

Глава 4

Надежность интернета

Связанные одной сетью

Практически каждый из нас ежедневно пользуется интернетом. Интернет — это сеть компьютеров и серверов, которые физически соединены каналами связи для передачи цифровой информации с одного сервера на другой.

Сигнал идет со скоростью света, поэтому совершенно неважно, где находятся серверы — в России, США или Австралии. Пройденные расстояния практически не влияют на скорость передачи. Мы все уже давно привыкли, что имейлы и WhatsApp доходят в считанные секунды, веб-страницы грузятся быстро, а наш голос и даже изображение передаются по скайпу в реальном времени. Но, если задуматься, где гарантия, что в любой момент любой сервер мира может связаться с любым другим?

В какой-то степени интернет можно сравнить с системой железных дорог. От любой станции можно добраться до любой другой. Но железные дороги спланированы централизованным образом, их план прошел множество инстанций. Интернет — совсем другое дело. Основные каналы связи (обычно это волоконно-оптические линии) принадлежат самым разным владельцам: компаниям и организациям, например крупным операторам телефонной и мобильной связи. Вместе они составляют так называемую *опорную сеть интернета*.

Большинство компаний, в том числе и многие интернет-провайдеры, заключают договоры на пользование каналами связи и платят аренду. Как только появляется выход к опорной сети, можно начинать строить собственную сеть, присоединять новые серверы и компьютеры. Возникают локальные сети, они соединяются друг с другом, образуют более крупные сети и так далее. И все эти гигантские сети сетей соединены центральной, опорной сетью. Отсюда и название интернет (Internet): *net* по-английски — *сеть*.

Ни один человек и ни одна компания в мире не отвечают за то, чтобы сервер, через который вы присоединились к интернету, был связан с другим сервером, скажем, на острове Кенгуру*. Но вся система по своей природе устроена так, что связь гарантирована. Интернет — гигантская международная технологическая и коммерческая конструкция, без которой мы уже не представляем своей жизни, — прекрасно обходится без правления и правительства. Если вдуматься, это просто поразительно!

Еще поразительнее то, что связь практически никогда не теряется, хотя в каналах связи случаются неполадки и неизбежные регулярные перегрузки. Может ли интернет, хотя бы временно, «развалиться на кусочки»? Может ли случиться так, что из-за сбоя где-то по дороге ваш сервер окажется полностью отрезанным от острова Кенгуру? На самом деле это очень сложный вопрос, на который нет однозначного ответа. При этом из опыта совершенно ясно, что интернет невероятно устойчив к помехам. Согласитесь: если ваш сервер и сервер получателя исправны, то информация всегда проходит через сеть безо всяких проблем.

Эта глава о том, как мы можем хотя бы частично понять и объяснить удивительную надежность интернета.

* Такой остров действительно есть, его население составляет 4,5 тысяч человек; кенгуру там гораздо больше!

Сети и помехи

Начнем с простого примера. Допустим, наш интернет состоит всего из трех компьютеров, которые соединены друг с другом как на рис. 4.1. Если все три канала связи работают, нет никаких проблем: все три компьютера могут обмениваться информацией.

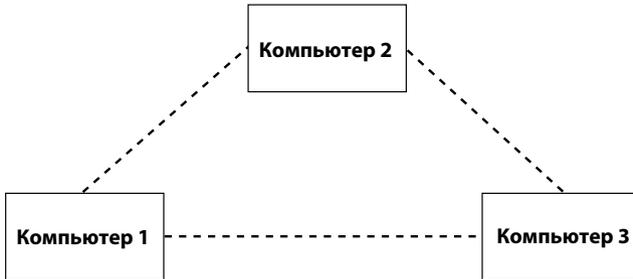


Рис. 4.1. Мини-интернет из трех компьютеров, соединенных каналами связи. Все три канала работают, все три компьютера могут обмениваться информацией

Теперь допустим, что в одном из каналов связи возникли помехи и передать по нему в данный момент ничего нельзя. Мы изобразили эту ситуацию на рис. 4.2. Сразу видно, что наш мини-интернет не распался. Несмотря на то что прямая связь между компьютерами 1 и 2 утеряна, они по-прежнему могут передавать друг другу информацию через компьютер 3. Заметит ли пользователь неполадку в канале? Скорее всего, нет. Поскольку сигнал идет со скоростью света, нет никакой разницы в скорости доставки информации — пойдет ли сигнал напрямую из Москвы в Нижний Новгород или даст кругалю через Сидней или Нью-Йорк.

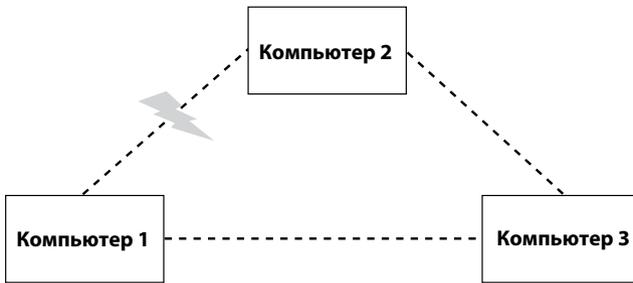


Рис. 4.2. Мини-интернет из трех компьютеров. Хотя канал связи между компьютерами 1 и 2 недоступен, они по-прежнему могут обмениваться информацией через компьютер 3

Чтобы развалить нашу маленькую сеть, нужно вывести из строя как минимум два, а то и все три канала связи, как показано на рис. 4.3.

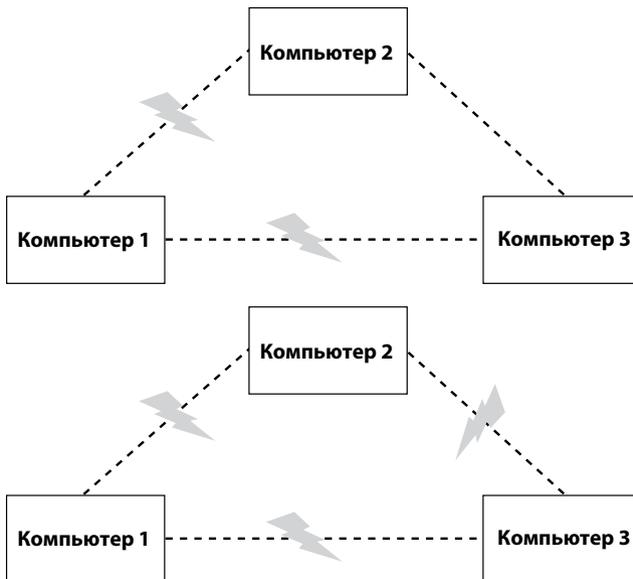


Рис. 4.3. Мини-интернет из трех компьютеров. Сверху: вышли из строя два канала связи, компьютер 1 оказался отрезанным от сети. Снизу: вышли из строя все три канала связи, связь между компьютерами полностью прервана

Насколько устойчива наша мини-сеть? Сосчитать это совсем нетрудно. Допустим, помехи в отдельных каналах связи возникают независимо друг от друга с какой-то вероятностью, скажем 40 %. На практике это означает, что в среднем в четырех из десяти случаев канал оказывается недоступным. Сорок процентов — многовато для реального интернета, но для примера подойдет.

Компьютер 1 может оказаться отрезанным от сети, как на рис. 4.3 сверху с вероятностью $0,4 \times 0,4 \times 0,6 = 0,096 (\times 100\%) = 9,6\%$. В аналогичную ситуацию могут попасть компьютеры 2 и 3 с той же долей вероятности. Наконец, надо добавить вероятность самой плохой ситуации, как на рис. 4.3 снизу, которая равна $0,4 \times 0,4 \times 0,4 = 0,064 (\times 100\%) = 6,4\%$. В результате получается, что наша сеть «развалится» с вероятностью $3 \times 9,6\% + 6,4\% = 35,2\%$.

Конечно, 35,2 % — довольно много, но мы взяли нереально большую вероятность помех. Самое интересное, что вероятность потери связи в сети *меньше*, чем вероятность потери связи в одном канале: 35,2 % меньше, чем 40 %. Сеть более устойчива, чем отдельно взятый канал!

Даже из нашего мини-примера понятно, откуда берется устойчивость сети. В сети компьютеры могут связаться друг с другом не одним, а несколькими способами, через другие компьютеры. Если один канал недоступен, можно найти альтернативный маршрут. Более того, этот эффект заметно усиливается при меньшей вероятности помех. Для примера мы приводим несколько результатов в табл. 4.1*.

* В приложении 1 к главе 4 для подготовленного читателя приведены формулы, которыми мы воспользовались при подготовке в табл. 4.1, в общем виде.

Таблица 4.1. Вероятности потери связи в мини-сети (правая колонка) при заданной вероятности потери связи в одном канале (левая колонка)

Вероятность помехи в канале	Вероятность потери связи в мини-сети
50%	50%
40%	35,2%
30%	21,6%
20%	10,4%
10%	2,8%
5%	0,725%
2%	0,12%
1%	0,03%

Мы видим, что значения в правой колонке убывают гораздо быстрее, чем в левой. Если вероятность недоступности канала 1% — величина вполне реальная на практике, — то наша скромная мини-сеть в 33 раза устойчивее, чем отдельный канал связи!

Конечно, наш мини-интернет очень далек от реальности. Что, если у нас не три компьютера, а целая сеть из десятков, сотен, тысяч машин? Скорость света по-прежнему позволит передавать информацию не напрямую, а по длинным цепочкам. Однако подсчет вероятностей значительно затруднится.

И вот тут опять понадобится математика! Задачи об устойчивости больших сетей требуют глубоких концепций и новых моделей на стыке комбинаторики и теории вероятностей. К счастью, необязательно вникать в длинные доказательства, чтобы понять основные идеи. О некоторых таких фундаментальных идеях, проникающих в самую суть устройства сетей и уже ставших классикой, мы расскажем в следующих разделах.

Случайные графы

Математическая теория, которая, в частности, позволяет ответить на вопрос об устойчивости больших сетей, возникла на рубеже 50–60-х годов XX века. Ее авторами стали два замечательных венгерских математика Пол Эрдеш и Альфред Реньи [22, 15].

Эрдеш — настоящий классик современной комбинаторики, теории чисел, теории вероятностей. Он написал более *полутора тысяч* статей, решил множество проблем и еще больше поставил задач, которые определили пути развития науки на долгие годы. Реньи — выдающийся венгерский специалист по теории вероятностей. Его именем назван математический институт в Будапеште, подобно тому как математический институт в Москве носит имя Владимира Андреевича Стеклова.

Пол Эрдеш

Пол Эрдеш (1913–1996) — очень необычная фигура в математике. Он написал около 1500 статей с 509 соавторами.

Практически вся его собственность умещалась в один чемодан. Деньги его совсем не интересовали. Он жил в дороге, ездил с одной конференции на другую или останавливался у коллег. Рассказывают, что он появлялся на пороге и говорил: «Мой мозг открыт». Затем он работал с хозяевами несколько дней, получал результаты для нескольких статей и ехал дальше, в следующий дом и к другим задачам.

Его соавтор Фэн Чжун написала в своих воспоминаниях: «Все 83 года своей жизни он был абсолютно верен себе. Его не соблазняли посты и деньги. Большинство из нас окружили себя множеством земных благ и обязательств. Каждая встреча с ним напоминала мне, что это все-таки возможно, вот так идти за своей мечтой, не обращая никакого внимания на мелочи жизни. Именно по этому качеству дяди Пола я скучаю больше всего».

В математике, как в искусстве или моде, есть индивидуальные стили и вкусы. Соавторы Эрдеша рассказывают, что ему удавалось найти задачи, подходящие именно для них. Так хорошо он понимал своих соавторов и столько разных задач у него было в запасе! Результаты Эрдеша — разнообразные и в огромном количестве — сильно повлияли на современную науку.

Среди математиков есть понятие *число Эрдеша*. Это количество соавторов, которые отделяют математика от Эрдеша. У соавторов Эрдеша число Эрдеша равно 1. У их соавторов, которые с Эрдешем не работали, число Эрдеша — 2. И так далее. Самое распространенное значение — 5, и очень редко у кого из математиков число Эрдеша равно 8 или больше.

У одного из нас число Эрдеша 3, а у другого 2. Мы оба работаем со случайными графами и вносим свой посильный вклад в решение пока нерешенных проблем.

Теория, основы которой заложили Эрдеш и Реньи, называется *теорией случайных графов*. А математическая модель, рассмотренная в их первых работах, носит имя *случайный граф Эрдеша — Реньи*. Конечно, эти графы не имеют ничего общего с князьями и баронами. Слово «граф» в данном случае имеет то же происхождение, что и слово «график», хорошо известное со школы. Граф — это просто рисунок.

Например, нашу мини-сеть из предыдущего раздела очень легко представить в виде графа. Мы это сделали на рис. 4.4 слева. Сеть изображена в виде трех *узлов* (компьютеров), которые соединены линиями (каналами связи). В математике узлы называются *вершинами* графа, а линии между ними — *ребрами*. Чтобы не затруднять восприятие абстрактными терминами, мы в основном будем пользоваться более интуитивными терминами «узлы» и «линии»*.

* В приложениях для подготовленного читателя мы будем придерживаться математической терминологии: «вершины» и «ребра».

В главе 7 мы вернемся к графам, когда будем обсуждать социальную сеть. В этом случае узлы — это люди, а линии между ними — «дружба» в социальной сети. Мы изобразили пример графа из главы 7 на рис. 4.4 справа.

Естественно, узлов у графа может быть и тысяча, и миллион, и миллиард...

Теория графов — это классическая область математики с огромным количеством приложений. В виде графа можно представить систему железных дорог, газопровод, последовательность операций на крупном производстве или слов в русской речи и многое другое.

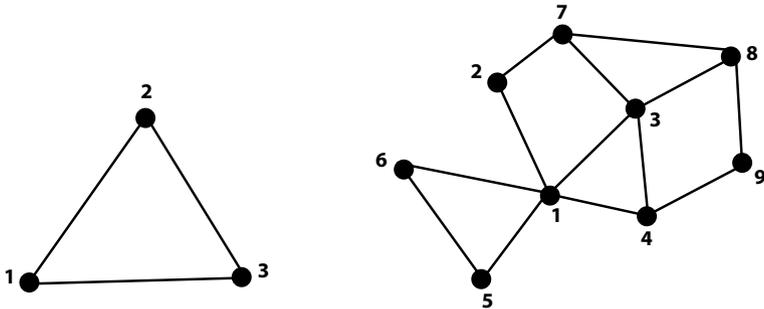


Рис. 4.4. Слева: мини-сеть в виде графа. Узлы — это компьютеры, а линии — каналы связи. Справа: социальная сеть из главы 7 в виде графа. Узлы — это люди, а линии — «дружба» в социальной сети

У случайных графов есть еще одна особенность. Нам неизвестно заранее, какие узлы связаны линией, а какие — нет. Линии между узлами могут существовать или нет с определенной вероятностью. Именно эту ситуацию мы обсуждали в примере с мини-сетью. При наличии помех канал связи становится недоступным, линия между узлами исчезает.

Случайный граф — естественная модель во многих ситуациях. Например, дружба в социальных сетях возникает непредсказуемым образом. В телекоммуникациях или электрических сетях на линиях связи могут случаться сбои.

Если попытаться смоделировать нейронную сеть мозга, то взаимодействия нейронов можно выявить только с определенной вероятностью.

В последнее время в связи с развитием интернета и социальных сетей и небывалой доступностью данных во всех областях — от энергоснабжения до биологии — интерес к теории случайных графов особенно вырос. Новые, очень сложные результаты появляются почти каждый день.

Что же говорит теория случайных графов об устойчивости сети?

Результат Эрдеша — Реньи

В связи с устойчивостью интернета нас интересует вопрос о *связности* случайного графа. Граф называется *связным*, если между двумя любыми его вершинами можно пройти по цепочке ребер, то есть все узлы связаны друг с другом. Оба графа на рис. 4.4 — связные. На рис. 4.5 мы сделали их *несвязными*, удалив по два ребра.

Эрдеш и Реньи задались вопросом: при какой вероятности помех сеть заданного размера остается связной? Результат получился поразительным! Оказывается, в больших сетях связность сохраняется даже при повышенной вероятности помех.

Например, возьмем сеть из 100 связанных между собой компьютеров. Получается, что каждый отдельный канал может быть недоступен с вероятностью аж 86 %, тем не менее сеть останется связной с вероятностью как минимум... 99 %! Эта ситуация изображена на рис. 4.6: 86 % из всех возможных линий отсутствует, однако сразу видно, что из любого узла можно добраться до любого другого.

А сеть из 1000 узлов — это и вовсе нечто фантастическое. Канал связи может быть недоступен с вероятностью 98%, а связность сохраняется с вероятностью 99,9%! Чем больше сеть, тем сильнее результат.

В табл. 4.2 мы приводим результаты для сетей разных размеров. Легко заметить, что число в самой правой колонке не что иное, как

$$\left(1 - \frac{1}{\text{размер сети}}\right) \times 100\%.$$

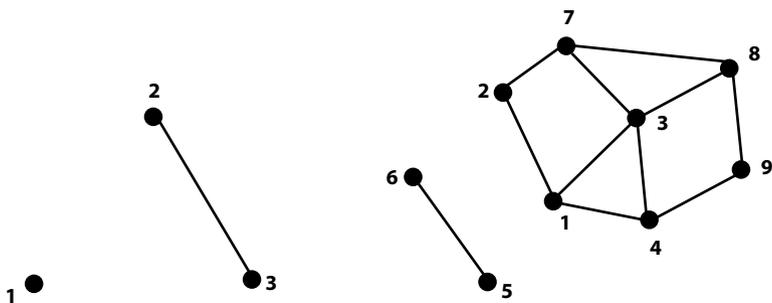


Рис. 4.5. Слева: мини-сеть в виде графа; каналы 1–2 и 1–3 недоступны; граф несвязный, из вершины 1 нельзя попасть в вершины 2 и 3. Справа: социальная сеть в виде графа; пользователь 1 незнаком с пользователями 5 и 6; граф несвязный; нет цепочки знакомых между пользователями 5, 6 и остальными

Таблица 4.2. Результат Эрдеша — Реньи

Размер сети	Вероятность помехи	Вероятность сохранения связности не меньше, чем
50	76,5%	98%
100	86,2%	99%
1000	97,9%	99,9%
10 000	99%	99,99%
100 000	99,97%	99,999%
1 000 000	99,996%	99,9999%

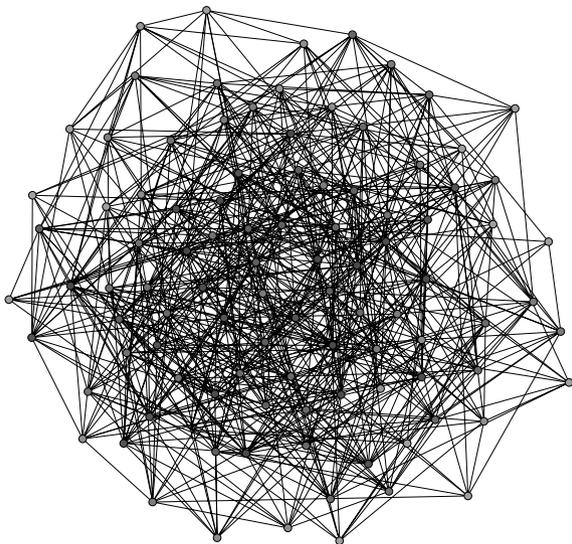


Рис. 4.6. Сеть из 100 компьютеров в виде графа.
Вероятность недоступности канала 86 %

Ниже во врезке приведена более общая математическая формулировка результата. Этот текст рассчитан на уровень средней школы, но если вы не хотите вдаваться в подробности, можете его пропустить.

Простейший вариант теоремы Эрдеша — Реньи

Символ $\ln(n)$ означает натуральный логарифм числа n . В данном случае нам важно только то, что если n увеличивается, то $\ln(n)$ тоже увеличивается, но очень медленно.

Теорема Эрдеша — Реньи. Допустим, сеть состоит из n узлов. Предположим, что связь между любыми двумя узлами недоступна с вероятностью $q(n)$ независимо от других связей в сети. Если

$$q(n) \leq 1 - \frac{3 \ln(n)}{n}, \quad (4.1)$$

то связность сети сохраняется с вероятностью не меньше, чем

$$1 - \frac{1}{n}. \quad (4.2)$$

В табл. 4.2 во втором и третьем столбце все значения умножены на 100%. Во втором столбце указаны значения, полученные с помощью формулы (4.1). Поскольку $\ln(n)$ растет гораздо медленнее, чем n , то допустимая вероятность помех увеличивается. В третьем столбце — значения (4.2), то есть $(1 - \frac{1}{n}) \times 100\%$.

Для многих приложений важно умение работать с сетями, в которых изначально присутствуют не все возможные связи. Например, таковы сети автомобильных дорог, социальные сети и тот же интернет.

Надежность сети, по сути, и есть та вероятность уничтожения отдельной связи в ней, начиная с которой общая связность маловероятна. Для описанной выше ситуации надежность исключительно высока, и это строго доказанный результат.

Фазовый переход

На самом деле теорема Эрдеша — Реньи несколько точнее и еще удивительнее, чем мы описали в предыдущем разделе. Эта теорема выявила интересное явление, которое физики называют *фазовым переходом*. Фазовый переход — это резкий скачок от одного состояния системы к совершенно другому*.

Самый знаменитый фазовый переход — изменение состояния воды в зависимости от температуры. При 0° Цельсия вода превращается в лед, а при 100° — в пар. 0° и 100° — *критические значения*, при которых состояние резко меняется.

* В приложении 2 к главе 4 мы приводим математическую формулировку теоремы Эрдеша — Реньи о фазовом переходе.

Нечто похожее происходит и с вероятностью связности сети, если изменять вероятность недоступности каналов. Оказывается, надежность сети меняется не постепенно, а очень резко. Если вероятность помехи меньше критического значения, то с подавляющей вероятностью связность сохраняется. Но стоит хотя бы немного пересечь критическую черту — и сеть почти наверняка распадется.

На рис. 4.7 показан пример сети из 100 компьютеров. Согласно результатам Эрдеша — Реньи, критическая вероятность помех в такой сети равна 95,4%. Слева вероятность помехи 95%, то есть меньше критического значения. Как видите, связность сети сохранилась. Мы неоднократно повторили эксперимент, но получить несвязную сеть нам так и не удалось. На рисунке справа вероятность помехи 96%. И что же? Одна точка оторвалась от сети, связность потеряна. Опять же как мы ни старались повторить эксперимент, связной сети мы не получили ни разу. Результат впечатляет тем, насколько тонкой оказывается грань между связностью и несвязностью!

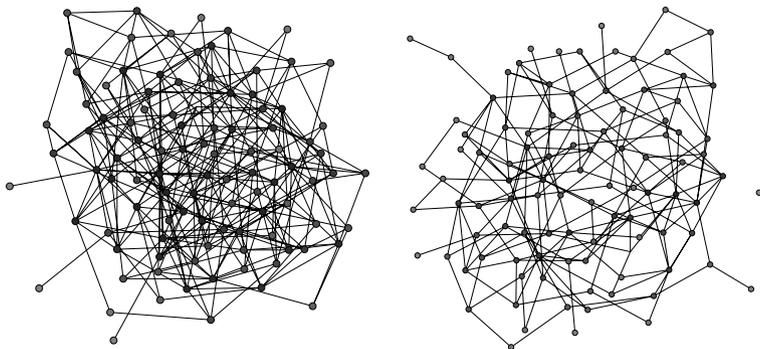


Рис. 4.7. Сеть из 100 компьютеров в виде графа. Слева: вероятность недоступности канала 95%; связность сохраняется. Справа: вероятность недоступности канала 96%; связность нарушена

Если вероятность помех в точности равна критической, то произойдет примерно то же самое, что и с водой и снегом при нуле градусов: может получиться и так и эдак. В нашем случае вероятность сохранения связности приблизительно составит 36,79 %.

Критическая вероятность недоступности канала для сети размера n вычисляется по формуле $\left(1 - \frac{\ln(n)}{n}\right) \times 100\%$. Для примера мы приводим несколько значений в табл. 4.3.

Таблица 4.3. Результат Эрдеша — Реньи: критическая вероятность помех. Если вероятность помех меньше критической, связность сети сохраняется, а если больше — разрушается

Размер сети	Критическая вероятность помехи
50	92,16 %
100	95,39 %
1000	99,31 %
10 000	99,91 %
100 000	99,99 %
1 000 000	99,999 %

Подобные фазовые переходы типичны для теории случайных графов. Эти результаты самые интересные, потому что многое говорят о природе сетей на практике. Состояние сети — это, как правило, две крайности. Социальная сеть либо становится популярной, либо умирает. Компьютерный вирус распространяется с огромной скоростью и размахом или сходит на нет в самом начале. И реальность, и математика подтверждают: среднего не дано. К сожалению, нам далеко не всегда известны критические значения и главное мы не знаем каким образом удержаться по правильную сторону от фазового перехода.

Как доказывается результат Эрдеша — Реньи

Глубокие математические доказательства часто строятся на очень простых интуитивных идеях. Результат Эрдеша — Реньи — блестящий пример данной закономерности*.

Математики заметили, что наиболее вероятный способ разрушить связность сети — отрезать один узел от всех каналов связи. Группу узлов отрезать гораздо труднее, потому что число каналов, которые связывают ее и остальную часть сети, относительно большое. Маловероятно, что все эти каналы недоступны. Тогда изначально сложный вопрос:

С какой вероятностью разрушится связность сети?

сводится к гораздо более простому вопросу:

С какой вероятностью хотя бы один из узлов потеряет все свои каналы связи?

Чтобы доказать, что эти вероятности приблизительно равны, понадобятся длинные и нетривиальные математические выкладки. Но доказать это можно, и усилия оправдываются, потому что второй вопрос гораздо проще первого.

Например, если у нас 100 узлов и вероятность помехи 0,96, то каждый узел может оказаться оторванным от всех 99 других узлов с вероятностью

$$(0,96)^{99} (\times 100\%).$$

Это очень специфическое выражение: число, близкое к единице, возведенное в большую степень. Такие выражения

* В приложении 3 к главе 4 мы приводим идею доказательства результата Эрдеша — Реньи в более точной математической формулировке. Подробно это доказательство рассматривается в книге Андрея «Модели случайных графов» [4].

хорошо известны в математике и относятся к так называемым *замечательным пределам*, из которых, по сути дела, и следует результат.

Что мы знаем и чего не знаем о надежности интернета

Результаты Эрдеша — Реньи полностью не решают проблемы устойчивости интернета. Их модель не очень похожа на реальный интернет. Например, в модели Эрдеша — Реньи число линий у разных узлов обычно близко к среднему. В интернете же разброс между серверами очень большой. У одних серверов сотни каналов связи, а у других — всего два-три.

В 2000 году журнал Nature опубликовал статью «Устойчивость к помехам и атакам в больших сетях» [9]. Авторы-физики, в том числе и весьма влиятельный и знаменитый ученый Ласло Барабаши, взяли данные небольшой части интернета и с помощью компьютерных экспериментов решили посмотреть, что произойдет, если по одному выводить из строя серверы.

Результаты получились нетривиальные. Если серверы выходят из строя случайным образом, например из-за помех, то связность сети долго сохраняется. Но если целенаправленно вывести из строя несколько серверов с самым большим количеством каналов связи, то сеть быстро распадется. Вывод звучал сенсационно: интернет надежный и в то же время довольно хрупкий. Он устойчив к помехам, но чувствителен к атакам!

Эта работа быстро приобрела широкую известность. Только инженеры, работающие в этой сфере, с недоумением пожимали плечами: «Не может быть, наши сети очень надежные!» Результаты Барабаши и соавторов вызвали волну

критики, особенно резкая критика прозвучала в вышедшей в 2005 году статье специалистов по телекоммуникациям [12].

Одним из главных аргументов критиков было то, что в интернете ключевыми являются не многоканальные серверы, а те, через которые проходит наибольшее количество информационного трафика. В интернете у некоторых серверов действительно много каналов связи, но зачастую эти серверы находятся на периферии. За трафик в основном отвечает плотная сеть узлов посередине — опорная сеть. Интуитивно понятно, что для того чтобы разрушить столь густую сеть маршрутов, нужно вывести из строя большое количество серверов. Скорее всего, в ближайшем будущем интернет не развалится!

Благодаря широкому взгляду физиков и специализированному анализу информатиков и инженеров мы знаем достаточно много об устойчивости интернета. Но вопрос пока не закрыт.

В идеале нам нужна строгая математическая модель, включающая в себя основные свойства интернета. И для этой модели необходимы результаты типа Эрдеша — Реньи, показывающие, что произойдет, если вывести из строя те или иные серверы или каналы связи. Только тогда у нас будут однозначно правильные, строго доказанные результаты. Продвижение в этих направлениях есть, но до точных ответов пока далеко. Работа продолжается.

Интернет в картинках

Если вы хотите посмотреть, как выглядит интернет, зайдите на сайт проекта Opte по адресу <http://www.opte.org/>.

Создатель проекта, программист и художник Баррет Лайон, задался целью, казалось бы, невыполнимой — нарисовать интернет! Причем именно в виде графа: серверы и каналы связи между ними.

Серверы принадлежат разным странам и компаниям. Собрать информацию обо всех каналах связи — колоссальная техническая задача. И это только начало. Не менее сложно изобразить все эти гигантские данные на одном понятном глазу рисунке.

Красочное изображение на черном фоне напоминает фейерверк. Линии разных цветов обозначают разные континенты, а плотные, белые, почти светящиеся магистрали между ними — опорная сеть. Рисунки OpTE известны по всему миру. Они получили множество наград и выставляются в Музее современного искусства в Нью-Йорке. Смотрите, это интернет!