиной явился неуместный оператор break; это отличный пример межуровневой взаимозависимости). Это явление похоже на фазовую синхронизацию слабосвязанных генераторов, когда случайные вариации в последовательности играют важную роль в стабильности системы [THOMPSON].

2.3. Уроки телефонных сетей

В 1970-1980 гг. Телефонные операторы конкурировали между собой, добавляя новые функции, которые привели к значительному усложнению сетей PSTN, особенно в инфраструктуре коммутации класса 5. Эта сложность обычно была связана с программами, а не оборудованием, поэтому имеет кривые затрат значительно хуже закона Мура. Таким образом, низкая рентабельность голосовых систем сегодня является следствием того, что расходы OPEX и CAPEX не снижаются, как можно было бы ожидать от простых аппаратных реализаций.

2.4. Расходы на обновление сложных систем

Следует учитывать стоимость обеспечения новых функций в сложной сети. Традиционная телефонная сеть практически не включает интеллектуальных оконечных устройств (телефонных аппаратов) и имеет очень эффективное ядро. В Internet используются интеллектуальные оконечные устройства, компьютеры с операционными системами, приложения и т. п., включающее плоскость управления и модули пересылки пакетов. Добавление новой службы Internet это лишь вопрос распространения приложений на несколько согласованных оконечных систем (настольных компьютеров), которые используют службу. Сравните это с добавлением новой телефонной услуги, когда требуется менять ядро.

3. Многоуровневая структура может быть вредной

Применительно к сетям имеется несколько базовых свойств многоуровневой или вертикальной интеграции. В общем случае определённый в этом контексте уровень реализует одну или несколько указанных ниже функций.

Контроль ошибок

Повышает надёжность «канала». Например, это может быть гарантированный транспорт.

Управление потоком данных

Предотвращает лавинную передачу медленным партнёрам (например, управлением потоком данных в ATM).

Фрагментирование

Делит большие блоки данных на более мелкие части, а затем собирает их обратно (например, TCP MSS).

Мультиплексирование

Позволяет организовать несколько сессий более высокого уровня через одно «соединение» нижележащего уровня (например, ATM PVC).

Организация соединений

Согласование с партнёром (например, трехэтапное согласование TCP, ATM ILMI).

Адресация, именование

Нахождение и поддержка идентификаторов, связанных с объектами (например, GOSSIP 2 NSAP [RFC1629]).

Такие многоуровневые системы имеют ряд концептуальных и структурных преимуществ. Однако в контексте сетей данных структурированная многоуровневая система предполагает, что функции каждого уровня выполняются до того, как блок данных протокола будет передан на следующий уровень. Это означает, что на каждом уровне оптимизация выполняется отдельно. Такое ограничение порядка противоречит эффективной реализации функций манипулирования данными. Можно считать ответственной за такой конфликт многоуровневую модель (например, TCP/IP и ISO OSI). Операции мультиплексирования и сегментирования фактически скрывают важные сведения, которые могут потребоваться нижележащим уровням для оптимизации их работы. Функции данного и нижележащего уровня могут частично перекрываться, например, при поэтапном (hop-hop) и сквозном восстановлении после ошибок. Вдобавок, разным уровням может требоваться одна информация (например, временная метка), уровню N могут потребоваться данные уровня N-2 (например, размер пакетов на этом уровне) и т. п. [WAKEMAN]. Связанное с этим и ещё более ироничное замечание приведено в классической статье Tennenhouse «Layered Multiplexing Considered Harmful» [TENNENHOUSE]: «Подход ATM к широкополосным сетям с настоящее время применяется в CCITT (и других местах) как базовый механизм для поддержки интеграции служб, адаптации скорости и контроля временных вариаций на нижних уровнях сетевой архитектуры. Этот документ рассматривает временные вариации (jitter), связанные с концепцией «среднего» и «верхнего» уровня, которые работают в оконечных системах и ретрансляторах мультисервисных сетей (multi-service network или MSN).»

В результате межуровневых зависимостей рост числа уровней может быстро привести к нарушению принципа простоты. Отраслевой опыт показывает, что рост числа уровней часто ведёт к усложнению и росту OPEX, как предсказывает принцип простоты. Это следствие указано в разделе 2 (п. 5) RFC 1925 [RFC1925]:

«Всегда можно объединить несколько отдельных проблем в одно сложное решение со взаимозависимостями. В большинстве случаев это плохая идея».

Таким образом, первый вывод заключается в том, что горизонтальное (в отличие от вертикального) разделение является более рентабельным и надёжным в долгосрочной перспективе.

3.1. Оптимизация может быть вредной

Следствием приведённых выше аргументов для многоуровневых систем является то, что оптимизация также может быть вредной. В частности, оптимизация вносит сложности, а также усиливает взаимозависимости между компонентами и уровнями.

Важное и связанное с этим влияние оптимизации описывается законом снижающейся доходности (Law of Diminishing Returns), который гласит, что рост одного фактора производства при сохранении других в конечном итоге ведёт к относительному снижению общей доходности после определённого момента [SPILLMAN]. Подразумевается, что попытка выжать эффективность выше определённого уровня лишь добавляет сложности и снижает надёжность.

3.2. Изобилие функций может быть вредным

Хотя добавление новых свойств (функций) может считаться полезным (и на деле часто встречается у производителей различного оборудования), с этим связаны и опасности, заключающиеся в росте сложности системы.

3.3. Развитие транспортной эффективности IP

Развитие транспортной архитектуры для протокола IP даёт хорошие примеры того, как снижение вертикальной интеграции ведёт к разным уровням эффективности. Иллюстрация представлена ниже.

```
| IP через ATM в сети SONET  -->
| IP через SONET в сети WDM  -->
| IP через WDM
\|/
Снижение сложности, CAPEX и OPEX
```

Ключевым моментом здесь является снижение числа уровней, ведущее к уменьшению расходов CAPEX и OPEX.

3.4. Уровни схождения

Схождение связано с описанными выше концепциями уровней. Конечным состоянием «аргумента конвергенции» является концепция «все через некий уровень» (Everything Over Some Layer или EOSL). Канал, DWDM, волокно, ATM, MPLS и даже IP были предложены как уровни схождения. Важно отметить, что в результате повышения операционных расходов при создании уровней, ожидается рост OPEX и в результате схождения. Это наблюдение согласуется с отраслевой практикой.

Есть много примеров отказа уровня схождения, из которых возможно наиболее ярким является IP в сети ATM. Прямым и наиболее очевидным следствием многоуровневости ATM является так называемый «налог на ячейки» (cell tax). Во-первых, отметим, что эффективность ATM зависит от распределения размеров пакетов. Предположим, что это типичный трафик Internet, где высока доля пакетов размером 40, 44 и 552 байта. Недавние данные [CAIDA] показывают, что около 95% сетей WAN и 85% пакетов относятся к TCP. Большую часть трафика составляют пакеты размером 40 - 44 байта. Рассмотрим теперь магистраль DS3 со включённым PLCP. Максимальная скорость здесь составляет 96000 ячеек в секунду. Если умножить эту скорость на число полезных байтов в ячейке и размер байта, получится $96000 * 48 * 8 = 36,864$ Мбит/с. Однако такая скорость нереальна, поскольку она предполагает плотную упаковку полезных данных (payload). Есть ещё 2 фактора, повышающие издержки ATM (налог на ячейки) - бесполезное наполнение и 8-байтовый заголовок SNAP. Этот заголовок вызывает большинство проблем (и с этим ничего не сделать), вынуждая расходовать на мелкие пакеты 2 ячейки, из которых вторая содержит в основном наполнение (это очень плохо в упомянутом выше случае передачи множества пакетов TCP Ack размером 40 - 44 байта). Это вызывает потерю ещё примерно 16% из пропускной способности 36,8 Мбит/с. В итоге пропускная способность составит для DS3:

скорость в линии DS3:	44,736
издержки PLCP	- 4,032
заголовки ячеек	- 3,840
заголовки SNAP и наполнение	- 5,900
=====	
	30,960 Мбит/с

В результате при скорости линии DS3 в 44,736 Мбит/с суммарные издержки составляют около 31%.

Другой взгляд на это основан на том, что трафик WAN включает в основном пакеты TCP ACK. IP в сети ATM требует:

данные IP (40 байтов в данном случае)
8 байтов SNAP
8 байтов AAL5
5 байтов заголовка ячейки
+ байты для заполнения последней ячейки

В ATM такие пакеты превращаются в 2 ячейки общим размером 106 байтов, которые передают 40 байтов информации. Следующим наиболее распространенным размером пакетов является диапазон 504-556 байтов, 636 байтов IP, TCP и 512 байтов данных TCP, а на третьем месте идут сообщения с размером данных более 1000 байтов.

Можно представить, что 87% полезных данных (556 байтов сообщения) больше, чем 37% (TCP ACK), но это не 95-98%, к которым привыкли клиенты, а преобладание TCP ACK снижает среднее значение.

3.4.1. Уровни транспортных протоколов

Многоуровневые модели протоколов часто называют «X over Y», где протокол Y служит для передачи блоков данных (возможно и блоков управления) протокола X через плоскость данных Y, т. е. Y является уровнем схождения (convergence layer). Примеры включают Frame Relay через ATM, IP через ATM, IP через MPLS. Хотя решения X over Y имели лишь незначительный успех [TENNENHOUSE, WAKEMAN], было несколько случаев достижения высокой эффективности. В частности, эффективное решение X over Y можно реализовать при наличии своего рода изоморфизма между X и Y (т. е. имеется небольшой уровень схождения). В таких случаях данные X и возможно трафик управления инкапсулируются и передаются по протоколу Y. Примеры включают Frame Relay через ATM, а также Frame Relay, AAL5 ATM и Ethernet через L2TPv3 [L2TPV3]. Фактором упрощения здесь служит отсутствие требования общей синхронизации для конечных точек и минимизация межсетевого взаимодействия в плоскости управления. Альтернативой является межсетевое взаимодействие для данных и управления протоколов X и Y. Межсетевое взаимодействие в плоскости управления рассматривается ниже.

3.5. Эффекты второго порядка

IP через ATM демонстрирует отличный пример непредвиденных эффектов второго порядка. В частности, классическое исследование Romanov и Floyd для TCP [ROMANOV] в сети ATM показывает, что для обеспечения разумной производительности нужны большие (больше окна TCP) буферы UBR, механизмы отбрасывания пакетов, такие как упреждающее отбрасывание (Early Packet Discard или EPD), повышают эффективность использования пропускной способности, а для высокой эффективности и беспристрастности в узких местах (bottleneck) могут потребоваться более сложные службы и стратегии отбрасывания, нежели FIFO+EPD, такие как очереди и учёт VC. Хотя все исследования явно показывают, что нужен размер буфера не меньше одного окна TCP, число реально требуемых буферов зависит от применяемого механизма отбрасывания и этот вопрос остаётся открытым.

Примеров проблем такого типа в многоуровневых системах можно найти очень много в реальных сетях. Рассмотрим влияние неявных предположений транспорта IP (протокол TCP) о нижележащих протоколах.

- Потери пакетов. TCP предполагает, что потеря пакетов говорит о перегрузке, но потери иногда возникают в результате повреждения пакетов в беспроводных каналах [RFC3115].
- Нарушение порядка доставки. TCP предполагает, что существенное нарушение порядка доставки говорит о перегрузке. Это верно не во всех случаях [FLOYD2001].
- Время кругового обхода. TCP измеряет время кругового обхода и предполагает, что отсутствие подтверждения в интервале, основанном на измеренном времени, говорит о потере пакета (следовательно, о перегрузке) [KARN].
- Контроль перегрузок в TCP неявно предполагает одинаковую обработки в сети всех пакетов одного потока, но это верно не всегда [HANDLEY].

3.6. Создание экземпляра модели EOSL с IP

Хотя протокол IP предлагается для транспортировки почти всего, базовое представление о том, что «Все через IP» (Everything over IP или EOIP) приведёт к росту эффективности OPEX и CAPEX, требует критического пересмотра. В частности, несмотря на возможность эффективной доставки многих протоколов по сетям IP (в частности, протоколов, которым не требуется восстанавливать синхронизацию между конечными точками, таким как Frame Relay, Ethernet, AAL5 ATM), принципы простоты и многоуровневости предполагают, что EOIP может не быть наиболее эффективной стратегией схождения для произвольных служб. Скорее наиболее эффективный уровень схождения для CAPEX и OPEX может быть существенно ниже (это также подсказывает принцип простоты).

Примером, где EOIP не обеспечивает наиболее эффективный для OPEX и CAPEX транспорт, является ситуация, где службе или протоколу требуется время восстановления как в сети SONET (например, 50 мсек). Несложно представить, что создание и эксплуатация сети IP с такими свойствами (если это возможно) будет стоить больше, чем просто создание сети SONET.

4. Исключение универсального межсетевого взаимодействия

Хотя реализаций функции универсального межсетевого взаимодействия (Universal Interworking Function или UIWF) было много, подходы IWF оказались проблематическими в масштабе больших сетей. Это отмечено в принципе минимального вмешательства (Principle of Minimum Intervention) [BRYANT]:

«Для минимизации области видимости информации и повышения эффективности потока данных через уровень инкапсуляции содержимое (payload) следует передавать, по возможности, без изменений».

4.1. Исключение межсетевого взаимодействия в плоскости управления

Это следствие лучше всего понятно в контексте пространства интегрированных решений. В этом случае архитектура и проектирование часто приходят к худшему из возможных вариантов. Это связано с тем, что интегрированные решения плохо работают на обоих краях спектра производительность-CAPEX-OPEX - протоколы с наименьшей потребностью в коммутации могут нести издержки самых дорогих, а протоколам с более строгими требованиями приходится уступать протоколам с иными потребностями. При добавлении к этому проблем межсетевого взаимодействия в плоскости управления существенно возрастает вероятность отказов. В результате функции межсетевого взаимодействия приходится ограничивать плоскостью данных и инкапсуляцией, применяя их на краю сети.

Как описано выше, модели межсетевого взаимодействия оказались успешными в тех случаях, когда имеется некий «изоморфизм» между уровнями. Компромисс, часто называемый «Integrated vs. Ships In the Night trade-off¹» исследовался в разное время на разных уровнях протоколов. В общем случае имеется несколько эффективных вариантов интегрированных решений. Multi-protocol BGP [RFC2283] является несколько иным, но заметным исключением. В этом случае плоскость управления не зависит от формата данных управления, т. е. не требуется преобразования данных плоскости управления в отличие от моделей межсетевого взаимодействия в плоскости управления, таких как взаимодействие ATM/IP, представляемое некоторыми производителями программных коммутаторов, и так называемое взаимодействие PNNI-MPLS SIN [ATMMPLS].

5. Фундаментальные различия в коммутации пакетов и каналов

Принято считать, что коммутация пакетов (packet switching или PS) по своей природе более эффективна, нежели коммутация каналов (circuit switching или CS) в первую очередь за счёт статистического мультиплексирования и возможности принимать решения о пересылке независимо на каждом этапе (hop-by-hop) [MOLINERO2002]. Кроме того, широко распространено мнение, что PS² проще коммутации каналов, поэтому будет более экономично при развёртывании и эксплуатации [MCK2002]. Однако при изучении этих и связанных с ними допущений возникает иная картина (см., например, [ODLYZKO98]). Упомянутые допущения рассматриваются в последующих параграфах.

5.1. Действительно ли PS эффективней, чем CS, по своей природе?

Хорошо известно, что коммутаторы пакетов эффективно используют дефицитную пропускную способность [BARAN]. Эта эффективность основана на статистическом мультиплексировании, присущем коммутации пакетов. Однако по-прежнему озадачивает то, что обычно считают низкой загрузкой магистралей Internet. Первый вопрос связан с текущей загрузкой магистралей Internet и её соотношением с загрузкой сетей дальней телефонной связи. Odlyzko Coffman [ODLYZKO,COFFMAN] сообщает, что средняя загрузка каналов в сетях IP составляет от 3% до 20% (корпоративные сети используют около 3%, коммерческие магистрали Internet - 15-20%). Средняя загрузка телефонных сетей дальней связи составляет около 33%. Кроме того, в 2002 г. средняя загрузка оптических сетей (все услуги) составляла около 11%, а средняя долгосрочная загрузка - около 15% [ML2002]. Возникает вопрос, чем обусловлены такие уровни загрузки, особенно с учётом допущения о большей эффективности PS по сравнению с CS. Причины, отмеченные Odlyzko и Coffman включают:

- (i) асимметричную природу трафика Internet с пиками и провалами, тогда как каналы симметричны и имеют фиксированную ёмкость (т. е. неизвестна матрица трафика или требуемая пропускная способность каналов);
- (ii) сложно предсказать рост трафика на канале, поэтому операторы склонны обеспечивать избыточную пропускную способность;
- (iii) падение цен при заказе большей пропускной способности делает более выгодным расширение полосы на значительную величину.

К другим статическим факторам относятся издержки протоколов, дискретность оборудования, возможности восстановления и временной интервал предоставления. Совместно они определяют необходимость избыточного выделения ресурсов (over-provision) [MC2001].

5.2. Действительно ли PS проще, чем CS?

Сквозной принцип можно считать утверждением о том, что сложности Internet находятся на краях (границах). Однако современные магистральные маршрутизаторы Internet очень сложны и сложность зависит от скорости линий. Поскольку относительная сложность коммутации каналов и пакетов, похоже, не поддаётся сравнительному анализу, вместо этого будет рассмотрено несколько артефактов коммутации каналов и пакетов как меры сложности. Среди показателей, на которые можно обратить внимание, находится сложность программ и оборудования, потребляемая мощность и плотность. Каждый из этих показателей рассматривается ниже.

5.2.1. Сложность программ и микрокода

Одним из показателей сложности программ и микрокода является число инструкций, требуемых для программирования устройства. Типичный образ программ для маршрутизатора Internet включает 8-10 миллионов инструкций (включая микрокод), а типичный транспортный коммутатор - примерно 8 миллионов инструкций [MCK2002].

Такое различие в сложности программ делает маршрутизаторы Internet ненадёжными и имеет дополнительные заметные эффекты второго порядка (например, долгая перезагрузка маршрутизатора). Другой точкой сравнения может служить коммутатор AT&T (Lucent) 5ESS класса 5, имеющий огромное число функций вызова и требующий примерно вдвое больше строк кода, нежели маршрутизатор ядра Internet [EICK].

Поскольку маршрутизаторы содержат программного кода не меньше, чем аппаратные системы, ещё одним эффектом сложности кода является то, что стоимость маршрутизаторов меньше выигрывает от закона Мура, чем содержащие меньше программного кода устройства. Это ведёт к компромиссу между пропускной способностью и аппаратной частью, отдавая преимущество пропускной способности.

¹Интеграция в сравнении с «кораблями в ночи».

²В оригинале ошибочно указано IP, см. <https://www.rfc-editor.org/errata/eid5608>. Прим. перев.

5.2.2. Сложность макроопераций

Линейные платы маршрутизаторов Internet должны выполнять много сложных операций, включая обработку заголовков пакета, поиск максимального соответствия префикса, генерацию сообщений ICMP об ошибках, обработку опций заголовка IP и буферизацию пакета для эффективного контроля перегрузок TCP (обычно это требует буфера, размер которого пропорционален произведению скорости линии на RTT, или примерно 250 мсек из данных пакета). Здесь не указана фильтрация маршрутов и пакетов, а также фильтрация QoS и VPN.

Транспортный коммутатор должен лишь отображать входные временные интервалы (time-slot) на выходные временные интервалы или интерфейсы, поэтому может быть значительно менее сложным.

5.2.3. Сложность оборудования

Одним из показателей сложности оборудования является число логических вентилях (транзисторов) на линейной плате [MOLINERO2002]. Плата высокоскоростного маршрутизатора Internet для линии OC192 POS содержит по меньшей мере 30 миллионов вентилях в микросхемах ASIC, хотя бы 1 CPU, 300 Мбайт пакетных буферов, 2 Мбайта таблиц пересылки и 10 Мбайт памяти состояний. Линейная плата сравнимого транспортного коммутатора имеет 7,5 миллионов логических вентилях без CPU, буфера пакетов, таблиц пересылки и памяти состояний. Скорей такая плата будет включать блок кадрирования SONET, микросхему для сопоставления входных временных интервалов с выходными и интерфейс к модулю коммутации (switch fabric).

5.2.4. Потребляемая мощность

Поскольку в транспортных коммутаторах традиционно применяется более простое оборудование, они потребляют меньшую мощность [PMC].

5.2.5. Плотность

Наиболее производительные транспортные коммутаторы имеют примерно в 4 раза большую пропускную способность по сравнению с маршрутизаторами IP [CISCO, CIENA] и примерно втрое меньшую стоимость в расчёте на Гбит/с. Оптические технологии (ООО) дополнительно увеличивают различия в сложности (например, перестраиваемые лазеры, коммутаторы MEMS, например, [CALIENT]), а мультиплексоры DWDM обеспечивают технологию создания транспортных коммутаторов с очень высокой производительностью и малой потребляемой мощностью.

Связанной с этим метрикой являются физические размеры. В общем случае транспортные коммутаторы имеют меньший размер в расчёте на Гбит/с.

5.2.6. Фиксированные и переменные расходы

Коммутация пакетов, по видимому, имеет высокие переменные затраты, а это означает, что отправка n -й части информации с использованием коммутации пакетов обходится дороже, чем при коммутации каналов. Большая часть этого связана с относительно статичной природой коммутации каналов (например, может использоваться заранее планируемое прибытие информации для снижения объёма операций, выполняемых при поступлении данных). В случае коммутации каналов не требуется буферизовать входные данные, выполнять обнаружение петель, определять следующий интервал пересылки (hop), менять поля в заголовках и т. п.. Кроме того, многие сети с коммутацией каналов сочетают более или менее статичную конфигурацию с отдельными (out-of-band) плоскостями управления (например, SS7), что существенно упрощает коммутацию в плоскости данных. Суть в том, что по мере роста скорости данных коммутация пакетов быстро усложняется, тогда как рост сложности коммутации каналов остаётся более или менее линейным.

5.2.7. QoS

Хотя компоненты полного решения для Internet QoS, включая контроль вызовов, эффективную классификацию пакетов и алгоритмы планирования широко исследовались и стандартизовались уже более 10 лет, сквозной сигнализации QoS для Internet до сих пор нет. С другой стороны, QoS с самого начала является частью инфраструктуры коммутации каналов. При этом QoS обычно развёртывается для задания дисциплин очередей, которые следует применять при недостатке пропускной способности для передачи трафика. В отличие от голосового трафика отбрасывание пакетов или значительные задержки для TCP могут оказывать более существенное влияние из-за наличия контуров обратной связи для перегрузок (в частности, TCP backoff/slow start).

5.2.8. Гибкость

Присущую Internet гибкость сложнее оценить количественно. Хотя эта гибкость привела к быстрому росту Internet. С ней сопряжены сравнительно высокие затраты на периферии, связанные с потребностью в квалифицированном персонале. Стандартное эмпирическое правило состоит в том, что в условиях предприятия достаточно одного человека в службе поддержки для обеспечения телефонной связи в группе, тогда как для поддержки сетевых коммуникаций в той же группе требуется десяток специалистов по компьютерным сетям [ODLYZKO98A]. Это явление описано также в [PERROW].

5.3. Относительная сложность

Вычислительную сложность коммутации каналов по сравнению с коммутацией пакетов сложно описать формально [PARK]. Поэтому в предшествующих разделах сложность описывалась с точки зрения наблюдаемых эффектов. С учётом этого становится ясно, что основной причиной роста отмеченной выше сложности является присущая архитектуре IP «поэтапная независимость» (hop-by-hop independence или HBHI). Это отличается от сквозных архитектур, таких как ATM или Frame Relay.

В [WILLINGER2002] это описано с точки зрения требований отказоустойчивости исходного проекта Internet и влияния этого на сложность сети. В частности рассматривается «спираль сложность-отказоустойчивость», где рост сложности ведёт к повышению чувствительности, требующему дополнительной отказоустойчивости (спираль).

Важным уроком этого раздела является то, что принцип простоты (Simplicity Principle) применительно к коммутации каналов и пакетов имеет решающее значение для контроля сложности (и, следовательно, свойств OPEX и CAPEX) в

пакетных сетях. Это подкрепляется наблюдением, говорящим о том, что хотя коммутация пакетов «моложе» и является менее зрелой по сравнению с коммутацией каналов, к коммутации пакетов присутствует тенденция усложнения линейных плат в то время как сложность коммутации каналов растёт линейно с повышением скорости и агрегированной пропускной способности.

5.3.1. NBNI и OPEX

NBNI требует подходить к сетям IP принципиально иначе, нежели к сетям с коммутацией каналов. В частности основным вопросом операционных расходов (OPEX) для сетей IP является то, что отладка крупномасштабных IP-сетей по-прежнему требует высокого уровня опыта и понимания из-за «поэтапной независимости», присущей пакетной архитектуре (ещё раз отметим, что такой независимости нет в сетях с коммутацией каналов, таких как ATM и Frame Relay). Например, может потребоваться обращение к большому числу маршрутизаторов для обнаружения проблемы внешней по отношению к вашей сети. Кроме того, используемые для диагностики инструменты отладки достаточно сложны и кое в чём примитивны. В сетях IP приходится иметь дело с людьми, у которых возникают проблемы с их DNS, почтовыми или новостными службами, а также с некоторыми новыми приложениями, которые обычно не проявляются в TDM, ATM и т. п. В случае IP проблему можно упростить за счёт автоматизации (отметим, что многие из упомянутых проблем обусловлены взаимодействием с клиентами). В общем случае имеется много внешних по отношению к сети переменных, которые влияют на OPEX.

Важно подчеркнуть, что количественные взаимосвязи между CAPEX, OPEX и присущей сетям сложностью пока изучены недостаточно. Фактически нет согласованных количественных показателей для описания сложности сетей, поэтому точные связи между CAPEX, OPEX и сложностью остаются неясными.

6. Миф об избыточности

Как отмечено в [MC2001] и других работах, большая часть сложностей, наблюдаемых сегодня в Internet, связана с ростом используемой пропускной способности. В результате желание сетевых инженеров сохранить загрузку сети ниже 50% получило название «избыточность» (over-provisioning). Однако такое использование термина является неверным. Неиспользуемая в современных магистралах Internet пропускная способность фактически является защитным свойством. В частности, это можно считать «защитой 1:1 на уровне IP» (1:1 protection at the IP layer). С этой точки зрения сеть IP, предназначенная для работы с загрузкой 50%, предоставляет избыточность не выше, чем обычная сеть SONET. Однако важным преимуществом такой IP-сети является малая задержка (скорость близка к скорости света) и почти нулевые потери пакетов [FRALEIGH]. Эти преимущества можно считать «побочным эффектом» защиты 1:1.

Имеются также системно-теоретические причины обеспечения защиты в стиле 1:1, среди которых наиболее заметным является то, что сети коммутации пакетов с контуром управления по основному каналу становятся нестабильными и склонными к осцилляциям и «синхронизации» при перегрузке. Сложные и нелинейные взаимодействия трафика ведут к влиянию перегрузки в одной части сети на другие участки. При потере пакетов протокола маршрутизации в результате перегрузки в сети или процессоре обработки маршрутов возникает несогласованность маршрутизации, которая может привести к возникновению петель и «чёрных дыр» или потере связности. Хотя статистическое мультимплексирование теоретически может обеспечить более высокую загрузку сети, на практике для поддержки согласованной производительности и достаточно стабильной работы сети динамика магистралей Internet благоприятствует защите 1:1 и её побочным эффектам для поддержания стабильности сети и малых задержек.

7. Миф о пяти девятках

Paul Baran в своей классической статье «SOME PERSPECTIVES ON NETWORKS--PAST, PRESENT AND FUTURE» говорит: «Компромисс между стоимостью и надёжностью системы предполагает, что самые надёжные системы можно построить из сравнительно ненадёжных (следовательно, недорогих) элементов, если речь идёт о надёжности системы при самой низкой общей стоимости» [BARAN77].

Сегодня это называют «мифом о пяти девятках». В частности, так называемая «надёжность 99,999» для элементов пакетной сети считается мифом по нескольким причинам. Во-первых, 80% незапланированных простоев вызваны ошибками людей или процессов [SCOTT] и для оптимизации остаётся лишь 20%. Таким образом, для повышения надёжности компонентов добавляется сложность (оптимизация зачастую ведёт к усложнению), которая является причиной 80% неплановых простоев. Это фактически сужает окно 20% (т. е. повышает вероятность отказов, связанных с людьми и процессами). Это явление характеризуется «спиралью сложность-отказоустойчивость» [WILLINGER2002], где рост сложности повышает чувствительность, что требует повышения отказоустойчивости (спираль). Вывод заключается в том, что, хотя системы, подобные Internet, могут обеспечивать надёжность 99,999, нежелательно (и, вероятно, невозможно) пытаться обеспечить такую надёжность любому отдельному элементу (особенно сложному).

8. Закон пропорциональности архитектурных компонент

Как отмечено в предыдущем параграфе, вычислительные сложности сетей с коммутацией пакетов, таких как Internet, сложно описать в формальных терминах. Однако интуитивно понятное высокоуровневое определение сложности архитектуры может состоять в том, что сложность архитектуры пропорциональна числу компонентов, а вероятность получения стабильной реализации архитектуры обратно пропорциональна числу компонентов. Как описано выше, компоненты включают дискретные элементы, такие как элементы оборудования, требования к площади и питанию, а также программы, микрокод и реализуемые протоколы. В более абстрактном виде это представлено ниже.

Пусть

- A - представление архитектуры A;
- |A| - число различных компонентов в пути доставки услуг архитектуры A;
- w - монотонно возрастающая функция;
- P - вероятность стабильной реализации архитектуры.

Тогда

$$\text{Complexity}(A) = O(w(|A|))$$

$$P(A) = O(1/w(|A|))$$

где $O(f) = \{g: N \rightarrow R \mid \text{существует } c > 0 \text{ и } n, \text{ такие что } g(n) < c \cdot f(n)\}$

Т. е. $O(f)$ включает набор функций g , для которого имеется константа c и число n , такие что $g(n)$ не больше $c \cdot f(n)$ для всех n . Т. е. $O(f)$ - набор всех функций, которое не растёт быстрее f без учёта постоянных факторов.

Интересно, что модель HOT (Highly Optimized Tolerance) [HOT] пытается охарактеризовать сложность в общих терминах (HOT является одной из попыток создать общую схему для изучения сложности и относится к семейству абстракций, называемому «новой наукой о сложностях» или «комплексными адаптивными системами»). Терпимость (tolerance) в семантике HOT означает, что «надёжность в сложных системах является ограниченной величиной, которая требует осторожного управления и защиты». Одной из целей модели HOT является характеристика распределений с «тяжёлым хвостом», таких как Complexity(A) в приведённом выше примере (другие примеры включают лесные пожары, отключение питания и распределение трафика в Internet). В частности, Complexity(A) пытается сопоставить предельную неоднородность частей системы (Internet) и влияние их организации на сильно структурированные сети с иерархией и множеством уровней.

8.1. Пути доставки услуг

Закон пропорциональности архитектурных компонентов (Architectural Component Proportionality Law или ACPL) гласит, что сложность архитектуры пропорциональна числу её компонентов. Следствием этого является минимизация числа компонентов в пути доставки услуг, которым может служить протокольный, программный или физический путь.

Это следствие является важным выводом из ACPL, поскольку путь между потребителем и желаемой услугой особо чувствителен к числу и сложности элементов пути. Это связано с тем, что сложность «сглаживания» при высоком уровне агрегирования [ZHANG] исчезает при приближении к краю, а также при сложных взаимодействиях с бэк-офисом и системами CRM. Примерами архитектур, которые не нашли применения из-за этого эффекта, включают CRM на основе TINA и архитектуры служб на базе CORBA/TINA. Основным уроком состоит в том, что единственными вариантами развёртывания этих систем были «Ограниченное развёртывание, такое как Starvision, где можно избежать проблем, связанных с расширением» или «Огромные инвестиции, как при сборке с нуля ORB операторского класса» [TINA]. Иными словами, эти системы имели сложный путь доставки услуг и были слишком сложны для развёртывания.

9. Заключение

В этом документе предпринята попытка систематизации известных принципов архитектуры Internet. В частности, описанный здесь принцип объединения лучше всего выражается принципом простоты (Simplicity Principle), который гласит, что сложность должна контролироваться, если нужна эффективная расширяемость сложного объекта. Идея о том, что простота сама может вести к некоей оптимизации, была общей в течение долгого времени и выражалась разными способами и в разных измерениях. Рассмотрим, например, «бритву Оккама» (Occam's Razor) - принцип сформулированный средневековым английским философом и монахом-францисканцем Уильямом Оккамом (William of Ockham, 1285-1349), гласящий: «Pluralitas non est ponenda sine neccesitate» или «plurality should not be posited without necessity» (Occam's Razor иногда называют принципом отказа от множественности или принципом простоты)¹. Возможно, более современная формулировка гласит, что простейшим объяснением явления является предпочитаемое природой. Другая формулировка той же идеи представлена в принципе сохранения простоты (KISS или Keep It Simple Stupid) и принципе наименьшего удивления (Principle of Least Astonishment), считающем наиболее подходящей систему, которая менее всего вызывает удивление. В [WILLINGER2002] приведено теоретическое обсуждение «отказоустойчивости за счёт простоты» (robustness through simplicity), а в обсуждении PSTN [KUNH87] отмечено, что во многих системах «возможен компромисс между простотой взаимодействия и взаимозависимостью».

Применительно к архитектуре сети с коммутацией пакетов некоторые следствия принципа простоты могут показаться «ересью». Например, подходы с высоким уровнем конвергенции будут явно менее эффективными, чем решения с меньшей конвергентностью. Иными словами, «оптимальный» уровень конвергентности может оказаться в стеке протоколов существенно ниже, чем принято считать. Кроме того, приведённый выше анализ ведёт к нескольким выводам, противоречащим традиционным представлениям о пакетных сетях. Возможно, наиболее важна убеждённость в том, что коммутация пакетов проще коммутации каналов. На основе этого убеждения возникло представление о том, что работа с пакетами дешевле работы с устройствами, поскольку пакет проще канала (устройства). Данный документ утверждает обратное. В частности, исследование описанных выше показателей показывает, что коммутация пакетов сложнее, чем коммутация каналов. Интересно, что этот вывод подтверждается тем фактом, что нормализованные показатели OPEX для сетей передачи данных обычно существенно выше, чем для голосовых (телефонных) сетей [ML2002].

Наконец, важным заключением этой работы является то, что для пакетных сетей масштаба современной сети Internet и крупнее нужно стремиться к максимально простым решениям для построения максимально эффективной экономически инфраструктуры. Это красноречиво изложено в [DOYLE2002]: «Развитие протоколов может привести к спирали отказоустойчивость-сложность-хрупкость, где сложность добавляется для отказоустойчивости, что ведёт к росту хрупкости и новому витку спирали». Это как раз то явление, предотвращению которого служит принцип простоты.

10. Вопросы безопасности

Документ не оказывает прямого влияния на безопасность имеющихся протоколов Internet. Однако соблюдение принципа простоты напрямую влияет на возможность внедрения защищённых систем. В частности, по мере роста сложности системы её труднее анализировать и моделировать, поэтому усложняется поиск и понимание последствий для безопасности, связанных с архитектурой, устройством и реализацией.

11. Благодарности

Многие идеи по сравнению сетей с коммутацией пакетов и каналов были предложены в разговорах с Nick McKeown, Scott Bradner, David Banister, Steve Bellovin, Steward Bryant, Christophe Diot, Susan Harris, Ananth Nagarajan, Andrew Odlyzko, Pete и Natalie Whiting, Lixia Zhang предоставили полезные замечания к ранней версии документа.

12. Литература

[АНУJA] "The Impact of Internet Policy and Topology on Delayed Routing Convergence", Labovitz, et. al. Infocom, 2001.

¹По русски часто это формулирует: «не умножать сущности сверх необходимого». Прим. перев.

- [ATMMPLS] "ATM-MPLS Interworking Migration Complexities Issues and Preliminary Assessment", School of Interdisciplinary Computing and Engineering, University of Missouri-Kansas City, April 2002
- [BARAN] "On Distributed Communications", Paul Baran, Rand Corporation Memorandum RM-3420-PR, <http://www.rand.org/publications/RM/RM3420>, August, 1964.
- [BARAN77] "SOME PERSPECTIVES ON NETWORKS--PAST, PRESENT AND FUTURE", Paul Baran, Information Processing 77, North-Holland Publishing Company, 1977,
- [BRYANT] "Protocol Layering in PWE3", Bryant et al, Work in Progress¹.
- [CAIDA] <http://www.caida.org>
- [CALLIENT] <http://www.caliient.net/home.html>
- [CARLSON] "Complexity and Robustness", J.M. Carlson and John Doyle, Proc. Natl. Acad. Sci. USA, Vol. 99, Suppl. 1, 2538-2545, February 19, 2002. <http://www.pnas.org/cgi/doi/10.1073/pnas.012582499>
- [CIENA] "CIENA Multiwave CoreDirector", <http://www.ciena.com/downloads/products/coredirector.pdf>
- [CISCO] <http://www.cisco.com>
- [CLARK] "The Design Philosophy of the DARPA Internet Protocols", D. Clark, Proc. of the ACM SIGCOMM, 1988.
- [COFFMAN] "Internet Growth: Is there a 'Moore's Law' for Data Traffic", K.G. Coffman and A.M. Odlyzko, pp. 47-93, Handbook of Massive Data Stes, J. Elli, P. M. Pardalos, and M. G. C. Resende, Editors. Kluwer, 2002.
- [DOYLE2002] "Robustness and the Internet: Theoretical Foundations", John C. Doyle, et. al. Work in Progress.
- [EICK] "Visualizing Software Changes", S.G. Eick, et al, National Institute of Statistical Sciences, Technical Report 113, December 2000.
- [MOLINERO2002] "TCP Switching: Exposing Circuits to IP", Pablo Molinero-Fernandez and Nick McKeown, IEEE January, 2002.
- [FLOYD] "The Synchronization of Periodic Routing Messages", Sally Floyd and Van Jacobson, IEEE ACM Transactions on Networking, 1994.
- [FLOYD2001] "A Report on Some Recent Developments in TCP Congestion Control, IEEE Communications Magazine, S. Floyd, April 2001.
- [FRALEIGH] "Provisioning IP Backbone Networks to Support Delay-Based Service Level Agreements", Chuck Fraleigh, Fouad Tobagi, and Christophe Diot, 2002.
- [GRIFFIN] "What is the Sound of One Route Flapping", Timothy G. Griffin, IPAM Workshop on Large-Scale Communication Networks: Topology, Routing, Traffic, and Control, March, 2002.
- [HANDLEY] "On Inter-layer Assumptions (A view from the Transport Area), slides from a presentation at the IAB workshop on Wireless Internetworking", M. Handley, March 2000.
- [HOT] J.M. Carlson and John Doyle, Phys. Rev. E 60, 1412-1427, 1999.
- [ISO10589] "Intermediate System to Intermediate System Intradomain Routing Exchange Protocol (IS-IS)".
- [JACOBSON] "Congestion Avoidance and Control", Van Jacobson, Proceedings of ACM Sigcomm 1988, pp. 273-288.
- [KARN] "TCP vs Link Layer Retransmission" in P. Karn et al., Advice for Internet Subnetwork Designers, Work in Progress.
- [KUHN87] "Sources of Failure in the Public Switched Telephone Network", D. Richard Kuhn, EEE Computer, Vol. 30, No. 4, April, 1997.
- [L2TPV3] Lau, J., et. al., "Layer Two Tunneling Protocol (Version 3) -- L2TPv3", Work in Progress².
- [MC2001] "U.S Communications Infrastructure at A Crossroads: Opportunities Amid the Gloom", McKinsey&Company for Goldman-Sachs, August 2001.
- [MCK2002] Nick McKeown, personal communication, April, 2002.
- [ML2002] "Optical Systems", Merrill Lynch Technical Report, April, 2002.
- [NAVE] "The influence of mode coupling on the non-linear evolution of tearing modes", M.F.F. Nave, et al, Eur. Phys. J. D 8, 287-297.
- [NEUMANN] "Cause of AT&T network failure", Peter G. Neumann, <http://catless.ncl.ac.uk/Risks/9.62.html#subj2>
- [ODLYZKO] "Data networks are mostly empty for good reason", A.M. Odlyzko, IT Professional 1 (no. 2), pp. 67-69, Mar/Apr 1999.
- [ODLYZKO98A] "Smart and stupid networks: Why the Internet is like Microsoft". A. M. Odlyzko, ACM Networker, 2(5), December, 1998.
- [ODLYZKO98] "The economics of the Internet: Utility, utilization, pricing, and Quality of Service", A.M. Odlyzko, July, 1998. <http://www.dtc.umn.edu/~odlyzko/doc/networks.html>
- [PARK] "The Internet as a Complex System: Scaling, Complexity and Control", Kihong Park and Walter Willinger, AT&T Research, 2002.
- [PERROW] "Normal Accidents: Living with High Risk Technologies", Basic Books, C. Perrow, New York, 1984.

¹Работа опубликована в RFC 3985. Прим. перев.²Работа опубликована в RFC 3931. Прим. перев.

- [PMC] "The Design of a 10 Gigabit Core Router Architecture", PMC-Sierra, http://www.pmc-sierra.com/products/diagrams/CoreRouter_lg.html
- [RFC1629] Colella, R., Callon, R., Gardner, E. and Y. Rekhter, "Guidelines for OSI NSAP Allocation in the Internet", RFC 1629, May 1994.
- [RFC1925] Callon, R., "The Twelve Networking Truths", RFC 1925, 1 April 1996.
- [RFC1958] Carpenter, B., Ed., "Architectural principles of the Internet", [RFC 1958](#), June 1996.
- [RFC2283] Bates, T., Chandra, R., Katz, D. and Y. Rekhter, "Multiprotocol Extensions for BGP4", [RFC 2283](#), February 1998.
- [RFC3155] Dawkins, S., Montenegro, G., Kojo, M. and N. Vaidya, "End-to-end Performance Implications of Links with Errors", BCP 50, RFC 3155, May 2001.
- [ROMANOV] "Dynamics of TCP over ATM Networks", A. Romanov, S. Floyd, IEEE JSAC, vol. 13, No 4, pp.633-641, May 1995.
- [SALTZER] "End-To-End Arguments in System Design", J.H. Saltzer, D.P. Reed, and D.D. Clark, ACM TOCS, Vol 2, Number 4, November 1984, pp 277-288.
- [SCOTT] "Making Smart Investments to Reduce Unplanned Downtime", D. Scott, Tactical Guidelines, TG-07-4033, Gartner Group Research Note, March 1999.
- [SPILLMAN] "The Law of Diminishing Returns:", W. J. Spillman and E. Lang, 1924.
- [STALLINGS] "Data and Computer Communications (2nd Ed)", William Stallings, Maxwell Macmillan, 1989.
- [TENNENHOUSE] "Layered multiplexing considered harmful", D. Tennenhouse, Proceedings of the IFIP Workshop on Protocols for High-Speed Networks, Rudin ed., North Holland Publishers, May 1989.
- [THOMPSON] "Nonlinear Dynamics and Chaos". J.M.T. Thompson and H.B. Stewart, John Wiley and Sons, 1994, ISBN 0471909602.
- [TINA] "What is TINA and is it useful for the TelCos?", Paolo Coppo, Carlo A. Licciardi, CSELT, EURESCOM Participants in P847 (FT, IT, NT, TI)
- [WAKEMAN] "Layering considered harmful", Ian Wakeman, Jon Crowcroft, Zheng Wang, and Dejan Sirovica, IEEE Network, January 1992, p. 7-16.
- [WARD] "Custom fluorescent-nucleotide synthesis as an alternative method for nucleic acid labeling", Octavian Henegariu*, Patricia Bray-Ward and David C. Ward, Nature Biotech 18:345-348 (2000).
- [WILLINGER2002] "Robustness and the Internet: Design and evolution", Walter Willinger and John Doyle, 2002.
- [ZHANG] "Impact of Aggregation on Scaling Behavior of Internet Backbone Traffic", Sprint ATL Technical Report TR02-ATL-020157 Zhi-Li Zhang, Vinay Ribeiroj, Sue Moon, Christophe Diot, February, 2002.

13. Адреса авторов

Randy Bush

E-Mail: randy@psg.com

David Meyer

E-Mail: dmm@maoz.com

14. Полное заявление авторских прав

Copyright (C) The Internet Society (2002). Все права защищены.

Этот документ и его переводы могут копироваться и предоставляться другим лицам, а производные работы, комментирующие или иначе разъясняющие документ или помогающие в его реализации, могут подготавливаться, копироваться, публиковаться и распространяться целиком или частично без каких-либо ограничений при условии сохранения указанного выше уведомления об авторских правах и этого параграфа в копии или производной работе. Однако сам документ не может быть изменён каким-либо способом, таким как удаление уведомления об авторских правах или ссылок на Internet Society или иные организации Internet, за исключением случаев, когда это необходимо для разработки стандартов Internet (в этом случае нужно следовать процедурам для авторских прав, заданных процессом Internet Standards), а также при переводе документа на другие языки.

Предоставленные выше ограниченные права являются бессрочными и не могут быть отозваны Internet Society или правопреемниками.

Этот документ и содержащаяся в нем информация представлены "как есть" и автор, организация, которую он/она представляет или которая выступает спонсором (если таковой имеется), Internet Society и IETF отказываются от каких-либо гарантий (явных или подразумеваемых), включая (но не ограничиваясь) любые гарантии того, что использование представленной здесь информации не будет нарушать чьих-либо прав, и любые предполагаемые гарантии коммерческого использования или применимости для тех или иных задач.

Подтверждение

Финансирование функций RFC Editor обеспечено Internet Society.

Перевод на русский язык

Николай Малых

nmalykh@protokols.ru