№13
Июнь 2020

MSK IX

Интернет изнутри

Будущее

Сможет ли Интернет, каким мы его знаем, удовлетворить потребности будущего?

Содержание:

Журнал «Интернет изнутри»

По всем вопросам пишите на info@internetinside.ru

Порядковый номер выпуска и дата его выхода в свет:
Выпуск №13, дата выхода: Июнь 2020 г.

Свидетельство о регистрации СМИ в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций.
Регистрационный номер:
ПИ № ФС77-71202 от 27.09.2017

Публикуется при поддержке АНО «ЦВКС «МСК-IX»

Главный редактор:
Андрей Робачевский

Зам. главного редактора:
Новикова Татьяна

Редакционная коллегия:
Воронина Елена
Платонов Алексей

Дизайн:
Ильина Наталья

Корректор:
Рябова Наталья

Будущее в настоящем

главный редактор,
Андрей Робачевский

Дорогой читатель!

Тема этого номера – **«будущее»** - достаточно условная. Технологии, о которых здесь пойдет речь, – искусственный интеллект, Интернет вещей и большие данные – это, по существу, технологии сегодняшнего дня. В то же время, области эти достаточно новые и стремительно развивающиеся, чей потенциал по-настоящему раскроется только завтра. Поэтому, не меняя жанр журнала на научно-фантастический, мы взглянули на них сегодняшними глазами и попытались определить тенденции будущего.

Журнал открывается статьей **Джеффа Хьюстона** с анализом новых технологий, которые, по его мнению, могут значительно повлиять на развитие Интернета. Непрерывное изменение остается, пожалуй, основной неизменной характеристикой Интернета и свидетельствует о его неослабевающей жизненной силе. Статьи **Глеба Пыжова** и **Юрия Демченко** расскажут о новых типах Интернета – Интернете вещей и Интернете больших данных. Хотя Интернет в этих областях играет роль поддержки, значение этих явлений для технологического развития общества и самого Интернета трудно переоценить. Я же в своей статье пытаюсь разобраться сможет ли Интернет удовлетворить эти потребности будущего.

Искусственный интеллект – завораживающая тема научной фантастики, а в реальности он только пошел в школу. О технологиях ИИ и его образовательной программе расскажет статья **Натальи Конновой** «Нейронные сети, или Как обучить искусственный интеллект». Кстати, к теме ИИ мы вернемся еще, в этот раз - на примере его использования в системах кибербезопасности.

Не оставили мы без внимания и наши стандартные темы. В разделе «Политика» **Мадина Кассенова** обсуждает будущее и реальность систем цифровой идентификации. «Технология в деталях» посвящена вопросу глубоководных коммуникационных кабелей (вот уж действительно - фантастика наяву), а **Павел Храмцов** приглашает вас в свою рубрику «Новости науки и техники».

Как всегда, нам очень интересно и важно знать ваше мнение. Что понравилось и что можно улучшить? Какие темы вы хотели бы увидеть в следующих выпусках?

Пишите нам по адресу info@internetinside.ru.

Новые коммуникационные технологии

Джефф Хьюстон (Geoff Huston)

Модели коммуникации продолжают развиваться, и это развитие определяет облик появляющихся технологий. В этой статье я бы хотел поговорить об эволюции технологий, на которых зиждется гигантская цифровая сеть, построенная нами за последние десятилетия, а начну я свой обзор с Интернета как такового.

В основе технологий Интернета – да и всей среды цифровой коммуникации в целом – лежит концепция пакетизации. Согласно этой модели пакет предписывает сети, как себя вести. И в этом плане тут в обозримом будущем вряд ли что-то изменится. IP произвел настоящий переворот в индустрии, где раньше безраздельно властвовала телефония. Вместо активной сети с коммутацией каналов, как в телефонной сети, архитектура IP построена на пассивной, по сути, сети, элементы которой просто коммутируют пакеты. Функция ответа на запрос услуги передана компьютеру, находящемуся на краю сети.

Но изменения – дело трудное, и мы несколько десятилетий пытались вернуться к старой модели сетевого сервиса. Мы изо всех сил старались сохранить ответ на запрос услуги в сети, увязав его с обработкой пакетов. Авторы некоторых подходов к качеству обслуживания (Quality of Service, QoS) пытались реализовать различное сетевое поведение для разных классов пакетных потоков в рамках одной и той же сетевой платформы. До сих пор некоторые технологии, такие как MPLS или варианты маршрутизации от источника, внешне эмулируют виртуальные. Прошлое не отпускает, и попытки возродить принципы коммутации каналов в пакетной сети до сих пор не прекращаются. Однако я бы не стал считать эти технологии новыми: мне они кажутся скорее шагом назад.

В то же время мы достигли невероятного прогресса в других аспектах сетевых технологий. Мы строим географически распределенные отказоустойчивые системы, которые обходятся без централизованного управления и контроля. Любой, кто изучает протокол междоменной маршрутизации BGP, который уже три десятилетия незаметно поддерживает Интернет, не может не изумиться дизайном, созданным с очевидным даром предвидения, распределенной системы, управляющей сетью на девять порядков величины больше сети начала 1990-х годов, для которой она была изначально разработана. Мы создали сеть нового типа: открытую и доступную. Создание новых приложений в телефонной сети было практически невозможно, а в Интернете именно это происходит все время. Весь мир сетевых технологий, от многообразия приложений до самых основ цифровой передачи данных, находится в непрерывном движении, а новые технологии появляются с ошеломляющей скоростью.

Какие новые технологии сыграют важную роль в последующие годы? Предлагаю вашему вниманию мой личный шорт-лист недавних технологических достижений, которые, с моей точки зрения, определят развитие Интернета в будущем.

Оптическая когерентность

Оптика долго не могла шагнуть дальше уровня обыкновенного электрического фонарика: либо в кабеле горит свет, либо он там не горит. Такой примитивный подход к оптическому кодированию, получивший название OOK (англ. on-off keying – включение-выключение), непрерывно дорабатывался и улучшался, пока не дорос до поддержки скоростей до 10 Гбит/с в оптике – но дальше инженеры уперлись в, казалось бы, непреодолимые ограничения применимости наших цифровых сигнальных процессов. Теперь мы обратились к принципам оптической когерентности, и здесь нас ждала вторая волна инноваций. Использование оптической когерентности само по себе далеко не ново – и уже сослужило нам отличную службу в других областях. Мы настроили аналоговые модемы для основной полосы частот голосовой связи так, что они стали выдавать 56 кбит/с на несущей частоте 3 кГц. А в мире радио аналогичный подход позволил системам 4G поддерживать скорости передачи данных вплоть до 200 Мбит/с.

В основе этого подхода лежит использование фазово-амплитудной модуляции для того, чтобы выжать из оптики максимум, вплотную приблизившись к теоретическому пределу Шеннона. Сейчас на рынке обычным делом стали оптические системы с пропускной способностью 100 Гбит на длину волны, уже появляются и 400-гигабитные. Очень вероятно, что в ближайшие годы мы увидим и терабитные системы, использующие фазово-амплитудную модуляцию высокой плотности в сочетании с индивидуально настроенной цифровой обработкой сигналов. И, как было и с другими оптическими системами, цены на единицу пропускной способности войдут в глубокое пике по мере наращивания объемов. Нынешний мир богат ресурсами связи, и через призму этого богатства открывается новый, зачастую неожиданный взгляд на сетевые архитектуры.

5G

А что можно сказать о радиосвязи? 5G, например, новая технология или нет? Лично я придерживаюсь мнения, что 5G не так уж и сильно отличается от 4G. Та действительно была

новаторской, потому что потребовала перехода от PPP-тунелей к собственной, непосредственной пересылке пакетов IP. А 5G – это та же самая 4G,только на более высоких частотах. В первых системах 5G используются несущие частоты 3,8 ГГц, но планируется охватить и миллиметровый диапазон с частотами до 24 ГГц, до 84 ГГц. Это неоднозначная перспектива, так как, с одной стороны, увеличение несущей частоты позволит выделять большие блоки частот (а следовательно, повысить пропускную способность радиосети), но, с другой стороны, высокая частота означает малую длину волны, а миллиметровый диапазон по своим характеристикам ближе к свету, чем к радио. На высоких частотах радиосигнал все время блокируется зданиями, стенами, деревьями и другими крупными предметами, создающими тень. В результате для достижения такого же качества покрытия требуется гораздо больше базовых станций. Если отвлечься от рекламных обещаний, неясно, есть ли для 5G-миллиметрового диапазона выгодная экономическая модель.

Поэтому я не стану включать 5G в свой список. Радио и мобильные сервисы были и останутся чрезвычайно важным аспектом Интернета, но 5G не представляет собой какого-то радикального прогресса в плане использования этих систем за пределами того, на что способна уже укоренившаяся технология 4G.

IPv6

Как-то даже неловко причислять IPv6 к «новым технологиям» в 2020 году. Первая спецификация IPv6 – RFC1883 – датируется 1995 годом, то есть самой технологии уже 25 лет. Но похоже, что спустя много лет нерешительности и даже отрицания проблем возможности IPv4 практически исчерпаны, а потому на сегодняшний момент IPv6 используется в четверти Интернета. Эта доля неизбежно будет увеличиваться. Трудно сказать, сколько времени это займет, но конечный результат, по сути, однозначен. Поэтому если говорить о «новом» в том плане, какие технологии будут широко внедряться в последующие годы, IPv6 точно следует отнести к этой категории, несмотря на почтенный возраст!

BBR

Разработанный Google алгоритм BBR (Bottleneck Bandwidth and Round-trip time) для TCP представляет собой революционный алгоритм управления. Он переосмысливает взаимосвязь между оконечными устройствами, сетевыми буферами и скоростью, позволяя оконечным системам эффективно использовать доступную пропускную способность сети, не испытывая негативного влияния внутренней буферизации в активных сетевых элементах. Алгоритмы контроля перегрузки на базе потерь сослужили нам хорошую службу в прошлом, но теперь, когда скорости передачи данных из конца в конец приближаются к сотням гигабит в секунду, консервативные алгоритмы контроля на основе потерь перестали быть практичными. BBR представляет собой совершенно новый подход и к управлению потоком, и к управлению скоростью: он пытается стабилизировать скорость потока на одном и том же уровне для большей части доступной пропускной способности сети. Это совершенно потрясающая технология.

QUIC

Сети и приложения с давних пор находятся в состоянии своеобразной холодной войны между собой. В мире, где повсеместно применяется сквозной контроль перегрузки TCP, сетевые ресурсы совместно используются несколькими активными клиентами, причем способ такого использования определяют сами клиенты. Для сетевых операторов это всегда было настоящим проклятием: они бы предпочли активно управлять ресурсами своих сетей и предоставлять клиентам четко определенный сервис. Для этой цели в сетях имеются разнообразные ограничители скорости на основе политик. Работают они так: по «сигнатуре» заголовка пакета определяется создавшее его приложение, и дальше политика определяет, как с пакетом поступить. Для этого требуется прозрачность содержимого каждого пакета IP, которая как раз предусмотрена в TCP.

QUIC – это форма инкапсуляции, в которой используется видимая внешняя оболочка пакета UDP, а внутренняя полезная нагрузка – TCP и контент – шифруется. При таком подходе параметры управления потоком TCP оказываются скрыты от сети и сетевых ограничителей, работающих на основе политик, но это еще не все: управление алгоритмом потока данных отбирается у общей операционной системы хоста и передается каждому отдельному приложению. Приложение получает больший контроль, а потому может независимо развиваться и корректировать свое поведение независимо от платформы.

Кроме того, устраняется необходимость в единообразной и равно неудобной для всех модели управления потоком данных, которая используется в платформенных TCP-приложениях операционных систем. При использовании QUIC приложение само может настраивать свои способы управления потоком так, чтобы оптимизировать поведение приложения в рамках текущего состояния сетевого пути.

DNS без резолверов

Я хотел было написать «DNS поверх HTTPS» (DoH), но не уверен, что DoH является такой уж новой технологией, поэтому сомневаюсь, что ее можно сюда отнести. Мы использовали HTTPS в качестве технологии для туннелирования брандмауэров и обеспечения приватности коммуникации практически с тех пор, когда появились брандмауэры и опасения слежки. Программные инструменты для туннелирования пакетов IP в сеансы HTTPS широко распространены и используются уже пару десятилетий, так что здесь ничего особенно нового нет. Инкапсуляция DNS внутри HTTPs – лишь незначительное развитие модели, в которой HTTPS используется для туннелирования всего.

В то же время сам HTTPS дает нам ряд дополнительных возможностей, которые недоступны старому доброму DNS

поверх TLS (части HTTPS, которая отвечает за создание безопасного канала). Я говорю о технологиях отправки данных по инициативе сервера (англ. server push) в веб. Например, веб-страница может ссылаться на кастомную страницу стиля, чтобы определить, как должно выглядеть ее визуальное оформление. Вместо того чтобы заставить клиента запросить страницу стиля отдельно, т.е. выполнить еще одно преобразование DNS и установить еще одно соединение, сервер может просто отправить этот ресурс клиенту вместе с использующей его страницей. Для HTTP запросы и ответы DNS выглядят точно так же, как любые другие транзакции с объектами данных, так что с точки зрения HTTP отправка ответа DNS без предваряющего его запроса DNS по тому же принципу мало чем отличается, скажем, от отправки таблицы стилей.

Однако для архитектуры имен в Интернете такое новшество значит очень, очень много. Дело вот в чем: вдруг имена, которые так пересылаются по HTTPS, были доступны только в контексте данной конкретной веб-среды и недоступны для любых других инструментов, включая обычные запросы DNS? Интернет можно определить как единое, взаимосвязанное и внутренне непротиворечивое пространство имен. При общении мы отправляем друг другу ссылки на ресурсы, например, их имена, и это возможно только в случае, если данное конкретное имя для меня и для вас означает один и тот же ресурс. Результат преобразования DNS остается тем же самым, независимо от того, какое приложение его запрашивает, в каком контексте и зачем. Но когда контент отправляет преобразованные имена клиентам, это делается просто для создания собственного контекста и среды, которые будут уникальными, т.е. отличаться от любого другого контекста имен. Вместо одного непротиворечивого пространства имен возникает множество фрагментированных и, возможно, перекрывающихся пространств, и нет никакого четкого способа разрулить потенциальные конфликты имен.

Движущими силами многих новых технологий являются скорость, удобство и настройка среды под каждого отдельного пользователя. С этой точки зрения появление DNS без резолвера практически неизбежно. Но обратной стороной медали при этом будет потеря связности Интернета, и неясно, послужит ли данная конкретная технология на благо Интернета или нанесет ему огромный вред. Что ж, так или иначе, время покажет!

Квантовые сети

В 1936 году, задолго до появления первых программируемых компьютеров современного типа, один британский математик (Алан Тьюринг, Alan Turing – прим. ред.) описал в качестве мысленного эксперимента универсальную вычислительную машину – и, что еще важнее, подразделил задачи на вычислимые (для которых решения можно достичь за конечное время) и невычислимые (при решении которых машина никогда не остановится). Так что можно сказать, что еще до появления первого физического компьютера было известно о существовании целого класса задач, решить которые на компьютере невозможно. В 1994 году аналогичный прорыв совершил Питер Шор (Peter Shor), разработав алгоритм разложения числа на простые множители за конечное время для квантового компьютера, который тогда был еще теоретической абстракцией. Возможности (и ограничения) этой новой машины для обработки данных были описаны задолго до того, как ее удалось построить. Сейчас квантовые компьютеры стали реальностью – новой и потенциально революционной технологией в компьютерном мире.

С этой технологией связана еще одна – квантовые сети, между которыми передаются квантовые биты (кубиты) информации. Здесь я, как и многие другие, ничего не могу сказать о том, что нас ждет: уготована ли квантовым сетям роль некоей эзотерической боковой ветви в эволюции цифровых сетей или же они получат всеобщее распространение и лягут в основу цифровых сервисов завтрашнего дня. Время покажет, а пока его прошло совсем немного.

Эволюция архитектуры

Почему техническая эволюция все никак не остановится? Почему мы все никак не можем сказать: «Все, ребята, дело сделано, пошли выпьем!» Мне кажется, что постоянное стремление модифицировать технические платформы Интернета происходит от эволюции архитектуры самого Интернета.

Первоначальная модель Интернета, по сути, была предназначена для того, чтобы соединять клиентов с сервисами. Теперь же можно построить архитектуру, в которой любой сервис может поддерживать выделенную сеть доступа, и клиенту потребуется подключиться к нужной сети, чтобы получить доступ к нужной услуге. Правда, мы уже попробовали эту модель в небольшом масштабе в 80-е годы прошлого века, и она повергла всех в ужас! Поэтому мы стали использовать Интернет как универсальную сеть для соединения всех клиентов со всеми сервисами. До тех пор, пока все сервисы и все серверы подключены к этой общей сети, клиент, подключившись к ней, получает доступ к ним ко всем. В 90-х годах это был революционный шаг, но с тех пор число пользователей непрерывно росло и в

Почему техническая эволюция все никак не остановится? Почему мы все никак не можем сказать: «Все, ребята, дело сделано, пошли выпьем!» Мне кажется, что постоянное стремление модифицировать технические платформы Интернета происходит от эволюции архитектуры самого Интернета.

конце концов переросло возможности серверной модели. Ситуация стала нестабильной. Популярные сервисы стали в каком-то смысле похожи на черные дыры в сети. Нам остро требовалось другое решение, и мы изобрели сети доставки контента (англ. content distribution networks, CDN). CDN использует выделенные сети для того, чтобы развернуть несколько равноправных точек доставки своих услуг по всему Интернету. Вместо того чтобы обращаться к любому сервису по единой глобальной сети, клиенту требуется всего лишь подключиться к сети доступа, которая «доведет» его до локальной агрегированной точки доступа CDN. Чем больше мы пользуемся локально доступными сервисами, тем меньше мы используем глобальную сеть.

Что это означает для технологий?

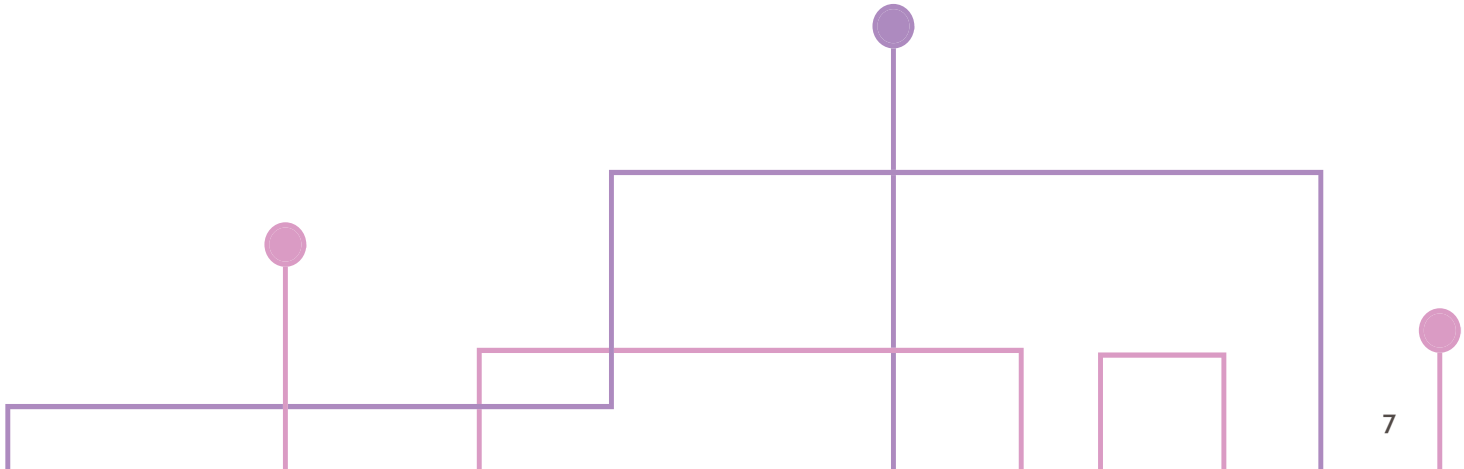
В результате ослабевает мотивация поддерживать Интернет единым и связным. Если большая часть цифровых сервисов для пользователей доступна в чисто локальной структуре, то кто же будет платить огромные деньги за глобальный транзит для доступа к крошечному рудименту оставшихся сервисов, доступных только удаленно? Нужен ли локальным сервисам доступ к уникальным инфраструктурным элементам глобальной сети? NAT – экстремальный пример того, что чисто локальные сервисы вполне функциональны при работе с чисто локальными адресами, и широкое использование локальных имен – еще одно этому доказательство. Я бы не стал прямо делать вывод, что рост популярности сетей доставки контента приведет к фрагментации Интернета, но фрагментация по самой своей сути подобна такому физическому явлению, как энтропия. А значит, нужны постоянные усилия, чтобы противостоять фрагментации. Если ослабить бдительность, если не прикладывать постоянные усилия к поддержанию единой глобальной системы уникальных идентификаторов, мы начнем двигаться в сторону сетей, которые смогут работать лишь локально.

Здесь проявляется еще одна тенденция: рост масштаба изоляции сервисов в приложениях. Примером может служить первый сценарий применения QUIC. QUIC использовался исключительно браузером Google Chrome при доступе к веб-серверам Google. Функции транспортного протокола, обычно являющиеся частью общего сервиса для приложений операционной системы, были перемещены в приложение. Старые концепции проектирования, которые предусматривают использование общего набора функций операционной системы вместо спроектированных на заказ функциональных возможностей приложений, больше не применяются. Развертывая более мощные оконечные системы и более быстрые сети, мы можем создавать приложения с высокой степенью кастомизации. Уже сейчас браузеры реализуют множество функций, которые раньше были доступны лишь операционным системам, а теперь по тому же пути пошли многие приложения. Дело не просто в желании более тонкого контроля над взаимодействием с конечным пользователем, хотя это тоже важное обстоятельство – дело еще и в том, что каждое приложение закрывает свое поведение и взаимодействие с пользователем от других приложений, от операционной системы хоста и от сети. Если теперь предположить, что деньги, которые вкладываются в развитие Интернета, – это деньги, получаемые из знания привычек и желаний конечного пользователя (а в случае Google, Amazon, Facebook, Netflix и многих других это чистая правда), то какой смысл этим приложениям раскрывать свою информацию третьим сторонам? На смену приложениям, опирающимся на богатый набор сервисов операционной системы и сети, приходит новая технологическая модель, которую можно характеризовать как «приложения-параноики». Такие приложения-параноики не только стремятся свести к минимуму зависимость от внешних ресурсов – они стараются как можно лучше замаскировать свое собственное поведение.

Жизнь – это непрерывное изменение

Появление этих новых технологий и их конкуренция с нынешними сервисами и инфраструктурой Интернета – пожалуй, самый обнадеживающий признак того, что Интернет жив, здоров и еще очень далек от того, чтобы кануть в пучину забвения. Мы до сих пор меняем основные элементы передачи данных, изменяем транспортные протоколы, дорабатываем инфраструктуру имен и адресов, пересматриваем модели доставки сервисов. И это – самый лучший признак того, что Интернет далек от того, чтобы окончательно устояться и больше не ставить важных технологических задач.

Источник: Internet Economics, https://www.potaroo.net/ispcol/2020-05/futuretech.html



Сможет ли Интернет, каким мы его знаем, удовлетворить потребности будущего?

Андрей Робачевский

Каждая неудачная сетевая архитектура обещала QoS, делая биты дороже; каждая успешная технология снижала общую стоимость развертывания и эксплуатации и делала биты дешевле.

Henning Schulzrinne

Успех Интернета поистине невероятен. Ключевые технологии, лежащие в его основе, и развитие самой инфраструктуры сети сетей радикально изменили то, как мы общаемся и ведем бизнес. Если бы Интернет внезапно исчез, социальное и экономическое благосостояние общества оказалось бы под серьезной угрозой. Несмотря на свою критическую роль в нашей жизни, дать точное и единственное определение Интернета невозможно. Дело в том, что Интернет как базовая технология, как коллективная коммуникационная платформа смешан со многими другими технологиями и приложениями: смартфонами, беспроводными сетями, облачными вычислениями, сетями доставки контента и стриминговыми услугами, социальными сетями, в конце концов. Каждая из которых так же важна и для многих составляет суть и определение Интернета. Но без фундаментальной коммуникационной платформы и ее технологий ни один из этих элементов работать не будет. Об этом нивидимом невооруженному глазу Интернете мы и поговорим в этой статье.

Что определило успех

Для ответа на этот вопрос необходимо обратиться к истории создания и развития Интернета. До середины 70-х, когда имя Интернет вошло в обиход, Сеть носила имя ARPANET. По сравнению с традиционными телефонными сетями, основанными на коммутации каналов, ARPANET использовала технологию коммутации пакетов, или дейтаграмм - данных ограниченного объема, заключенных в «конверты» с указанием источника и получателя. Поскольку каждый пакет обрабатывался независимо, сети не требовалось хранить информацию о соединениях между оконечными компьютерами и потоках данных между ними. Этот подход позволил существенно упростить архитектуру сети и повысить ее надежность. Узел сети мог выйти из строя - и его функцию без особых проблем мог подхватить другой узел. Другой особенностью являлось то, что асинхронная пакетная передача больше соответствовала характеру работы операционных систем разделения времени, таких как Unix, обеспечивавших функционирование компьютеров, чем синхронная передача данных, как в случае голосовой связи.

Однако в то время для обмена данными между компьютерами, или хостами, использовался протокол NCP (Network Control Protocol), предтеча сегодняшнего TCP/IP. Ограничением NCP являлось то, что это, по существу, был транспортный протокол, который не был хорошо приспособлен для работы с разнообразными технологиями – например, цифровой радио- и спутниковой связью. Более того, он был предназначен для работы с одной сетью – ARPANET – и не предоставлял возможности адресовать другие сети и подключенные к ним компьютеры.

Поэтому следующим фундаментальным изменением архитектуры Сети явились разработка и внедрение многоуровневой модели протоколов семейства TCP/IP, основанных на концепции обмена данными между независимыми сетями, различными по своей архитектуре и используемым технологиям. Окончательный перход Интернета на эту модель произошел в 1983 году.

Однако топология Интернета 80-х была достаточно проста и неадаптивна. Использовавшийся протокол маршрутизации (Exterior Gateway Protocol - EGP) хотя и поддерживал обмен маршрутами между независимыми сетями, был более приспособлен к иерархическим древовидным топологиям – опорной сети NSFNET с подключенными к ней региональными и немногочисленными зарубежными сетями. Пришедший ему на смену в начале 90-х протокол BGP снял эти ограничения и доказал прекрасную масшабируемость и производительность в системе произвольной связности между независимыми сетями (в терминах междоменной маршрутизации называемых автономными системами, AS), каковой сегодня и является Интернет.

Оглядываясь назад и оценивая аспекты, которые оказались ключевыми в феноменальной эволюции Интернета, можно сказать, что его фундамент держится на трех опорах:

- Единая платформа пакетной передачи с общим знаменателем в виде протокола IP, позволяющая использовать сквозную адресацию подключенных устройств и обладающая минимальными требованиями совместимости между сетями (собственно, соответствие стандарту IP). На уровне IP Интернет представляет собой однородную глобальную сеть, несмотря на разнородность и независимость составляющих его отдельных сетей.

- Уровневая архитектура протоколов, прозрачно обеспечивающая поддержку IP в гетерогенных сетях различных технологий нижнего уровня - мобильные и спутниковые сети, Ethernet, DSL и т.п., - а также независимое развитие протоколов и соответствующих им приложений верхнего уровня - UDP, TCP, FTP, SMTP, HTTP и т.п. Эта модульность, обеспечиваемая совместимыми "строительными блоками", позволяет конструировать сети различного назначения, топологии и архитектуры с заведомой гарантией их совместимости с глобальным Интернетом.

- Децентрализованная и распределенная глобальная система маршрутизации, позволяющая составляющим Интернет сетям независимо принимать решения о связности (с какими сетями соединяться) и политике маршрутизации (каким маршрутам отдавать предпочтение). Данная модель также свела к минимуму необходимость координации - для того, чтобы стать полноценным участником Интернета (и его глобальной системы маршрутизации), необходимо договориться хотя бы с одной сетью (и подключиться к ней).

Удивительно, что эти основные технологические нововведения остались относительно неизменными на протяжении более трех десятилетий, несмотря на колоссальные изменения в характере использования Интернета, доказывая их необыкновенную масштабируемость и адаптивность.

Новые потребности

До недавнего времени не вызывало особых сомнений, что Интернет по-прежнему сможет соответствовать требованиям развивающейся цифровой экосистемы. Наиболее тревожные аспекты, как, например, гарантия качества услуг, ограниченное адресное пространство и рост таблицы маршрутизации, находили свое решение без радикальнных изменений архитектуры и основных технологий. Тем самым еще раз подтверждая адаптивные возможности платформы и ее неиссякающий инновационный потенциал.

И в то же время все более слышны голоса, что текущая архитектура может не справиться с растущими потребностями, и простыми «заплатками» делу не поможешь - нужен качественно новый подход, архитектура и протоколы. И действительно, при всех преимуществах, недостатки текущей модели и ограничения протоколов также заметны.

Возьмем, например, качество услуг. Лет двадцать назад трудно было себе представить, что в Интернете можно будет смотреть видео. Сегодня трансляция видео высокого

Оглядываясь назад и оценивая аспекты, которые оказались ключевыми в феноменальной эволюции Интернета, можно сказать, что его фундамент держится на трех опорах:

✓ **Единая платформа пакетной передачи с общим знаменателем в виде протокола IP**, позволяющая использовать сквозную адресацию подключенных устройств и обладающая минимальными требованиями совместимости между сетями (собственно, соответствие стандарту IP). На уровне IP Интернет представляет собой однородную глобальную сеть, несмотря на разнородность и независимость составляющих его отдельных сетей.

✓ **Уровневая архитектура протоколов**, прозрачно обеспечивающая поддержку IP в гетерогенных сетях различных технологий нижнего уровня - мобильные и спутниковые сети, Ethernet, DSL и т.п., - а также независимое развитие протоколов и соответствующих им приложений верхнего уровня - UDP, TCP, FTP, SMTP, HTTP и т.п. Эта модульность, обеспечиваемая совместимыми "строительными блоками", позволяет конструировать сети различного назначения, топологии и архитектуры с заведомой гарантией их совместимости с глобальным Интернетом.

✓ **Децентрализованная и распределенная глобальная система маршрутизации**, позволяющая составляющим Интернет сетям независимо принимать решения о связности (с какими сетями соединяться) и политике маршрутизации (каким маршрутам отдавать предпочтение). Данная модель также свела к минимуму необходимость координации - для того, чтобы стать полноценным участником Интернета (и его глобальной системы маршрутизации), необходимо договориться хотя бы с одной сетью (и подключиться к ней).

разрешения через Интернет - реальность. Однако для этого потребовалось создание дополнительных, подчас параллельных сетей и архитектур - таких, как сети распределения контента CDN.

Или - Интернет вещей. Многие считали (и считают), что каждая вещь может и должна стать полноправным членом Интернета. Стандартизованный в 1998 году протокол нового поколения IPv6 призван обеспечить громадное пространство для адресации невообразимого количества узлов. То есть в теории каждая из вещей в обозримом будущем может быть адресована и может использовать IPv6 для передачи данных. Однако в реальности многие из этих «вещей» настолько ограничены в своих возможностях, что поддержка накладных расходов IPv6 не представляется возможной. Вместо этого они используют специально оптимизированные протоколы (например, Z-wave, Zigbee, BLE), а при необходимости глобальной связности используют шлюз. Проблема решена, хотя и с нарушением элегантности принципа сквозной связности.

Но инновации в области беспроводных коммуникаций, распределенных вычислений и новых приложений виртуальной реальности и удаленного присутствия выдвигают новые требования. Сможет ли Интернет в рамках сегодняшней архитектуры по-прежнему обеспечить фундаментальную коммуникационную инфраструктуру. Или нужно задуматься о разработке Интернета нового поколения, который придет на смену сегодняшнему, как когда-то Интернет сменил традиционные телефонные сети?

Для того, чтобы ближе подойти к ответу на эти вопросы, давайте посмотрим на новые требования более внимательно и проанализируем предлагаемые решения. В этой связи интересным представляется взгляд на будущие потребности, представленный в отчете фокусной группы МСЭ-Т под заголовком «Новые услуги и возможности Сети 2030: Описание, анализ целевых показателей производительности и GAP-анализ» (New Services and Capabilities for Network 2030: Description, Technical Gap and Performance Target Analysis, https://www.itu.int/en/ITU-T/focusgroups/net2030/Documents/Deliverable_NET2030.pdf). Этот анализ подкреплен более техническим документом, предаставленным на обсуждение в IETF, - «Проблемы уровня передачи» (Forwarding Layer Problem Statement, https://datatracker.ietf.org/doc/draft-bryant-arch-fwd-layer-ps/). Целью исследования является ответ на конкретные вопросы о том, какие типы сетевой архитектуры и вспомогательных механизмов подходят для таких новых сценариев.

Перспективных сценариев несколько. Остановимся на основных из них.

Голографический тип связи

Этот сценарий предполагает, что мы будем двигаться к голографическому обществу, в котором пользователи будут удаленно взаимодействовать с физическим миром через сеть. В промышленности цифровая модель близнецов позволит управлять реальными объектами с помощью цифровых копий. Телеприсутствие выйдет на новый уровень, и совместная работа географически распределенных групп станет намного ближе к физическим встречам, которые могут проводиться без затрат времени и ущерба окружающей среде. Трехмерные медицинские сканы полностью приобретут трехмерный вид вместо набора двумерных сечений органов. Или другой пример - виртуальный оркестр и концерты. Представьте себе инструментальный ансамбль, в котором голографические трехмерные проекции музыкантов в натуральную величину собираются вместе и выступают вживую на сцене перед аудиторией. Фантастика? Кто знает, может быть, потребность в таких приложениях сегодня уже более реальна, чем в 2019 году. В любом случае, легко представить, что эта технология выведет доставку сообщений на совершенно новый уровень.

Тактильный Интернет для удаленных операций

В качестве примеров этого класса приложений были предложены два случая. Первый - это дистанционное управление промышленностью, которое включает мониторинг и контроль операций промышленной инфраструктуры в режиме реального времени. Второй включает дистанционную роботизированную хирургию. Дистанционная роботизированная хирургия в комплексе операционных систем сегодня является стандартной практикой, однако есть потребность расширить спектр случаев ее применения.

Интегрированная спутниково-наземная сеть

Решающим фактором в области интегрированной спутниково-наземной связности является активное развертывание огромного числа недорогих спутниковых созвездий на низкой околоземной орбите (Low Earth Orbit, LEO). Спутники LEO имеют ряд свойств, которые делают их привлекательными, но, пожалуй, наиболее важным является то, что они сочетают в себе глобальное покрытие с низкой задержкой. Восходящая линия связи с кластером LEO-спутников должна постоянно изменять точку присоединения к кластеру, поскольку спутники, которые формируют кластер, быстро перемещаются по небу как относительно Земли, так и относительно спутников на других орбитах.

ManyNets

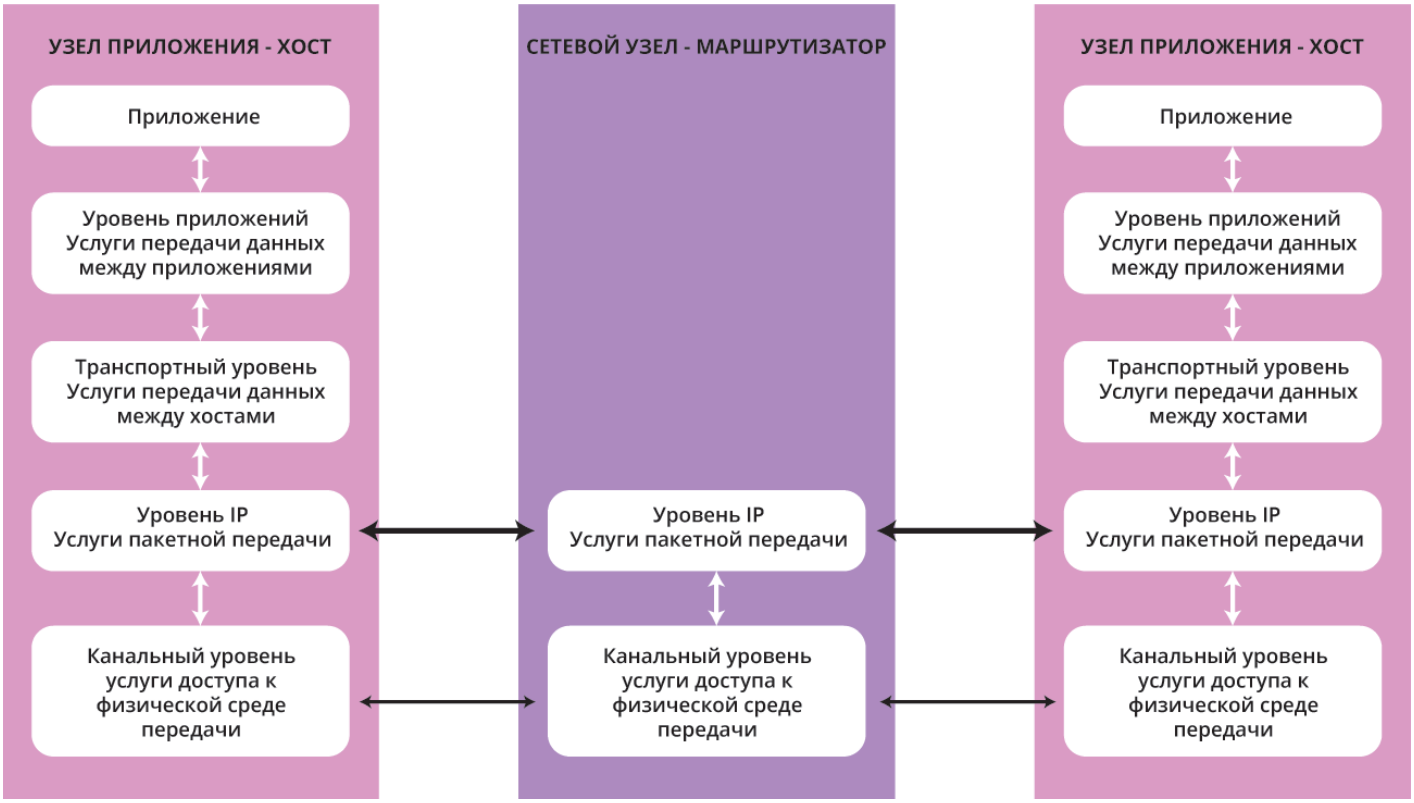
Концепция ManyNets - и не то чтобы сценарий приложений будущего, и не то чтобы требование. ManyNets сводится к наблюдению, что сети, возможно, оптимизированные для конкретного использования и использующие различные протоколы, должны беспрепятственно взаимодействовать. Постойте, подумаете вы, это ли не концепция самого Интернета? Но похоже, что авторы отчета думают иначе, поскольку требование поддержки ManyNets является для них доказательством, что Интернет нужно в корне переосмыслить. Как и зачем - остается во многом загадкой.

Сетевые модели будущего

При рассмотрении предложенных сценариев логично возникает вопрос - сможет ли завтрашний Интернет путем эволюционного развития поддержать эти амбициозные сценарии, или требуется фундаментально новая коммуникационная система?

Рис. 1. Эталонные сетевые модели Интернета и сети 2030.

Эталонная модель Интернета



Эталонная модель сети 2030

По мнению авторов отчета – не сможет. Для обеспечений требований приложений будущего предлагается новая модель, существенно отличающаяся от эталонной модели Интернета – модели TCP/IP. Вместо определения стека протоколов модель рассматривает «сервисы», используя упрощенную трехуровневую схему, см. рис. 1.

В рамках новой архитектуры различают два сетевых сервиса – базовый и составной.

БАЗОВЫЙ СЕТЕВОЙ СЕРВИС – сервис, требующий специальной поддержки на некоторых или всех узлах сетевой системы, которые предоставляют услугу между двумя или

более узлами прикладной системы. Например, в Интернете таким базовым сервисом является передача IP-пакетов от одного узла к другому. Но в новой модели базовые сервисы, предоставляемые сетью, гораздо более существенны. Различаются сервисы своевременной доставки, скоординированной доставки и доставки с определенными параметрами качества. Рассмотрим их подробнее.

- Своевременная доставка и доставка к определённому времени (In-time and on-time services)

Сервисы in-time — это услуги, которые обеспечивают доставку пакетов с необходимой максимальной задержкой. Пакеты могут быть доставлены в любое время, но не позже крайнего срока. Типичными приложениями, которым могут потребоваться услуги этого типа, являются мультимедийные приложения, поддерживающие возможности буферизации.

Сервисы on-time — это услуги, которые обеспечивают поступление данных в рамках определенного временного окна. Подобно сервисам in-time, они налагают максимальную задержку, которую нельзя превышать. Кроме того, они поддерживают минимальную задержку. Пакет должен быть доставлен не позднее верхней границы временного окна, но также не ранее нижней границы временного окна. Такие услуги обычно должны работать с высокоточной временной шкалой (например, микросекунды) и обеспечивать детерминированные задержки. Например, в промышленных приложениях контроллеру может потребоваться отправить команды на серию устройств. Каждое устройство должно получать команду в определенное время для работы в квазисинхронизированном режиме.

В этой связи, конечно, возникает вопрос, каким образом эта инфраструктура будущего будет воплощена в жизнь? Двадцатилетний опыт внедрения IPv6, который, кстати, не меняет фундаментальных основ Интернета и всего, что на его базе построено, наглядно показывает, что силы рынка гораздо больше подвластны экономическим законам, чем техническим спецификациям. Особенно, если технология обладает низкой совместимостью с существующими и требует внедрения на глобальном уровне.

- Скоординированные сервисы обеспечивают гарантированную доставку нескольких взаимозависимых или связанных потоков. Эти потоки могут обладать различными видами зависимостей или отношений. Доставка с использованием этих услуг гарантирует сохранение зависимостей и ограничений, наложенных на потоки услугами или приложениями верхнего уровня.

- Качественный коммуникационный сервис предполагает возможность сети различать части содержимого пакетов. При перегрузке, вместо того, чтобы отбрасывать пакет целиком, как это происходит сегодня, этот механизм позволяет сети отбрасывать лишь менее значимые части или части с более низким приоритетом. Таким образом, сеть позволяет избежать критических потерь, обеспечивая повторную передачу только существенно необходимых данных.

СОСТАВНОЙ СЕТЕВОЙ СЕРВИС — это сервис, который может состоять из одного или нескольких базовых сервисов. Составные сервисы – это своего рода «комплексное меню», объединяющее несколько базовых услуг для удовлетворения требований определенных приложений. В общем случае составные сервисы не определяют какую-либо новую сетевую услугу на уровне сетевого узла.

Примером составного сервиса может служить сервис тактильных коммуникаций, представляющий собой набор нескольких скоординированных сервисов и сервисов своевременной доставки:

- тактильный канал обратной связи;
- при необходимости, каналы для дополнительных информационных потоков, которые связаны с тактильной обратной связью и должны быть синхронизированы с ней;
- управляющий канал.

Подобным же образом составные услуги голографической связи обеспечивают набор каналов с различными характеристиками (для обмена данными, метаданными, управляющими данными), каждый из которых, в свою очередь, представляет собой определенный базовый сервис.

Роль Интернета в цифровой инфраструктуре будущего

На первый взгляд, анализ сценариев показывает, что три опоры, о которых я говорил в начале статьи, требуют пересмотра. Все сценарии, за исключением, пожалуй, загадочного ManyNets, требуют от сетевой инфраструктуры высокоточных гарантий параметров передачи, абсолютной защищенности и значительной пропускной способности. Это, в свою очередь, требует значительно более широкого спектра услуг от базовой сетевой инфраструктуры, чем просто пакетной передачи без гарантий. Взгляните хотя бы на базовые сервисы, о которых я говорил выше, не говоря уже о составных услугах, которые могут порождать самые причудливые комбинации требований. А с учетом того, что именно эта «элементарность» сетевой функции и послужила залогом успешного развития Сети до сегодняшнего дня,

эволюционно Интернет развиться до требуемого состояния, похоже, не сможет.

В этой связи, конечно, возникает вопрос, каким образом эта инфраструктура будущего будет воплощена в жизнь? Двадцатилетний опыт внедрения IPv6, который, кстати, не меняет фундаментальных основ Интернета и всего, что на его базе построено, наглядно показывает, что силы рынка гораздо больше подвластны экономическим законам, чем техническим спецификациям. Особенно, если технология обладает низкой совместимостью с существующими и требует внедрения на глобальном уровне.

Однако, может быть, не так все плохо. Да, требования ряда сценариев по точности передачи значительно превышают архитектурные возможнности глобального Интернета. Но, во-первых, присмотревшись, обнаруживаешь, что многие из этих сценариев относятся скорее к интранетам - инфраструктуре под единым контролем, где все это, при желании, можно обеспечить, даже с использованием современных разработок (см., например, работу в IETF по детерминистическим сетям, https://datatracker.ietf.org/wg/detnet/about/, работы IEEE в области сетей, чувствительных ко времени (Time-Sensitive Networking, TSN), или такие технологии как программно-определяемые сети - SDN). Например, для тактильной связи требование сверхнизкой задержки в несколько миллисекунд сталкивается с физическими ограничениями из-за распространения сигнала, которое не может превышать скорость света (300 км в одну сторону за одну миллисекунду или 200 км в оптическом волокне).

Ну а во-вторых, там, где требуется междоменная связность, решение подчас лежит в изменении топологии и оптимизации связности, как это произошло с видеостримингом реального времени - приложение, немыслимое 20 лет назад.

Наверняка для обеспечения возможных приложений 2030 и более далекого будущего потребуются новые протоколы и новые подходы. Но разумнее всего разрабатывать их в виде технологических «строительных блоков», совместимых с технологиями Интернета и его тремя опорами. Не все из этих «блоков» будут использованы, но наиболее подходящие и соответствующие реальным требованиям будут применены и внедрены участниками рынка. Потому что, в конечном итоге, именно участники рынка – сетевые операторы, производители и, конечно, пользователи - определяют будущее Интернета.

Умные вещи века

Глеб Пыжов*

«Умные» вещи не смогли стать лучшими друзьями человека – слишком много вопросов вызывает их весьма вольное обращение с нашими самыми интимными данными, слишком многое они о нас знают, а кому рассказывают, понятно далеко не всегда. Зато бизнес-преимущества устройств IoT оценили уже давно: сегодня более 85% компаний в мире используют или внедряют решения Интернета вещей. В России цифры немного скромнее - IoT применяют чуть более половины российских предприятий. Больше всего решений Интернета вещей в транспорте и логистике, меньше всего – в самых, казалось бы, известных в этом отношении отраслях: в ЖКХ и в медицине. Причины, по которым Интернет вещей стал развиваться именно этим путем, мы и попробуем исследовать в нашей статье.

В 1990 году Джон Ромки (John Romkey) подключил свой тостер к компьютеру, добавив в конструкцию тостера специальный чип. Но надежды человечества на кофеварки, с которыми можно поговорить по душам, и холодильники, следящие за диетой хозяина лучше, чем личный врач, пока что не сбылись. Более того: уже через 20 лет после появления самого понятия «Интернет вещей» выяснилось, что тут ни вещами в привычном нам понимании, ни, собственно, Интернетом и не пахнет.

«Умные» вещи не смогли стать лучшими друзьями человека – слишком много вопросов вызывает их весьма вольное обращение с нашими самыми интимными данными, слишком многое они о нас знают, а кому рассказывают, понятно далеко не всегда. Зато бизнес-преимущества устройств IoT оценили уже давно: сегодня более 85% компаний в мире используют или внедряют решения Интернета вещей. В России цифры немного скромнее - IoT применяют чуть более половины российских предприятий. Больше всего решений Интернета вещей в транспорте и логистике, меньше всего – в самых, казалось бы, известных в этом отношении отраслях: в ЖКХ и в медицине. Причины, по которым Интернет вещей стал развиваться именно этим путем, мы и попробуем исследовать в нашей статье.

Что такое Интернет вещей?

На вопрос о том, что же такое Интернет вещей, российский «Яндекс» и американский Google отвечают очень по-разному, причем в первую очередь - в количественном отношении (см. рис. 1).

Разница в количестве ответов составляет несколько порядков. Но различия здесь не только количественные, но и качественные. Это видно, например, по уровню вакансий в этой области (см. рис. 2).

Ну а теперь от комикса перейдем сначала к теории, а потом и к практике. Начнем с определений. Существует несколько различных формулировок, которые в разные годы использовали как профессионалы, так и массовые пользователи. Все эти определения - попытка выразить различные взгляды и подходы к постоянно меняющемуся ландшафту новой реальности. У нас в Ассоциации Интернета вещей принято считать, что наиболее точными и емкими являются определения, которые утверждены МСЭ. Итак, Интернет вещей это, с одной стороны:

> Концепция вычислительной сети физических предметов («вещей»), оснащённых встроенными технологиями для взаимодействия друг с другом или с внешней средой, рассматривающая организацию таких сетей как явление, способное перестроить экономические и общественные процессы, исключающее из части действий и операций необходимость участия человека;

но также и:

> Сеть сетей уникально идентифицируемых объектов, осуществляющих интеллектуальное взаимодействие без человеческого вмешательства через IP-подобные соединения.

Почему «Интернет» и откуда в нем вещи?

Считается, что термин «Internet of things» ввел в обиход Кевин Эштон (Kevin Ashton), основатель исследовательского центра Auto-ID при Массачусетском технологическом институте (MIT). В 1997-1999 годах Кевин Эштон работал ассистентом бренд-менеджера в компании Procter & Gamble – и именно эта, казалось бы, не имеющая прямого отношения к высоким технологиям должность и привела Эштона к идеям Интернета вещей. В 1999 году Кевин Эштон готовил презентацию для своего руководства, в которой рассказывалось о системах сквозной маркировки и прослеживаемости продукции. Идеи использования RFID для управления цепочкой поставок заинтересовала Эштона и привела в MIT, где он и основал свой исследовательский центр. Центр открылся как исследовательский проект, спонсируемый отраслью, с целью создания глобальной системы открытых стандартов, которая повсеместно использовала бы RFID. Центр и его лаборатории и сегодня продолжают свою работу под названием Auto-ID Labs.

Само же явление «Интернет вещей» Кевин Эштон объяснял так:

Рис. 1. Что же такое Интернет вещей по мнению российского «Яндекс» и американского Google поисковиков.

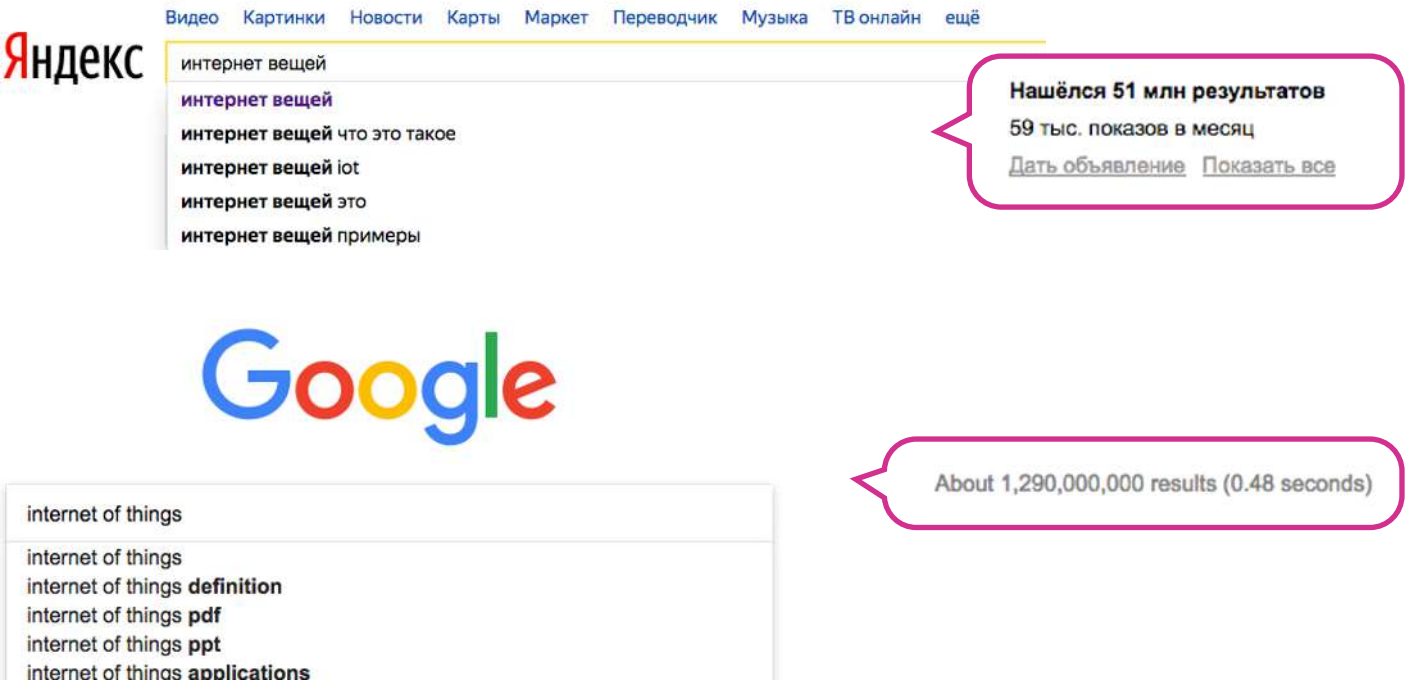
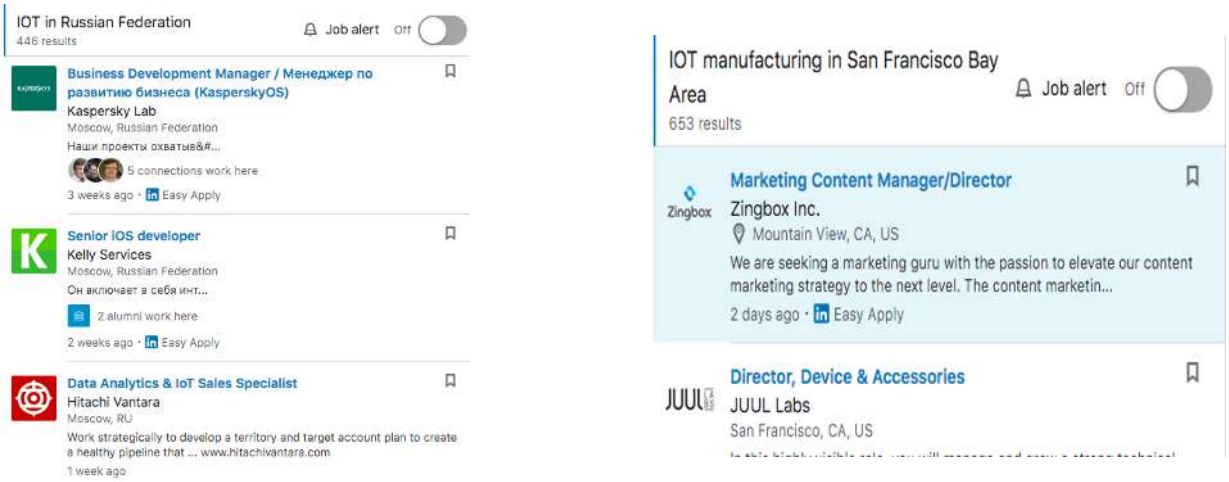


Рис. 2. Уровень вакансий в этой области.



Само же явление «Интернет вещей» Кевин Эштон объяснял так:

«Если бы у нас были компьютеры, которые знают об окружающих нас вещах на основе данных, которые собираются без нашей помощи, - мы смогли бы посчитать и отследить вообще все вокруг, и это снизило бы уровень отходов, потерь и затрат. Мы бы знали, когда и какие вещи нуждаются в замене, ремонте или апгрейде, и знали бы, откуда они взялись».

«Если бы у нас были компьютеры, которые знают об окружающих нас вещах на основе данных, которые собираются без нашей помощи, - мы смогли бы посчитать и отследить вообще все вокруг, и это снизило бы уровень отходов, потерь и затрат. Мы бы знали, когда и какие вещи нуждаются в замене, ремонте или апгрейде, и знали бы, откуда они взялись».

Обратите внимание: Интернет в 1999 году уже есть и вовсю развивается, но речь идет вовсе не о связи с помощью Всемирной сети: Эштон говорит о применении средств радиочастотной идентификации для взаимодействия физических предметов между собой и с внешним окружением.

Но с момента появления термина и самой концепции Интернета вещей в массовом сознании циркулировала мысль исключительно о бытовом применении этой технологии. Возможность управлять своим домом через Интернет, давать задания кофеварке и тостеру с помощью компьютера и писать электронные письма системе садового полива очень нравилась обывателю, и журналисты с удовольствием публиковали статьи о том, как прекрасен «умный дом». Причем все эти варианты домашней автоматизации были далеко не новы, и новой там была лишь идея объединения устройств и «вещей» в единую вычислительную сеть, обслуживаемую интернет-протоколами, и рассмотрение «Интернета вещей» как особого явления.

На рубеже 2009 года произошло еще одно событие, которое аналитики считают «настоящим рождением «Интернета вещей»: количество устройств, подключенных к Глобальной сети, превысило численность населения Земли, тем самым «Интернет людей» стал «Интернетом вещей».

Но вот тут-то и случилось то, чего следовало ожидать даже раньше. Хакерам понадобилось несколько лет для того, чтобы понять, как объединять устройства «Интернета вещей», подключенные именно через Глобальную сеть, в мощные ботнеты. Дело в том, что наличие «ума» у кофеварки или охранной системы, как ни странно, не предполагает у них сколь-либо серьезной защиты от кибератак. Для такой защиты нужны серьезные и дорогие модули, а это крайне нерентабельно: ведь никто не будет переплачивать за киберзащиту кофеварки. Этим-то и воспользовались хакеры, в результате чего в 2016 году мир столкнулся с ботнетом, мощнее которого никто не видывал – речь идет о Mirai (https://ru.wikipedia.org/wiki/Mirai_(ботнет)).

Нет, атаки ботнетов, состоявших из устройств Интернета вещей, предпринимались и ранее, но именно Mirai заставил задуматься, а так ли уж нужен Интернет в Интернете вещей? Ответ – нет, и сегодня индустриальный IoT, да и другие серьезные направления развития Интернета вещей стараются минимизировать свою связь с Глобальной сетью.

Кому с Интернетом вещей жить хорошо?

Интернет вещей сегодня в той или иной степени коснулся практических всех отраслей экономики. Конечно, уровень внедрения решений IoT в разных отраслях не одинаков и во многом зависит от того, насколько экономически эффективным будет это внедрение. Давайте посмотрим, как выглядит сегодня рынок IoT в мире и в России.

И тем не менее, посмотрим, с чем российский рынок IoT подошел к середине 2020 года. Самый высокий рост, по результатам прошлого года, показали транспорт и логистика – 12%. На втором месте по уровню роста – устройства для маркировки и прослеживаемости товаров: 10%. Высокими темпами внедрение решений IoT идет в обрабатывающих отраслях. А вот в энергетике, ЖКХ, муниципальных услугах решения IoT внедряются не спеша – это довольно непростая среда для ведения бизнеса, и затрагиваются здесь не только экономические, но и социальные интересы, так что сопротивляемость внедрению, например, устройств объективного учета потребления коммунальных услуг очень высокая.

В конце 2019 года более 50% российских компаний уже внедрили или же планировали в течение 2020 года закончить внедрение решений с использованием технологий Интернета вещей. Такие данные в своем отчете «Возможности и тенденции Интернета вещей: углубленный анализ российского рынка» (Russia Internet-of-Things Market 2019-2023 Forecast) предоставила в декабре 2019 года аналитическая компания IDC. Этот показатель ниже, чем у мировых компаний – но отставание обусловлено, скорее, не технологическими, а экономическими и частично регуляторными причинами. Общий объем рынка IoT в России составил в 2019 году 3,7 миллиарда долларов, но этот рынок, по мнению аналитиков IDC, не консолидирован, на нем очень много игроков и нет ярко выраженного лидера. Общая же динамика внедрения проектов IoT на российском рынке, по мнению IDC, в ближайшие пять лет могла бы составить 19,7%, причем прогноз по российскому рынку оказался даже выше мирового.

Почему «могла бы»? Потому что сегодня уже очевидно, что такого роста в 2020 году точно не будет – пандемия внесла свои очень серьезные коррективы в развитие мировой экономики в целом и Интернета вещей в частно-

сти. И если 2019 год в отношении Интернета вещей можно было назвать «годом осознанности», то 2020, похоже, станет годом освобождения от иллюзий.

И тем не менее, посмотрим, с чем российский рынок IoT подошел к середине 2020 года. Самый высокий рост, по результатам прошлого года, показали транспорт и логистика – 12%. На втором месте по уровню роста – устройства для маркировки и прослеживаемости товаров: 10%. Высокими темпами внедрение решений IoT идет в обрабатывающих отраслях. А вот в энергетике, ЖКХ, муниципальных услугах решения IoT внедряются не спеша – это довольно непростая среда для ведения бизнеса, и затрагиваются здесь не только экономические, но и социальные интересы, так что сопротивляемость внедрению, например, устройств объективного учета потребления коммунальных услуг очень высокая.

Интересно развивается IoT в сельском хозяйстве. С одной стороны, сельское хозяйство – одна из наиболее устойчивых отраслей экономики: производство еды будет востребовано всегда и в любых условиях. С другой стороны, это достаточно консервативная отрасль, где есть как объективные, так и субъективные препятствия к внедрению Интернета вещей. Тут и огромные площади, которые необходимо покрывать сетью датчиков и сенсоров, и привычка работать «по-старинке», и элементарное недоверие к современным устройствам IoT. При этом именно в сельском хозяйстве появляются подчас самые изящные и востребованные решения на основе IoT.

Например, агрохолдинг «Красава» в Пермском крае только за первый месяц после внедрения инновационных решений на основе IoT сумел увеличить производительность предприятия на 20%. Проект по внедрению IoT предложил «ЭР-Телеком Холдинг» – крупнейший оператор и интегратор промышленного Интернета вещей. Так, вся сельскохозяйственная техника агрохолдинга была оснащена IoT-трекерами, контролирующими местоположение транспорта. В зданиях агрокомплекса было реализовано решение по сбору информации с приборов учета энергоресурсов в режиме реального времени. А на цистерны временного хранения продукции были установлены беспроводные IoT-датчики, контролирующие актуальные данные по объему отгруженного молока и соответствие нормам. Помимо этого, в зоне отгрузки продукции была размещена система видеонаблюдения, позволяющая в режиме онлайн отслеживать действия персонала. Все эти, казалось бы, простые и очевидные решения, интегрированные профессионалами, показали свою высокую эффективность.

Еще одно популярное платформенное решение, которое базируется на технологиях IoT и больших данных, – платформа «Агросигнал». Платформа способна рассчитать и площадь обработки земли, и объем собранного урожая, и расход топлива, и скорость, с которой движется по полю сельхозтехника. «Агросигнал» работает от Краснодара до Забайкалья и обрабатывает данные, поступающие с 4 млн гектаров. Системой пользуются более 200 российских сельхозпредприятий: и крупные агрохолдинги, и небольшие компании.

Очевидно, что Agro IoT имеет все возможности для того, чтобы в перспективе стать одним из флагманов промышленного Интернета вещей. А вот то направление, которое казалось всем наиболее «продвинутым», а именно «умный дом», развивается вовсе не такими высокими темпами, как можно было бы себе представить. В 2017-2018 годах это направление Интернета вещей не росло совсем, в 2019 году показало небольшой рост (в районе 3%). Перспективы его в ближайшую пару лет вообще не ясны: все же для большинства людей использование «ручных» IoT-устройств является чем-то вроде игры, а из этого следует, что «умные» замки, системы освещения и стиральные машины в условиях возможного экономического кризиса и падения доходов населения наверняка не войдут в список услуг первой и даже второй необходимости.

В целом же, по оценке Ассоциации Интернета вещей, общий рост российского рынка Интернета вещей составил в 2019 году 8,3%, и уже тогда основной рост пришелся на IT- и бизнес-услуги – они заняли почти 40% российского рынка внедрений IoT. Далее идут инвестиции в оборудование – серверные мощности и устройства: датчики, сенсоры для сбора данных и т.д.

В 2020 году эта тенденция будет только нарастать: все, кто будет заниматься внедрением IoT, сосредоточатся на менее ресурсоемких решениях, не требующих инвестиций в инфраструктуру и оборудование. В этом отношении весьма интересны облачные решения, которые как раз стали появляться в последнее время как ответ на сложившуюся почти экстремальную ситуацию.

Приведем живой и свежий пример. В апреле 2020 года компания АПРОТЕХ («дочка» «Лаборатории Касперского» и ИТЭЛМА) представила программно-аппаратный комплекс, предназначенный для сбора и обработки «сырых» промышленных данных (шлюз промышленного Интернета вещей — IIoT GateWay). Это сквозной цифровой сервис, который построен на открытой облачной платформе и собирает данные о состоянии оборудования, параметрах окружающей среды и даже условиях работы персонала. Зачем он нужен? А затем, чтобы можно было собрать и обработать именно объективные, «сырые» данные – ведь почти 85% промышленного оборудования по всему миру остаётся неподключенным именно потому, что уверенности в полной достоверности данных, которые в итоге ложатся в основу анализа, нет. Есть опасность и перехвата данных, и их намеренного искажения, и т.д. Поэтому решение, гарантирующее безопасность передаваемых данных, рынку необходимо, особенно сегодня: сводятся к минимуму действия человека в производственном контуре, а экспертная оценка данных может проводиться в онлайн-режиме небольшой группой специалистов или с помощью технологий машинного обучения. Решение, кстати, уже внедрено и на предприятиях Группы компаний ЧТПЗ.

Особенности национальных стандартов

«Зоопарк стандартов» - так до недавнего времени можно было назвать то, что происходило в области стандартов как в России, так и за рубежом. Технологии Интернета вещей, собственно, тем и хороши, что передавать данные можно

с помощью самых разных протоколов, и каждый может выбрать то, что подходит для его случая. Но лавинообразный рост числа стандартов Интернета вещей во всем мире грозил превратить мир IoT в настоящую Вавилонскую башню – с сопоставимыми по масштабам последствиями.

Эта проблема была замечена многими специалистами, и одной из главных задач в области регулирования Интернета вещей стала разработка единых стандартов, которые бы применялись во всем мире (или хотя бы на территории нескольких стран или даже одной страны). Важно было не допустить ситуации, когда каждый разработчик и каждый интегратор будет предлагать решение, основанное только на его собственных стандартах. Это удалось, и здесь стоит сказать о значительной роли, которую в этом процессе сыграла Ассоциация Интернета вещей и ее члены.

В нашей стране существуют два вида регулирования технической деятельности, в том числе и IoT: нормативно-правовое и нормативно-техническое. Если НПР это прерогатива государства, то НТР законом передано рынку. И здесь Ассоциация Интернета вещей активно участвует в работе по созданию экосистемы нормативно-технического регулирования, вместе с бизнес-сообществом и организациями, агрегирующими запросы рынка, такими как ТК 194 «Кибер-физические системы»(Технический комитет Росстандарта). Основная цель такой работы – выработка национальных стандартов, которые обеспечат совместимость, мультивендорность и мультиплатформенность для всего Интернета вещей.

Так, с апреля 2019 года действует предварительный национальный стандарт «Протокол беспроводной передачи данных на основе узкополосной модуляции радиосигнала» (NB-Fi), который был утвержден Федеральным агентством по техническому регулированию и метрологии (Росстандарт). Этот стандарт был разработан АИВ совместно с ТК 194.

Этими же участниками в 2019 году разработан проект предварительного национального стандарта «Информационные технологии. Интернет вещей. Протокол обмена для высокоемких сетей с большим радиусом действия и низким энергопотреблением», который определяет сетевой протокол и системную архитектуру сети LoRaWAN (Long Range Wide Area Networks).

Другая важная часть работы Ассоциации Интернета вещей – участие в международных организациях, таких как технические консорциумы и комитеты по стандартизации. Участвовать в этой работе просто необходимо: это залог совместимости отечественных решений с международными стандартами, а также реальная возможность продвижения отечественных стандартов на мировой рынок. Например, стандарт LoRaWan получил российские региональные параметры именно благодаря успешному взаимодействию рабочей группы АИВ LoRaWAN с LoRa Alliance, а упомянутый ранее проект предварительного национального стандарта был поддержан международными экспертами во главе со специалистами LoRa Alliance в качестве полноценного протокола семейства LoRaWAN и одобрен к использованию в качестве региональной спецификации для российского рынка LoRaWAN RU.

Кроме того, Ассоциация Интернета вещей - коллективный участник Консорциума промышленного Интернета вещей (IIC, https://www.iiconsortium.org/). В работе Консорциума участвуют и многие члены АИВ.

Текущая парадигма национальной стандартизации в области Интернета вещей – это не определение границ и барьеров, как это делается в нормативно-правовом регулировании, а создание комфортной среды для развития рынка, построение технологического стека, которым смогут пользоваться все.

Вкалывают роботы – счастлив человек?

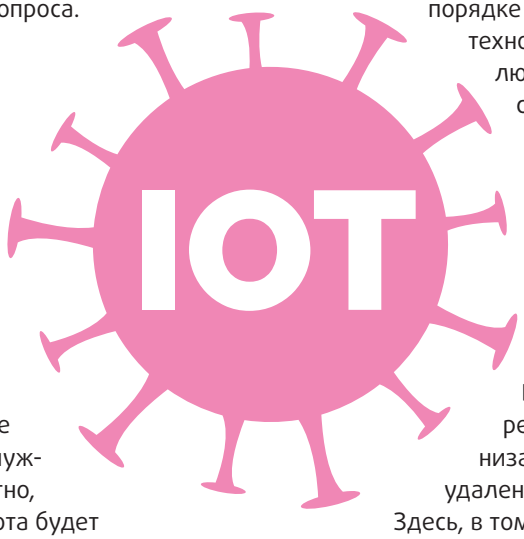
Одна из основных целей человечества на протяжении всей его истории – это сделать так, чтобы ничего не делать. В народных сказках практически всех народов мира обязательно присутствует сюжет о том, как некоему доброму молодцу удалось получить власть над чем-то волшебным (печью, котом, пони, рыбой – список можно продолжать до бесконечности), и теперь эти волшебные животные и предметы выполняют за хозяина всю работу. А он лежит на печи да ест калачи.

Похоже, что Индустрия 4.0 (https://ru.wikipedia.org/wiki/Четвертая_промышленная_революция) нас всех к этой идеальной картине мира приблизила. На дорогах уже появились беспилотные автомобили, и не за горами будущее, в котором умение водить машину перейдет в разряд экзотических навыков. Целые заводские цеха уже давно работают практически без участия людей, картины заводской проходной, через которую утром по гудку проходят тысячи угрюмых людей, тоже остались в далеком прошлом. Да те же лекции в вузах – сегодня, в эпоху массовой самоизоляции, выяснилось, что лектору совсем не обязательно надрывать горло каждый день по многу раз – достаточно записать нужный цикл лекций, а студенты будут его слушать в Интернете сколько угодно раз. В общем, картина прекрасная во всем, если бы не одно «но».

Возьмем, к примеру, тот же транспорт, в котором внедрение технологий Интернета вещей идет чуть ли не самыми высокими темпами. Безлюдные технологии пришли сюда и начали вытеснять живых людей с их водительских мест. Пионерами здесь стали дальние перевозки, а также машины, работающие на горнодобывающих карьерах. И там, и там маршрут совершенно понятен и относительно прост, внештатные ситуации вполне поддаются алгоритмизации, а водители, потерявшие работу, пройдут переподготовку и займутся чем-то другим. Но тут обнаружилось, что те самые дальнобойщики, которых заменяют бездушные роботы, создавали множество рабочих мест на всем пути следования. Ведь живому дальнобойщику надо поесть, сходить, пардон, в туалет, сделать покупки, ему может понадобиться лекарство, новая футболка или стакан кофе. И те, кто живут в населенных пунктах вдоль трасс, прекрасно к этому приспособились и давно уже создали развитую инфраструктуру, удовлетворяющую всем потребностям дальнобойщиков – от кафе и туалетов до живого человеческого общения. Но роботам-то все это не нужно. И если чуть сгустить краски, то картина выглядит довольно апокалиптически: бездушный робот-трак едет по стране,

оставляя за собой опустевшие города и поселки и тысячи голодных и озлобленных людей: всех-то на другие специальности не переучишь, да и кто этим будет заниматься в таком масштабе.

Свою лепту в историю развития Интернета вещей в мире прямо на наших глазах вносит и пандемия. С одной стороны, по мнению множества специалистов, серьезное развитие IoT в ближайшие год-два, скорее всего, станет невозможным: серьезных инвестиций в инфраструктуру Интернета вещей ожидать сейчас не стоит. С другой стороны, в мире, где самоизоляция становится фактически условием выживания, безлюдные сценарии обсуждаются все чаще и чаще, и то, что совсем недавно было просто хайпом, может стать реальным инструментом для спасения экономики.

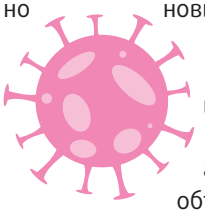


Но и здесь есть оборотная сторона вопроса. К примеру, технологии IoT уже давно разработаны для горнодобывающей промышленности, их эффективность получила практическое подтверждение, и постепенно эти самые безлюдные технологии внедряются на все большем числе карьеров, заменяя собой людей. Соответственно, рабочие места сокращаются, а эффективность производства и его безопасность повышаются. И это было прекрасно ровно до того момента, пока огромное количество людей не оказалось в вынужденном бездействии, и вполне вероятно, что после завершения карантина работа будет далеко не у всех. Стоит ли в такой ситуации внедрять новые технологии и сокращать рабочие места – или же разумнее будет подождать пока с заменой человека на устройство IoT?

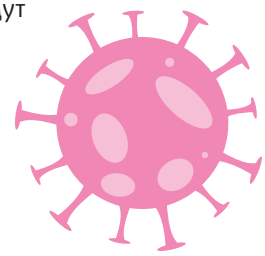
В общем, идеальной картины, где роботы вкалывают, а Емеля лежит на печи и ест калачи, никак не получается. Куда деть всех этих освободивших «человеков», никто пока не придумал – в мире «умных» гаджетов для человека вполне может не остаться места.

Интернет вещей на фоне Covid-19: жертва или лекарство?

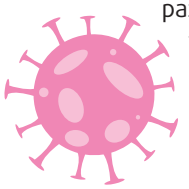
Как уже говорилось раньше, пандемия весьма сурово обошлась с Интернетом вещей. По сути, направление встало: проекты не развиваются, инвестиции заморожены, порваны многие снабженческие и технологические цепочки. Большинство решений IoT в 2020 году будут основаны именно на сервисной модели: ожидать инвестиций в долгосрочные проекты и создание сложных решений не приходится - останутся только те инвестиции, которые показывают быстрые результаты. Разумеется, циклы инвестиций, которые были начаты ранее, будут завершены, но новых денег на развитие IoT пока не будет.



Зато явно в фаворе оказался «человеческий IoT» - инвестировать будут (и уже начали) в медицину, безопасность, а также в ускорение создания систем объективного учета и контроля. Это как раз те направления, которые будут востребованы в мире еще очень долго.

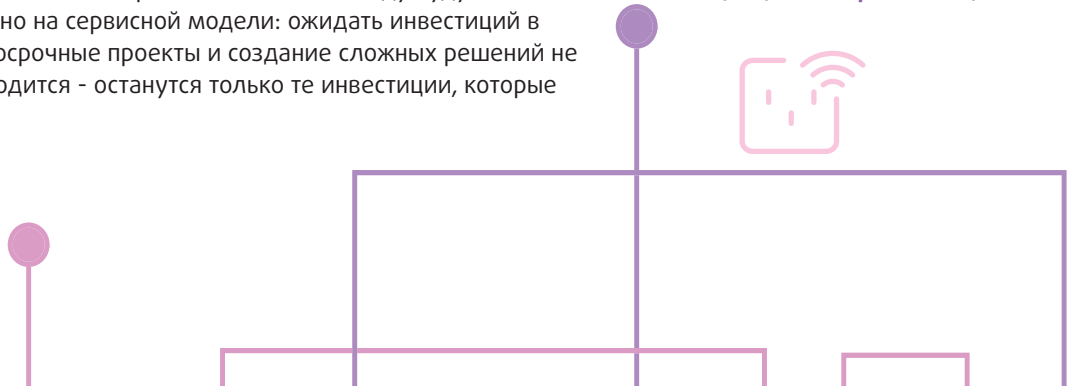


Но Интернет вещей может сыграть и серьезную социальную роль. Люди вынужденно перешли на удаленную работу – и нравится им это или нет, им всем приходится в авральном порядке осваивать цифровые технологии. Если не поддержать этих людей – причем с проблемами столкнулись все, от обычных клерков до руководителей предприятий, – результаты могут быть весьма печальными. И здесь поддержку как раз оказало государство: Минпромторг России вместе с Ассоциацией Интернета вещей разработал Витрину технических решений для организации удаленной работы. Здесь, в том числе, представлены апробированные продукты, предназначенные для удаленного мониторинга и управления производственным оборудованием и логистическими процессами, организации удаленных рабочих мест и формирования всей необходимой виртуальной инфраструктуры.



Разумеется, драматические события, разворачивающиеся на наших глазах, окажут (и уже оказывают) серьезное влияние на российский и всепланетарный Интернет вещей. Но в отрасли работают креативные люди – и можно быть уверенными, что Интернет вещей выстоит, вырастет и станет одной из главных движущих сил мировой экономики. Как, собственно, случалось в предыдущие годы со всей IT-отраслью, которая в эпоху кризисов всегда росла и крепла.

*** Автор Глеб Пыжов является заместителем директора Ассоциации Интернета вещей**



Интернет больших данных

Юрий Демченко

Большие данные выросли из концепции и броского термина в набор технологий, архитектур и приложений, которые, совмещенные с современной аналитикой, предоставляют практически неограниченные возможности для оптимизации процессов, а также открывают новые возможности для научных исследований. Даная статья предлагает краткий структурированный обзор развития этой области и тенденций создания инфраструктуры обмена и обработки данных, которые можно определить как будущий Интернет (больших) данных. Статья приводит примеры стандартизации, проектов и инициатив для обеспечения доступности, обнаружимости и совместимости данных.

1. Вступление

Большие данные (БД) являются устоявшимся термином и как концепция обозначают данные чрезвычайно большого объема, недоступные обработке традиционными программными средствами с реалистичными сроками. К характеристикам БД относят их объем, скорость генерации и обработки, а также степень разнообразия типов и природы данных. От БД неотделимы технологии их обработки и анализа, широко используемые практически во всех областях деятельности человека и частью бизнес-процессов современных фирм. Большие данные являются результатом современных бизнес-процессов, промышленных и технологических процессов, а также социальной активности человека в соцсетях. На уровне вычислительных и коммуникационных технологий/процессов данные представляют собой основной продукт, цель и инструмент работы большинства приложений и процессов. Большие данные, совмещенные с современной аналитикой, предоставляют практически неограниченные возможности для оптимизации процессов, а также открывают новые возможности для научных исследований. Широкое использование данных в бизнесе привели к появлению так называемых предприятий, управляемых (и зависимых от) данными (Agile Data Driven Enterprise), то же самое происходит с современной наукой, которая становится также зависимой от данных и все больше использует методы исследований на основе интенсивного анализа данных (Data Driven Science and Research).

Развитие технологий и инфраструктуры больших данных привело к появлению науки о данных (Data Science) и сильно стимулировало развитие облачных вычислений (Cloud Computing), которые совместно смогли обеспечить необходимые ресурсы и платформу/ресурсы для обработки, хранения и передачи/обмена больших данных. В свою очередь, развитие инфраструктуры больших данных и облачных вычислений привело к значительным изменениям в том, как работает и развивается современный Интернет. Многие ветераны-исследователи Интернета, а также профессиональные организации в области телекоммуникационных технологий (таких как ITU-T, TMForum) видят тенденцию эволюции современной телекоммуникационно-информационной инфраструктуры от Интернета к Интернету вещей и к Интернету данных. Большинство современных приложений в явном или неявном виде используют элементы распределенной и глобальной инфраструктуры данных, по сути, Интернета данных. Мы можем ожидать множество инноваций и приложений на этом технологическом уровне, которые должны обеспечить возможность глобального обмена, интеграции и даже торговли данными при условии соблюдения суверенности, доверительности и защиты персональных данных. Множество инициатив существует в этой области в Европе и в мире, включая поддержку Европейской рамочной программы Horizon 2020.

2. Свойства больших данных

Первые определения больших данных были предложены в 2011 году. Определение аналитической фирмы IDC можно рассматривать как достаточно консервативное: «Новое поколение технологий и архитектур, предназначенных для экономически эффективного извлечения полезной значимой информации из большого количества разнообразных данных посредством эффективного сбора, обнаружения и анализа» (IDC, 2011)[1].

Аналитическая фирма Gartner предложила определение больших данных как информационных активов большого объема, высокой скорости, большого разнообразия, которые требуют новых эффективных форм обработки информации для улучшенного понимания и принятия решений (Gartner, 2011). Такое определение фактически ссылается на «основные 3V» больших данных (Volume, Velocity, Variety - Объем, Скорость, Разнообразие), указывает на новые методы аналитики, необходимость учитывать целевое применение БД.

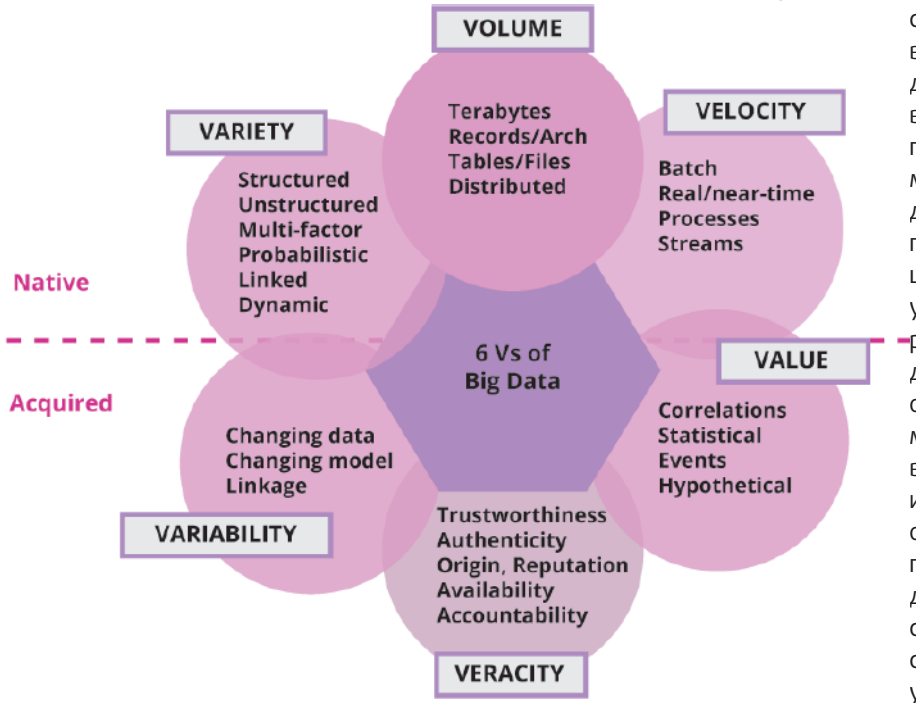
Однако наиболее популярным стало определение от Forrester Research (2013): «Большие данные представляют массивный объем структурированных и не структурированных данных, которые настолько велики, что их сложно обрабатывать, используя традиционные базы данных и программные средства»[2]. Эффективность такого определения, появившегося в то время, когда появились первые средства для работы с большими данными и наметились определенные тенденции в области разработки и применения БД, в основном связана с разработкой нереляционных баз данных (NoSQL), средств хранения БД и масштабируемых средств массивной обработки, таких как MapReduce, Hadoop, Spark.

Изначально БД характеризовались тремя основными свойствами, так называемыми 3V. В дальнейшем к базовым 3V добавились свойства больших данных, которые они приобретают в процессе обработки и которые являются основополагающими для определения требований к системам обработки больших данных: Value, Variability, Veracity (Значимость, Изменчивость, Доверительность).

Рисунок 1 представляет графически базовые и приобретенные свойства БД, которые составляют 6V больших данных; также на рисунке перечислены основные характеристики, которые относятся к соответствующим свойствам.

Однако быстрое развитие и внедрение технологий БД в бизнес и промышленность было уже хорошо подготовлено развитием систем и платформ для научных исследований, создавших предпосылки для реализации новых методов научных исследований, которые были определены ученым

Рис. 1. 6V больших данных (BDAF, 2014).



в области компьютерных наук Джимом Греем (Jim Gray) как четвертая парадигма научных исследований: «Средства и технологии для такой наукоемкой науки, управляемой данными, являются настолько иными, что имеет смысл отличать науку, управляемую данными, от вычислительной/компьютерной науки как новую четвертую парадигму научных исследований»[3]. (Grey, 2010) Третья парадигма научных исследований уже использовала вычислительные технологии, но в основном для моделирования и симуляции научных теорий и моделей, разработанных классическими теоретическими (аналитическими) и экспериментальными научными методами. К моменту написания книги Грея[4] наука уже стояла перед задачами, которые имели дело с большими объемами данных и требовали новой инфраструктуры для обработки экспериментальных данных, включая обеспечение сотрудничества большого количества всемирно/глобально распределенных групп ученых.

3. Примеры применения больших данных в науке и промышленности

Давайте рассмотрим примеры использования больших данных в науке и технике, которые создали предпосылки и стимулировали развитие технологии БД.

3.1. Большой адронный ускоритель LHC и инфраструктура больших данных LCG

В первую очередь применение БД относится к глобальному международному проекту по обнаружению частицы Хиггса, бозона, для чего в 2008 году было закончено строительство величайшего научного инструмента - Большого адронного ускорителя (Large Hadron Collider, LHC) в ЦЕРНе в Швейцарии. Строительство ускорителя заняло несколько лет и потребовало не только сотрудничества ученых со всего мира, но и решения новых задач сбора, хранения, распространения и обработки экспериментальных данных, производимых датчиком столкновения пучков протонов ALICE, разогнанных до околосветовой скорости. Фактически само столкновение частиц продолжается в течение 10 мс, скорость производства данных составляет единицы PB (Petabyte) в секунду. Суммарное количество данных, производимых в экспериментах в течение месяца, составляет более 10 PB, данные должны быть собраны/зарегистрированы, первично обработаны, сохранены в центрах хранения данных первого уровня и распространены до уровня региональных научных центров обработки данных по всему миру. Обработка всего объёма экспериментальных данных не могла быть выполнена средствами одного вычислительного центра в то время (да и сейчас это тоже невозможно). Таким образом, случай LHC демонстрирует пример двух важнейших свойств больших данных: большой объем и большая скорость, - и технические задачи, которые сопряжены с этим. Как известно, работа ускорителя и глобальное сотрудничество исследователей всего мира увенчались успешным открытием бозона в 2013 году.

Для создания инфраструктуры обработки данных, производимых ускорителем, международное сообщество финансировало несколько проектов по созданию Worldwide LHC Grid (WLCG), которые включали проект EGEE (Enabling Grids for E-sciencE), финансированный европейскими рамочными программами FP6-FP7 в 2004-2010, проект Open Science Grid в США, финансированный странами-участниками проекта. В настоящее время инфраструктура европейского грида поддерживается ассоциацией «Европейская грид-инициатива» (EGI – European Grid Initiative)[5] совместно с координационными органами во всех европейских странах. EGI предоставляет вычислительные ресурсы (как суперкомпьютеров, так и облачных центров данных) практически для всех наукоемких европейских научных проектов, а также продолжает развитие инфраструктурных сервисов

грид, которые в настоящее время работают на основе стандартных облачных платформ.

Результатом проекта EGEE и соответствующего международного сотрудничества было не только создание собственно инфраструктуры для обработки научных данных, но и фактически ускорение появления таких новых направлений в области вычислительных и информационных технологий как Web Services (включая SOAP- и REST-протоколы), вычислительные гриды (Computer Grids) и модель управления вычислительными задачами в распределенных центрах данных, федеративная модель управления идентификацией пользователей и федеративный доступ к сетевым вычислительным ресурсам и данным.

Многие идеи по доступу и интеграции данных, изначально предложенные в компьютерных грид-системах, были развиты в современной архитектуре облачных вычислений (Cloud Computing) и федеративного доступа к распределенным сервисам. Это также стимулировало широкую международную стандартизацию, целью которой является обеспечение эффективного доступа к распределенным вычислительным ресурсам, включая обеспечение безопасности данных. Другие направления стандартизации включают разработку методов и форматов описания, идентификации и поиска данных, таких как метаданные, постоянные/глобальные идентификаторы (PID – Persistent Identifiers), доверительные распределенные вычислительные модели.

3.2. Большие данные в промышленности и бизнесе

Промышленные процессы и машины в настоящее время полностью компьютеризованы и автоматизированы, что приводит к генерированию огромного количества данных, которые необходимо сохранять и обрабатывать для оптимизации процессов и - в конечном итоге - для увеличения безопасности и уменьшения затрат. Промышленный Интернет вещей (Industrial Internet of Things)[6] в настоящее время включает в себя и опирается на такие технологии как облачные вычисления, граничные вычисления (Edge Computing), автоматизация и промышленные роботы, искусственный интеллект и цифровые двойники (Digital Twins)[7], в основе которых лежит работа с большим объемом разнообразных данных, требующих быстрой обработки для управления процессами или принятия решений. Перечисленный комплекс технологий предъявляет свои требования к инфраструктуре больших данных, которые включают кроме общих требований к хранению, обработке и доступу к большим данным дополнительные требования по защищенному обмену данными, стандартизации описания данных, качеству данных, хранению истории данных и пр.

В случае авиационной промышленности сбор и хранение данных о работе всех систем самолета является основой прогнозного мониторинга, который общепризнано позволяет существенно снизить стоимость обслуживания самолета и повысить безопасность. Известно, что четырехмоторный самолет Boeing 747 Jumbo производит 640 Тбайт данных за время трансатлантического перелета, и все это должно быть умножено на более чем 25 тысяч полетов каждый день. В случае инцидента данные о необходимых системах должны быть извлечены и обработаны, используя необходимые методы машинного анализа данных. Качество анализа и диагностики зависит в большой степени от цифровой модели самолета (по сути, цифрового двойника) и качества данных. Характерная ситуация в расследовании авиационных инцидентов требует обмена данными между различными организациями и фирмами, часто конкурирующими, и это требует создания соответствующей инфраструктуры для доверительного обмена данными и безопасной доверительной среды обработки, которая бы защищала как входные и выходные данные, так и алгоритмы, которые также могут представлять коммерческую тайну.

Классическим примером использования больших данных являются поисковые системы. Поисковая система Google, которая является абсолютным доминантом на рынке поисковых систем, использует все наилучшие системы и алгоритмы для обработки БД с целью предоставления качественных результатов (и в наименьшее время), ожидаемых пользователями. Результатом исследований в Google и других фирмах в области поисковых систем и инфраструктуры БД стало появление открытого программного обеспечения для обработки больших данных, такого как Hadoop, MapReduce, HBase, Pig, Hive, которые составляют основу современных технологий больших данных. Все указанные приложения позволяют работать с глобально распределенными данными в масштабе Интернета.

Кроме основного бизнеса предоставления поисковых результатов, поисковые системы и Google, в частности, предоставляют платные целевые объявления/рекламу, которая является одной из основных форм зарабатывания денег на поисковом бизнесе. Таргетированная реклама (англ. targeted advertising) включается в страницу поисковых результатов на основе контекста запроса, истории запросов пользователя или дополнительно других данных о пользователе, собираемых поисковой системой. Но критически важным критерием для эффективности такой рекламы является ее включение в ответную страницу в течение 40 мс после получения запроса, в противном случае внимание пользователя будет переключено на собственно результаты поиска. Оптимальное решение этой задачи является одной из проблем как алгоритмов обработки данных, так и построения самой инфраструктуры данных.

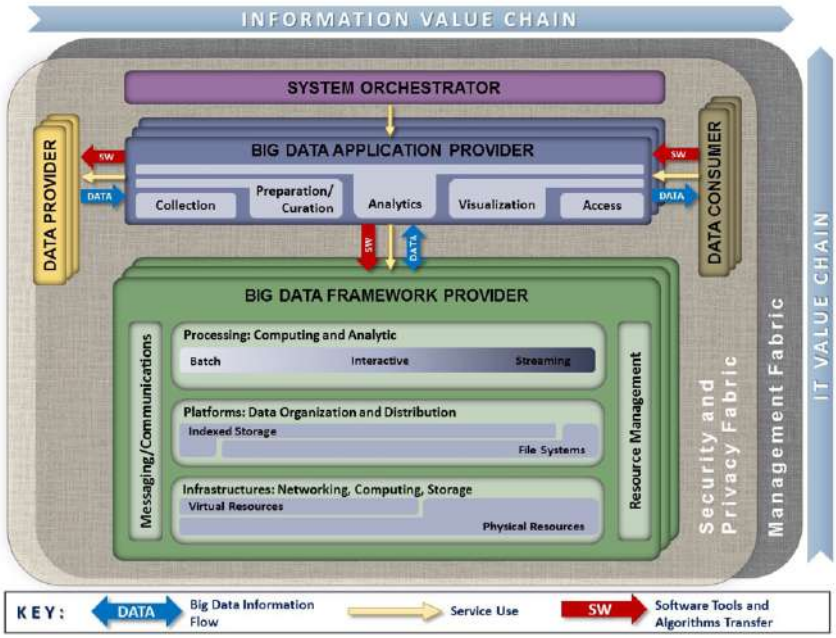
3.3. Другие примеры использования больших данных в науке и промышленности

Исследования генома и генетические исследования в общем являются одним из примеров использования различных аспектов больших данных, включая обработку данных секвенирования генома, хранение эталонного генома и доступ к нему, а также хранение индивидуальных геномов при массовых исследованиях. Согласно различным оценкам, полное секвенирование индивидуального генома человека производит около 250 Гбайт исходных данных[8] (собственно объем генома оценивается в 4,23 Гбайт[9]), последующая обработка с целью диагностики может потребовать доступа к геномным базам данных и производить дополнительно большой объем данных. Эффективное

использование этих данных для научных исследований и в медицинских целях ставит задачу создания специальной инфраструктуры геномных данных, которая бы позволяла семантический поиск по шаблонам и доступ к необходимым частям генома.

В Европе создана специальная организация ELIXIR[10], которая координирует европейские проекты по исследованию генома и поддерживает инфраструктуру для доступа к базам данных геномов, включая эталонный геном человека, приложения для геномных исследований, вычислительные ресурсы. Одной из задач инфраструктуры ELIXIR является распределение и обслуживание реплик базы данных геномов от центральной базы данных до национальных баз данных и тех, которые используются в больших исследовательских центрах, включая контроль доступа, качество и защиту данных.

Рис. 2. Эталонная архитектура систем больших данных NIST (NBDRA, 2020).



4. Эталонная архитектура систем больших данных и другие стандарты

Работа с распределенными и разнообразными данными больших объемов с самого начала потребовала стандартизации на уровне инфраструктуры, описания и идентификации данных, что стало предметом работы многих организаций по стандартизации, как международных, так и национальных или отраслевых, общее направление которых можно определить как создание Интернета (больших) данных. В этом разделе мы рассмотрим эталонную архитектуру больших данных (NIST Big Data Reference Architecture, NBDRA), предложенную NIST (National Institute of Standards and Technologies), а также работы Research Data Alliance (RDA), направленные на создание стандартов описания данных, определения метаданных, а также определение постоянных идентификаторов PID (Persistent Identifier) данных с соответствующей инфраструктурой.

4.1. Эталонная архитектура систем больших данных NIST

Стандарт NIST SP1500[11] включает три главных компонента: определение концепции/парадигмы больших данных, определение эталонной архитектуры больших данных, определение специалиста по данным (Data Scientist) как новую профессию (NBDRA, 2020).

Рисунок 2 изображает предложенную архитектуру, которая включает главные функциональные компоненты, сгруппированные по отношению к основным функциональным ролям в экосистеме больших данных: провайдер инфраструктуры и платформ больших данных (Big Data framework provider), провайдер приложений больших данных (Big Data application provider), системный интегратор (system orchestrator), а также провайдеры (data provider) и потребители (data consumer) данных. Платформа и инфраструктура больших данных включает, по сути, облачные инфраструктурные сервисы, платформу управления данными, включая реляционные и нереляционные (SQL/NoSQL) базы данных, а также платформы и средства для анализа больших данных, такие как Hadoop и Spark, с соответствующими компонентными сервисами.

Группа функций, относящаяся к провайдеру приложений больших данных, включает основные функции процесса обработки или жизненного цикла данных: сбор данных, подготовка, анализ, визуализация и презентация результатов конечному потребителю, которым может быть как человек, так и процесс. NBDRA также включает общие функции управления ресурсами и процессами, инфраструктуру безопасности и обмена сообщениями.

NBDRA построена в двух координатах: цепочка создания значимости данных от сбора данных до целевого использования и инфраструктурные уровни обработки данных от общих компьютерных или облачных ресурсов до специализированных платформ для анализа данных.

Более широкое определение архитектуры систем больших данных должно также включать определение формальных моделей больших данных, включая форматы и идентификаторы данных, метаданные, семантические модели, которые важны для межсистемного и межплатформенного обмена, интеграции и совместимости данных (BDAF, 2014).

4.2. Фабрики данных и постоянные идентификаторы данных PID

Research Data Alliance (RDA)[12] создан в 2012 году как совместная инициатива Европейской комиссии (European Commission, EC), Национального научного фонда США (National Science Foundation, NSF), Австралийского государственного департамента инноваций (Australian Government's Department of Innovation) и NIST для координации решения вопросов обмена научными данными и построения

социальной и технической инфраструктуры для обмена научными данными и их эффективного использования. В настоящее время RDA представляет собой форум, координирующий инициативы исследовательских групп по разработке рекомендаций и определению существующих практик обмена научными данными в различных областях науки. RDA не является органом стандартизации, но рекомендации RDA находят последующую стандартизацию в NIST, IEEE, ISO.

Группа Data Fabric Interest Group (IG-DF)[13] в RDA ставит своей задачей определение общих компонентов и сервисов, характерных для типичных процессов создания и использования данных в ежедневной практике научных и промышленных лабораторий, с целью создания условий использования и интеграции данных из различных областей для реализаций междоменных и межотраслевых исследований. Глобальной целью IG-DF является создание Интернета данных как инфраструктуры для обмена научными данными в процессе всех этапов обработки данных с той же простотой, как это происходит с обычными данными в Интернете. На начальном этапе своей работы IG-DF детально исследовала опыт создания и развития Сети и привлекала экспертов, имеющих опыт разработки интернет-стандартов и активно работающих с IETF (Internet Engineering Task Force).

Результатом работы IG-DF и соответствующих рабочих групп стала разработка регистра типов данных, директории метаданных, постоянных идентификаторов данных PID (Persistent Identifier). Предполагается, что PID сможет выполнять такую же роль для идентификации, поиска и доступа к данным, какую IP и доменные адреса играют для доступа ко всевозможным ресурсам в Интернете.

4.3. Общие требования к инфраструктуре больших данных

Для решения задач эффективной обработки данных с целью извлечения максимальной пользы из доступных данных и оптимизации процессов, которые зависят от данных и управляются ими, будущая инфраструктура больших данных должна обладать рядом свойств (соответствовать ряду требований), таких как:

- широкая автоматизация всех процессов обработки и управления данными, включая предоставление ресурсов и сервисов по требованию (provisioning on demand), характерных для облачных сервисов;
- преобразование всех процессов управления инфраструктурой и процессами в цифровую форму (digitalisation), постоянный мониторинг и регистрация этапов преобразования данных;
- возможность повторного использования данных, включая повторное использование для других целей и смешивание с открытыми данными, одним из важных источников которых являются социальные сети и медиа;
- глобальный доступ к данным для сотрудничающих проектов и групп исследователей, в первую очередь касающийся научных данных и научного сотрудничества;
- улучшенная безопасность инфраструктуры и защиты данных, включая персональные данные, на основе современных механизмов и сервисов безопасности, федеративного управления идентификацией пользователей и сервисов.

Существенно, чтобы будущая инфраструктура больших данных (фактически будущий Интернет данных) обеспечивала доверительную среду для обмена данными и поддерживала суверенность данных в том смысле, чтобы владелец/собственник данных сохранял контроль над использованием данных, что является принципиально важным для промышленных данных.

5. Другие инициативы в направлении построения Интернета больших данных

Вопрос обмена данными является важным как для науки, так и для промышленности. Одной из таких инициатив является GO FAIR Data Initiative[14] по внедрению принципов FAIR (Findable, Accessible, Interoperable, Reusable) для обеспечения обнаружимости, доступности, совместимости и повторного использования в первую очередь научных данных, изначально предложенная европейским научным сообществом и в настоящее время поддерживаемая всемирно, в частности, RDA и рядом проектов, финансируемых в рамках европейской рамочной программы Horizon 2020 EOSC (European Open Science Cloud)[15] и FAIRsFAIR[16].

Важной инициативой по определению принципов доверительного обмена промышленными данными и созданию элементов инфраструктуры обмена промышленными данными является создание International Data Spaces Association (IDSA)[17], которая на данный момент объединяет более 100 организаций от промышленности, науки и университетов. Эталонная архитектура, предложенная IDSA, предполагает использование современных облачных технологий и определяет ключевые функциональные компоненты для обеспечения доверительного обмена и использования промышленных данных на основе принципа суверенности данных (Data Sovereignty)[18].

6. Заключение

Эффективное использование БД является важным движущим фактором современных технологий. Большинство приложений БД используют данные из разных источников, включающих как исходные данные (например, производимые системой датчиков или социальными медиа), так и специальные наборы данных, подготовленных и обработанных другими приложениями, или данные, используемые в качестве эталонных (такие как данные геномов, звездная база данных или наборы данных для обучения алгоритмов искусственного интеллекта). Эффективный доступ к таким данным, которые в большинстве являются распределенными, вызывают необходимость построения инфраструктуры доступа и обмена данными или, другими словами, будущего Интернета данных, на создание

которого направлено много проектов и инициатив во всем мире. Создание будущего открытого Интернета данных будет способствовать дальнейшему развитию технологий и приложений, основанных на данных.

Создание Интернета данных по принципу Интернета, а также поддерживания FAIR-принципов обмена данными создаст условия, при которых больше организаций будут готовы обмениваться своими данными посредством открытых сервисов обмена данными или рынков данных.

Ссылки

1. *"A new generation of technologies and architectures designed to economically extract value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis". (IDC, 2011)*

2. *"Big Data: a massive volume of both structured and unstructured data that is so large that it's difficult to process using traditional database and software techniques". Forrester Research (2013)*

3. *"The techniques and technologies for such data-intensive science are so different that it is worth distinguishing data-intensive science from computational science as a new, fourth paradigm for scientific exploration"* (из книги Джима Грея *"The Fourth Paradigm: Data-Intensive Scientific Discovery").*

4. Фактически книга Джима Грея была издана после его смерти на основе его записок и работ группой его коллег из *Microsoft Research*.

5. https://www.egi.eu/

6. https://en.wikipedia.org/wiki/Industrial_internet_of_things

7. https://en.wikipedia.org/wiki/Digital_twin

8. *Karen Y. He, Dongliang Ge, Max M. He, Big Data Analytics for Genomic Medicine, [online]* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5343946/

9. *Lander, E., Linton, L., Birren, B. et al. Initial sequencing and analysis of the human genome. Nature 409, 860–921 (2001).* https://doi.org/10.1038/35057062 *[online]* https://www.nature.com/articles/35057062

10. https://elixir-europe.org/

11. *NIST Special Publication 1500. NIST Big Data Interoperability Framework [online]* https://bigdatawg.nist.gov/V3_output_docs.php

12. *Research Data Alliance (RDA) [online]* https://www.rd-alliance.org/

13. *Data Fabric Interest Group (IG-DF)* https://www.rd-alliance.org/group/data-fabric-ig.html

14. *GO FAIR Foundation [online]* https://gofairfoundation.org/

15. *EOSC Secretariat [online]* https://www.eoscsecretariat.eu/

16. *FAIRsFAIR Project: Fostering FAIR data practices in Europe [online]* https://www.fairsfair.eu/

17. *International Data Spaces Association (IDSA)* https://www.internationaldataspaces.org/

18. *IDSA Reference Architecture Model (IDSA RAM) [online]* https://www.internationaldataspaces.org/

Литература

(Grey, 2010) Grey, J. The Fourth Paradigm: Data-Intensive Scientific Discovery. [online] http://research.microsoft.com/en-us/collaboration/fourthparadigm/

(Forrester, 2013) The Forrester Wave: Big Data Predictive Analytics Solutions, Q1 2013. [online] Available at http://www.forrester.com/pimages/rws/reprints/document/85601/oid/1-LTEQDI

(IDC, 2011) Big Data: What it is and Why You Should Care, IDC White paper. [online] Available at http://sites.amd.com/us/Documents/IDC_AMD_Big_Data_Whitepaper.pdf

(Gartner, 2011) Big Data Definition, Gartner. [online] Available at http://www.gartner.com/it-glossary/big-data/

(Hadoop, 2020) Apache Hadoop. Apache Foundation. [online] http://hadoop.apache.org/

(MapReduce, 2017) MapReduce Tutorial. Apache Foundation. [online] http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/DownstreamDev.html

(NBDRA, 2020) NIST Special Publication NIST SP 1500: NIST Big Data Interoperability Framework (NBDIF). [online] http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-6r2.pdf

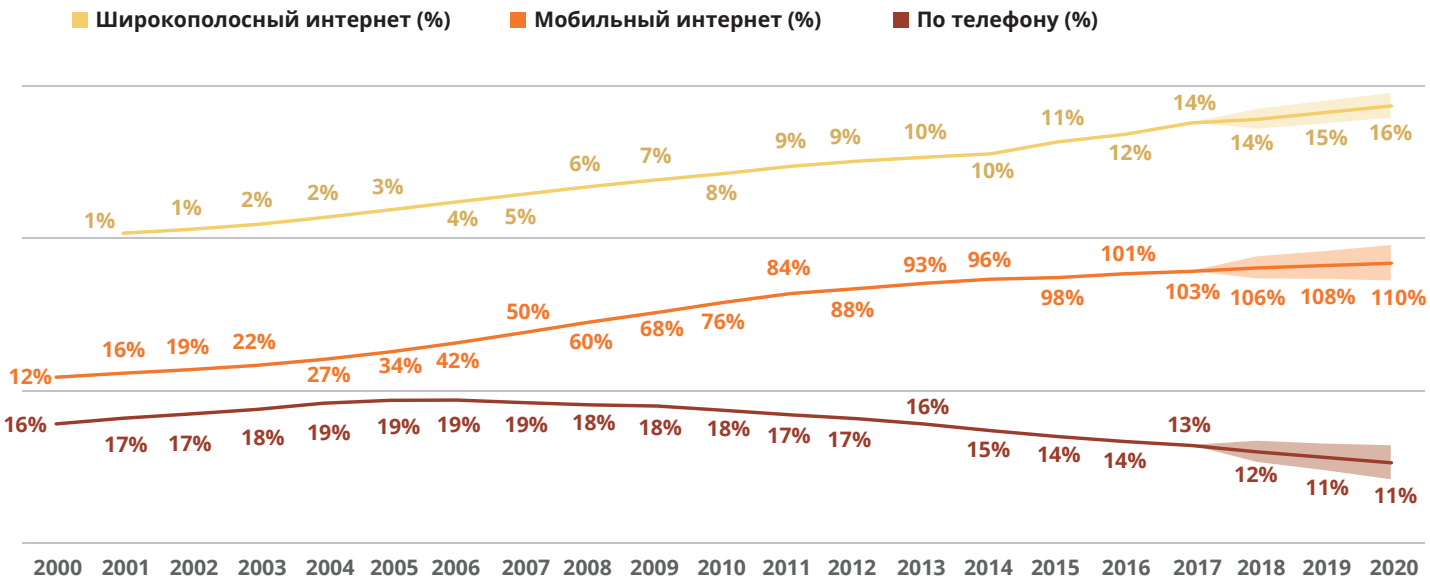
[NBDIF, 2020] NIST Special Publication 1500. NIST Big Data Interoperability Framework [online] https://bigdatawg.nist.gov/V3_output_docs.php

ITU Journal: ICT Discoveries, Special Issue 2 "Data for Goods", November 2018 [online] https://www.itu.int/en/journal/002/Documents/ITU2018-12.pdf

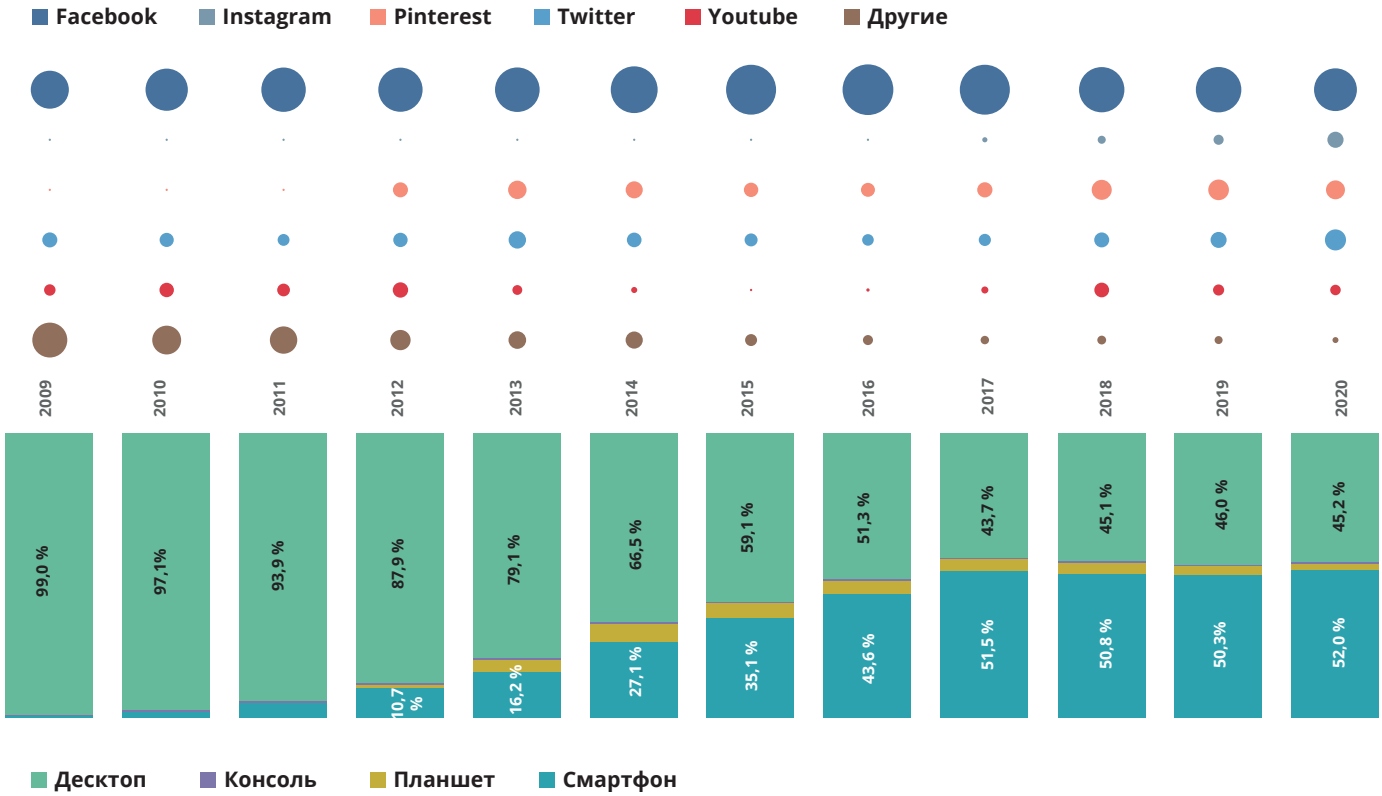
РОСТ ИСПОЛЬЗОВАНИЯ ИНТЕРНЕТА

Источник: https://public.tableau.com **Автор:** Thiago Suzin

КОЛИЧЕСТВО ИНТЕРНЕТ-ПОЛЬЗОВАТЕЛЕЙ С 2000 ПО 2020 ГОД

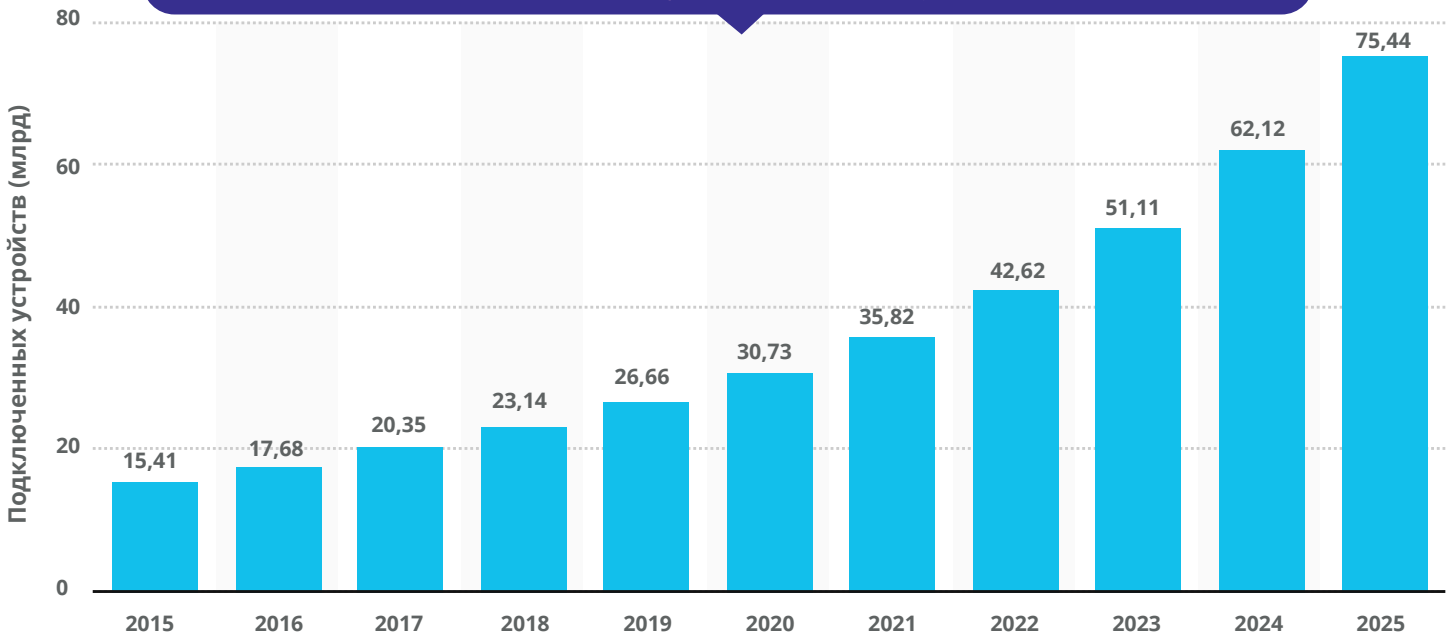


РЫНОЧНАЯ ДОЛЯ (ИСПОЛЬЗОВАНИЕ СОЦИАЛЬНЫХ МЕДИА И УСТРОЙСТВ) С 2009 ПО 2020 ГОД



РОСТ ИНТЕРНЕТА ВЕЩЕЙ, УСТАНОВЛЕННЫХ В МИРЕ С 2015 ПО 2025 ГОД

Источник: Statista, https://www.statista.com/)



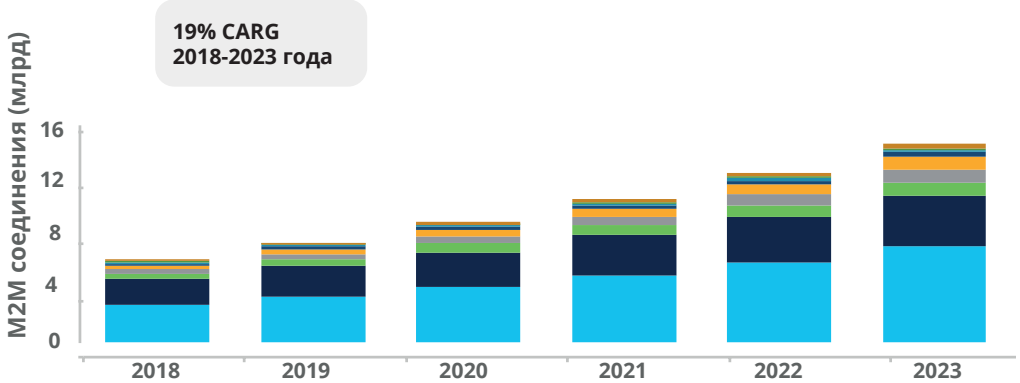
ТРЕБОВАНИЯ ПРОПУСКНОЙ СПОСОБНОСТИ СЕТИ (MBPS) ДЛЯ РАЗЛИЧНЫХ ПРИЛОЖЕНИЙ ДОМАШНЕГО ПОЛЬЗОВАТЕЛЯ

- UHD Камеры (безопасность)
- UHD вещание
- VR вещание
- Беспилотные транспортные средства (Диагностика)
- Облачные игры
- UHD IP видео
- 8K-экран Wall TV
- HD VR
- UHD VR

Краткосрочная перспектива

Будущее

РОСТ ИНТЕРНЕТА ВЕЩЕЙ В РАЗЛИЧНЫХ ОТРАСЛЯХ ПРОМЫШЛЕННОСТИ С 2018 ПО 2023 ГОД



- Подключенные дома (20% CARG)
- Подключенные города (26% CARG)
- Розничная торговля (11% CARG)
- Работа (15% CARG)
- Подключенные машины (30% CARG)
- Энергия (24% CARG)
- Здоровье (19% CARG)
- MFG и цепочки поставок (8% CARG)
- Другие (27% CARG)

Источник: Cisco Annual Internet Report, 2018-2023, https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

На глубине

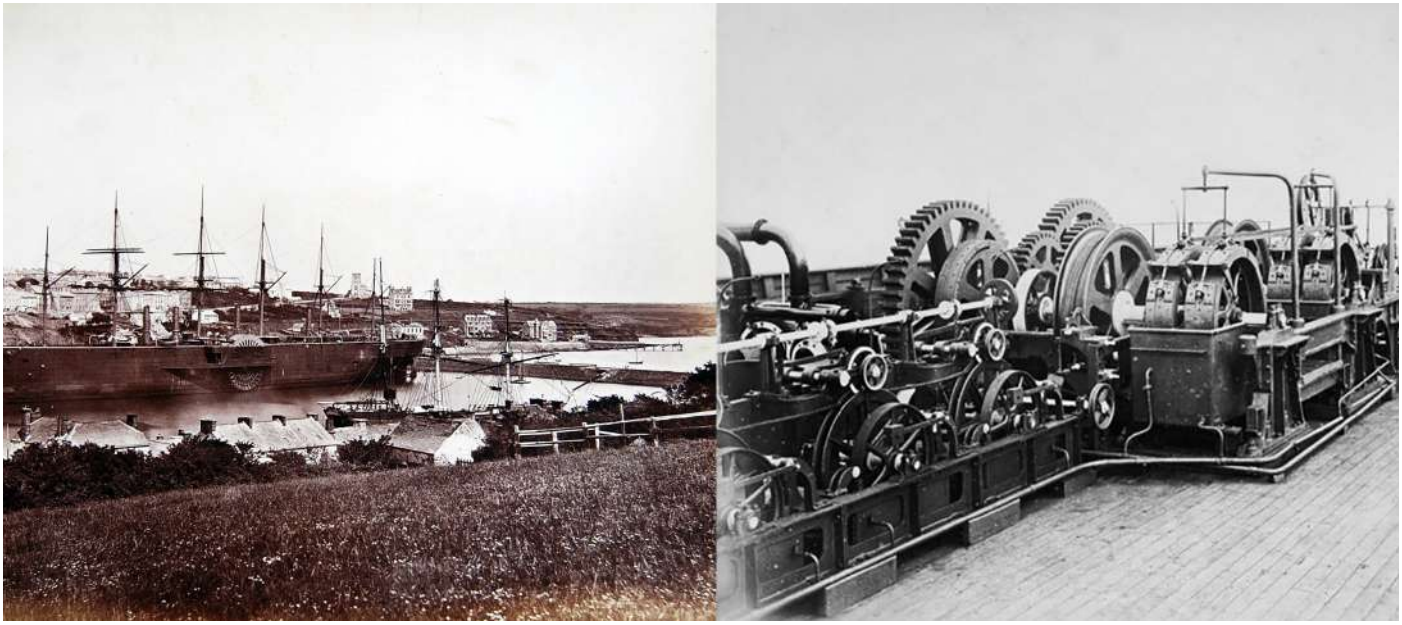
Джефф Хьюстон (Geoff Huston)

В начале этого года я был на встрече новозеландской группы сетевых операторов (New Zealand Network Operators' Group, NZNOG'20). И лично для меня одним из самых интересных докладов было выступление Беатти Лейн-Дэвис (Beatty Lane-Davis) из Cisco о современном состоянии технологии подводных кабелей. Есть что-то особенное в проектировании суперсовременного технологического продукта, который сбрасывают с корабля за борт и который потом без сучка и задоринки работает в морской пучине следующие двадцать пять лет, а то и больше! В создании такого инфраструктурного чуда задействованы и современная физика, и технологии кораблестроения, и интересные инженерные решения.

Краткий экскурс в историю подводных кабелей глазами австралийца

После нескольких неудачных попыток, 5 августа 1856 года, американская компания «Атлантик Телеграф» завершила прокладку первого в истории трансатлантического телеграфного кабеля (см. рис. 1). С конструктивной точки зрения он был прост и незамысловат: семижильный медный проводящий сердечник в обмотке из трех слоев новомодного чудо-материала под названием гуттаперча (резина по-нынешнему). Поверх гуттаперчи шла обмотка из просмоленной пеньки, а потом еще спиральная обмотка из 18-жильных стальных тросов. (Вот как выглядел такой кабель: https://soo.yaplakal.com/pics/pics_original/3/9/9/13434993.jpg - прим. редакции) Прослужил этот кабель недолго, так как главный инженер-электрик компании, доктор Уилдмен Уайтхауз (Wildman Whitehouse), предпочитал бороться с затуханием сигнала повышением напряжения (в противоположность методу Уильяма Томпсона (William Thompson), будущего лорда Кельвина, который увеличивал чувствительность зеркальных гальванометров, служивших записывающими устройствами). Двух киловольт постоянного тока кабель не пережил: изоляцию пробило, и все было кончено.

Рис. 1. Пароход Great Eastern Исамбарда Кингдома Брунела: с него в 1866 году проложили первый «долгоиграющий» трансатлантический кабель.



В последующие годы технология шагнула вперед: появились релейные усилители, позволявшие передавать сигнал на более дальние расстояния, а технология обработки сигналов непрерывно совершенствовалась, расширяя возможности подобных систем. Телеграф уступил место телефонии, электронные лампы – транзисторам, а полимеры вытеснили резину, но по сути конструкция осталась той же самой: медный сердечник, герметичная изоляция и стальная обмотка для защиты кабеля на мелководье.

В Австралии первая телеграфная система, введенная в строй в 1872 году, состояла из 3200-километровой наземной линии, которая шла с юга на север через весь континент, от Аделаиды до Дарвина (см. рис. 2). Потом

Рис. 2. Установка первого столба трансконтинентальной австралийской наземной телеграфной линии в Дарвине, 1870 г.

она короткими подводными сегментами соединялась с Сингапуром и дальше с Великобританией через Индию (см. https://en.wikipedia.org/wiki/Australian_Overland_Telegraph_Line – прим. редакции).

Такие кабели передавали только телеграфный сигнал. Трансокеанские телефонные системы поначалу использовали радиосвязь, и лишь через много десятилетий развитие электроники дошло до такого уровня, чтобы можно было передавать по кабелю речь.

В Австралии одна из первых таких систем появилась в 1962 году. COMPAC поддерживала 80 голосовых каналов по 3 кГц и соединяла Австралию через Новую Зеландию, острова Фиджи и Гавайские острова с западным побережьем Канады, дальше сигнал передавался по радиорелейной связи на восточное побережье, а оттуда в Великобританию шел кабель CANTAT. Для усиления сигнала использовались подводные ламповые реле. COMPAC проработал до 1984 года, когда вступил в строй кабель ANZCAN.

ANZCAN был проложен через Тихий океан примерно тем же маршрутом: «мокрые» сегменты шли от Сиднея до острова Норфолк, дальше на Фиджи, оттуда на Гавайи и дальше в Канаду. Это уже была аналоговая система на 14 МГц, с твердотельными релейными усилителями через каждые 13,5 км.

В 1995 году ей на смену пришла система кабелей PACRIM – опять аналоговая, но уже на 2x560 МГц. COMPAC прослужила 22 года, ANZCAN – 11 лет. PACRIM не проработала и двух, после чего ее срочно пришлось менять на оптоволоконные подводные кабели пропускной способностью 2,5 ГГц: с появлением Интернета возможности PACRIM стали просто смешны.

Сейчас в эксплуатации по всему миру находится почти 400 подводных кабелей общей протяженностью 1,2 млн км. На сайте Telegeography есть карта кабелей, показывающая все это богатство (см. рис. 3).

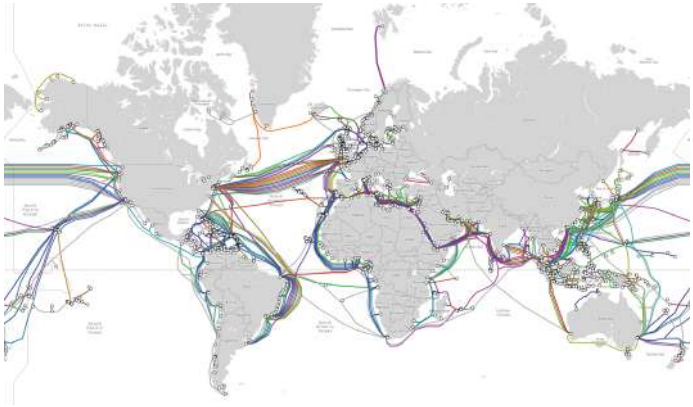
Владельцы

Первые кабельные системы были чудовищно дорогими по сравнению с размерами обслуживаемых экономик – и в строительстве, и в обслуживании. Поэтому для большинства потенциальных пользователей цена услуги оказывалась запредельно высока. Например, отправка телеграммы из 30 слов в Великобританию по только-только построенной трансконтинентальной линии стоила столько, сколько средний австралиец зарабатывал за три недели. Потому и пользовались телеграфом в основном власти и пресса.

Вдобавок к дороговизне телеграф изобиловал ошибками. Дело в том, что на наземном телеграфе имелись «промежуточные станции», где операторы вручную записывали поступающие сообщения и дальше вбивали их, опять же вручную, для передачи по следующему кабельному сегменту. С учетом того, сколько телеграфистов в Азии знало английский, неудивительно, что перевиралось едва ли не каждое третье слово в сообщении.

Удивительно, пожалуй, то, что несмотря на все трудности система прижилась – и со временем становилась все надежнее и дешевле.

Рис. 3. Карта подводных кабелей с сайта Telegeography.com.



Большинство подобных проектов финансировалось государствами, а каждый кабель прокладывала и обслуживала своя компания. В конце XIX века их существовала масса: British Indian Submarine Telegraph Company, Eastern Extension Australasia and China Telegraph Company, British Australia Telegraph Company и так далее и тому подобное. Связь с государством была очевидна, особенно в те моменты, когда оно заявляло свои права: например, подчинив себе во время Первой мировой войны все кабельные линии, ведущие из Британии. Появление модели национальных телефонных операторов в первой половине прошлого века было зеркалом государственной принадлежности кабельных систем.

В таких рамках была разработана модель владения кабелем через консорциум: для прокладки кабеля образовывалось частное акционерное общество, которое собирало деньги на прокладку, залезая в долги к традиционным банковским учреждениям. Фактически компания принадлежала национальным операторам связи, которые покупали пропускную способность кабеля в пропорции, соответствующей их долевому участию в компании.

Как правило, покупка пропускной способности кабеля принимает форму неотъемлемого права пользования (англ. indefeasible right of use, IRU), которое дает владельцу IRU исключительный доступ к пропускной способности кабеля в течение фиксированного времени (обычно 15-25 лет, в зависимости от ожидаемого срока службы кабеля). Типовое IRU включает в себя и обязательство оплачивать соответствующую долю эксплуатационных расходов.

В годы, когда крупнейшими заказчиками подводных кабельных систем были национальные операторы связи, затраты по каждому IRU обычно делились поровну между двумя операторами на концах сегмента IRU. (Это было частью сбалансированного режима денежных расчетов между национальными операторами, при котором затраты на общую инфраструктуру, предназначенную для соединения национальных служб связи, делились поровну между соединяемыми сторонами.) Эта модель возникла в эпоху национальных монополий, но постепенное дерегулирование индустрии связи далеко не сразу подорвало систему консорциумов, и она оставалась неизменной еще не один десяток лет. В частности, модель деления расходов пополам между совладельцами одного IRU обосновывалась тем, что операторы сотрудничают в плане капитальных и текущих расходов на строительство и эксплуатацию оборудования – и при этом конкурируют в плане услуг.

Одним из важных элементов такого бюрократического стиля владения было то, что консорциум устанавливал цену пропускной способности кабеля по своему усмотрению. Делалось это для того, чтобы не допустить демпинга и сохранить высокую рыночную стоимость кабеля. В результате же получалась классическая ситуация ценового сговора с рационированием товара, при которой пропускная способность кабеля выводилась на рынок мелкими порциями, так, чтобы спрос всегда превышал доступное предложение в течение срока службы кабеля и цены оставались раздутыми.

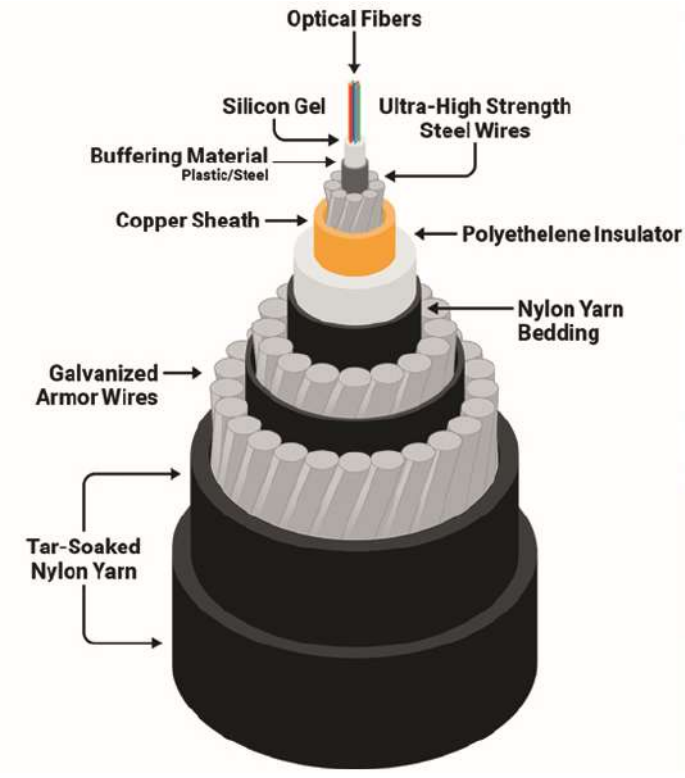
Интернет-бум в начале девяностых совпал по времени с крупномасштабным дерегулированием на многих телекоммуникационных рынках. Поэтому права на прокладку кабелей во многих странах получили сторонние компании, а на рынке подводных кабелей появились «оптовики». Тогда же возник такой феномен как кабельная система в чьем-то единоличном владении. Поначалу на этих рынках доминировали оптовики, обслуживающие зародившийся сектор ISP, такие как Global Crossing, но относительно быстро к ним присоединились крупные провайдеры контента, такие как Google, Facebook и другие. Главным отличием такой модели является то, что цены теперь никто не диктует и они могут отражать реальный баланс спроса и предложения.

Кабели

Базовая физическая конструкция подводного кабеля до сих пор практически не изменилась. Разве что носитель из медного стал оптоволоконным, в центральной части добавился стальной элемент жесткости, а вокруг носителя появилась гелевая обмотка для борьбы с истиранием.

Получившийся сигнальный комплект заключен в медную оплетку, осуществляющую электропередачу, затем идет водонепроницаемый изоляционный слой (полиэтиленовая смола), затем защитные слои, подбор которых зависит от конкретного сегмента кабеля. Чем меньше глубина, на которой будет прокладываться кабель, и чем больше интенсивность судоходства в районе, тем больше коли-

Рис. 4. Кабель в разрезе.



чество защитных элементов, призванных защитить его от повреждения, если его кто-то случайно зацепит (см. рис. 4).

Как правило, кабель прокладывается прямо на донный грунт, но иногда на участках с оживленным судоходством кабель в стальной оплетке укладывается в специально прокопанную траншею, а если дно представляет собой каменный монолит, в некоторых случаях в нем вырубают канаву.

Техника укладки кабелей ничуть не изменилась. Весь «мокрый» сегмент погружается на судно-кабелеукладчик, проверяется с начала до конца, а затем судно выходит в море и укладывает весь кабель в один проход. Скорость и местоположение судна определяется со всей тщательностью, чтобы не подвергнуть кабель чрезмерным нагрузкам при укладке. Корабль проходит весь маршрут прокладки кабеля за один прием, не останавливаясь, и по пути

Рис. 5. Погрузка кабеля на судно.

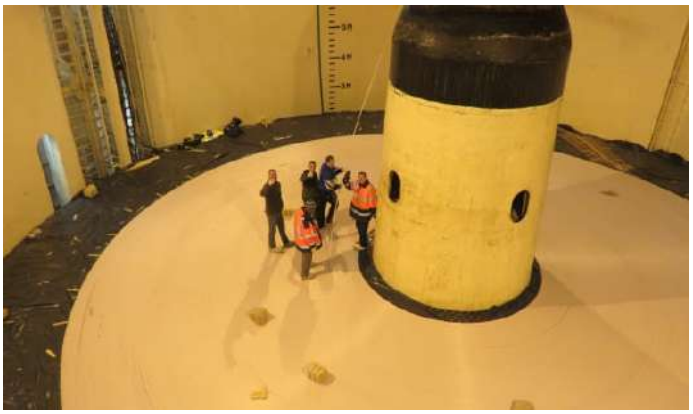


Рис. 6. Судно-кабелеукладчик.

Рис. 7. Релейный усилитель.

укладывает кабель на дно – в среднем на глубине 3600 м, но иногда она доходит и до 8 км. Кабель находится под натяжением вплоть до 8 км за кораблем.

Ремонт кабелей тоже проблематичен. На шестикилометровой глубине требуется около 20 часов, чтобы сбросить вниз захват, подцепить кабель и подтянуть один его конец к поверхности. Глубже 6 км такая операция вообще невозможна, поэтому в более глубоких впадинах кабели не ремонтируют, а просто делают врезку по обе стороны впадины. Так или иначе, неисправность кабеля на очень большой глубине устранять долго, дорого и сложно.

В ранних кабелях были только соединения «точка-точка», но коммерческие перспективы использования одной кабельной системы для соединения множества оконечных точек были так велики, что появились разветвители. Простейший вариант оптического разветвителя – разделение физических волокон в сердечнике. В наши дни чаще используются реконфигурируемые оптические мультиплексоры с вводом-выводом (ROADM). Эти устройства позволяют добавлять и/или удалять отдельные или несколько длин волн, несущих каналы данных, из транспортного волокна без необходимости преобразовывать сигналы по всем каналам WDM в электронные сигналы и обратно в оптические сигналы. Основные преимущества использования ROADM – не нужно планировать распределение всей полосы пропускания заранее, так как ROADM позволяют перенастраивать полосу по мере надобности. Такая перенастройка выполняется в любой момент и никак не влияет на трафик, уже проходящий через ROADM.

Подводная система обычно называется «мокрым» (wet) сегментом, и эти системы соединяются с поверхностными на кабельных станциях. На станциях находится оборудование, обеспечивающее электропитание кабеля. Используется постоянный ток, причем в системах питания

Рис. 8. Кабельный разветвитель.

кабелей большой длины как правило подается 10 кВ постоянного тока с обоих концов. Кроме того, на кабельной станции обычно устанавливается оконечное оборудование для длины волны и оборудование контроля линии.

Оптические повторители

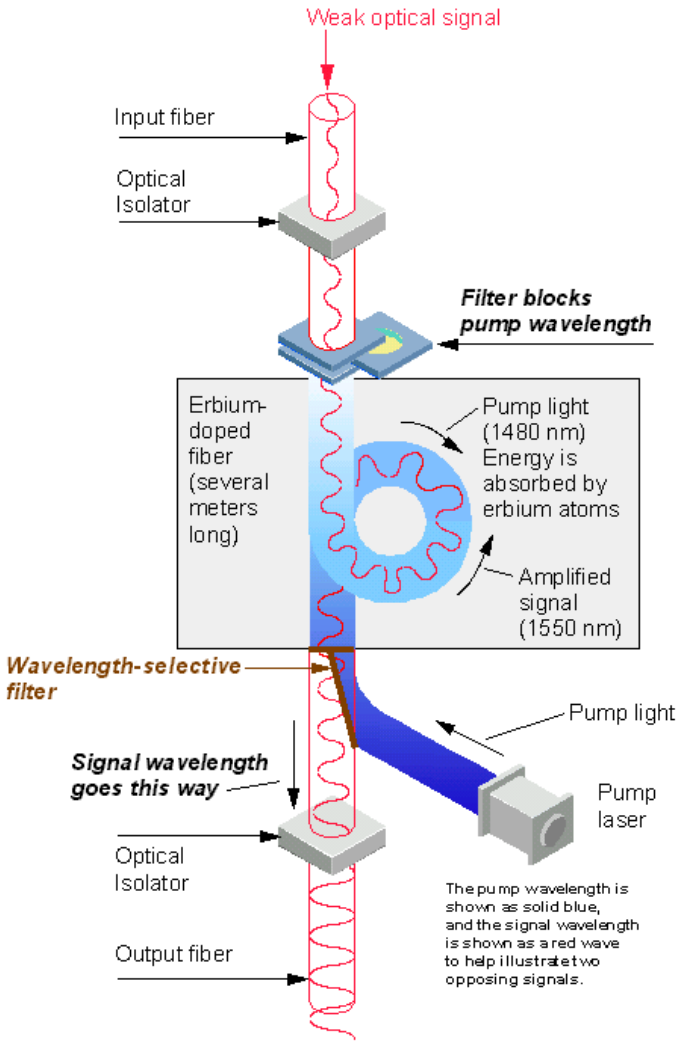
Название «оптические повторители», пожалуй, сейчас не совсем точно отражает суть дела. Ранние электрические повторители работали в режиме обычного повторения сигнала: ресивер преобразовывал аналоговый сигнал на входе в цифровой, затем опять в аналоговый - и пускал его по следующему сегменту кабеля.

Современные оптические повторители представляют собой фотонные усилители, которые работают на полной мощности на дне морском и рассчитаны на 25-летний срок службы. Световой поток накачки (с длиной волны 980 нм или 1480 нм) поступает в относительно короткий сегмент волокна, легированного эрбием, смешиваясь с входным сигналом. Ионы эрбия вызывают усиление входящего потока света с длинами волн порядка 1550 нм. Энергия накачки заставляет ионы эрбия перейти в возбужденное состояние (с повышенной энергией), и когда в систему входит фотон полезного сигнала, ион возвращается в исходное состояние и испускает квант избыточной энергии – еще один фотон, частота которого равна частоте исходного фотона. Усиленный таким образом сигнал по направлению и фазе совпадает с исходным. Подобные устройства называются EDFA-усилителями (см. рис. 9) (англ. Erbium Doped Fibre Amplifiers – усилители на оптоволокне, легированном эрбием). Технология EDFA произвела настоящую революцию в конструкции подводных кабелей. Весь «мокрый» сегмент, включая повторители, как будто выносит передаваемый сигнал за скобки. Число используемых лямбд, кодирование-декодирование сигнала, полная пропускная способность кабеля – все это теперь регулируется на берегу: на кабельных станциях по оба конца кабеля. Таким образом удалось продлить срок службы оптических систем, так как можно извлечь дополнительную пропускную способность из существующих кабелей, просто заменив оборудование на концах кабеля, а сам кабель не трогать: он как лежал себе под водой, так и лежит, просто теперь по нему передается больше информации.

Подводные оптические повторители рассчитаны на весь срок службы кабеля и не требуют последующего вмешательства. Их конструкция включает в себя элемент резервирования: если повторитель выходит из строя, пропускная способность кабеля несколько проседает, но некритично.

Устройства EDFA отличаются ошибкой усиления по всему диапазону рабочих частот, поэтому необходимо добавлять пассивный фильтр к усиленному сигналу, чтобы получить более плоский спектр мощностей. Благодаря этому совокупная сумма линейных усилителей дает максимальную производительность сигнала по всему диапазону частот в кабеле. На значительном расстоянии и этого оказывается мало, потому в кабелях могут использоваться активные элементы, так называемые корректоры уровня усиления (англ. Gain Equalisation Unit) (см. рис. 10). Количество

Рис. 9. EDFA-усиление.



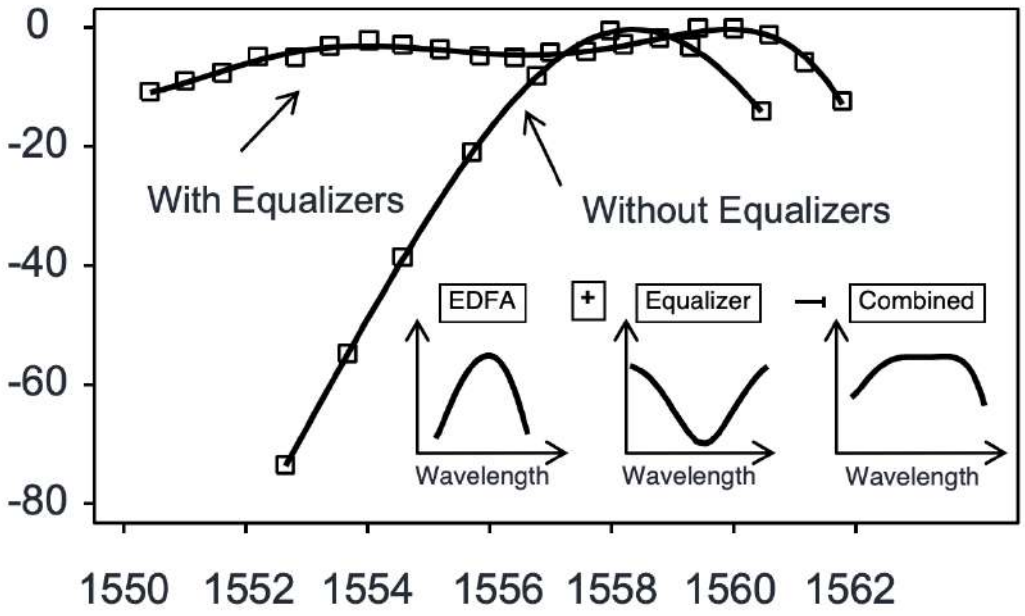
таких корректоров, расстояние между ними и параметры коррекции настраиваются для каждой кабельной системы отдельно.

В наземных системах управлять усилением можно динамически, и по мере добавления или удаления каналов усилители перенастраиваются так, чтобы выдавать оптимальный уровень. В подводных усилителях такого динамического управления нет, и они настраиваются на насыщение, т.е. всегда «на максимум». Чтобы избежать перегрузки подсвеченных каналов, все неиспользуемые каналы в спектре заняты сигналом «простой».

На повторители приходится значительная часть общей стоимости кабельной системы, поэтому приходится решать, разместить ли повторители поближе друг к другу – например, через каждые 60 км или около того – или же растянуть их до 100 км, чтобы сэкономить на общем числе повторителей в системе. Оборотная сторона медали в том, что чем больше вы готовы потратить на кабельную систему, тем выше будет ее пропускная способность.

Главная мысль здесь в том, что подводная кабельная система – это не конструктор, где все компоненты стандартны и соединяются между собой по единым конструкторским правилам, а индивидуальный продукт, где каждый компонент настраивается так, чтобы вся система в целом давала оптимальные показатели для конкретной среды примене-

Рис. 10. Коррекция уровня усиления EDFA.



ния. В сущности, каждый проект, связанный с прокладкой подводных кабелей, во многом разрабатывается с нуля.

Пропускная способность кабеля и кодирование сигнала

Самые ранние подводные оптоволоконные кабельные системы были разработаны в 80-х годах прошлого века и введены в строй в конце того же десятилетия. Эти первые кабельные системы были коаксиальными, и для регенерации и усиления сигнала в них использовалось электрооборудование. Усилители, как правило, располагались через каждые 40 км кабеля. Первое, что сделали для повышения пропускной способности кабеля, – это применили систему, которая давно и прочно прижилась в радиосвязи: частотное разделение каналов, или мультиплексирование с разделением по частоте (англ. frequency division multiplexing, FDM). Первые оптические кабели с электроусилением использовали FDM для создания нескольких голосовых контуров в одном и том же несущем коаксиальном кабеле. Такие кабели поддерживали 560 Мб/с всего, разделенные примерно на 80 тысяч голосовых контуров.

Несущую способность коаксиального кабеля в таких гибридных оптоэлектрических системах планировалось удвоить, но в этот момент появились полностью оптические системы EDFA, которые впервые были применены под водой в 1994 году. Такие кабели использовали тот же самый тип оптического мультиплексирования с разделением по частоте: каждый оптический кабель делился на несколько каналов с той или иной длиной волны (или «лямбд», по буквенному обозначению длины волны) в аналоговой общей структуре, получившей название WDM (Wave Division Multiplexing – мультиплексирование по длине волны, или спектральное уплотнение каналов).

Несущую способность коаксиального кабеля в таких гибридных оптоэлектрических системах планировалось удвоить, но в этот момент появились полностью оптические системы EDFA, которые впервые были применены под водой в 1994 году. Такие кабели использовали тот же самый тип оптического мультиплексирования с разделением по частоте: каждый оптический кабель делился на несколько каналов с той или иной длиной волны (или «лямбд», по буквенному обозначению длины волны) в аналоговой общей структуре, получившей название WDM (Wave Division Multiplexing – мультиплексирование по длине волны, или спектральное уплотнение каналов).

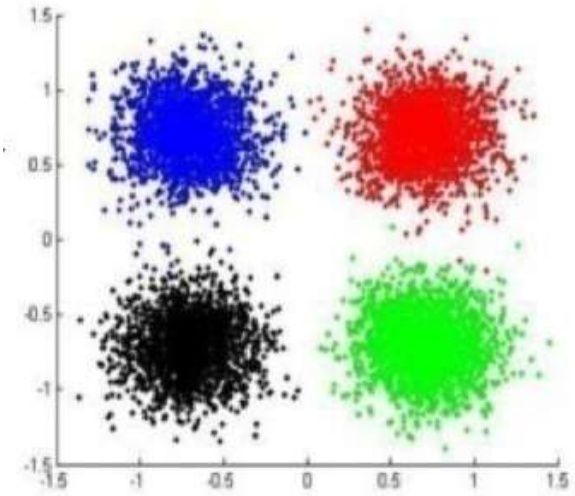
С повышением несущей частоты сигнала на больших расстояниях стало играть заметную роль такое явление как хроматическая дисперсия. Дело в том, что на разных частотах свет в оптическом волокне движется чуть-чуть с разной скоростью. В результате этого прямоугольный

импульс на входе превращается на выходе в сглаженный, а в какой-то момент искажения становятся столь велики, что цифровой процессор обработки сигналов (ЦПОС или DSP – digital signal processor) на выходе уже не может надежно декодировать сигнал. Для борьбы с хроматической дисперсией используются сегменты волокна с отрицательной дисперсией, легированные диоксидом германия для компенсации хроматической дисперсии. Идеальным, правда, такое решение не назовешь: хотя и можно спроектировать систему компенсации дисперсии на средней частоте полосы, по краям на длинных расстояниях дисперсия все равно будет значительной.

В полностью оптических системах первого поколения использовалась простейшая технология цифровой модуляции «включение-выключение» (one/off keying, OOK). Техника кодирования сигнала OOK применялась в системах WDM для скоростей сигнала до 10 Гбит/с на лямбду (этот потолок был достигнут в 2000 году), но для кабелей большей пропускной способности на длинных дистанциях этот метод непригоден из-за сочетания хроматической и поляризационной дисперсии.

В этот момент в цифровых процессорах обработки сигналов стали использовать методы модуляции когерентных

Рис. 11. Фазово-амплитудное созвездие для QPSK.



С помощью таких DSP становится возможно модулировать сигнал в каждой лямбде путем фазовой модуляции сигнала. Метод квадратурной фазовой манипуляции (Quadrature Phase Shift Keying, QPSK) определяет четыре точки сигнала, отделенные друг от друга 90-градусным фазовым смещением, что позволяет закодировать 2 бита в одном символе. Сочетание QPSK и кодирования по 2-точечной поляризации позволяет иметь по 2 бита на символ.

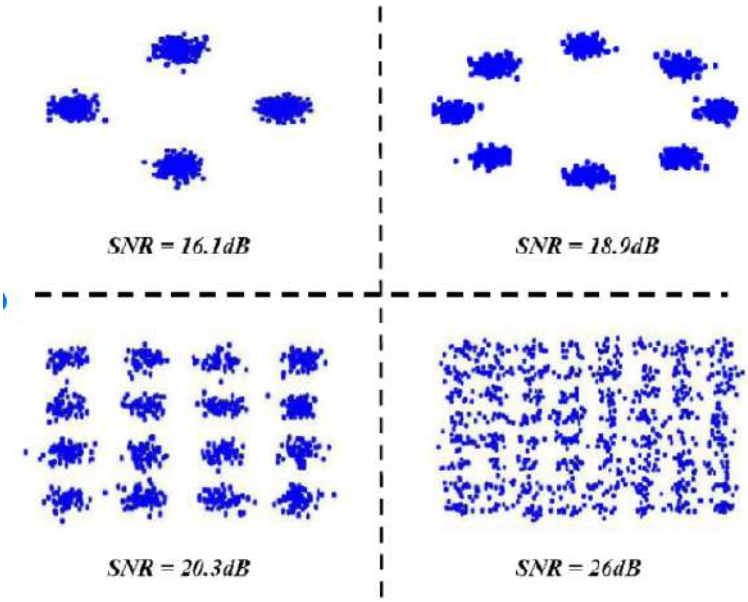
радиочастот в сочетании со спектральным уплотнением каналов. Это стало возможно благодаря появлению усовершенствованных методов цифровой обработки сигналов, обогащенных наработками из мира радио, где приемник может обнаруживать быстрые изменения фазы несущего сигнала на входе в дополнение к изменениям амплитуды и поляризации.

С помощью таких DSP становится возможно модулировать сигнал в каждой лямбде путем фазовой модуляции сигнала. Метод квадратурной фазовой манипуляции (Quadrature Phase Shift Keying, QPSK) определяет четыре точки сигнала, отделенные друг от друга 90-градусным фазовым смещением (см. рис. 11), что позволяет закодировать 2 бита в одном символе. Сочетание QPSK и кодирования по 2-точечной поляризации позволяет иметь по 2 бита на символ. На практике получается, что в C-диапазоне система на оптическом носителе на 5 ТГц, использующая QPSK и DWDM, может быть настроена так, чтобы иметь общую пропускную способность примерно в 25 Тбит/с, в предположении, что соотношение сигнал/шум будет приемлемым. Еще одно преимущество – столь высоких скоростей можно достичь с помощью гораздо более скромных компонентов. Канал на 100G состоит из восьми носителей по 12,5G.

Амплитудная модуляция позволяет еще больше доработать этот метод кодирования. Возможности QPSK расширяет 8QAM, добавляя к кодированию QPSK еще четыре точки, так как появляются дополнительные фазы по 45 градусов с половинной амплитудой. 8QAM позволяет кодировать группы по 3 бита на символ, но предъявляет более высокие требования к отношению сигнал/шум – 4 дБ. 16QAM, как можно заключить из названия, определяет 16 отдельных точек в фазово-амплитудном пространстве, позволяя кодировать по 4 бита на символ, ценой повышения минимально приемлемого соотношения сигнал/шум еще на 3 дБ (см. рис. 12). Практическим лимитом количества точек кодирования в фазово-амплитудном пространстве является показатель отношения сигнал/шум для самого кабеля, так как чем сложнее кодирование, тем выше требования к декодеру.

Еще одна методика, которая помогает извлечь плотный сигнал из шумного аналогового носителя, – использование кодов прямой коррекции ошибок (Forward Error Correcting,

Рис. 12. Созвездия адаптивной модуляции для QPSK, 8PSK, 16QAM и 64QAM.



Амплитудная модуляция позволяет еще больше доработать этот метод кодирования. Возможности QPSK расширяет 8QAM, добавляя к кодированию QPSK еще четыре точки, так как появляются дополнительные фазы по 45 градусов с половинной амплитудой. 8QAM позволяет кодировать группы по 3 бита на символ, но предъявляет более высокие требования к отношению сигнал/шум – 4 дБ. 16QAM, как можно заключить из названия, определяет 16 отдельных точек в фазово-амплитудном пространстве, позволяя кодировать по 4 бита на символ, ценой повышения минимально приемлемого соотношения сигнал/шум еще на 3 дБ. Практическим лимитом количества точек кодирования в фазово-амплитудном пространстве является показатель отношения сигнал/шум для самого кабеля, так как чем сложнее кодирование, тем выше требования к декодеру.

FEC) в цифровом сигнале. Коды FEC позволяют обнаруживать и исправлять небольшое число ошибок на кадр FEC, жертвуя для этого частью пропускной способности.

В области FEC сейчас вершиной является полярный код, позволяющий каналу практически достичь предела Шеннона – теоретического максимума скорости передачи при заданной пропускной способности и заданном отношении сигнал/шум.

Сейчас можно и выгодно развертывать кабельные системы средней и большой дальности пропускной способностью под 50 Тбит/с в одном оптическом волокне. Но это не тот предел, которого можно достичь в плане пропускной способности таких систем.

Проектировщикам кабельных систем доступны два диапазона частот. Самый распространенный – это C-диапазон, охватывающий длины волн с 1530 нм по 1565 нм. Соседний с ним L-диапазон простирается от 1570 нм до 1610 нм. В аналоговых терминах каждый диапазон составляет примерно от 4,0 до 4,8 ТГц. Если задействовать оба диапазона с кодированием DWDM и QPSK, получатся кабельные системы, которые смогут передавать где-то 70 Тбит/с на волокно на расстояниях до 7500 км.

Подобная пропускная способность дается немалой ценой, так как устройства EDFA работают либо в C-диапазоне, либо в L-диапазоне – поэтому, если нам нужны оба, потребуется вдвое больше EDFA-усилителей (а значит, и вдвое больше мощности на кабельных станциях). Возможно, в какой-то момент конструкторы кабельных систем обнаружат, что дешевле увеличить число пар волокон в кабеле, чем использовать все более сложные механизмы кодирования – хотя имеются и ограничения общей мощности, которую можно «вкачать» в дальние подводные кабели, так что, как правило, подобные системы большой дальности содержат не более восьми пар волокон.

Совершенно другой принцип оптического усиления используют т.н. рамановские, или ВКР-усилители (англ. Fibre Raman Amplifier, FRA). В них используется вынужденное комбинационное рассеяние (ВКР), известное также как эффект Рамана (Stimulated Raman Scattering, SRS). Активной средой является нелегированное оптоволокно, а мощность преобразуется в оптический сигнал за счет нелинейного оптического процесса – комбинационного рассеяния света, открытого индийским физиком Ч. Раманом (и независимо от него Л. И. Мандельштамом и Г. С. Ландсбергом в МГУ. Поскольку Нобелевскую премию за это открытие получил один Раман, термин «эффект Рамана» весьма непопулярен среди физиков на постсоветском пространстве: у нас это называют просто «комбинационным рассеянием света» – прим. редакции). При столкновении фотона с молекулой стекла, сопровождающемся обменом энергией, электрон в молекуле возбуждается до виртуального состояния, а при обратном переходе в вибрационное состояние происходит вынужденная эмиссия. К преимуществам FRA относятся переменное усиление длин волн, совместимость с установленным одномодовым волокном и пригодность для продления EDFA. Для FRA-усиления требуются лазеры с очень высокой мощностью накачки и сложное регулирование усиления, но сочетание EDFA и FRA может дать снижение средних требований к мощности на сегменте.

FRA-усилители работают в очень широком диапазоне сигналов (1280 нм – 1650 нм).

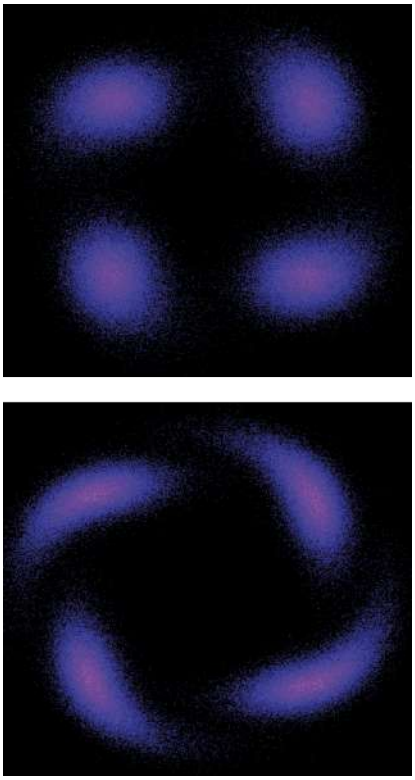
Однако все это верно лишь в предположении, что стекло является пассивной средой, подверженной лишь линейным искажениям (или шумам): если увеличить длину кабеля вдвое, то затухание сигнала, уровень хроматической дисперсии и т.д. вырастут тоже вдвое. Но если вкачать в стекло большой объем энергии в виде фотонов, оно начинает вести себя нелинейно, и чем выше уровень энергии, тем более выраженной становится эта нелинейность.

При накачке оптических систем лазером возникает оптический эффект Керра: высокая интенсивность света вызывает изменение показателя преломления стекла, что в свою очередь может привести к нестабильности модуляции оптических систем, особенно фазовой. Прохождение сильных потоков света сквозь стекло может вызвать в нем акустические вибрации, приводящие к возникновению крупных фононов низкой частоты, а эти фононы вызывают рассеяние Мандельштама-Бриллюэна. В результате возникает ремодуляция светового луча, а кроме того, изменяются характеристики усиления и поглощения света.

Что получается? Для передачи сигнала высокой плотности по кабелю требуется большая мощность, а большая мощность приводит к нелинейному искажению сигнала (см. рис. 13), особенно фазовому. Приходится балансировать энергобюджет, хроматическое и фазовое искажение, а также пропускную способность кабеля. В современных кабельных системах иногда для оптимизации общей пропускной способности кабеля применяются разные алгоритмы кодирования для разных лямбд.

Начиная с 2010 года конструкторы кабельных систем по большому счету предоставляют DSP разбираться с дисперсией и отказываются от использования сегментов компенсации дисперсии в самом кабеле. DSP теперь выполняет компенсацию обратной связи. Благодаря этому повышается когерентность сигнала, пусть и ценой повышения сложности DSP. Кроме того, конструкторы наращивают диаметр стеклянного сердечника в SPF-волокне. Увеличение эффективного размера стеклянного сердечника снижает выраженность нелинейных эффектов в стекле, накачанном энергией, что позволяет увеличить полезную пропускную способность оптоволоконных систем с толстым сердечником в десять раз. Кроме того, разрабатываются более мощные DSP, которые могут выполнять больше функций. В определенной степени конструкторы при этом опираются на закон Мура. Наращивание числа элементов в микросхеме позволяет загрузить в чип DSP больший функционал, не выходя за рамки требований к мощности и охлаждению. DSP могут не только работать с компенсацией обратной связи, но и выполнять адаптивное кодирование и декодирование. Например, DSP может переключаться между двумя кодировками PSK для каждой пары символов. Пусть, например, каждый следующий символ был закодирован попеременно по 8QAM и 16QAM: тогда в результате мы получим по 7 бит на символ, что сократит значительное приращение между различными уровнями кодирования PSK. DSP также может тестировать различные фазово-амплитудные точки и сокращать их использование для символов с высокой вероятностью ошибки. Это так называемое вероятностное формирование созвездий (англ. Probabilistic Constellation Shaping, PCS). Если добавить к нему прямую коррекцию ошибок,

Рис. 13. Нелинейное искажение сигнала, модулированного по QPSK.



Что получается? Для передачи сигнала высокой плотности по кабелю требуется большая мощность, а большая мощность приводит к нелинейному искажению сигнала, особенно фазовому. Приходится балансировать энергобюджет, хроматическое и фазовое искажение, а также пропускную способность кабеля. В современных кабельных системах иногда для оптимизации общей пропускной способности кабеля применяются разные алгоритмы кодирования для разных лямбд.

Эволюция подводных кабельных систем ни в коем случае не близится к концу, и все время возникают новые идеи о том, как сделать их мощнее и дешевле. В последние тридцать лет пропускная способность оптики каждое десятилетие увеличивалась примерно в сотню раз, и было бы глупо полагать, что такой бурный рост вдруг резко остановится – но нельзя и не признать, что поддержание его на прежнем уровне потребует немалого технологического новаторства в дальнейшем.

отведя на нее примерно 30% полосы, можно получить в системе большой диапазон пригодных для использования уровней пропускной способности.

Кстати о пропускной способности современных кабельных систем: мощнейшим на сегодняшний момент является кабель MAREA, идущий от испанского Бильбоа до Вирджиния-Бич в США. Его рабочая пропускная способность составляет 208 Тбит/с.

Что дальше?

Эволюция подводных кабельных систем ни в коем случае не близится к концу, и все время возникают новые идеи о том, как сделать их мощнее и дешевле. В последние тридцать лет пропускная способность оптики каждое десятилетие увеличивалась примерно в сотню раз, и было бы глупо полагать, что такой бурный рост вдруг резко остановится – но нельзя и не признать, что поддержание его на прежнем уровне потребует немалого технологического новаторства в дальнейшем.

Замечу, что до сих пор основные усилия были сосредоточены на том, чтобы выжать максимум из одной оптоволоконной пары. Проблема здесь в том, что для этой цели мы загоняем систему в крайне неэффективный энергетически режим, поскольку значительная часть затрачиваемой энергии уходит в оптический шум, который потом еще и приходится отфильтровывать. Альтернативным вариантом является использование нескольких волокон в составе одного сердечника при гораздо более низкой мощности. Так можно повысить и пропускную способность системы, и ее энергоэффективность.

Будут продолжать совершенствоваться и DSP, но следует ожидать изменений в системах, отправляющих сигнал в кабель. Точно так же, как в векторных системах DSL используется предварительная компенсация запускаемого сигнала для того, чтобы скомпенсировать искажения в медном контуре, возможно, удастся использовать предварительное искажение в возбудителях лазеров – или, возможно, даже в сегментах EDFA – для того, чтобы вывести подводные кабели еще на более высокий уровень производительности.

Источник: Internet Economics, https://www.potaroo.net/ispcol/2020-02/subsea.html

Нейронные сети, или Как обучить искусственный интеллект

Коннова Н. С.

Машинное обучение является не единственным, но одним из фундаментальных разделов искусственного интеллекта, занимающимся вопросами самостоятельного извлечения знаний интеллектуальной системой в процессе ее работы. Основу машинного обучения составляют так называемые искусственные нейронные сети, принцип работы которых напоминает работу биологических нейронов нервной системы животных и человека.

Технологии искусственного интеллекта все активнее используются в самых различных областях – от производственных процессов до финансовой деятельности. Популярные услуги техгигантов, такие как «Яндекс.Алиса», Google Translate, Google Photos и Google Image Search (AlexNet, GoogleNet) основаны на работе нейронных сетей.

Что такое нейронные сети и как они работают – об этом пойдет речь в статье.

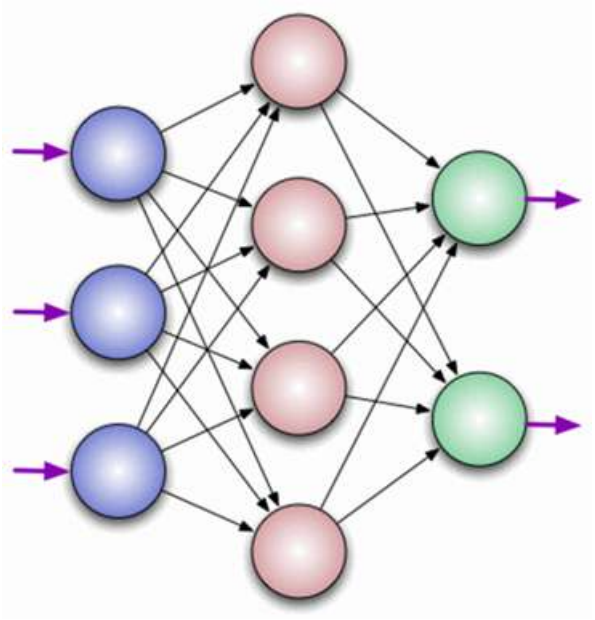
Введение

У искусственного интеллекта множество определений. Под этим термином понимается, например, научное направление, в рамках которого ставятся и решаются задачи аппаратного или программного моделирования тех видов человеческой деятельности, которые традиционно считаются творческими или интеллектуальными, а также соответствующий раздел информатики, задачей которого является воссоздание с помощью вычислительных систем и иных искусственных устройств разумных рассуждений и действий, равно как и свойство систем выполнять «творческие» функции. Искусственный интеллект уже сегодня прочно вошел в нашу жизнь, он используется так или иначе во многих вещах, которые нас окружают: от смартфонов до холодильников, - и сферах деятельности: компании многих направлений, от IT и банков до тяжелого производства, стремятся использовать последние разработки для повышения эффективности. Данная тенденция будет лишь усиливаться в будущем.

Машинное обучение является не единственным, но одним из основных разделов, фундаментом искусственного интеллекта, занимающимся вопросами самостоятельного извлечения знаний интеллектуальной системой в процессе ее работы. Зародилось это направление еще в середине XX века, однако резко возросший интерес направление получило с существенным развитием области математики, специализирующейся на нейронных сетях, после недавнего изобретения эффективных алгоритмов обучения таких сетей. Идея создания искусственных нейронных сетей была перенята у природы. В качестве примера и аналога была использована нервная система как животного, так и человека. Принцип работы искусственной нейронной сети соответствует алгоритму работы биологических нейронов.

Нейрон – вычислительная единица, получающая информацию и производящая над этой информацией какие-либо вычисления. Нейроны являются простейшей структурной единицей любой нейронной сети, из их множества, упорядоченного некоторым образом в слои, а в конечном счете – во всю сеть, состоят все сети. Все нейроны функционируют примерно одинаковым образом, однако бывают некоторые частные случаи нейронов, выполняющих специфические функции. Три основных типа нейронов:

Рис. 1. Основные типы нейронов.



Входной – слой нейронов, получающий информацию (синий цвет на рис. 1).

Скрытый – некоторое количество слоев, обрабатывающих информацию (красный цвет на рис. 1).

Выходной – слой нейронов, представляющий результаты вычислений (зеленый цвет на рис. 1).

Нейрон – вычислительная единица, получающая информацию и производящая над этой информацией какие-либо вычисления.

Помимо самих нейронов существуют синапсы. Синапс – связь, соединяющая выход одного нейрона со входом другого. Во время прохождения сигнала через синапс сигнал может усиливаться или ослабевать. Параметром синапса является вес (некоторый коэффициент, может быть любым вещественным числом), из-за которого информация, передающаяся от одного нейрона к другому, может изменяться. При помощи весов входная информация проходит обработку - и после мы получаем результат.

На рис. 2 изображена математическая модель нейрона. На вход подаются числа (сигналы), после они умножаются на веса (каждый сигнал – на свой вес) и суммируются. Функция активации высчитывает выходной сигнал и подает его на выход.

Сама идея функции активации также взята из биологии. Представьте себе ситуацию: вы подошли и положили палец на конфорку варочной панели. В зависимости от сигнала, который при этом поступит от пальца по нервным окончаниям в мозг, в нейронах вашего мозга будет принято решение: пропускать ли сигнал дальше по нейронным связям, давая посыл мышцам отдернуть палец от горячей конфорки, или не пропускать сигнал, если конфорка холодная и палец можно оставить. Математический аналог функции активации имеет то же назначение. Таким образом, функция активации позволяет проходить или не проходить сигналам от нейрона к нейронам в зависимости от информации, которую они передают. Т.е. если информация является важной, то функция пропускает ее, а если информации мало или она недостоверна, то функция активации не позволяет ей пройти дальше.

В таблице 1 представлены некоторые виды активационных функций нейрона. Простейшей разновидностью функции активации является пороговая. Как следует из названия, ее график представляет из себя ступеньку. Например, если поступивший в нейрон суммарный сигнал от предыдущих нейронов меньше 0, то в результате применения функции активации сигнал полностью «тормозится» и дальше не проходит, т.е. на выход данного нейрона (и, соответственно, входы последующих нейронов) подается 0. Если же сигнал >= 0, то на выход данного нейрона подается 1.

Функция активации должна быть одинаковой для всех нейронов внутри одного слоя, однако для разных слоев могут выбираться разные функции активации. В искусственных нейронных сетях выбор функции обуславливается

Рис. 2. Модель искусственного нейрона.

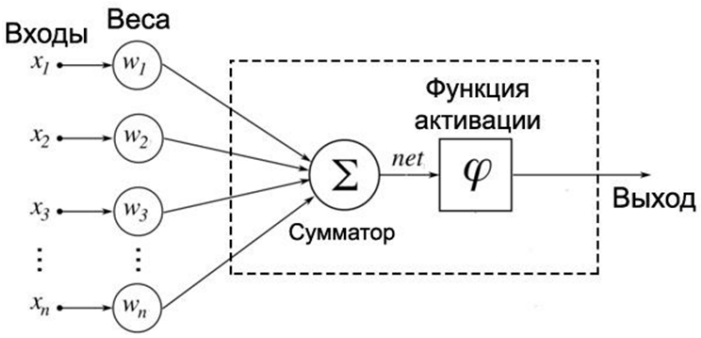


Таблица 1. Популярные разновидности активационных функций искусственного нейрона

Название	Математическое представление	Особенность
Пороговая	$\theta(h)=\begin{cases}0, h<0\\1, h\geq 0\end{cases}$	Может применяться в системах на базе логических элементов
Линейная	$\theta(h)=h$	(-∞;∞) – область применения
Сигмоидная	$\theta(h)=\frac{1}{1+e^{-h}}$	Является дифференцируемой на всей оси абсцисс
Гиперболический тангенс	$\theta(h)=\frac{e^{h}-e^{-h}}{e^{h}+e^{-h}}$	Может принимать положительные и отрицательные значения

областью применения и является важной исследовательской задачей для специалиста по машинному обучению.

Нейронные сети набирают все большую популярность и область их использования также расширяется. Список некоторых областей, где применяются искусственные нейронные сети:

1. **Ввод и обработка информации.** Распознавание текстов на фотографиях и различных документах, распознавание голосовых команд, голосовой ввод текста.

2. **Безопасность.** Распознавание лиц и различных биометрических данных, анализ трафика в сети, обнаружение подделок.

3. **Интернет.** Таргетинговая реклама, капча, составление новостных лент, блокировка спама, ассоциативный поиск информации.

4. **Связь.** Маршрутизация пакетов, сжатие видеоинформации, увеличение скорости кодирования и декодирования информации.

Помимо перечисленных областей, нейронные сети применяются в робототехнике[1], компьютерных и настольных

играх, авионике, медицине, экономике, бизнесе и в различных политических и социальных технологиях, финансово-экономическом анализе[2].

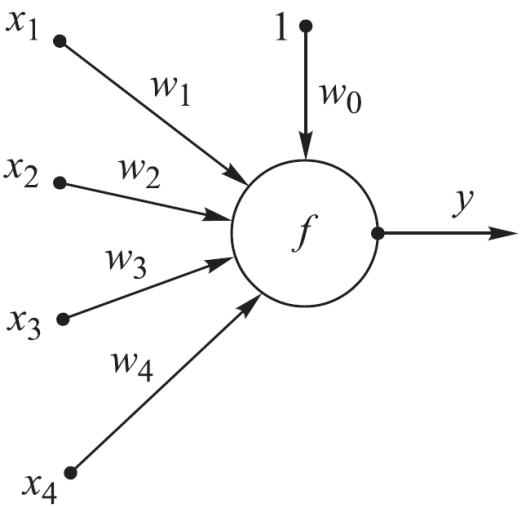
Архитектуры нейронных сетей

Архитектур нейронных сетей большое количество, и со временем появляются новые. Рассмотрим наиболее часто встречающиеся базовые архитектуры.

Персептрон

Персептрон – это простейшая модель нейросети, состоящая из одного нейрона. Нейрон может иметь произвольное количество входов (на рис. 3 изображен пример вида

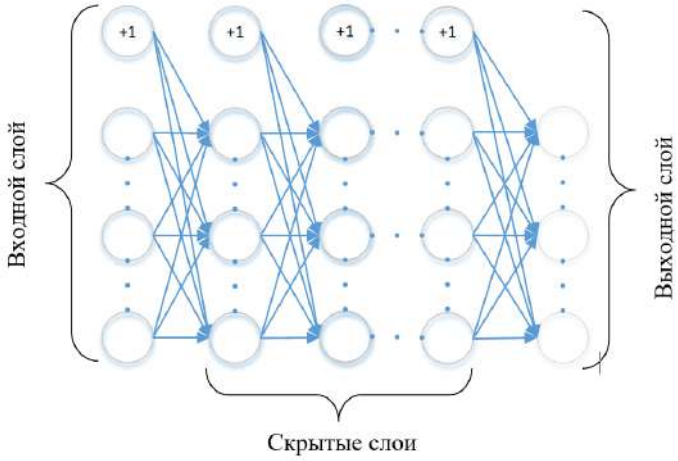
Рис. 3. Пример архитектуры персептрона[3].



персептрона с четырьмя входами), а один из них обычно тождественно равен 1. Этот единичный вход называют смещением, его использование иногда бывает очень удобным. Каждый вход имеет свой собственный вес. При поступлении сигнала в нейрон, как и было описано выше, вычисляется взвешенная сумма сигналов, затем к сигналу применяется функция активации и сигнал передается на выход. Такая простая на первый взгляд сеть, всего из одного-единственного нейрона, способна, тем не менее, решать ряд задач: выполнять простейший прогноз, регрессию данных и т.п., а также моделировать поведение несложных функций. Главное для эффективности работы этой сети – линейная разделимость данных.

Для читателей, немного знакомых с таким разделом как логика, будут знакомы логическое И (умножение) и логическое ИЛИ (сложение). Это булевы функции, принимающие значение, равное 0, всегда, кроме одного случая, для умножения, и, наоборот, значения, равные 1, всегда, кроме одного случая, для сложения. Для таких функций всегда можно провести прямую (или гиперплоскость в многомерном пространстве), отделяющие 0 значения от 1. Это и называют линейной разделимостью объектов. Тогда подобная модель сможет решить поставленную задачу классификации и сформировать базовый логический элемент. Если же объекты линейно неразделимы, то сеть из

Рис. 4. Архитектура многослойного персептрона.



одного нейрона с ними не справится. Для этого существуют более сложные архитектуры нейронных сетей, в конечном счете, тем не менее, состоящие из множества таких персептронов, объединенных в слои.

Многослойный персептрон

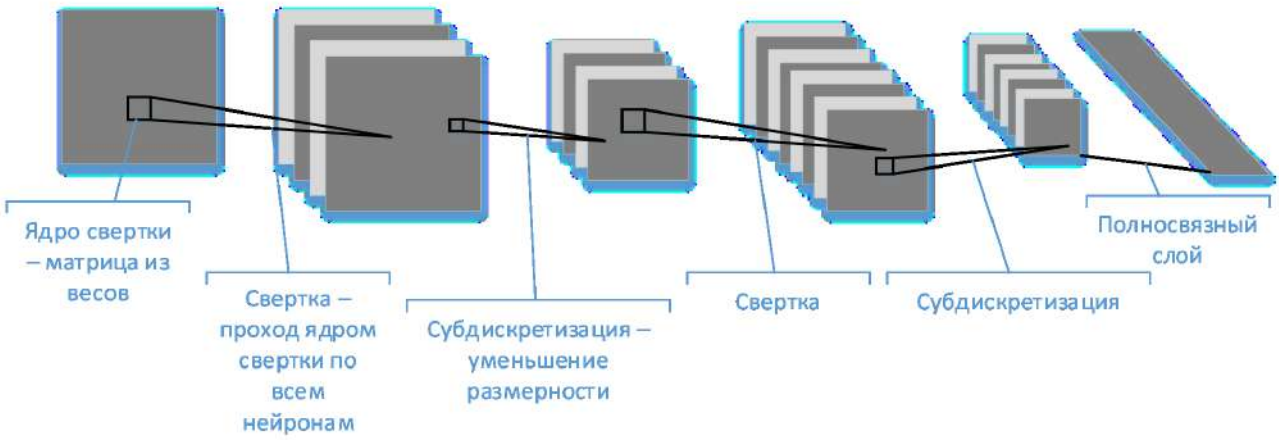
Слоем называют объединение нейронов, на схемах, как правило, слои изображаются в виде одного вертикального ряда (в некоторых случаях схему могут поворачивать, и тогда ряд будет горизонтальным). Многослойный персептрон является обобщением однослойного персептрона. Он состоит из некоторого множества входных узлов, нескольких скрытых слоев вычислительных нейронов и выходного слоя (см. рис. 4).

Отличительные признаки многослойного персептрона:

- Все нейроны обладают нелинейной функцией активации, которая является дифференцируемой.
- Сеть достигает высокой степени связности при помощи синаптических соединений.
- Сеть имеет один или несколько скрытых слоев[4].

Такие многослойные сети еще называют глубокими. Эта сеть нужна для решения задач, с которыми не может справиться персептрон. В частности, линейно неразделимых задач. В действительности, как следствие из теоремы Колмогорова, именно многослойный персептрон является единственной универсальной нейронной сетью, универсальным аппроксиматором, способным решить любую задачу. Возможно, не самым эффективным способом, который могли бы обеспечить более узкопрофильные нейросети, но все же решить. Именно поэтому, если исследователь не уверен, нейросеть какого вида подходит для решения стоящей перед ним задачи, он выбирает сначала многослойный персептрон.

Рис. 5. Архитектура сверточной нейронной сети.



Сверточная нейронная сеть (convolution neural network)

Сверточная нейронная сеть – сеть, которая обрабатывает передаваемые данные не целиком, а фрагментами. Данные последовательно обрабатываются, а после передаются дальше по слоям. Сверточные нейронные сети состоят из нескольких типов слоев: сверточный слой, субдискретизирующий слой, слой полносвязной сети (когда каждый нейрон одного слоя связан с каждым нейроном следующего – полная связь). Слои свертки и подвыборки (субдискретизации) чередуются и их набор может повторяться несколько раз (см. рис. 5). К конечным слоям часто добавляют персептроны, которые служат для последующей обработки данных[5].

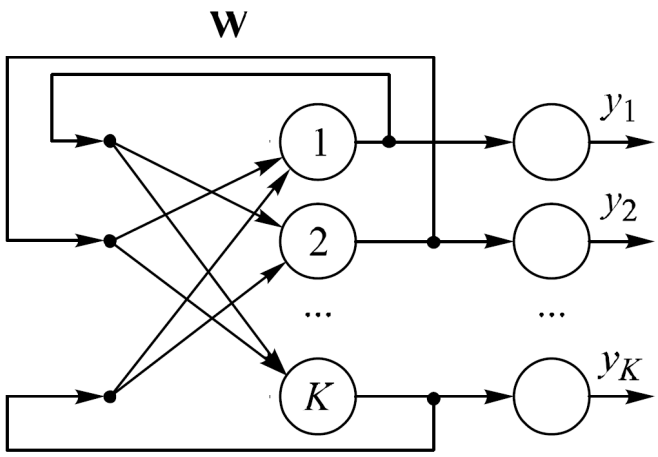
Название архитектура сети получила из-за наличия операции свёртки, суть которой в том, что каждый фрагмент изображения умножается на матрицу (ядро) свёртки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения. Необходимо это для перехода от конкретных особенностей изображения к более абстрактным деталям, и далее – к ещё более абстрактным, вплоть до выделения понятий высокого уровня (присутствует ли что-либо искомое на изображении).

Сверточные нейронные сети решают следующие задачи:

- **Классификация.** Пример: нейронная сеть определяет, что или кто находится на изображении.
- **Детекция.** Пример: нейронная сеть определяет, что/кто и где находится на изображении.
- **Сегментация.** Пример: нейронная сеть может определить каждый пиксель изображения и понять, к чему он относится.

Описанные сети относятся к сетям прямого распространения. Наряду с ними существуют нейронные сети, архитектуры которых имеют в своем составе связи, по которым сигнал распространяется в обратную сторону.

Рис. 6. Архитектура рекуррентной нейронной сети[3].



Рекуррентная нейронная сеть

Рекуррентная нейронная сеть – сеть, соединения между нейронами которой образуют ориентированный цикл. Т.е. в сети имеются обратные связи. При этом информация к нейронам может передаваться как с предыдущих слоев, так и от самих себя с предыдущей итерации (задержка). Пример схемы первой рекуррентной нейронной сети (НС Хопфилда) представлен на рис. 6.

Характеристики сети:

- Каждое соединение имеет свой вес, который также является приоритетом.
- Узлы подразделяются на два типа: вводные (слева) и скрытые (1, 2, ... K).
- Информация, находящаяся в нейронной сети, может передаваться как по прямой, слой за слоем, так и между нейронами.

Такие сети еще называют «памятью» (в случае сети Хопфилда – автоассоциативной; бывают также гетероассоциативные (сеть Коско – развитие сети Хопфилда) и другие). Почему? Если подать данной сети на вход некие «образцы» –

последовательности кодов (к примеру, 1000001, 0111110 и 0110110) - и обучить ее на запоминание этих образцов, настроив веса синапсов сети определенным образом при помощи правила Хебба[3], то затем, в процессе функционирования, сеть сможет «узнавать» запомненные образы и выдавать их на выход, в том числе исправляя искаженные поданные на вход образы. К примеру, если после обучения такой сети я подам на вход 1001001, то сеть узнает и исправит запомненный образец, выдав на выходе 1000001. Правда, эта нейронная сеть не толерантна к поворотам и сдвигам образов, и все же для первой нейронной сети своего класса сеть весьма интересна.

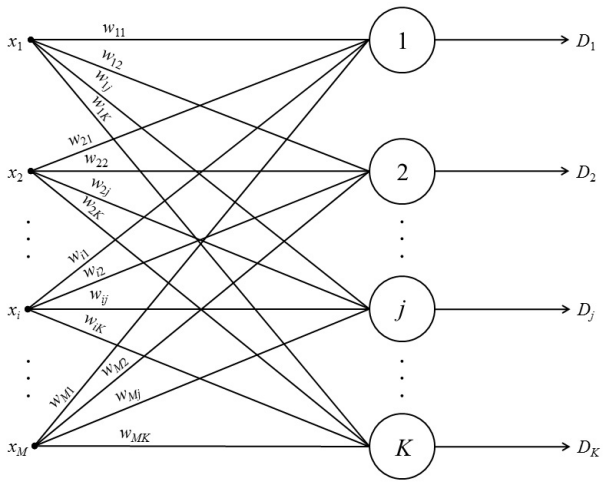
Таким образом, особенность рекуррентной нейронной сети состоит в том, что она имеет «области внимания». Данная область позволяет задавать фрагменты передаваемых данных, которым требуется усиленная обработка.

Информация в рекуррентных сетях со временем теряется со скоростью, зависящей от активационных функций. Данные сети на сегодняшний день нашли свое применение в распознавании и обработке текстовых данных. Об этом речь пойдет далее.

Самоорганизующаяся нейронная сеть

Пример самоорганизующейся нейронной сети – сеть Кохонена. В процессе обучения осуществляется адаптация сети к поставленной задаче. В представленной сети (см. рис. 7) сигнал идет от входа к выходу в прямом направлении. Структура сети имеет один слой нейронов, которые не имеют коэффициентов смещения (тождественно

Рис. 7. Архитектура самоорганизующейся нейронной сети.



единичных входов). Процесс обучения сети происходит при помощи метода последовательных приближений. Нейронная сеть подстраивается под закономерности входных данных, а не под лучшее значение на выходе. В результате обучения сеть находит окрестность, в которой находится лучший нейрон. Сеть функционирует по принципу «победитель получает все»: этот самый лучший нейрон в итоге на выходе будет иметь 1, а остальные нейроны, сигнал на которых получился меньше, 0. Визуализировать это можно так: представьте, что правый ряд нейронов, выходной, на

Рис. 8. Процесс обучения нейронной сети.

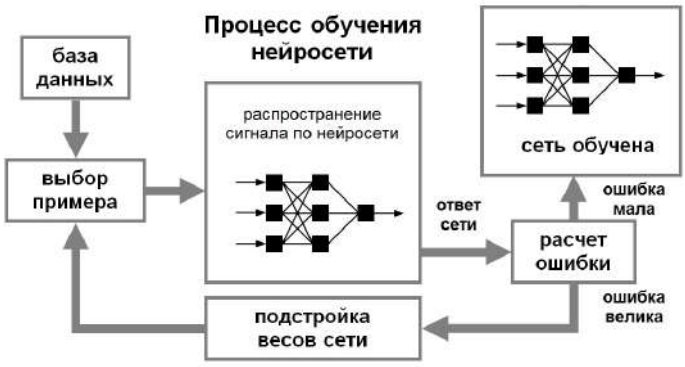


рис. 7 – это лампочки. Тогда после подачи на вход сети данных после их обработки на выходе «зажжется» только одна лампочка, указывающая, куда относится поданный объект. Именно поэтому такие сети часто используются в задачах кластеризации и классификации.

Алгоритмы обучения нейронных сетей

Чтобы получить решение поставленной задачи с использованием нейронной сети, вначале требуется сеть обучить (рис. 8). Процесс обучения сети заключается в настройке весовых коэффициентов связей между нейронами.

Алгоритмы обучения для нейронной сети бывают с учителем и без.

- **С учителем:** предоставление нейронной сети некоторой выборки обучающих примеров. Образец подается на вход, после происходит обработка внутри нейронной сети и рассчитывается выходной сигнал, сравнивающийся с соответствующим значением целевого вектора (известным нам «правильным ответом» для каждого из обучающих примеров). Если ответ сети не совпадает с требуемым, производится коррекция весов сети, напрямую зависящая от того, насколько отличается ответ сети от правильного (ошибка). Правило этой коррекции называется правилом Видроу-Хоффа и является прямой пропорциональностью коррекции каждого веса и размера ошибки, производной функции активации и входного сигнала нейрона. Именно после изобретения и доработок алгоритма распространения этой ошибки на все нейроны скрытых слоев глубоких нейронных сетей эта область искусственного интеллекта вернула к себе интерес.

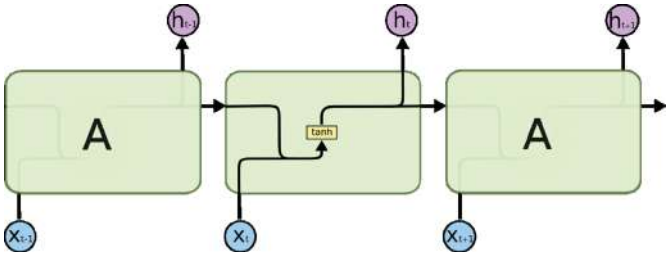
- **Без учителя:** алгоритм подготавливает веса сети таким образом, чтобы можно было получить согласованные выходные векторы, т.е. предоставление достаточно близких векторов будет давать похожие выходы. Одним из таких алгоритмов обучения является правило Хебба, по которому настраивается перед работой матрица весов, например, рекуррентной сети Хопфилда.

Рассмотрим теперь, какие архитектуры сетей являются наиболее востребованными сейчас, каковы особенности использования нейросетей на примерах известных крупных проектов.

Рекуррентные сети на примерах Google и «Яндекса»

Уже упоминавшаяся выше рекуррентная нейронная сеть (RNN) – сеть, которая обладает кратковременной «памятью», из-за чего может быстро производить анализ последовательностей различной длины. RNN разбивает

Рис. 9. Модули рекуррентной нейронной сети[7].



потоки данных на части и производит оценку связей между ними.

Сеть долговременной краткосрочной памяти (Long short-term memory – LSTM) – сеть, которая появилась в результате развития RNN-сетей. Это интересная модификация RNN-сетей, позволяющая сети не просто «держать контекст», но и умеющая «видеть» долговременные зависимости. Поэтому LSTM подходит для прогнозирования различных изменений при помощи экстраполяции (выявление тенденции на основе данных), а также в любых задачах, где важно умение «держать контекст», особенно хорошо подвластное для данной нейросети.

Форма рекуррентных нейронных сетей представляет собой несколько повторяющихся модулей. В стандартном виде эти модули имеют простую структуру. На рисунке 9 изображена сеть, которая имеет в одном из слоев гиперболический тангенс в качестве функции активации.

Рис. 10. Модуль LSTM-сети[7].

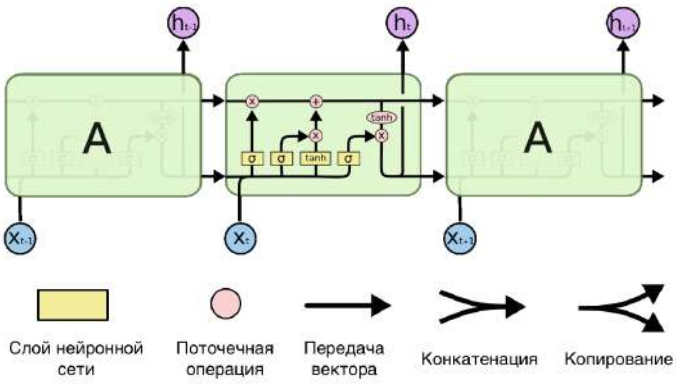
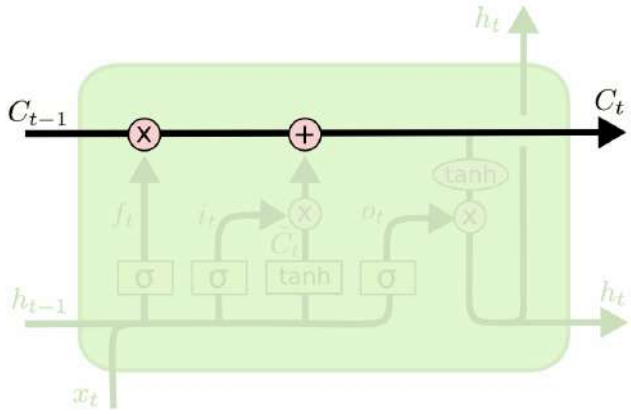
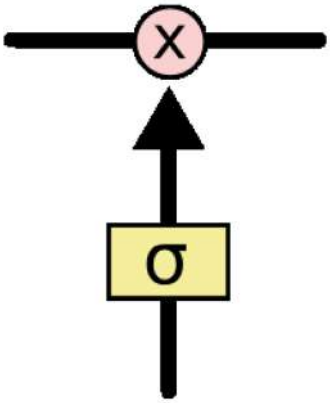


Рис. 11. Клеточное состояние LSTM-модуля[7].



LSTM также имеет цепочечную структуру. Отличие состоит в том, что сеть имеет четыре слоя, а не один (рис. 10). Главным отличием LSTM-сети является ее клеточное состояние (или состояние ячейки), которое обозначается горизонтальной линией в верхней части диаграммы и по которой проходит информация (рис. 11). Это самое состояние ячейки напоминает ленту конвейера: она проходит напрямую

Рис. 12. Вентиль[7].

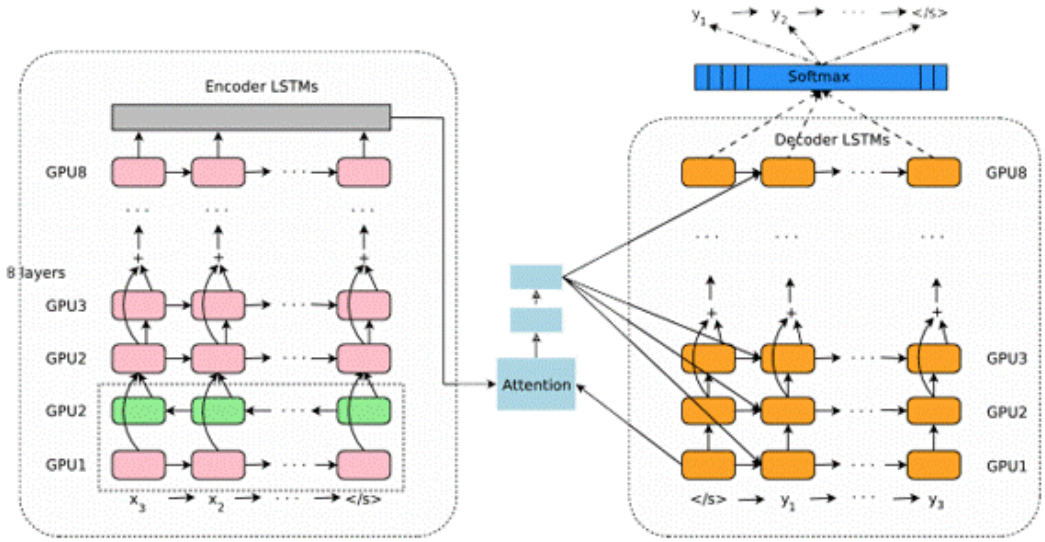


через всю цепочку, участвуя в некоторых преобразованиях. Информация может как «течь» по ней, не подвергаясь изменениям, так и быть подвергнута преобразованиям со стороны нейросети[7]. (Представьте себе линию контроля на конвейере: изделия продвигаются на ленте перед сотрудником контроля, который лишь наблюдает за ними, но если ему что-то покажется важным или подозрительным, он может вмешаться).

LSTM имеет способность удалять или добавлять информацию к клеточному состоянию. Данная способность осуществляется при помощи вентилей, которые регулируют процессы взаимодействия с информацией. Вентиль – возможность выборочно пропускать информацию. Он состоит из сигмоидного слоя нейронной сети и операции поточечного умножения (рис. 12).

Сигмоидный слой (нейроны с сигмоидальной функцией активации) подает на выход числа между нулем и

Рис. 13. Рекуррентная двунаправленная нейронная сеть Google-переводчика.



единицей, описывая таким образом, насколько каждый компонент должен быть пропущен сквозь вентиль. Ноль – «не пропускать вовсе», один – «пропустить все». Таких «фильтров» в LSTM-сети несколько. Для чего сети нужна эта способность: решать, что важно, а что нет? Представьте себе следующее: сеть переводит предложение и в какой-то момент ей предстоит перевести с английского языка на русский, скажем, местоимение, прилагательное или причастие. Для этого сети необходимо, как минимум, помнить род и/или число предыдущего существительного, к которому переводимое слово относится. Однако как только мы встретим новое существительное, род предыдущего можно забыть. Конечно, это очень упрощенный пример, тем не менее, он позволяет понять трудность принятия решения: какая информация еще важна, а какую уже можно забыть. Иначе говоря, сети нужно решить, какая информация будет храниться в состоянии ячейки, а затем – что на выходе из нее будет выводить.

Примеры использования нейронных сетей

Мы рассмотрели основные типы нейронных систем и познакомились с принципами их работы. Остановимся теперь на их практическом применении в системах Google Translate, «Яндекс.Алиса», Google Photos и Google Image Search (AlexNet, GoogleNet).

LSTM в Google Translate

Google-переводчик в настоящее время основан на методах машинного обучения и использует принцип работы LSTM-сетей. Система производит вычисления, чтобы понять значение слова или фразы, основываясь на предыдущих значениях в последовательности (контексте). Такой алгоритм помогает системе понимать контекст предложения и верно подбирать перевод среди различных вариантов. Еще несколько лет назад Google Translate работал на основе статистического машинного перевода – это разновидность перевода, где перевод генерируется на основе статистических моделей (бывают: по словам, по фразам, по синтаксису и т.д.), а параметры этих моделей являются результатами анализа корпусов текстов двух выбранных для перевода языков. Эти статистические модели обладали большим быстродействием, однако их эффективность ниже. Общий прирост качества перевода после внедрения системы на основе нейросетей в Google-переводчик составил, казалось бы, не очень много – порядка 10%, однако для отдельных языковых пар эффективность перевода достигла 80-90%, вплотную приблизившись к оценкам качества человеческого перевода. «Платой» за такое качество перевода является сложность построения, обучения и перенастройки системы на основе нейросети: она занимает недели. Вообще для современных глубоких нейронных сетей, использующихся в таких крупных проектах, в порядке вещей обучение, занимающее дни и недели.

Рассмотрим LSTM-сеть переводчика (GNMT) подробнее. Нейронная сеть переводчика имеет разделение на два потока (см. рис. 13). Первый поток нейронной сети (слева) является анализирующим и состоит из восьми слоев. Данный поток помогает разбивать фразы или предложения на несколько смысловых частей, а после производит их анализ. Сеть позволяет читать предложения в двух направлениях, что помогает лучше понимать контекст. Также она занимается формированием модулей внимания, которые позволяют второму потоку определять ценности отдельно взятых смысловых фрагментов.

Второй поток нейронной сети (справа) является синтезирующим. Он позволяет вычислять самый подходящий вариант для перевода, основываясь на контексте и модулях внимания (показаны голубым цветом).

В системе нейронной сети самым маленьким элементом является фрагмент слова. Это позволяет сфокусировать вычислительную мощность на контексте предложения, что обеспечивает высокую скорость и точность переводчика. Также анализ фрагментов слова сокращает риски неточного перевода слов, которые имеют суффиксы, префиксы и окончания[8].

LSTM в «Яндекс.Алисе»

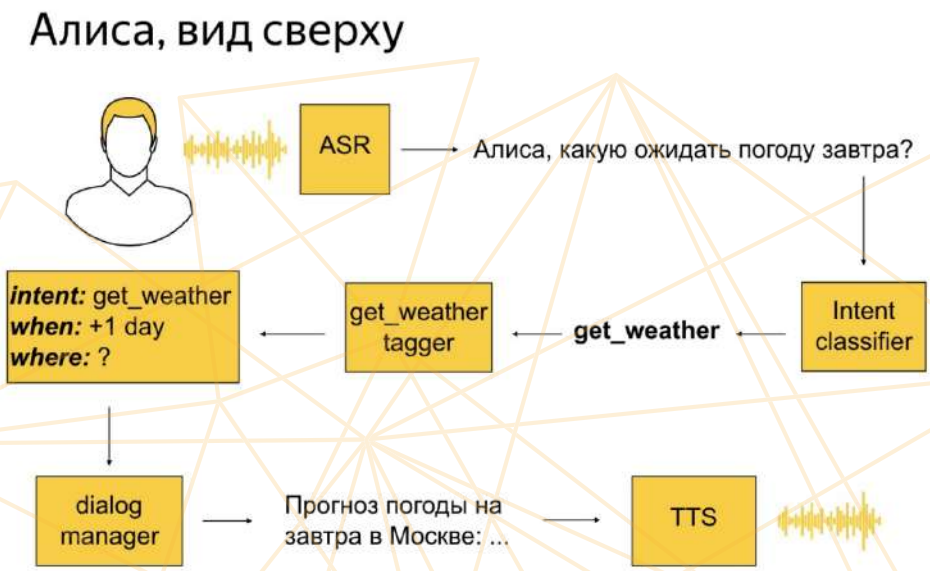
Компания «Яндекс» также использует нейросеть типа LSTM в продукте «Яндекс.Алиса». На рисунке 14 изображено взаимодействие пользователя с системой. После вопроса пользователя информация передается на сервер распознавания, где она преобразуется в текст. Затем текст поступает в классификатор интентов (интент (от англ. intent) – намерение, цель пользователя, которое он вкладывает в запрос), где при помощи машинного обучения анализируется. После анализа интентов они поступают на семантический теггер (tagger) – модель, которая позволяет выделить полезную и требующуюся информацию из предложения. Затем полученные результаты структурируются и передаются в модуль dialog manager, где обрабатывается полученная информация и генерируется ответ.

LSTM-сети применяются в семантическом теггере. В этом модуле используются двунаправленные LSTM с attention. На этом этапе генерируется не только самая вероятная гипотеза, потому что в зависимости от диалога могут потребоваться и другие. От состояния диалога в dialog manager происходит повторное взвешивание гипотез. Сети с долговременной памятью помогают лучше вести диалог с пользователем, потому что могут помнить предыдущие сообщения и более точно генерировать ответы, а также они помогают разрешать много проблемных ситуаций. В ходе диалога нейронная сеть может убирать слова, которые не несут никакой важной информации, а остальные анализировать и предоставлять ответ[9]. Для примера на рис. 14 результат работы системы будет следующим. Сначала речь пользователя в звуковом формате будет передана на сервер распознавания, который выполнит лексический анализ (т.е. переведет сказанное в текст и разобьет его на слова – лексемы), затем классификатор интентов выполнит уже роль семантического анализа (понимание смысла того, что было сказано; в данном случае – пользователя интересует погода), а семантический теггер выделит полезные элементы информации: теггер выдал бы, что завтра – это дата, на которую нужен прогноз погоды. В итоге данного анализа запроса пользователя будет составлено внутреннее представление запроса – фрейм, в котором будет указано, что интент – погода, что погода нужна на +1 день от текущего дня, а где – неизвестно. Этот фрейм попадет в диалоговый менеджер, который хранит тот самый контекст беседы и будет принимать на основе хранящегося контекста решение, что делать с этим запросом. Например, если до этого беседа не происходила, а для станции с «Яндекс.Алисой» указана геолокация г. Москва, то менеджер получит через специальный интерфейс (API) погоду на завтра в Москве, отправит команду на генерацию текста с этим прогнозом, после чего тот будет передан в синтезатор речи помощника. Если же Алиса до этого вела беседу с пользователем о достопримечательностях Парижа, то наиболее вероятно, что менеджер учтет это и либо уточнит у пользователя информацию о месте, либо сразу сообщит о погоде в Париже. Сам диалоговый менеджер работает на основе скриптов (сценариев обработки запросов) и правил, именуемых form-filling. Вот такой непростой набор технических решений, методов, алгоритмов и систем кроется за фасадом, позволяющим пользователю спросить и узнать ответ на простой бытовой вопрос или команду.

Свёрточные нейронные сети, используемые для классификации изображений, на примере Google Photos и Google Image Search

Задача классификации изображений – это получение на вход начального изображения и вывод его класса (стол, шкаф, кошка, собака и т.д.) или группы вероятных классов, которая лучше всего характеризует изображение. Для людей это совершенно естественно и просто: это один из первых навыков, который мы осваиваем с рождения. Компьютер лишь «видит» перед собой пиксели изображения, имеющие различный цвет и интенсивность. На сегодня задача классификации объектов, расположенных на изображении, является основной для области компьютерного зрения. Данная область начала активно развиваться с 1970-х, когда компьютеры стали достаточно мощными, чтобы оперировать большими объёмами данных. До развития нейронных сетей задачу классификации изображений

Рис. 14. Процесс работы сервиса «Яндекс.Алиса»[9].



решали с помощью специализированных алгоритмов. Например, выделение границ изображения с помощью фильтра Собеля и фильтра Шарра или использование гистограмм градиентов[10]. С появлением более мощных GPU, позволяющих значительно эффективнее обучать нейросеть, применение нейросетей в сфере компьютерного зрения значительно возросло.

В 1988 году Ян Лекун предложил свёрточную архитектуру нейросети, нацеленную на классификацию изображений[11]. А в 2012 году сеть AlexNet, построенная по данной архитектуре, заняла первое место в конкурсе по распознаванию изображений ImageNet, имея ошибку распознавания равную 15,3%[12]. Свёрточная архитектура нейросети использует некоторые особенности биологического распознавания образов, происходящих в мозгу человека. Например, в мозгу есть некоторые группы клеток, которые активируются, если в определённое поле зрения попадает горизонтальная или вертикальная граница объекта. Данное явление обнаружили Хьюбель и Визель в 1962 году[13]. Поэтому, когда мы смотрим на изображение, скажем, собаки, мы можем отнести его к конкретному классу, если у изображения есть характерные особенности, которые можно идентифицировать, такие как лапы или четыре

Рис. 15. Архитектура свёрточной сети AlexNet[14].

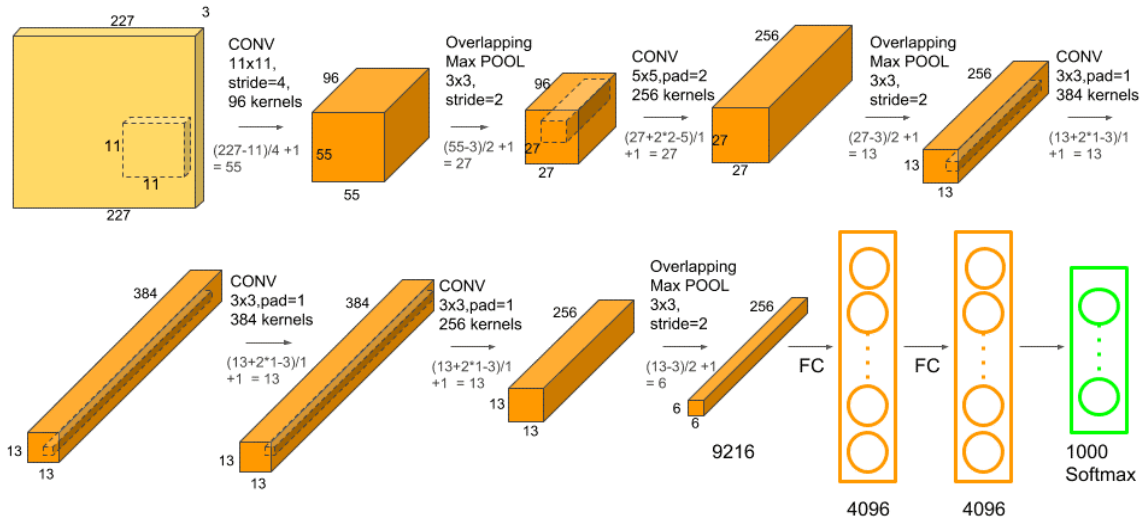
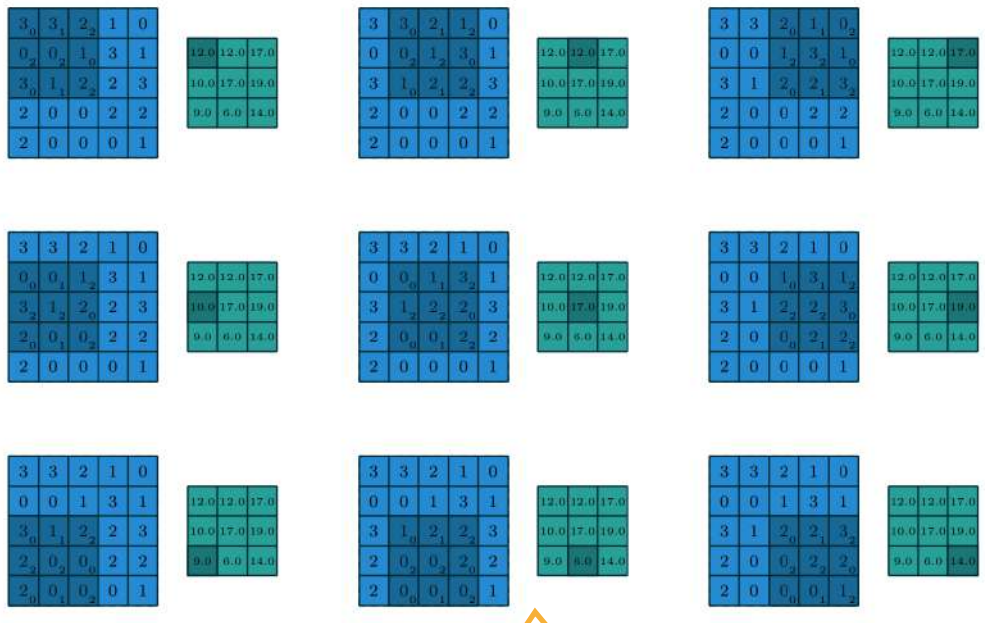


Рис. 16. Операция свёртки[15].



Исходная матрица A обозначена голубым цветом (имеет бОльшую размерность), ядро свертки – матрица B - обозначена темно-синим цветом и перемещается по матрице A; результирующая матрица изображена зеленым, при этом соответствующая каждому шагу свертки результирующая ячейка отмечена темно-зеленым.

ноги. Аналогичным образом компьютер может выполнять классификацию изображений через поиск характеристик базового уровня, например, границ и искривлений, а затем - с помощью построения более абстрактных концепций через группы сверточных слоев.

Рассмотрим подробнее архитектуру и работу свёрточной нейросети (convolution neural network) на примере нейросети AlexNet (см. рис. 15). На вход подаётся изображение 227x227 пикселей (на рисунке изображено желтым), которое необходимо классифицировать (всего 1000 возможных классов). Изображение цветное, поэтому оно разбито по трём RGB-каналам. Наиболее важная операция – операция свёртки.

Операция свёртки (convolution) производится над двумя матрицами A и B размера $[a \times b]$ и $[c \times d]$, результатом которой является матрица C размера

$$\left[\left(\frac{a-c}{s}+1\right)\times\left(\frac{b-d}{s}+1\right)\right]$$

, где s – шаг, на который сдвигают матрицу B. Вычисляется результирующая матрица следующим образом:

$$c_{i,j} = \sum_{u=0}^{c-1}\sum_{v=0}^{d-1} a_{i+u,j+v} b_{u,v}$$

см. рис. 16.

Рисунок 16. Операция свёртки[15]: Исходная матрица A обозначена голубым цветом (имеет большую размерность), ядро свертки – матрица B - обозначена темно-синим цветом и перемещается по матрице A; результирующая матрица изображена зеленым, при этом соответствующая каждому шагу свертки результирующая ячейка отмечена темно-зеленым.

Чем больше результирующий элемент $\mathbf{c}_{i,j}$, тем более похожа область матрицы **A** на матрицу **B**. **A** называют изображением, а **B** – ядром или фильтром. Обычно шаг смещения **B** равен **s = 1**. Если увеличить шаг смещения, можно добиться уменьшения влияния соседних пикселей друг на друга в сумме (уменьшить рецептивное поле восприятия нейросети), в подавляющем большинстве случаев шаг оставляют равным 1. На рецептивное поле также влияет и размер ядра свёртки. Как видно из рис. 16, при свёртке результат теряет в размерности по сравнению с матрицей A. Поэтому, чтобы избежать данного явления, матрицу A дополняют ячейками (padding)[15]. Если этого не сделать, данные при переходе на следующие слои будут теряться слишком быстро, а информация о границах слоя будет неточной.

Таким образом, свёртка позволяет одному нейрону отвечать за определённый набор пикселей в изображении. Весовыми коэффициентами являются коэффициенты фильтра, а входными переменными – интенсивность свечения пикселя плюс параметр смещения.

Данный пример свёртки (см. рис. 16) был одномерным и подходит, например, для черно-белого изображения. В случае цветного изображения операция свёртки происходит следующим образом: в один фильтр включены три ядра, для каждого канала RGB соответственно, над каждым каналом операция свёртки происходит идентично примеру выше, затем три результирующие матрицы поэлементно складываются - и получается результат работы одного фильтра – карта признаков.

Эта на первый взгляд непростая с математической точки зрения процедура нужна для выделения тех самых границ и искривлений – контуров изображения, при помощи которых, по аналогии с мозгом человека, сеть должна понять, что же изображено на картинке. Т.е. на карте признаков будут содержаться выделенные характеристические очертания

Рис. 17. Первая реализация блока Inception[16].

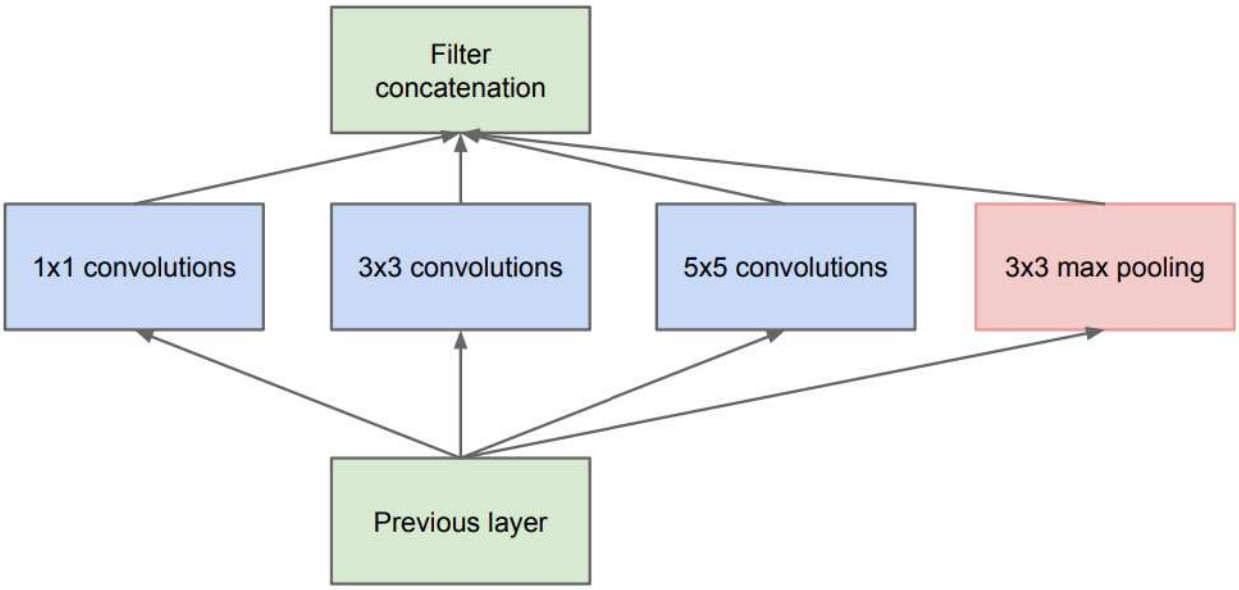
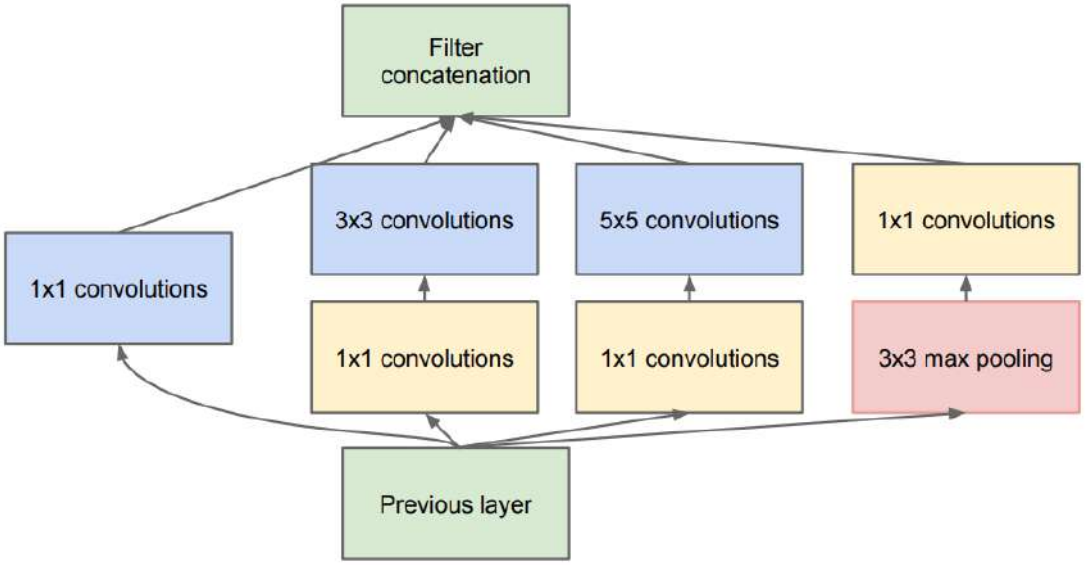


Рис. 18. Вторая реализация блока Inception[16].



детектированных объектов. Эта карта признаков будет являться результатом работы чередующихся слоев свертки и пулинга (от англ. pooling, или иначе – subsampling layers). Помните, в разделе описания концепции архитектуры сверточных сетей мы говорили, что эти слои чередуются и их наборы могут повторяться несколько раз (см. рис. 5)? Для чего это нужно? После свертки, как мы знаем, выделяются и «заостряются» некие характеристические черты анализируемого изображения (матрицы пикселей) путем увеличения значений соответствующих ядру ячеек матрицы и «занулению» значений других. При операции субдискретизации, или пулинга, происходит уменьшение размерности сформированных отдельных карт признаков. Поскольку для сверточных нейросетей считается, что информация о факте наличия искомого признака важнее точного знания его координат на изображении, то из нескольких соседних нейронов карты признаков выбирается максимальный и принимается за один нейрон уплотнённой карты признаков меньшей размерности. За счёт данной операции, помимо ускорения дальнейших вычислений, сеть становится также более инвариантной к масштабу входного изображения, сдвигам и поворотам изображений. Данное чередование слоев мы видим и для данной сети AlexNet (см. рис. 15), где слои свертки (обозначены CONV) чередуются с пулингом (обозначены эти слои как POOL), при этом следует учитывать, что операции свертки тем ресурсозатратнее, чем больше размер ядра свертки (матрицы B на рис. 16), поэтому иногда, когда требуется свертка с ядром достаточно большой размерности, ее заменяют на две и более последовательных свертки с ядром меньших размерностей. (Об этом мы поговорим далее.) Поэтому на рис. 15 мы можем увидеть участок архитектуры AlexNet, на котором три слоя CONV идут друг за другом.

Самая интересная часть в сети – полносвязный слой в конце. В этом слое по вводным данным (которые пришли с предыдущих слоев) строится и выводится вектор длины N, где N – число классов, по которым мы бы хотели классифицировать изображение, из них программа выбирает нужный. Например, если мы хотим построить сеть по распознаванию цифр, у N будет значение 10, потому что цифр всего 10. Каждое число в этом векторе представляет собой вероятность конкретного класса. Например, если результирующий вектор для распознавания цифр это [0 0.2 0 0.75 0 0 0 0 0.05], значит существует 20% вероятность, что на изображении «1», 75% вероятность – что «3», и 5% вероятность – что «9». Если мы хотим сеть по распознаванию букв латинского алфавита, то N должно быть равным 26 (без учета регистра), а если и буквы, и цифры вместе, то уже 36. И т.д. Конечно, это очень упрощенные примеры. В сети AlexNet конечная размерность выхода (изображена на рис. 15 зеленым) - 1000, а перед этим идут еще несколько полносвязных слоев с размерностями входа 9216 и выхода 4096 и обоими параметрами 4096 (изображены оранжевым и отмечены FC – full connected - на рисунке).

Способ, с помощью которого работает полносвязный слой, – это обращение к выходу предыдущего слоя (который, как мы помним, должен выводить высокоуровневые карты свойств) и определение свойств, которые больше связаны с определенным классом. Поэтому в результате работы окончания сверточной сети по поданному на вход изображению с лужайкой, на которой, скажем, собака гоняется за птичкой, упрощенно можно сказать, что результатом может являться следующий вывод:
[dog (0.6), bird (0.35), cloud (0.05)].

Свёрточная нейронная сеть GoogleNet

В 2014 году на том же соревновании ImageNet, где была представлена AlexNet, компания Google показала свою свёрточную нейросеть под названием GoogleNet. Отличительная особенность этой сети была в том, что она использовала в 12 раз меньше параметров (почувствуйте масштабы: 5 000 000 против 60 000 000) и её архитектура структурно была не похожа на архитектуру AlexNet, в которой было восемь слоёв, причём данная сеть давала более точный результат – 6,67% ошибки против 16,4%[16].

Разработчики GoogleNet пытались избежать простого увеличения слоёв нейронной сети, потому что данный шаг привёл бы к значительному увеличению использования

Ещё один полезный аспект в архитектуре GoogleNet, по мнению разработчиков, состоит в том, что сеть интуитивно правильно обрабатывает изображение: сначала картинка обрабатывается в разных масштабах, а затем результаты агрегируются. Такой подход больше соответствует тому, как подсознательно выполняет анализ окружения человек.

памяти и сеть была бы больше склонна к перенасыщению при обучении на небольшом и ограниченном наборе примеров.

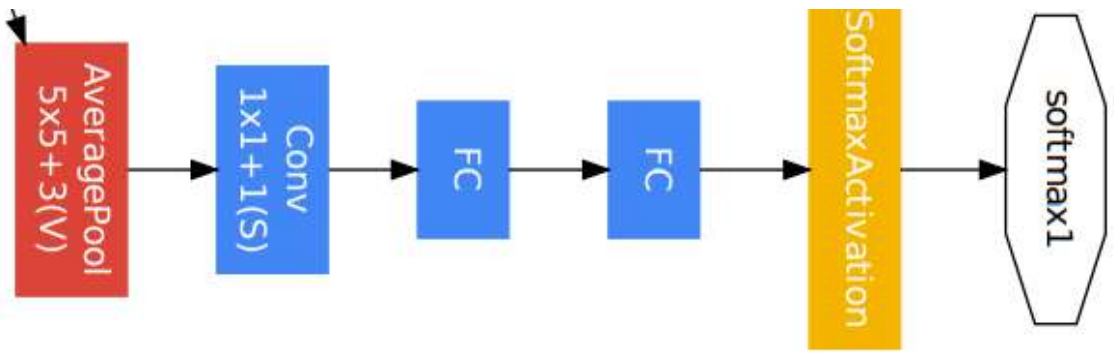
Проанализируем приём, на который пошли разработчики компании Google для оптимизации скорости и использования памяти. В нейросети GoogleNet свёртка с размером фильтра 7×7 используется один раз в начале обработки изображения, далее максимальный размер – 5×5. Так как количество параметров при свёртке растёт квадратично с увеличением размерности ядра, нужно стараться заменять одну свёртку на несколько свёрток с меньшим размером фильтра, вместе с этим пропустить промежуточные результаты через ReLU (функция активации, возвращающая 0, если сигнал отрицателен, и само значение сигнала, если нет), чтобы внести дополнительную нелинейность. Например, свёртку с размером фильтра 5×5 можно заменить двумя последовательными операциями с размером ядра 3×3. Такая оптимизация позволяет строить более гибкие и глубокие сети. На хранение промежуточных свёрток тратится память, и использовать их более разумно в слоях, где размер карты признаков небольшой. Поэтому в начале сети GoogleNet вместо трёх свёрток 3×3 используется одна 7×7, чтобы избежать избыточного использования памяти.

Схему архитектуры сети целиком мы приводить здесь не будем, она слишком громоздка и на ней трудно что-либо разглядеть, кому это будет интересно, могут ознакомиться с ней на официальном ресурсе[16]. Отличительной особенностью нейросети от компании Google является использование специального модуля – Inception (см. рис. 17). Данный модуль, по своей сути, является небольшой локальной сетью. Идея данного блока состоит в том, что на одну карту признаков накладывается сразу несколько свёрток разного размера, вычисляющихся параллельно. В конце все свёртки объединяются в единый блок. Таким образом, можно из исходной карты признаков извлечь признаки разного размера, увеличив эффективность сети и точность обработки признаков. Однако при использовании данной реализации (см. рис. 17) нужно выполнить колоссальный объём вычислений, между тем, при включении данных модулей друг за другом размерность блока будет только расти. Поэтому разработчики во второй версии добавили свёртки, уменьшающие размерность (см. рис. 18).

Жёлтые свёртки размера 1×1 введены для уменьшения глубины блоков, и благодаря им, значительно снижается нагрузка на память. GoogleNet содержит девять таких Inception-модулей и состоит из 22 слоёв.

Из-за большой глубины сети разработчики также столкнулись с проблемой затухания градиента при обучении (см. описание процедуры обучения нейросети: коррекция весов осуществляется в соответствии со значением ошибки,

Рис. 19. Вспомогательный модуль GoogleNet, использующийся при обучении сети[16].



производной функции активации – градиента – и т.д.) и ввели два вспомогательных модуля классификатора (см. рис. 19)[16]. Данные модули представляют собой выходную часть малой свёрточной сети и уже частично классифицируют объект по внутренним характеристикам самой сети. При обучении нейронной сети данные модули не дают ошибке, распространяющейся с конца, сильно уменьшиться, так как вводят в середину сети дополнительную ошибку.

Ещё один полезный аспект в архитектуре GoogleNet, по мнению разработчиков, состоит в том, что сеть интуитивно правильно обрабатывает изображение: сначала картинка обрабатывается в разных масштабах, а затем результаты агрегируются[16]. Такой подход больше соответствует тому, как подсознательно выполняет анализ окружения человек.

В 2015 году разработчики из Google представили модифицированный модуль Inception v2, в котором свёртка 5×5 была заменена на две свёртки 3×3, по причинам, приведённым выше, вдобавок потери информации в картах признаков при таком действии происходят незначительные, так как соседние ячейки имеют между собой сильную корреляционную связь[17]. Также, если заменить свёртку 3×3 на две последовательные свёртки 3×1 и 1×3, то это будет на 33% эффективнее, чем стандартная свёртка, а две свёртки 2×2 дадут выигрыш лишь в 11%[17]. Данная архитектура нейросети давала ошибку 5,6%, а потребляла ресурсов в 2,5 раза меньше.

Стоит отметить, что точного алгоритма построения архитектуры нейросети не существует. Разработчики составляют свёрточные нейросети, основываясь на результатах экспериментов и собственном опыте.

Нейронную сеть можно использовать для точной классификации изображений, загруженных в глобальную сеть, либо использовать для сортировки фотографий по определённому признаку в галерее смартфона.

Свёрточная нейронная сеть имеет следующие преимущества: благодаря своей структуре, она позволяет значительно снизить количество весов в сравнении с полносвязанной нейронной сетью, её операции легко поддаются распараллеливанию, что положительно сказывается на скорости обучения и на скорости работы, данная архитектура лидирует в области классификации изображений, занимая первые места на соревновании ImageNet. Из-за того, что для сети более важно наличие признака, а не его место на изображении, она инвариантна к различным сдвигам и поворотам сканируемого изображения.

К недостаткам данной архитектуры можно отнести следующие: состав сети (количество её слоёв, функция активации, размерность свёрток, размерность pooling-слоёв, очередность слоёв и т.п.) необходимо подбирать эмпирически к определённому размеру и виду изображения; сложность обучения: нейросети с хорошим показателем ошибки должны тренироваться на мощных GPU долгое время; и, наверное, главный недостаток – атаки на данные нейросети. Инженеры Google в 2015 году показали, что к картинке можно подмешать невидимый для человеческого зрения градиент, который приведёт к неправильной классификации[18].

Список литературы

1. *Vlasov A., Yudin A. Distributed control system in mobile robot application: general approach, realization and usage // Communications in Computer and Information Science. 2011. T. 156 CCIS. P. 180–192.*

2. Яковлев В.Л., Яковлева Г.Л., Власов А.И. Нейросетевые методы и модели при прогнозировании краткострочных и долгосрочных тенденций финансовых рынков // В сборнике: VI Всероссийская конференция «Нейрокомпьютеры и их применение», 2000. С. 372–377.

3. Интеллектуальные технологии на основе искусственных нейронных сетей / М.А. Басараб, Н.С. Коннова. – Москва: Издательство МГТУ им. Н.Э. Баумана, 2017. – 53 с.

4. Солдатова О.П. Нейроинформатика. Курс лекций. СГАУ, 2013. 130 с.

5. *Y. LeCun, Y. Bengio. Convolutional Networks for Images, Speech, and Time-Series // The Handbook of Brain Theory and Neural Networks, MIT Press, 1995.*

6. *Understanding LSTM Networks* [Электронный ресурс] URL: https://colah.github.io/posts/2015-08-Understanding-

LSTMs/ (Дата обращения 26.03.2020).

7. LSTM – сети долгой краткосрочной памяти [Электронный ресурс] URL: https://habr.com/ru/company/wunderfund/blog/331310/ (Дата обращения 30.03.2020).

8. Как работает нейросеть Google Translate [Электронный ресурс] URL: https://www.cossa.ru/152/196086 (Дата обращения 26.03.2020).

9. Как устроена Алиса. Лекция Яндекса [Электронный ресурс] URL: https://habr.com/ru/company/yandex/blog/349372/ (Дата обращения 26.03.2020).

10. Д.А. Хрящёв. Об одном методе выделения контуров на цифровых изображениях // Вестник АГТУ, 2010. № 2. С. 1-7.

11. *LeCun Y., Bottou L., Bengio Y., Haffner P. GradientBased Learning Applied to Document Recognition // Proceedings of the IEEE. 1998. 46 p.*

12. *Krizhevsky A., Sutskever I., Hinton G. ImageNet Classification with Deep Convolutional Neural Networks // Advances in neural information processing systems, 25 (2). 2012.*

13. *Hubel D., Wiesel T. Brain mechanisms of vision // Scientific American, 241(3). 1979.*

14. *Nayak S. Understanding AlexNet* [Электронный ресурс] URL: https://www.learnopencv.com/understanding-alexnet/ (Дата обращения 30.03.2020).

15. *Dumoulin V., Visin F. A guide to convolution arithmetic for deep learning* [Электронный ресурс] URL: https://arxiv.org/pdf/1603.07285.pdf (Дата обращения 30.03.2020).

16. *Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015. P. 1–10.*

17. *Szegedy C., Vanhoucke V., Ioffe S., Shlens J. Rethinking the Inception Architecture for Computer // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016* [Электронный ресурс] URL: https://arxiv.org/pdf/1512.00567.pdf (Дата обращения 30.03.2020).

18. *Ian J. Goodfellow, Shlens J., Szegedy C. Explaining and harnessing adversarial examples // Published as a conference paper at ICLR. 2015* [Электронный ресурс] URL: https://arxiv.org/pdf/1412.6572.pdf (Дата обращения 30.03.2020).

Системы цифровой идентификации лиц: будущее и реальность

Мадина Касенова

Будущее социально-экономического развития современных государств невозможно представить без цифровых решений, оптимизирующих множество производственных процессов, документооборот, упрощающий доступ к разнообразным услугам, и т.д.[1] «Цифровой тренд» затрагивает в том числе проблемы электронной/цифровой идентификации физических лиц, которые в настоящее время по-разному решаются государствами, предлагаются международными организациями (межправительственными и неправительственными), а также и компаниями частного сектора. В эпоху цифровых технологий привычные удостоверения личности (паспорт, водительские права, социальная карточка и проч.), выдаваемые компетентными органами государства, нередко маркируются цифровыми метками, позволяющими установить личность человека в онлайн-режиме, и наряду с этим используются новые цифровые средства, посредством которых осуществляется онлайн-идентификация физических лиц.

Комплекс проблем, связанных с идентификацией физических лиц, с очевидностью проявился в реальных условиях пандемии коронавируса (Covid-19), когда государства вынуждены предпринимать беспрецедентные меры санитарно-карантинных ограничений и вводить административные запреты. Так, к известным и широко применяемым способам идентификации различных объектов относится применение «кодирования с немедленным обратным ответом», т.н. QR-code (Quick Response code)[2]. Этот способ цифровой идентификации («QR-код») некоторые государства распространили для учета их передвижения. К примеру, в рамках общих ограничительных мер для предотвращения распространения коронавирусной инфекции в ряде регионов России введен разрешительный порядок передвижения лиц в соответствующей местности, предусматривающий получение цифровых удостоверений и, в том числе, QR-кода[3].

Цифровая идентификация физических лиц, как правило, осуществляется в конкретной сфере социальной жизни и имеет целевую направленность, что обусловливает, с одной стороны, разнообразие применяемых технологических способов и средств идентификации, с другой стороны, определяет формирование различных систем идентификации лиц как на национальном, так и на международном уровнях. Соответственно, задачей будущего является не только обеспечение устойчивого и безопасного использования идентификационных систем и их развития, но также содействие взаимному признанию государствами действующих систем, включая их трансграничную совместимость.

Принимая во внимание существующий опыт ряда зарубежных стран, свидетельствующий о разнообразных подходах формирования и использования систем идентификации физических лиц, в настоящей статье представлен достаточно краткий обзор систем идентификации, функционирование которых имеет нормативно-правовые основания: регламентировано нормативным актом государства, предложено соответствующим документом международной межправительственной организацией.

1. Соединеные Штаты Америки

1.1. Конгресс США в 2005 году принял общефедеральный нормативный акт – «Закон о подлинных удостоверениях личности» (The USA Real ID Act), реализация которого до настоящего времени не завершена, поскольку изначально для этого документа (Акт «Real ID») был установлен режим «отложенного исполнения» (deferred enforcement), предусматривающий несколько этапов. Следует пояснить ключевые моменты, объясняющие цель принятия Акта «Real ID» и причины такой поэтапной реализации.

Во-первых, в силу федеративного устройства США, конституционно компетенция между Федерацией и штатами (и Округом Колумбия) определена и разграничена[4]. Во-вторых, установление правил выпуска и порядка выдачи официальных документов, удостоверяющих личность, а также установление видов и форм таких документов относится к компетенции штата (и Округа Колумбия)[5], что приводит к отсутствию единообразия выдаваемых документов, удостоверяющих личность. В-третьих, к документам, удостоверяющим личность, относится водительское удостоверение (Driver's License), которая зачастую имеет большее значение по сравнению с иными документами. Поскольку выдача водительского удостоверения отнесена к компетенции штата, то не только порядок, но и форма удостоверения может отличаться от штата к штату.

Сказанное объясняет, что Акт «Real ID» был принят с целью

ЦИФРОВАЯ ИДЕНТИФИКАЦИЯ ФИЗИЧЕСКИХ ЛИЦ

Задачей будущего является не только обеспечение устойчивого и безопасного использования идентификационных систем и их развития, но также содействие взаимному признанию государствами действующих систем, включая их трансграничную совместимость.

законодательного установления основных минимальных стандартов, которым должны отвечать документы, удостоверяющие личность, выдаваемые компетентными органами штатов и федеральными органами. При этом, существующая «множественность» форм и видов удостоверяющих документов объективно определила невозможность их единовременной замены (равно как и введение некоего «единого федерального» документа) и это, в свою очередь, обусловило необходимость поэтапной реализации Акта «Real ID».

Акт «Real ID» устанавливает основные минимальные стандарты, которым должны отвечать документы, удостоверяющие личность. Предусмотрено, что компетентные органы штатов и федеральные органы обязаны обеспечить оформление, выпуск и выдачу таких удостоверений личности, которые отвечают требованиям основных минимальных стандартов. Согласно Акту «Real ID», основными минимальными стандартами для водительского удостоверения, которое выдается компетентными органами штатов, являются наличие маркировки удостоверения звездочкой (или нанесение машиночитаемого штрих-кода), наличие цифровой фотографии получателя лицензии (цифрового изображения лица получателя лицензии). Важно отметить, что иные документы, удостоверяющие личность, должны не только соответствовать основным минимальным стандартам, но также быть совместимыми с водительским удостоверением.

Соответственно, только те документы, удостоверяющие личность, которые отвечают и основным минимальным стандартам, и совместимы с водительским удостоверением, в соответствии с Актом «Real ID», рассматриваются как «подлинные» (действительные) документы, на основании которых может быть подтверждена личность человека. Помимо этого, согласно Акту «Real ID», компетентные органы вправе отказать в признании удостоверяющих лицо документов, выданных в других государствах, если таковые не отвечают требованиям основных минимальных стандартов.

Поскольку идентификация лиц осуществляется на основании «подлинных» (действительных) документов, только такие лица получают право доступа на объекты федерального значения, включая возможность доступа на борт воздушных судов, как внутреннего, так и международного авиасообщения.

Контроль и мониторинг реализации Акта «Real ID» начиная с 20 января 2014 года осуществляет Департамент внутренней национальной безопасности США (US Department of Homeland Security, DHS). В этих целях DHS, inter alia, проверяет своевременность направления данных о выданных компетентными органами штатов документах; взаимодействует с Департаментом транспортных средств (US Department of Motor Vehicle, DMV) на предмет обеспечения «совместимости» выдаваемых документов с водительской лицензией; определяет единые стандарты сбора и хранения информации для формирования единой национальной базы данных Real ID.

В силу существующего «многообразия» действующих удостоверений личности, выдаваемых органами штатов, DHS предусмотрел, что уровень соответствия удостоверяющих документов основным минимальным стандартам Акта «Real ID», а также их совместимость с водительской лицензией должны быть достигнуты во всех штатах до 1 октября 2020 года, однако этот срок был продлен до 1 октября 2021 года в связи с пандемией коронавируса.

Таким образом, идентификация лиц в США основывается на законодательном акте федерального уровня – Акте «Real ID», создающем систему удостоверений личности, коррелирующих параметрам соответствия основным минимальным стандартам и совместимых с водительским удостоверением, что позволяет формировать единую национальную идентификационную базу данных Real ID, мониторинг которой осуществляет Департамент внутренней безопасности.

1.2. Для идентификации лиц в США применяются технологии распознавания лиц (Facial Recognition Technologies). Однако в настоящее время в США нет действующего федерального законодательного акта, регулирующего порядок и основания использования технологий распознавания лиц. На федеральном уровне существуют лишь акты рекомендательного характера. Так, два органа федерального уровня, а конкретно, Федеральная торговая комиссия (Federal Trade Commission, FTC) и Департамент по телекоммуникациям и информации министерства торговли США (NTIA), приняли Руководящие принципы порядка применения технологии распознавания лиц (в 2012 и в 2016 году соответственно). Названные органы исходили из того, что сбор и анализ идентификационных данных лиц, осуществляемый посредством использования систем распознавания лиц, невозможен без предварительного согласия идентифицируемых лиц.

Отсутствие федерального законодательного регулирования использования систем распознавания лиц не препятствует их «практическому» применению, в частности, Федеральным бюро расследований (Federal Bureau of Investigation, FBI), Управлением по иммиграции и таможне (Immigration and Custom Enforcement, ICE), Управлением транспортной безопасности (Transport Security Administration, TSA). На уровне штатов системы распознавания лиц также применяются правоохранительными органами, несмотря на то, что конституции большинства штатов гарантируют право на неприкосновенность частной жизни, гражданские права и свободы.

Широкому использованию правоохранительными органами систем распознавания лиц способствовало применение программного обеспечения, разработанного компанией Clearview AI, которое основано на соответствующей выборке данных из основных платформ социальных сетей, иных идентификационных баз данных, позволяющих аккумулировать фотографии людей, загруженные на такие платформы.

1.3. Своеобразным «ответом» на существующие проблемы, связанные с нарушением конфиденциальности частной жизни, прав и свобод человека, отсутствием правовых оснований для «слежки и наблюдения за передвижением людей без их согласия» и т.д., стало внесение в Конгресс США в ноябре 2019 года «законопроекта, ограничивающего

использование технологии распознавания лиц федеральными органами государственной власти, а также в иных целях» (The Bill to limit the use of facial recognition technology by Federal agencies, and for other purposes)[6], кратко именуемого «Facial Recognition Technology Warrant Act of 2019» (Законопроект Facial Recognition 2019).

Законопроект Facial Recognition 2019 предусматривает, что технологии распознавания лиц для правоохранительных целей могут применяться исключительно на основании судебного ордера. Такой ордер может быть выдан максимум на 30 дней (с возможностью его 30-дневного продления), и правоохранительные органы обязаны минимизировать сбор, хранение и распространение информации, касающейся лиц, на которых не распространяется действие ордера. Если деятельность по наблюдению осуществлялась с нарушением ордера или если ордер был получен необоснованно, то идентифицируемое лицо (лицо, находящееся под наблюдением) вправе ходатайствовать об уничтожении собранной информации до начала разбирательства в суде.

Законопроект Facial Recognition 2019, кроме того, возлагает обязанность на судей сообщать в Административную канцелярию судов США (Administrative Office of the United States Courts) информацию относительно любого ходатайства о выдаче ордера (или отказа в выдаче ордера). В свою очередь, Административная канцелярия судов США несет ответственность за ведение соответствующего реестра, а также подотчётна в своей деятельности Сенату США и Палате представителей США. Административная канцелярия судов США должна публиковать ежегодный публичный отчет с изложением информации по ходатайствам о выдаче ордеров, с указанием статуса и личности запрашивающего должностного лица правоохранительного органа; личности судьи; условий, на которых ордер был выдан; используемой технологии распознавания лиц (оборудование, камеры, данные, используемые для идентификации), включая данные, связанные с адекватной или ошибочной идентификации лиц, сделанных в течение года.

Законопроект Facial Recognition 2019 предусматривает, что для ограничения ошибочных идентификаций лиц и мониторинга эффективности систем распознавания лиц руководители правоохранительных органов и учреждений, использующие системы распознавания лиц, совместно с Национальным институтом стандартов и технологий США (National Institute of Standards and Technology, NIST) обязаны проводить периодическое тестирование эффективности применяемых систем распознавания лиц, в том числе на предмет проверки «частоты» ошибок идентификации лиц, искажений показателей распознавания лиц на основе расы, пола и иных характеристик физических лиц.

1.4. Нельзя не упомянуть также еще один законопроект, обсуждаемый в настоящее время в Конгрессе США, именуемый «Акт об этическом использовании распознавания лиц» (The Ethical Use of Facial Recognition Act)[7] от 12 февраля 2020 года (Законопроект Ethical Facial Recognition 2020).

Законопроект Ethical Facial Recognition 2020 направлен на обеспечение защиты конфиденциальности частной жизни лиц и предотвращение использования «стремительно развивающихся технологий распознавания лиц и методов сбора данных, которые повышают риск избыточного наблюдения за лицами и чрезмерную полицейскую деятельность». Законопроект Ethical Facial Recognition 2020 предполагает введение моратория на использование технологии распознавания лиц федеральным правительством до тех пор, пока Конгресс США не примет законодательство, определяющее конкретные виды использования идентификационных данных лиц, а также соответствующий порядок и процедуру применения систем распознавания лиц. Законопроект Ethical Facial Recognition 2020 также предусматривает запрещение использования средств федерального бюджета, которые могут быть доступны органам власти (как федерации, так и штатов) и которые государственные органы могут инвестировать для закупки соответствующих технологий распознавания лиц. Примечательно, что целый ряд правил Законопроекта Ethical Facial Recognition 2020, к примеру, в части порядка использования данных из систем распознавания лиц правоприменительными органами только на основании ордера, выданного судом, коррелирует положениям Законопроекта Facial Recognition 2019.

Обобщая сказанное, отметим, что при отсутствии в настоящее время действующего федерального законодательства, регулирующего порядок применения технологий распознавания лиц, важен сам факт существования рассмотренных выше законопроектов, а также текущий процесс подготовки принятия этих законопроектов. Рассмотрение названных выше законопроектов в Конгрессе США уже имеет практический эффект, поскольку в ряде штатов США в настоящее время введен запрет на использование системы распознавания лиц в правоохранительных целях.

2. Система идентификации в Европейском союзе

Европейский союз (Евросоюз или ЕС), будучи международной межправительственной международной организацией интеграционного типа, обладает специфической внутриорганизационной структурой. Речь идет о «первичной» и «вторичной» правовой основе, которая обеспечивает единство функционирования интеграционного объединения ЕС. Равно как и в США, в Евросоюзе невозможно одномоментно «ввести» некую единую систему идентификации, поскольку каждая страна-член ЕС является суверенным государством. Акты вторичного законодательства ЕС, прежде всего директивы и регламенты, принятие которых направлено на создание основополагающего нормативно-правового фундамента взаимодействия государств-членов, не могут не учитывать различия, в том числе, в плане технологического развития стран-членов ЕС применительно к системам и средствам идентификации.

Практически во всех странах-членах ЕС используются системы цифровой идентификации и в них применяются национальные идентификационные карты (eID), выпущенные не только для граждан соответствующих стран-членов ЕС, но также и для иных категорий физических лиц, легитимно находящихся на территории этих стран. Несмотря на то, что лица, проживающие в странах-членах ЕС, обладают своими eID, в практическом плане нередки случаи, когда лица ЕС сталкиваются с препятствиями использования eID в

рамках ЕС (при устройстве на работу не по месту постоянного проживания, при пересечении границ в рамках ЕС и т.д.).

Решить комплекс вопросов, в том числе, связанных с упрощением идентификации лиц в странах-членах ЕС, обеспечить совместимость применяемых странами-членами ЕС систем идентификации в рамках Евросоюза, расширить возможность доступа к товарам и услугам для лиц в рамках Евросоюза и т.д. был призван акт вторичного законодательства, принятый в форме регламента, т.е. законодательного акта прямого действия. Речь идет о Регламенте (ЕС) № 910/2014 об электронной идентификации и доверительных услугах по электронным сделкам на внутреннем рынке (Regulation (EU) 910/2014 on electronic identification and trust services for electronic transactions in the internal market) от 23.07.2014 г. (Регламент eIDAS)[8], который применяется с 1 июля 2016 г.

Регламент eIDAS направлен на повышение доверия к электронным транзакциям на внутреннем рынке Евросоюза путем обеспечения общей основы для безопасного электронного взаимодействия между гражданами, предприятиями и государственными органами, для повышения эффективности государственных и частных онлайн-услуг, электронного бизнеса и электронной торговли в Евросоюзе.

С одной стороны, Регламент eIDAS исходит из того, что государства-члены остаются свободными в использовании или внедрении средств для целей электронной идентификации для доступа к онлайн-услугам; обладают возможностью решать, следует ли привлекать частный сектор к предоставлению такого рода средств; не обязаны уведомлять Европейскую комиссию ЕС (Еврокомиссия) о своих электронных системах идентификации. Государства-члены принимают самостоятельно решение о том, уведомлять ли Еврокомиссию обо всех, некоторых или ни одной из электронных схем идентификации, используемых на национальном уровне для доступа, по крайней мере, к государственным онлайн-службам или конкретным услугам.

С другой стороны, необходимо принимать во внимание, что в большинстве случаев граждане ЕС не могут использовать свои электронные идентификационные данные, полученные в стране-члене ЕС, для аутентификации в другом государстве-члене ЕС, поскольку национальные электронные идентификационные системы и средства их страны не признают иные системы и средства идентификации других государств-членов ЕС. Существующий «электронный барьер» не позволяет поставщикам услуг в полной мере пользоваться преимуществами внутреннего рынка. Сказанное делает понятным, что принятие Регламента eIDAS было вызвано необходимостью обеспечить механизм и условия взаимного признания электронных средств идентификации, которые способствовали бы трансграничному предоставлению многочисленных услуг на внутреннем рынке и позволили бы предприятиям работать на трансграничной основе, не сталкиваясь со многими препятствиями во взаимодействии с государственными органами. Соответственно, ключевым принципом, из которого исходит Регламент eIDAS, является взаимное признание государствами-членами ЕС электронных систем и средств идентификации, существующих в государствах-членах ЕС.

В рамках Евросоюза и стран, входящих в Европейское экономическое пространство, Еврокомиссия является центральным органом, координирующим все аспекты реализации Регламента eIDAS.

Еврокомиссия концентрирует свои усилия на решении вопросов, связанных с координацией управления идентификацией в ЕС на основе децентрализованной идентификации (Decentralised Identity, DID) или суверенной идентификации (Self-Sovereign Identity, SSI).

Фундаментом для возможности решения таких вопросов являются нормативные положения Регламента eIDAS, создающие «уровни доверия» взаимного признания государствами-членами электронных систем и средств идентификации. Регламент eIDAS предусматривает три различных уровня доверия (низкий, значительный и высокий):

- низкий уровень доверия означает, что степень уверенности в идентификации лица ограничена, поскольку технические средства не вполне обеспечивают контроль за возможностями лица неверно использовать «свою личность» или изменить «свою личность» - и существуют соответствующие риски;
- значительный уровень доверия означает, что степень уверенности в идентификации лица обеспечены техническими спецификациями, стандартами и процедурами, а также техническими средствами контроля, которые снижают риск неправомерного использования или изменения личности лица;
- высокий уровень доверия означает, что степень уверенности в идентификации лица не только обеспечена техническими спецификациями, стандартами и процедурами, а также техническими средствами контроля, которые снижают риск неправомерного использования или изменения личности, но также существуют технические возможности предотвратить злоупотребление или изменение личности лица.

Страны-члены Евросоюза вправе самостоятельно решать вопрос о том, разрешают ли они использование электронных средств идентификации для онлайн-доступа в своей юрисдикции к своим общедоступным услугам, если уровень доверия электронных средств идентификации другой страны-члена ЕС «ниже» или «равен» уровню доверия, предусмотренного Регламентом eIDAS. Этот подход подчеркивает, что управление идентичностью в цифровой среде базируется на децентрализованной идентификации, когда лица (физические/юридические) имеют возможность создавать и контролировать свою личность, не полагаясь на какую-либо централизованную власть.

В этой связи следует отметить, что Регламент eIDAS:

а) предусматривает гарантии для лиц (физических/

юридических) в отношении их права использовать национальную электронную систему идентификации (eID) для доступа к государственным услугам в других странах ЕС, где доступны eID;

б) создает (в рамках ЕС и Европейской экономической зоны) европейский внутренний рынок для трансграничных электронных трастовых услуг и гарантирует, что они обладают таким же юридическим статусом и порождают те же правовые последствия, что и традиционные офлайновые трастовые услуги. В этом смысле принципиально важно, что Регламент eIDAS определяет содержательные характеристики понятия «электронный документ» (electronic document), означающего любой контент, хранящийся в электронной форме, в частности, текстовую, звуковую, визуальную или аудиовизуальную запись, что распространяется на «программные блоки» в блокчейне[9].

Услуги идентификации помогают проверить личность физических и юридических лиц, взаимодействующих в Интернете, а также подлинность удостоверительных документов, предъявленных в электронной форме. В этой связи на внутреннем рынке Евросоюза при заключении сделок в электронной форме (в т.ч. трансграничных), согласно Регламенту eIDAS, предусмотрено использование следующих средств электронной идентификации: электронные подписи (Electronic signatures), электронная печать (Electronic seals), штамп времени (Time stamps), электронная зарегистрированная служба доставки (Electronic registered delivery), аутентификация сайта (Website authentication).

Практический аспект применения средств электронной идентификации, предусмотренных Регламентом eIDAS, можно проиллюстрировать двумя примерами. Первый пример. Студент из Германии хочет продолжить обучение в университете Испании. Идентифицируя себя по своему немецкому eID, студент отправляет на веб-сайт университета заявку о зачислении, которую он подписывает своей электронной подписью (E-signature). Через электронную зарегистрированную службу доставки (E-registered delivery) студент направляет в адрес университета Испании имеющиеся документы об образовании в «оцифрованной» форме. В свою очередь, также через электронную зарегистрированную службу доставки, университет подтверждает факт получения заявки, соответствующих документов об образовании и их регистрацию, а также уведомляет об этом студента.

Другой пример. Гражданин Италии переезжает во Францию для работы по контракту. За соответствующий период времени итальянский гражданин должен подать налоговую декларацию во Франции. Налоговую декларацию итальянский гражданин может подать посредством регистрации на веб-сайте французской налоговой службы по своему итальянскому eID и подписать декларацию своей электронной подписью. Французский налоговый орган заверит электронной печатью (E-seal) уплату налоговых платежей и отправит заверенную налоговую декларацию в адрес итальянского гражданина через электронную зарегистрированную службу доставки (E-registered delivery).

В настоящее время большинство стран-членов уже ЕС обменялись уведомлениями о тех национальных системах идентификации, которые дают возможность получить широкий трансграничный доступ к услугам общего пользования в рамках Евросоюза.

В перспективе «завтрашнего дня» объем электронных транзакций будет умножаться, и в этом плане нормативно-правовая основа, созданная Регламентом eIDAS, способна обеспечить единую для ЕС основу для безопасного электронного взаимодействия между лицами (физическими и юридическими) и государственными органами, что позволит повысить эффективность государственных и частных онлайн-услуг, электронного бизнеса и электронной торговли в ЕС.

3. Международные межправительственные организации

Наряду с компетентными органами государств, в разработке и мониторинге систем идентификации физических лиц принимают участие международные межправительственные организации, в числе которых отметим следующие.

3.1. Управление Верховного комиссара ООН по делам беженцев (The United Nations High Commissioner for Refugees, UNHCR). Названный орган (Управление UNHCR) в рамках системы ООН осуществляет деятельность, связанную с оказанием помощи беженцам. С учетом того, что на начало 2020 года в мире насчитывалось более 25,9 млн беженцев и 3,5 млн лиц, запросивших убежище, едва ли стоит объяснять значение мероприятий Управления UNHCR, связанных с решением проблемы идентификации беженцев, что, прежде всего, сопрягается с необходимостью выдачи официального удостоверения личности беженца.

Идентификация беженцев отличается спецификой по следующим параметрам. Во-первых, многие беженцы не имеют удостоверений личности, когда они оказываются в принимающем государстве, поскольку их документы потеряны, уничтожены, нечитаемые и т.д. Во-вторых, беженцы могли вообще не иметь изначально удостоверений личности, поскольку в их стране не существует системы регистрации или унифицированной системы учета, либо они были дискриминированы (усечены в правах и т.д.) и по этим основаниям не внесены в соответствующие официальные системы идентификации страны происхождения. В-третьих, по общему правилу, власти страны пребывания беженцев связываются с властями страны «происхождения беженца» для проверки личности беженца без получения согласия беженцев, для минимизации потенциальных негативных рисков, которые могут последовать от конкретного беженца к органам страны, предоставляющей убежище.

Обеспечение идентификации беженцев требует мониторинга и согласованности на разных уровнях, включая управление рисками и формирование личности беженца в его потенциально новом «статусе». Международная стандартизация выдачи удостоверения личности беженцев требует комплексных мероприятий, сбора информации и доказательств как из страны происхождения беженца, так и из страны, предоставляющей убежище.

Государства, принимающие беженцев, взаимодействуют с Управлением UNHCR для получения технической помощи на основании документа «Регистрация населения и использование экосистемы идентификации Управления Верховного комиссара ООН» (UNHCR's Population Registration and Identity Management Ecosystem, PRIMES)[10]. Этот документ (PRIMES UNHCR) государства применяют для национальной системы цифровой идентификации беженцев.

Инструментарий PRIMES UNHCR нацелен на соответствие стандартным цифровым параметрам для идентификации беженцев, и в некоторых случаях Управление UNHCR может осуществлять часть функций «от имени принимающего государства». Так, к началу 2019 года биометрическая идентификация лиц (более 8 млн человек в 63 странах мира) была осуществлена по системе PRIMES UNHCR. Беженцы, которые зарегистрированы по биометрической идентификации лиц по системе PRIMES UNHCR, могут получить удостоверение личности, выданное Управлением UNHCR, что, в свою очередь, дает возможность беженцам получать доступ к различным услугам в стране пребывания.

3.2. Международная организация гражданской авиации (International Civil Aviation Organization, ICAO) является международной межправительственной организацией, обладающей статусом специализированного учреждения ООН, предметная сфера которой связана с обеспечением безопасного и надлежащего развития международной гражданской авиации[11].

В стремительно растущем и развивающемся объеме цифровых транзакций, увеличении потока авиаперевозок (внутренних и международных) и т.д. едва ли можно представить себе будущее, в котором авиапассажиры будут предъявлять «обычный» паспорт. По последним данным ICAO, в настоящее время 139 государств выпустили в обращение более одного миллиарда электронных паспортов (e-Passport), т.е. паспортов, которые содержат биографические и биометрические данные лица, хранящиеся в чипе/микросхеме (IC/Integrated Circuit). Применение e-Passport упрощает формальности при идентификации пассажиров, повышает безопасность пограничного контроля и т.д., что в целом оптимизирует весь цикл перевозки пассажиров в аэропортах.

Сказанное делает понятным значение Программы идентификации путешественников (Traveler Identification Program,TRIP), которая была инициирована Международной организацией гражданской авиации и продвижение которой ICAO осуществляет в сотрудничестве с целым рядом иных международных организацией, включая неправительственные[12]. В рамках ICAO на постоянной основе действует Техническая консультативная группа по машиночитаемым проездным документам (Technical Advisory Group on Machine Readable Travel Documents, TAG/MRTD), компетенция которой связана с разработкой спецификаций проездных документов, в том числе связанных с новыми технологиями. Деятельность TAG/MRTD осуществляется посредством различных рабочих групп, в числе которых следует назвать Рабочую группу по новым технологиям (New Technologies Working Group, NTWG), а также действующую в ее рамках специализированную подгруппу по цифровым проездным идентификационным документам (Digital Travel Credentials, DTC).

Специализированная подгруппа DTC объединяет представителей государств, а ее деятельность поддерживается международной организацией по стандартизации (International Organization for Standardization, ISO). Мандат специализированной подгруппы DTC направлен на разработку технических спецификаций для формирования международных политик по выпуску «цифровых идентификационных форм документов», использование которых облегчит прохождение проверок на транспорте, будет содействовать безопасному и ускоренному прохождению контрольно-пропускных пунктов авиапассажирами и т.д.

Специализированная подгруппа DTC, в частности, предложила создать эталон идентификационного документа (digital travel credential DTC), выпущенного как токен (token), который: а) является производным и совместимым с e-Passport (т.е. содержит данные, извлеченные из официально выданного e-Passport) и/или б) может быть выпущен в некоторых случаях «параллельно» с e-Passport, либо «заменять» e-Passport.

Эталон идентификационного документа DTC предполагает наличие в нем идентификационных данных лица (фотография лица (изображение), данные держателя документа, «метки безопасности» и др.), т.е. тех же данные, что и e-Passport, что в практическом плане означает атрибутивную совместимость всех генераций эталона идентификационного документа DTC и выпущенных e-Passport. Окончательное утверждение идентификационного документа DTC, совместимого с e-Passport, было запланировано к концу 2020 года, однако пандемия Covid-19, как представляется, внесет свои коррективы.

Логично возникают вопросы следующего свойства: зачем нужно внедрять эталон идентификационного документа DTC, если есть e-Passport, а идентификационный документ DTC, по сути, выполняет те же функции, что и e-Passport, в контексте подтверждения личности путешественника? Кратко и оставляя в стороне детали, ответить на эти вопросы можно следующим образом. Во-первых, идентификационный документ DTC является своеобразным «следующим шагом» стандартизации цифровых проездных документов, а это, безусловно, будет развивать непрерывность, связанность и полноту «цикла» перевозки пассажиров в целом. Во-вторых, многообразие существующих и формирующихся систем электронного удостоверения личности в различных государствах не должно препятствовать аэропортам и авиакомпаниям использовать автономные системы, позволяющие, в частности, упрощать заблаговременную регистрацию пассажиров на рейс, расширять возможности самообслуживания пассажиров в терминалах, ускорять проверку пассажиров при прохождении ими формальностей в контрольно-пропускных пунктах выезда и въезда. В-третьих, оптимизировать оценку безопасности и/или рисков на основе «совмещения» идентификационных данных пассажиров.

* * *

Параметры определения подлинной (действительной) идентичность физических лиц (цифровая идентичность), как представляется, приобретают решающее значение, а это одновременно порождает необходимость диверсификации регулирования применения систем идентификации, формирующихся и действующих в различных сферах социальной жизни.

Ссылки

1. Указ мэра Москвы от 10.04.2020 № 42-УМ «О внесении изменений в указ Мэра Москвы от 5 марта 2020 г. N 12-УМ». СПС «КонсультантПлюс»

2. Посредством QR-кода можно зашифровать любую информацию (числительные данные, текстовые и буквенные фрагменты, ссылки на интернет-сайты и т.д.). Для QR-кода существует лишь ограничения числа символьных обозначений (не более 4000 символов). Подробнее о QR-code см., например: https://www.qrcode.com/en/

3. См., например, указ мэра Москвы от 05.03.2020 № 12-УМ (ред. от 19.03.2020) «О введении режима повышенной готовности». СПС «Консультант Плюс»

4. Автономия американских штатов в рамках Федерации основана на 10-й поправке к Конституции США, которая гласит: «все полномочия, не делегированные Соединенным Штатам Конституцией или не запрещенные им штатами, принадлежат, соответственно, штатам или народу». СПС «Консультант Плюс»; а также см., например, URL: http://hrlibrary.umn.edu/education/all_amendments_usconst.htm

5. Далее по тексту статьи при упоминании штата также имеется в виду и Округ Колумбия.

6. *The Bill to limit the use of facial recognition technology by Federal agencies, and for other purposes* https://www.jurist.org/news/2019/11/bill-to-restrict-police-use-of-facial-recognition-introduced-in-us-senate/

7. *The Ethical Use of Facial Recognition Act* https://www.congress.gov/bill/116th-congress/senate-bill/3284/text

8. *Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC.* Официальный журнал Европейского союза. Право. 257/73, 2014

9. Регламент eIDAS подробно определяет четыре десятка терминов, связанных со средствами идентификации.

10. *UNHCR's Population Registration and Identity Management Ecosystem, PRIMES* https://www.unhcr.org/blogs/data-millions-refugees-securely-hosted-primes/

11. Подробнее см., например, Международное право = Volkerrecht / Вольфганг Граф Витцтум, М. Боте, Р. Дольцер и др.; пер. с нем. Т. Бекназара, А. Насыровой, Н. Спица; науч. ред. Т.Ф. Яковлева; пред., сост. В. Бергманн. М.: «Инфотропик Медиа», 2011. СПС «Консультант Плюс». *Digital travel credentials: Looking beyond the book.* https://www.unitingaviation.com/news/security-facilitation/digital-travel-credentials-beyond-the-book/

12. Так, в рамках Симпозиума ICAO в рамках Программы идентификации пассажиров 2020 (TRIP2020) запланировано провести первый совместный форум ICAO/Интерпол по обмену данными о пассажирах. Сроки проведения Симпозиума ICAO перенесены в связи с пандемией Covid-19 на сентябрь 2020 г. https://www.icao.int/Meetings/TRIP-Symposium-2020/Pages/default.aspx. Деятельность ICAO осуществляется в тесном взаимодействии с международной неправительственной организацией по стандартизации (*International Organization for Standardization, ISO*). https://www.icao.int/Newsroom/Pages/ICAO-technical-co-operation-programme-achieves-ISO-certification.aspx

Веб-ориентированная обманная система для выявления хакеров

Андрей Вишневский, Петр Ключарёв

Хакерские атаки наносят существенный вред современной цифровой экономике. Традиционные методы защиты не всегда эффективны, а зачастую не работают вообще. Для обнаружения актуальных компьютерных атак все более широкое применение находят технологии искусственного интеллекта, как, например, построение поведенческих портретов посетителей веб-сайтов, анализ больших данных для поиска закономерностей между категориями веб-контента и содержанием вредоносного кода, глубинный анализ сетевого трафика для выявления неочевидных признаков атак и повышение интерактивности обманных систем с помощью сочетания машинного обучения и теоретико-игровых моделей. Обсуждаемая в данной статье обманная система генерирует веб-сайты, способные ввести нападающего в заблуждение и завлечь его в ловушку, убедив, что перед ним настоящий информационный ресурс. Текст, изображения и ссылки на этих веб-сайтах подбираются таким образом, чтобы по списку посещенных страниц легко было отличить поискового робота от человека и обычного пользователя от хакера. Опыт использования системы показывает, что подобные подходы позволяют существенно повысить качество работы средств защиты информации.

В настоящее время Интернет стал ареной, на которой непрерывно идет информационное противостояние. Цена информации высока, и поэтому те, кто защищает компьютерные ресурсы, стараются как можно лучше изучить приемы нападающих. Этого они добиваются введением в контур систем защиты информации обманных систем (иногда также используется русскоязычный термин «ложные системы» или англоязычный – «honeypots»), целью которых является вовлечение нарушителя в своего рода «игру», увеличивая тем самым время, необходимое на обход средств защиты. Работа обманных систем заключается в том, что они эмулируют уязвимости программного обеспечения, которых в реальности не существует.

Для проверки ряда гипотез нами была создана обманная система, классифицирующая посетителей веб-сайтов на основе запрошенных ими веб-страниц. Нами была поставлена цель – при помощи обманной системы отделить атакующих как от поисковых роботов, так и от обычных пользователей, и составить белые и черные списки IP-адресов.

На наш взгляд, актуальные черные и белые списки IP-адресов важны для современных средств защиты информации. Песочницы (виртуальные или физические машины, эмулирующие компьютеры потенциальных жертв) и другие инструменты динамического анализа обнаруживают вредоносную активность по обращениям к IP-адресам с плохой репутацией. Антивирусы и средства статического анализа могут извлекать IP-адреса из сканируемых файлов и выдавать вердикт на основе сведений об этих адресах.

Межсетевые экраны, системы по обнаружению и предотвращению вторжений и другие инструменты для отслеживания сетевого взаимодействия могут в своей работе опираться на списки известных IP-адресов источников атак. Черные и белые списки IP-адресов весьма востребованы в индустрии информационной безопасности. Их можно купить либо составить самостоятельно. Чтобы накопить черные и белые списки сетевых адресов своими силами, мы заложили в обманную систему следующую логику. В черный список должны были попадать IP-адреса атакующих обманную систему. Сетевые адреса поисковых роботов, которые безопасно считывают контент, должны были сохраняться в белом списке. Для классификации посетителя обманная система определяла три поведенческих признака:

- Во-первых, рассчитывалось число обращений посетителя веб-сайта к несуществующим страницам.
- Во-вторых, определялось число веб-сайтов, на которые зашли с одного и того же IP-адреса.
- В-третьих, фиксировалось число таких переходов по ссылкам, которые показывали, что посетитель понимает смысл текста на сайте.

Для такой проверки были созданы специальные веб-сайты, контент, размещенный на которых, представлял собой иллюстрированные самоучители по иностранным языкам с тестами на проверку знаний посетителя.

Поисковые роботы, как правило, считывают контент

веб-сайтов с одного и того же IP-адреса и не могут сравниться в понимании семантики текста с человеком. Хакеры же в свою очередь используют автоматизированные сканеры уязвимостей, которые проверяют наличие определенных страниц на множестве сайтов. В предложенной обманной системе таких уязвимых страниц не было.

В качестве примера ниже приведены обнаруженные запросы атакующих к несуществующим потенциально уязвимым страницам входа в административные панели WordPress и Bitrix (атакующие собирают ссылки на страницы авторизации для последующего перебора паролей):

<IPaddress1> - - [31/Jan/2017:02:06:43 +0100] "GET /admin/ HTTP/1.1" 404 445 "-" "Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; MRSPUTNIK 2, 4, 0, 270; .NET CLR 1.1.4322; .NET CLR 2.0.50727; .NET4.0C)"

<IPaddress2> - - [27/Jun/2017:06:09:01 +0200] "GET /bitrix/admin/index.php HTTP/1.1" 404 517 "http://italian-cards.ru/bitrix/admin/index.php" "Mozilla/5.0 (compatible; uCrawlr/1.0 ; +https://blog.ucoz.ru/upolicy)"

На каждом веб-сайте в разработанной обманной системе был заранее известен список выложенных файлов. В ходе анализа файлов журнала веб-сервера обращения к несуществующим файлам легко выявлялись и считались признаками аномального поведения. Все веб-сайты в обманной системе были созданы по одному шаблону и отличались только доменным именем и языком контента. Доменные имена были зарегистрированы в один день, поэтому поисковые роботы, зашедшие на один из сайтов, как правило, заходили и на остальные. Тесты, содержавшиеся на сайтах, были написаны на чистом HTML, чтобы упростить анализ журналов веб-сервера. На рисунке 1 показан пример тестового задания и реакции на неверный ответ.

В результате эксперимента была зафиксирована активность роботов известных поисковых систем и поведение автоматизированных хакерских утилит. Были выявлены пути к файлам, в которых злоумышленники ожидали найти уязвимости. Поведение реальных пользователей

Рис. 1. Пример тестов на знание английского языка.

отличалось от активности автоматизированных утилит как по количеству ошибок в тестах, так и по языковым предпочтениям.

Все собранные IP-адреса были размечены по предполагаемым странам-источникам с помощью сервиса геолокации Maxmind. Это позволило выявить страны, из которых чаще всего осуществлялись атаки на наши сайты-сенсоры в доменной зоне .ru, и страны, из которых проводились наиболее мощные атаки.

Эволюция технологии отслеживания веб-приложениями поведенческих признаков

Поведенческие признаки веб-сайты начали собирать уже достаточно давно, но с каждым годом методы выявления аномальной активности становятся все более изощренными и узкоспециализированными. В начале 2010-х поведенческие портреты посетителей веб-сайтов строились на основе анализа интересов пользователя и его манере просмотра веб-контента. Если пользователь посещал сайты определенных банков, политических организаций и интернет-магазинов, это становилось дополнительным положительным признаком при аутентификации, т.е. повышало доверие к нему[1].

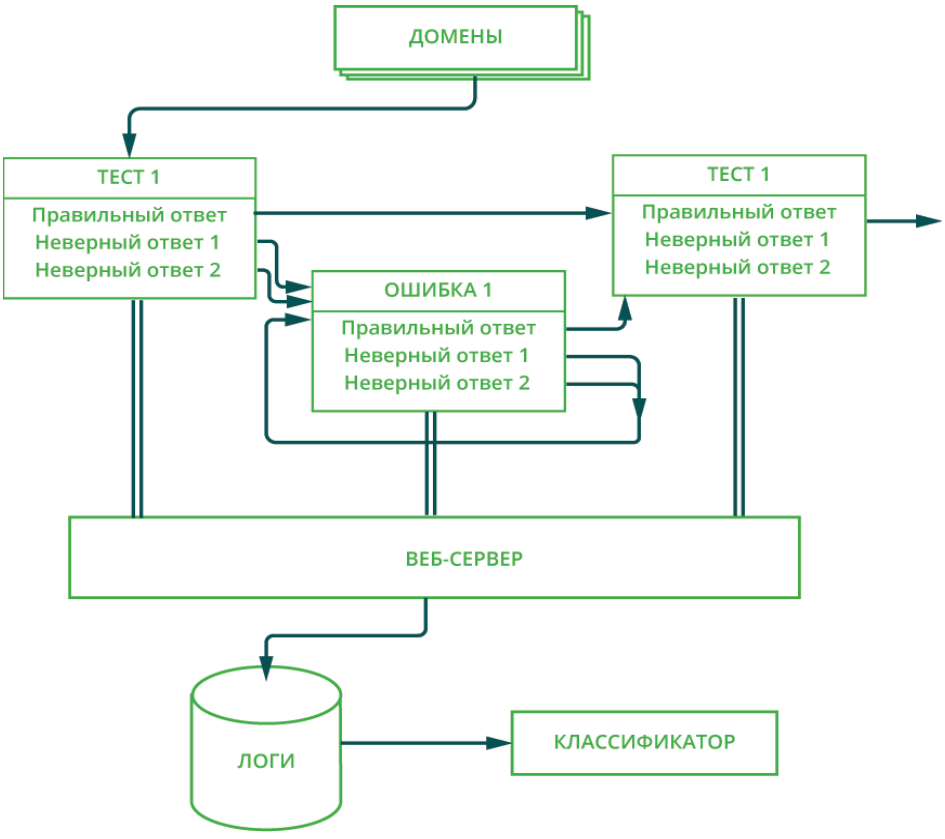
С развитием технологий искусственного интеллекта списки посещенных веб-страниц стали обрабатываться с помощью алгоритмов машинного обучения, которые могут распознать, какие шаблоны поведения характерны для хакеров, а какие – для обычных пользователей[2]. Успешное применение такого статистического обучения защитных систем спровоцировало интерес к анализу больших данных о поведении пользователей Интернета. В частности, по полученным от антивирусной компании Symantec историям веб-запросов 160 тысяч пользователей компьютеры были классифицированы на подверженные риску заражения и находящиеся в безопасности. Классификация опиралась на количество веб-запросов пользователя, время наибольшей активности, разнообразие посещаемых веб-сайтов и их категории, тип устройства, с которого пользователь выходил в Интернет, популярность посещенных веб-страниц, устойчивость списка посещаемых интернет-ресурсов. Было показано, что чем больше различных сайтов посещает пользователь и чем чаще он посещает веб-сайты с агрессивным контентом, тем выше риск заражения его компьютера[3].

Затем для выявления аномального поведения стали использоваться тонкие методы сравнения интересов конкретного пользователя с интересами большинства[4]. В погоне за точностью идентификации подозрительного посетителя были предложены эффективные методы деанонимизации браузеров, выполняющихся в режиме конфиденциального просмотра (например, режим инкогнито в браузере Chrome и режим приватного просмотра в браузере Mozilla Firefox), и методы обнаружения пользователей браузера Tor[5,6], опубликованные в 2015, 2016 годах. В качестве признаков на вход нейронной сети подавали скорость передачи данных, время отклика, джиттер и число потерянных сетевых пакетов.

Современные обманные системы собирают информацию об атаках на веб-приложения, сохраняют вредоносные файлы, которые злоумышленники пытаются распространить, эксплуатируя различные уязвимости (например, PHP-инъекции и SQL-инъекции)[7]. Разработаны высокоинтерактивные обманные системы, замаскированные под SSH-серверы, поведение которых определяется теоретико-игровыми моделями и машинным обучением[8]. Такие ловушки обучаются на обнаруженных опасных объектах и предшествовавших действиях самой защитной системы.

В последние годы активно стало развиваться научное направление по сочетанию машинного обучения и теории игр в обманных системах. Начало этой ветви исследований положили аналитические расчеты оптимальных стратегий для злоумышленника и обманной системы. Например, в статье[9] авторы теоретически вывели оптимальные стратегии для игровой модели, в которой злоумышленник владеет бот-сетью, т.е. множеством подконтрольных ему зараженных машин, а обманная система либо разрешает, либо блокирует атаку.

Рис. 2. Структурная схема обманной системы.



В 2012 году в работе[10] была предложена теоретико-игровая модель размещения обманных систем в социальных сетях. Обманные системы в этой модели базируются на поддельных месторасположениях на карте в основанной на геопозиционировании социальной сети. Злоумышленники стремятся набрать бонусы в такой социальной сети. Для этого они отмечаются в тех местах, которые они на самом деле не посещали. Был предложен подход к обнаружению таких нечестных пользователей, опирающийся на размещение поддельных мест, где можно отметить свое присутствие.

В работе[11] того же года с помощью теоретико-игрового подхода была выведена стратегия выбора информационных ресурсов, под которые рационально маскировать обманные системы.

В 2015 году в работе[12] была описана теоретико-игровая модель, в которой стратегии представлены в виде графов атак. Графы атак содержат все известные векторы угрозы, с помощью которых можно скомпрометировать систему. С помощью предложенной модели можно выбирать оптимальное множество ловушек внутри сети, такое, что атакующий будет стремиться атаковать наименее ценный информационный ресурс.

В 2016 году была предложена[13] обманная система из поддельных профилей в социальных сетях для того, чтобы обнаруживать взломанные анкеты. Для выработки стратегии были использованы сигнальные игры и расчет равновесия по Нэшу.

В 2017 году в работе[14] была предложена идея мимикрии обманной системы. В рамках предложенной модели сторона защиты имеет в распоряжении реальные сервисы, обманные системы и реальные системы, замаскированные под обманные. Защитная система стремится отличить легитимного пользователя от атакующего. В качестве потенциально уязвимых сервисов был использован FTP-сервер.

Таким образом, в ранее опубликованных исследованиях сказано о том, что для составления поведенческого портрета веб-пользователя использовались данные с сетевого уровня, сведения о веб-браузере и его расширениях, а также история запросов, интересы и характер взаимодействия пользователя с веб-сайтом. В упомянутых статьях сказано, что такие поведенческие признаки полезно собирать для обнаружения хакерских атак. Следовательно, на этих признаках можно строить обманные системы, накапливающие черные и белые списки IP-адресов.

Существующие применяемые на практике обманные системы, как правило, обнаруживают атаки на один информационный ресурс, опираясь на события, связанные только с этим защищаемым ресурсом. Кроме того, существующие обманные системы не используют контент при анализе поведения потенциальных атакующих. Наше исследование показало, как с помощью достаточно простой обманной системы с жестко запрограммированной логикой можно собирать актуальные сетевые адреса атакующих, которых нет в базах знаний наиболее известных и авторитетных зарубежных интернет-порталов, посвященных информационной безопасности. Для этого был разработан оригинальный метод, описанный в следу-

ющем разделе. Следует также отметить, что накопленные с помощью этого метода данные можно использовать, в том числе, и для проверки более сложных и интеллектуальных средств защиты информации.

Структура веб-ориентированной обманной системы

Нами была разработана и введена в эксплуатацию обманная система, в состав которой входил набор из двадцати доменных имен в зоне .ru, привязанных к сайтам на одном выделенном сервере. На сервере был установлен HTTP-сервер Apache и выложены HTML-файлы с текстом и изображениями. Текст страниц на каждом из двадцати веб-сайтов соответствовал названию домена и был посвящен изучению одного из иностранных языков. Например, в доменном имени сайта-самоучителя по корейскому языку присутствовала подстрока korean. Полные имена сайтов, как и сетевые адреса посетителей веб-сайтов, не публикуются, чтобы не разглашать персональные данные.

На каждом сайте была создана главная страница и страницы с тестами на знание иностранных языков. В тестах нужно было выбрать из трех слов то, которое соответствовало изображению. Пример показан на рисунке 1. Клик по неверному ответу приводил на страницу с ошибкой. Переход на страницу с ошибкой так же, как и выбор правильного ответа, фиксировался в файле журнала веб-сервера Apache. Программа, написанная на языке программирования Python, анализировала полученные журналы HTTP-запросов и классифицировала IP-адреса посетителей. На рисунке 2 показана схема разработанной обманной системы.

Принцип работы

Поскольку все внутренние ссылки на веб-сайтах в обманной системе вели на существующие страницы, попытки доступа к несуществующим страницам считались атаками. У обычного пользователя не было возможности кликнуть по ссылке так, чтобы была зафиксирована атака. Ложное срабатывание возможно было, только если пользователь неумышленно ошибется при ручном вводе названия веб-страницы.

Распознавание поисковых роботов было основано на количестве веб-сайтов, на которые зашли с одного и того же IP-адреса. Доменные имена сайтов в обманной системе были выбраны так, как будто это сайты для изучения арабского, китайского, английского, финского, французского, немецкого, греческого, итальянского, японского, корейского, литовского, норвежского, польского, португальского, испанского, турецкого, украинского языков, а также иврита, фарси и хинди.

Хотя людей, владеющих двумя языками, достаточно много[15,16], чаще всего они знают диалекты одного языка или государственные языки соседних стран. Например, в Европейском союзе людей, которые говорят одновременно на трех иностранных языках, не более 10%[17]. Поэтому решение обманной системы о том, что запрос исходит от поискового робота, принимался, если с IP-адреса было посещено хотя бы два подконтрольных сайта.

Способ сбора подозрительных IP-адресов

Анализатор разбивал файл журнала веб-сервера на группы запросов от каждого отдельного IP-адреса, затем вычислял число ошибок в тестах, число посещенных доменов среди двадцати подконтрольных обманной системе и определял наличие обращений к веб-страницам, отсутствовавшим на сайте.

При классификации поведения с помощью теста допускалось, что человек может ошибаться. Предполагалось, что обычный пользователь может допустить в тесте не более трех ошибок. Если допущено четыре и больше ошибок, считалось, что это похоже на поведение поискового робота, который переходит по всем ссылкам подряд.

Число допустимых ошибок было выбрано эмпирически на основании ручной проверки зафиксированных событий. Результаты классификации не изменялись либо изменялись незначительно при допуске больше трех ошибок, как видно из рисунка 3. Однако если усилить критерий до двух или одной допустимой ошибки, то десятки сетевых запросов от реальных пользователей неверно классифицировались как взаимодействие с поисковыми роботами.

Описание и результаты эксперимента

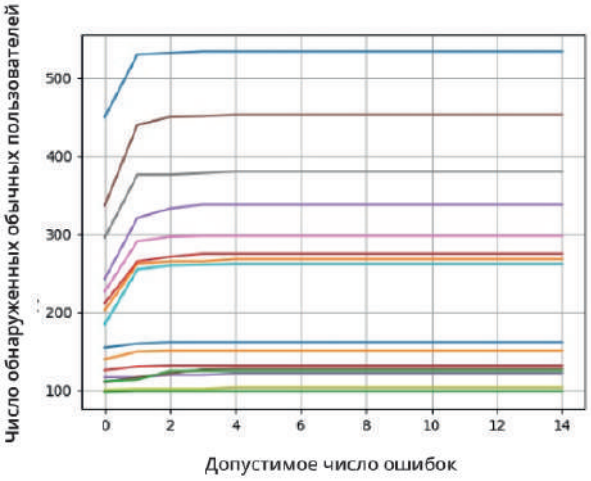
Журналы веб-сервера накапливались с января по декабрь 2017-го года. В них были зафиксированы запросы атакующих к потенциально уязвимым веб-страницам. Статистика геолокации IP-адресов показала основные тренды атак на домены в зоне .ru. Геоданные были получены с помощью сервиса Maxmind[18].

На рисунке 4 показаны распределения запросов к развернутым веб-сайтам без учета активности поисковых роботов. По горизонтальной оси отмечены коды стран, из которых, согласно базе данных Maxmind, исходили атаки. По вертикальной оси отложено число уникальных IP-адресов, с которых приходили запросы к ловушкам. Гистограммы запросов атакующих и обычных пользователей показаны разным цветом и текстурой на графиках. Общее число запросов к каждому сайту доходило до нескольких тысяч, но среди них нас интересовали только несколько десятков запросов, которые исходили со стороны злоумышленников и обычных пользователей.

Можно заметить, что распределение стран-источников атак отличается от распределения стран, из которых заходили обычные пользователи. Интерес русскоязычных пользователей к изучению иврита можно объяснить культурными связями, так же, как и востребованность материалов по немецкому языку среди жителей Нидерландов. С другой стороны, атаки с российских и украинских IP-адресов были распределены равномерно по всем сайтам, за исключением корейского языка.

Чаще всего атаки исходили с украинских IP-адресов, а самые продолжительные атаки осуществлялись с американских IP на домен, в название которого входила подстрока «korean». Вероятно, на тот момент активно исследовались корейские веб-ресурсы в связи с напряженной политической обстановкой в тихоокеанском регионе.

Рис. 3. Зависимости числа посетителей, классифицированных как обычные пользователи, от числа допустимых ошибок на каждом из двадцати опубликованных веб-сайтов.



Различными цветами на графике обозначены данные, собранные с разных сайтов.

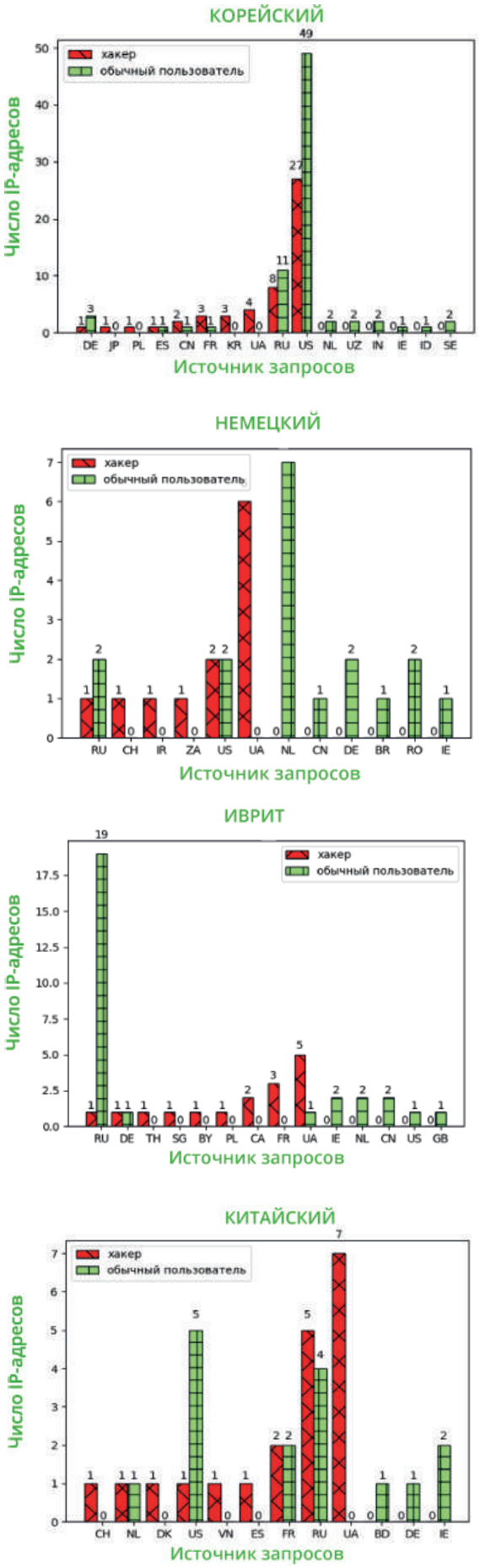
Классифицированные IP-адреса по двум странам из общего списка были проверены вручную и просканированы в авторитетном репутационном сервисе Virustotal[19]. Портал Virustotal объединяет в себе экспертные знания о вредоносных файлах, ссылках и IP-адресах, собранные в 67-ми крупнейших антивирусных компаниях мира. В инфраструктуры этих компаний входят в том числе и обманные системы, на основе которых и публикуются данные о подозрительных IP-адресах. Это крупнейший мировой агрегатор признаков вредоносных файлов и компьютерных атак, который принадлежит информационному гиганту – Google.

Среди вручную подтвержденных источников атак Virustotal предупредил об опасности только в 21% случаев, т.е. приблизительно четыре из пяти IP-адресов, с которых осуществлялись попытки подбора паролей, несанкционированной авторизации и эксплуатации уязвимостей в сайтах, расположенных в доменной зоне .ru, Virustotal посчитал безопасными.

На основании ручной проверки по двум странам были перепроверены результаты классификации. Точность классификации составила 94-97%, а полнота составила 97-98%. На точность негативно повлияли ложные срабатывания на безопасные обращения поисковых роботов к несуществующим страницам. Полноту обнаружения атак снизили неожиданные опасные сетевые запросы, в которых эксплуатация осуществлялась не через путь к запрашиваемой веб-странице, а через поле User-agent. Таким образом, созданная обманная система оказалась способной успешно обнаруживать источники неизвестных атак.

Описанная в статье обманная система позволяет с высокими показателями качества определять источники хакерских атак, которые не были известны крупным и авторитетным зарубежным репутационным сервисам. Лежащий в ее основе подход может позволить повысить качество отечественных средств информационной безопасности.

Рис. 4. Статистика атак на сайты об изучении корейского, китайского, немецкого языков и иврита.



Ссылки

1. Olejnik L., Castelluccia C. Towards Web-Based Biometric Systems Using Personal Browsing Interests. 2013 International Conference on Availability, Reliability and Security. Regensburg. 2013. P. 274-280. https://hal.inria.fr/hal-00917046/document

2. Zhong J., Yan C., Yu W., Zhao P., Wang M. A Kind of Identity Authentication Method Based on Browsing Behaviors. 2014 Seventh International Symposium on Computational Intelligence and Design. Hangzhou. 2014. P. 279-284. https://pdfs.semanticscholar.org/63co/ca25fb308a90dcc265fod3d5c8c087fedooe.pdf

3. Davide Canali, Leyla Bilge, and Davide Balzarotti. On the effectiveness of risk prediction based on users browsing behavior. In Proceedings of the 9th ACM symposium on Information, computer and communications security (ASIA CCS '14), https://www.nortonlifelock.com/content/dam/nortonlifelock/docs/research-papers/on-the-effectiveness-of-risk-prediction-based-on-users-browsing-behavior-en.pdf

4. Yu S., Guo S., Stojmenovic I. Fool Me If You Can: Mimicking Attacks and Anti-Attacks in Cyberspace. in IEEE Transactions on Computers. 2015. V. 64. N 1. P. 139-151. https://www.researchgate.net/publication/273396276_Fool_Me_If_You_Can_Mimicking_Attacks_and_Anti-Attacks_in_Cyberspace

5. Zhao B., Liu P. Private Browsing Mode Not Really That Private: Dealing with Privacy Breach Caused by Browser Extensions. 2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, Rio de Janeiro. 2015. P. 184-195. https://www.researchgate.net/publication/314033298_Private_Browsing_Mode_Not_Really_That_Private_Dealing_with_Privacy_Breach_Caused_by_Browser_Extensions

6. Ishitaki T., Oda T., Barolli L. A Neural Network Based User Identification for Tor Networks: Data Analysis Using Friedman Test. 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA), Crans-Montana. 2016. P. 7-13. https://pdfs.semanticscholar.org/1883/c16498aca6do6f6c9f2ad9d3b45784fa64e2.pdf

7. Glastopf Web Honeypot Project, Lukas Rist, Glastopf honeypot, https://www.honeynet.org/projects/old/glastopf/

8. Wagener, Self Adaptive High Interaction Honeypots Driven by Game Theory, в Conference: Stabilization, Safety, and Security of Distributed Systems, 11th International Symposium, SSS 2009, Lyon, France, 2009. http://www.foo.be/papers/sss09_wagener_state_dulaunoy_engel.pdf

9. Hayatle O., Otrok H.,Youssef A. A game theoretic investigation for high interaction honeypots // Proceedings of the 2012 IEEE International Conference on Communications (ICC), Ottawa, Canada. 2012. Pp. 6662-6667.

10. Pelechrinis K., Krishnamurthy P., Zhang K. Gaming the game: Honeypot venues against cheaters in location-based social networks [Электронный ресурс] // CoRR., 2012. URL: https://arxiv.org/pdf/1210.4517.pdf (дата обращения: 01.10.2017)

11. Pibil R., Lisy V., Kiekintveld C., Bosansky B., Pechoucek M. Game theoretic model of strategic honeypot selection in computer networks // Grossklags J., Walrand J. (Eds.) Decision and Game Theory for Security. GameSec 2012. Lecture Notes in Computer Science. 2012. Vol. 7638. Pp 201-220. DOI: 10.1007/978-3-642-34266-0_12

12. Kiekintveld C., Lisy V., Pibil R. Game-Theoretic Foundations for the Strategic Use of Honeypots in Network Security // Jajodia S., Shakarian P., Subrahmanian V., Swarup V., Wang C. (Eds.) Cyber Warfare. Advances in Information Security. Vol. 56. Pp. 81-101.

13. Mohammadi A., Manshaei M. H., Moghaddam M. M., Zhu Q. A Game-Theoretic Analysis of Deception over Social Networks Using Fake Avatars // Proceedings of the Decision and Game Theory for Security - 7th International Conference, GameSec 2016. 2016. Vol. 9996. Pp. 382-394.

14. Shi L., Zhao J., Jiang L., Xing W., Gong J., Liu X. Game theoretic simulation on the mimicry honeypot // Wuhan University Journal of Natural Sciences. 2016. Vol. 21. Pp. 69–74.

15. Ansaldo, A. I., Marcotte, K., Scherer, L., & Raboyeau, G. (2008). Language therapy and bilingual aphasia: Clinical Implications of psycholinguistic and neuroimaging research. Journal of Neurolinguistics, 21, 539-557.

16. De Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. Applied Linguistics, 13 (1), 1-24. https://www.researchgate.net/publication/249237191_A_Bilingual_Production_Model_Levelt's_'Speaking'_Model_Adapted

17. Europeans and their languages // Special Eurobarometer 386, 2012. URL: http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_386_en.pdf (дата обращения: 06.04.2018)

18. Maxmind. Базы данных GeoIP. URL:https://www.maxmind.com/ru/home (дата обращения: 14.03.2018)

19. VirusTotal. URL: https://www.virustotal.com/ (дата обращения: 16.03.2018)

Ротация ключей DNSSEC – как это делать при карантине, опыт и обсуждение в ICANN

Павел Храмцов

Большинство из наших читателей знакомо с термином «DNSSEC» и технологией, которая означается этим термином. Главная задача DNSSEC – обеспечить надежный обмен достоверной информацией в рамках Системы доменных имен Интернета, проще говоря, DNS.

Для обеспечения этой достоверности используются криптографические ключи, которые нужно время от времени менять.

Для корня системы DNS используется два ключа: ключ подписи ключей (Key Signing Key) и ключ подписи ресурсных записей зоны DNS (Zone Signing Key).

Процедура замены ключей называется процедурой ротации. KSK принято ротировать редко. Для зоны корня системы DNS это сделали только один раз, и то с определенными сложностями и задержками. Подробно об этом мы писали в ноябре 2017 года.

ZSK меняют гораздо чаще – четыре раза в год на протяжении последних десяти лет (ознакомиться с процессом ротации ключей можно в прямом смысле слова «воочию» на сайте PTI (IANA).

Но вот в 2020 году что-то с этой рутинной процедурой не задалось. Сначала 11 февраля 2020 года при подготовке к церемонии ротации ZSK во время проверки доступности пакетов с разделяемыми секретами сломался сейф. Сломался не виртуально, а натурально. Его не смогли физически открыть. Церемонию пришлось переносить с 12 февраля на более поздний срок. Ее провели 16 февраля 2020.

Но как оказалось, это были цветочки. Грянула эпидемия. Самолеты не летают, люди не перемещаются. Следовательно, необходимое для проведения церемонии ротации ключей количество криптоофицеров в американскую Вирджинию может просто (по разным объективным причинам) не приехать.

Церемония назначена на 23 апреля 2020 года. Но есть большие сомнения, что губернатор штата Вирджиния Ральф Нортэм (Ralph Northam) отменит свои распоряжения от 24 марта 2020 и 30 марта 2020, которые вводят существенные ограничения на перемещения и собрания до 10 июня 2020. Ограничения введены практически такие же, как в Москве и Московской области, с тем отличием, что бегать и ловить рыбу не запрещается (ознакомиться с этими распоряжениями можно на сайте губернатора). Правда, ситуация в Вирджинии, похоже, налаживается, согласно статистике Департамента здравоохранения штата.

В общем, не очень понятно, что теперь со всем этим - я имею в виду ротацию ключей - делать.

Директор PTI Ким Дейвис (Kim Davis) обратился к сообществу с письмом, в котором предлагает три варианта решения:

1. Провести церемонию, как и было запланировано.
2. Перенести церемонию на другую дату и провести ее только с участием американских криптоофицеров.
3. Провести церемонию по специальной процедуре (disaster recovery procedure), которая не требует физического присутствия криптоофицеров и проводится силами персонала PTI.

Сообщество высказалось в том духе, что любой из способов, отличающийся от первого, может вызвать сомнения в «прозрачности» процедуры ротации ключей.

Церемония была придумана для того, чтобы показать независимость ICANN от каких-либо государственных влияний. А тут, «понимаешь», «красная кнопка» в чистом виде!

Решение еще не принято, но PTI активно работает над приемлемой реализацией варианта №3.

На этом фоне сразу становятся актуальными рассуждения о децентрализованном корне - и на горизонте появляется Blockchain.

В общем, поживем – увидим. Ждать осталось не так и долго. Характерно, что на вебинаре, который прошел 9 апреля и был посвящен техническим вопросам работы DNS во время кризиса, о процедуре ротации не было сказано ни слова.

Инструменты удаленной работы и учебы (Zoom, Skype и все-все-все)

Рассуждая по поводу надежности инфраструктуры открытых ключей, Джефф Хьюстон (Geoff Huston) написал: «Нам нужна безопасная и вызывающая доверие

инфраструктура. Мы должны быть уверены, что сервис, с которым мы контактируем, подлинный, что транзакция защищена от соглядатаев, и что мы не оставляем следов». Это высказывание верно и в более широком контексте. Особенно сейчас, когда удаленная работа стала нормой, а физическое присутствие в офисах, учебных аудиториях и на конференциях если и возможно, то только по острой необходимости.

Существует много разновидностей инструментов организации работы в удаленном режиме. Мы коснемся только одного из них – видео-конференц-связи, или сокращенно ВКС.

ВКС – это не изобретение сегодняшнего дня. Инструменты для ее проведения существовали задолго до Эпидемии. Это и Skype, и TrueConf, и Zadarma, и, конечно, Google Hangouts. Однако самой популярной платформой стал в эпоху пандемии Zoom.us.

Раньше речь в контексте инструментов удаленной работы шла об Экономии: экономии на офисах, экономии на перелетах, экономии времени и о прочих других экономиях, т.е. о повышении эффективности работы.

Сейчас же акценты сместились. В настоящее время подчас просто нет другой формы производственного общения. Нет ВКС – нет работы, нет бизнеса, нет образования, да и много чего другого нет.

В этой связи надежность и безопасность ВКС в контексте, о котором говорил Хьюстон, как никогда важны.

Самый простой и быстрый способ реализовать ВКС – это обратиться к онлайн-сервису. Не нужно собственное оборудование, не нужен специально обученный персонал. Завел корпоративный аккаунт - и Voi la! Тем более, во время кризиса это все еще и бесплатно!

Но, как и следовало ожидать, простых решений, которые удовлетворяли бы современным требованиям информационной безопасности, не бывает. Это лишний раз показала история с zoom.us.

По уверениям главы и основателя Zoom Эрика Юна (Eric Yuan), сервис, рассчитанный на десять миллионов пользователей, во время пандемии обслуживает 200 миллионов. Это привело к проблемам с безопасностью – в Сеть утекли записи телеконференций, а кроме того, стало понятно, что несмотря на все уверения, Zoom не использует сквозное шифрование.

Это привело к запрету использования Zoom в Google, SpaceX, министерстве образования Нью-Йорка, Сенате США и правительстве Тайваня.

Любопытно, что в России никаких официальных запретов на Zoom не введено. И им продолжают пользоваться в различных сферах, хотя отечественные разработки, подобные Zoom, есть.

Насколько чувствительная информация попадает в эту систему и кто и как ей воспользуется, покажет время. А пока системные администраторы активно продвигают zoom.us, а ответственные за информационную безопасность с ним борются.

Вавилонская башня Universal Acceptance

В 2010 году был принят документ RFC-5891, который определил правила использования доменных имен сети Интернет на национальных языках в приложениях. Идея применения имен доменов на национальных языках проистекает из того факта, что на английском языке читает и пишет далеко не все население Земли.

Согласно некоторым источникам, например, сайту ethnologue.com, самым популярным языком в мире считается английский (см. рис. 1.)

На нем разговаривают почти 1,3 миллиарда человек. Но, например, на китайском разговаривает примерно столько же, сколько и на английском.

Всего в мире 7,7 миллиарда жителей, т.е. английский язык понимают только 17% жителей планеты. Считалось и считается до сих пор, что это тормозит проникновение новых технологий вообще и Интернета в частности в жизнь и быт населения Земли. Внедрение же идентификаторов на национальных языках эту проблему должно решить.

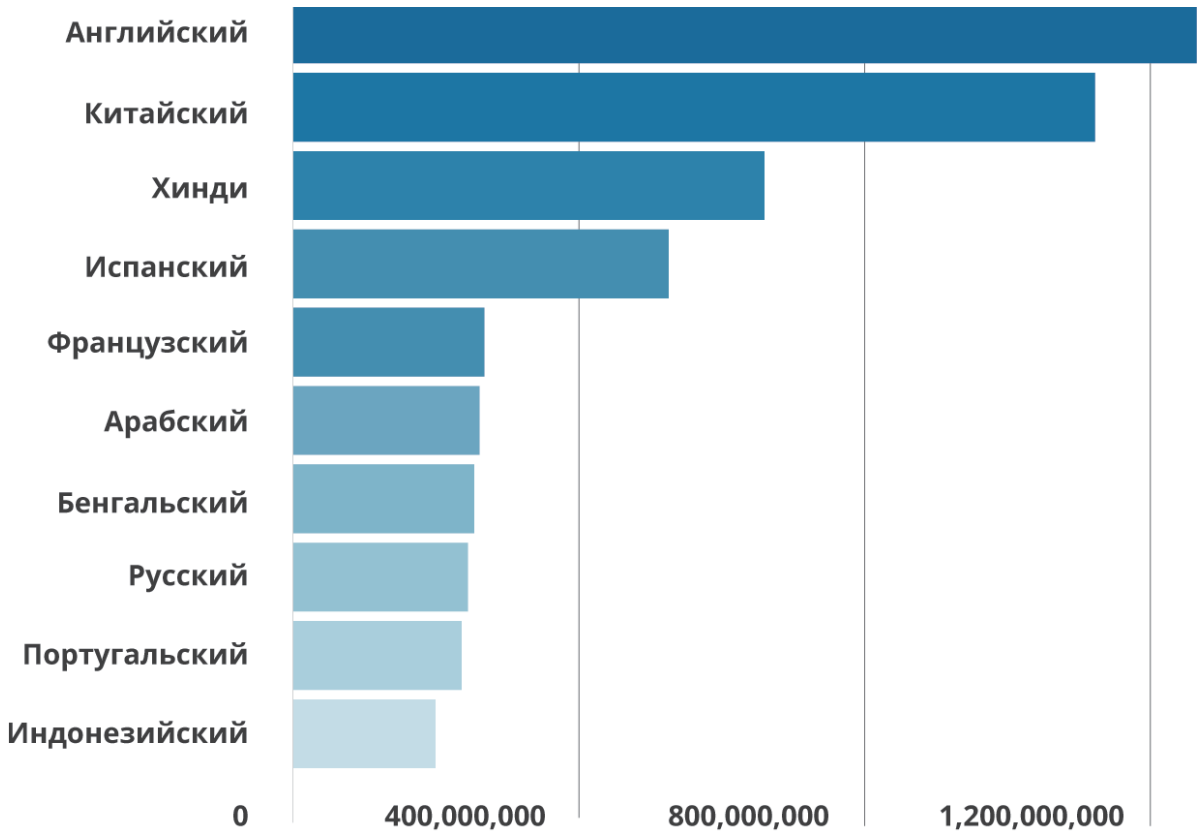
Сам по себе тезис достаточно спорный, но, тем не менее, в конце концов, ведь придумали Unicode и смогли его внедрить в современных операционных системах, что существенно, как считается, облегчило внедрение цифровых технологий в делопроизводство на национальных языках.

Именно этим посылом и руководствовались инициаторы программы внедрения международных доменных имен (Internationalized Domain Names, IDNs) в корпорации ICANN.

Важной вехой внедрения IDN стала программа ICANN по созданию доменов верхнего уровня на национальных языках (IDN ccTLD Fast Track Process). Что любопытно, эту программу в свое время выделили из общего процесса внедрения IDN, т.к. складывалось впечатление, что IDN глобально и в целом не внедрят никогда. Именно поэтому в названии программы для национальных доменов появилось словосочетание «Fast Track Process».

Совет директоров ICANN принял эту программу в 2009 году, а в 2010 году были запущены первые национальные домены. Российский IDN (.рф) стартовал одним из первых и долгое время был самым успешным доменом верхнего уровня на национальном языке. В свои наилучшие времена, в декабре 2011 года, количество зарегистрированных доменов второго уровня в .рф достигало 937 211 штук. Все ждали одного миллиона доменов, но увы!

Рис. 1. Популярность языков мира.



Впоследствии количество регистраций и продлений доменов стало падать. В настоящее время в .рф зарегистрировано 735 214 имен. В итоге домен .рф отдал пальму первенства среди IDN-доменов китайскому национальному домену .中国, в котором зарегистрировано 1,7 миллиона доменных имен.

Многие считают, что главную роль в падении регистраций сыграло повышение стоимости регистрации в июле 2017 года с 70 рублей до 120 рублей за домен. Однако следует отметить, что снижение количества регистраций и продления регистраций началось гораздо раньше, и произошло это по совершенно другим причинам.

Падение интереса к кириллическому .рф в первую очередь объясняется невозможностью идентификации всего множества интернет-сервисов.

Внедрение IDN решает только часть задачи перевода адреса информационного интернет-ресурса (не путать с IP-адресом) на национальный язык, а именно доменного имени. Другие компоненты этого адреса остаются в рамках прежней кодировки. Чаще всего это Latin1.

Наиболее выпукло проблема обозначилась в адресах электронной почты. Невозможность реализации электронной почты в домене .рф с самого начала ограничило возможности его применения.

Такая ситуация была не только с доменом .рф, но и с другими IDN-доменами. Китайцы решали ее своими сугубо «китайскими» методами – строили соответствия между «обычными» адресами и адресами с использованием IDN, например.

В 2007 году появился первый информационный документ, который рассматривал возможность интернационализации электронной почты. Вопрос заключался не только в том, чтобы «понимать» адреса, но и в том, чтобы эти адреса правильно обрабатывались в рамках почтовых протоколов.

К 2012 году был стандартизован протокол SMTP для поддержки обмена почтовыми сообщениями Интернационализированной электронной почты, в 2013 году появился стандарт заголовков такой почты и в том же году были стандартизированы дополнения к протоколам IMAP и POP3. Фактически, в 2013 году было завершено формирование корпуса технических документов, которые были необходимы для появления технического решения. Само решение получило название Email Address Internationalization (EAI).

Однако реального внедрения в массовый сервис пришлось ждать до 5 августа 2014 года, когда Google объявил о внедрении стандартов EAI в свой сервис Gmail.

Несмотря на это, EAI до сих пор не получило широкого распространения ни в корпоративных решениях, ни в публичных сервисах.

Для продвижения идеи использования полностью интернационализованных адресов, как имен доменов, так и адресов электронной почты, в 2015 году была образована UASG - Universal Acceptance Steering Group. В нее вошли 120 компаний, включая Afilias, Apple, CNNIC, Eco, i2 Coalition,

Группа дала определение понятию «Универсального признания» (Universal Acceptance):

«Universal Acceptance (UA) - это состояние, в котором все действительные доменные имена и адреса электронной почты принимаются, проверяются, хранятся, обрабатываются и отображаются правильно и последовательно.

Программное обеспечение и онлайн-сервисы поддерживают Universal Acceptance, когда они соответствуют требованиям RFC, перечисленным в таблице 1, для всех доменов и имен электронной почты».

Таблица 1. RFCs для UA

Название	Ссылка
Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework	https://tools.ietf.org/html/rfc5890
Overview and Framework for Internationalized Email	https://tools.ietf.org/html/rfc6530
SMTP Extension for Internationalized Email	https://tools.ietf.org/html/rfc6531
Internationalized Email Headers	https://tools.ietf.org/html/rfc6857
Internationalized Delivery Status and Disposition Notifications	https://tools.ietf.org/html/rfc6858
IMAP Support for UTF-8	https://tools.ietf.org/html/rfc6855
Post Office Protocol Version 3 (POP3) Support for UTF-8	https://tools.ietf.org/html/rfc6856
Post-Delivery Message Downgrading for Internationalized Email Messages	https://tools.ietf.org/html/rfc6857
Simplified POP and IMAP Downgrading for Internationalized Email	https://tools.ietf.org/html/rfc6858
PRECIS framework overview	https://tools.ietf.org/html/rfc8264
PRECIS usernames and passwords	https://tools.ietf.org/html/rfc8265
Unicode security mechanisms UTS#39	http://unicode.org/reports/tr39/

ICANN, Google, Microsoft, NIXI, THNIC и Yandex, государственные и инициативные группы.

Группа дала определение понятию «Универсального признания» (Universal Acceptance):

«Universal Acceptance (UA) - это состояние, в котором все действительные доменные имена и адреса электронной почты принимаются, проверяются, хранятся, обрабатываются и отображаются правильно и последовательно.

Программное обеспечение и онлайн-сервисы поддерживают Universal Acceptance, когда они соответствуют требованиям RFC, перечисленным в таблице 1, для всех доменов и имен электронной почты.»

Если с интернационализацией доменных имен, в общем, проблем нет (браузеры и прочие приложения - как на десктопах, так и на мобильных устройствах - сейчас в целом поддерживают «правильное» отображение IDN), то с почтой все не так радужно.

Поэтому UASG выпустила техническое описание EAI в качестве введения для разработчиков. Такое описание существует и для русскоязычной аудитории.

В этом документе достаточно понятно разбирается архитектура электронной почты, роль каждого из компонентов этой архитектуры в почтовом обмене и особенности реализации этих компонентов с учетом стандартов EAI. Завершается документ обширным перечнем литературы по теме EAI.

Следует заметить, что UA – это не только о национальных скриптах. Это и о допустимой длине имен, и о конвертации Unicode в октеты, и о фишинге при использовании различных, схожих до степени смешения разных скриптов, и еще о многих проблемах, которые неизбежно появляются при усложнении технологий.

Вообще говоря, и сама проблема со скрещиванием доменных имен с именами почтовых аккаунтов во многом связана с подходом по интернационализации доменных имен, в которых был применен «костыль» Punycode. Такое решение (Punycode) понятно с точки зрения «перелопачивания» всей системы стандартов DNS, но оно же породило сложности при реализации EAI.

В общем, внедрение UA – это процесс долгий. Не зря, обозначая роль Координационного Центра доменов RU/РФ в работе UASG, директор КЦ Андрей Воробьев назвал эту группу **«действующей на постоянной основе»**. Видимо, в обозримом будущем окончательно свои задачи это сообщество не выполнит.

Новости доменной индустрии

ДОМЕН .РФ – НОВАЯ ЭРА РУНЕТА!

12 мая 2020 года российский Интернет отметил знаменательный день: ровно 10 лет назад России был делегирован первый кириллический домен верхнего уровня .рф. Это событие стало фактически началом новой эры не только для Рунета, но и для всего мирового Интернета.

История .рф

В первые дни мая 2010 года были делегированы четыре первых IDN верхнего уровня (Internationalized Domain Names, интернационализованные доменные имена – доменные имена, которые содержат символы национальных алфавитов): свои домены на национальных алфавитах тогда получили три страны Ближнего Востока - Египет, ОАЭ, Саудовская Аравия – и Россия.

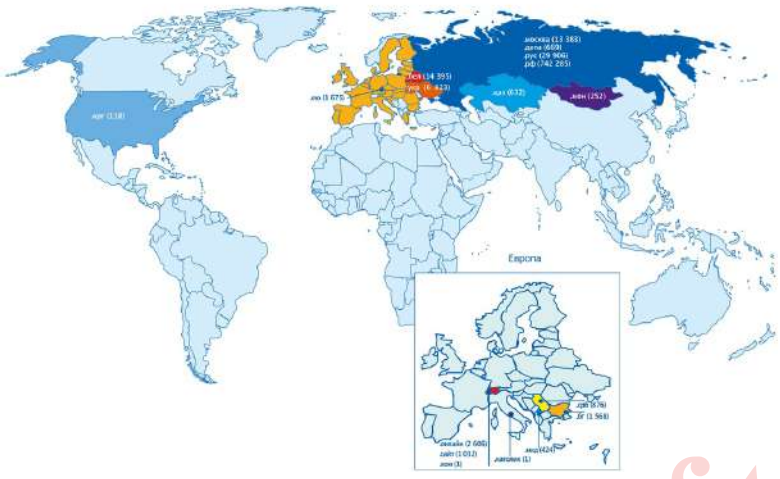
Национальной регистратурой домена .рф стал Координационный центр доменов .RU/.РФ – именно Координационный центр еще в 2008 году являлся инициатором создания национального домена, использующего символы кириллицы на русском языке. А в июне 2008 года на X Всемирном конгрессе русской прессы создание кириллических доменных имен в сети Интернет поддержал президент РФ Дмитрий Медведев. «Мы должны сделать все от нас зависящее, чтобы мы смогли добиться присвоения доменных имен на кириллице. Это символическая значимость русского языка, и у нас неплохие шансы добиться соответствующего решения, – сказал Медведев. – Мы должны использовать все технологические возможности для укрепления русскоязычного пространства».

Эксперты Координационного центра участвовали в процессе подготовки, согласования и утверждения процедур ICANN по внедрению IDN. Российская заявка, поданная КЦ, была одной из первых заявок на создание IDN ccTLD сразу после утверждения ICANN соответствующей процедуры. И уже в ноябре 2009 года было принято положительное решение о делегировании домена .рф. В начале 2010 года была открыта предварительная регистрация в новой доменной зоне. 12 мая 2010 года записи о новой доменной зоне были внесены в корневую зону Интернета - и этот день считается официальным днем рождения первого в мире домена верхнего уровня на кириллице, домена .рф. Тогда же заработали первые сайты в новой доменной зоне – это были президент.рф и правительство.рф.

11 ноября 2010 года стартовала открытая регистрация в домене .рф, которая стала сенсацией: в первый день открытой регистрации было зарегистрировано более 240 тысяч кириллических доменных имен. Домен .рф практически сразу вошел в топ-20 самых крупных национальных доменов Европы, и за прошедшие десять лет ни разу не покидал этот рейтинг.

Домен .рф в мировом доменном пространстве

Домен .рф стал примером для многих стран. На сегодняшний день существует 17 доменов верхнего уровня, записанных кириллическим шрифтом. Девять из них являются национальными IDN-доменами, а восемь – общими доменами верхнего уровня. Свои кириллические домены верхнего уровня зарегистрировала большая часть стран, где применяется кириллица. Это Россия, Беларусь, Украина, Сербия, Болгария, Македония, Казахстан, Монголия, а также Евросоюз.



732 998
доменов в зоне .РФ

В настоящее время в доменной зоне .рф зарегистрировано более 730 тысяч доменных имен. При этом делегировано более 667 тысяч имен, что составляет 91% от общего количества доменов. По количеству зарегистрированных доменных имен, а также по количеству доменов, используемых для практических нужд, домен .рф уже много лет занимает лидирующее место среди мировых IDN ccTLD.

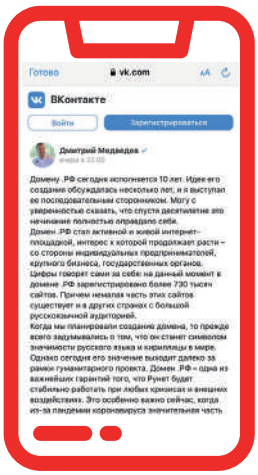
«Домен .рф – это гордость России. Это не просто показатель уровня использования Интернета в стране, это еще возможность объединить всех русскоговорящих

людей в мире и развивать это сообщество, поддерживать многоязычие и культурное разнообразие в Сети», - отметил директор Координационного центра доменов .RU/.РФ Андрей Воробьев.

Юбилей .рф на фоне пандемии – как это было

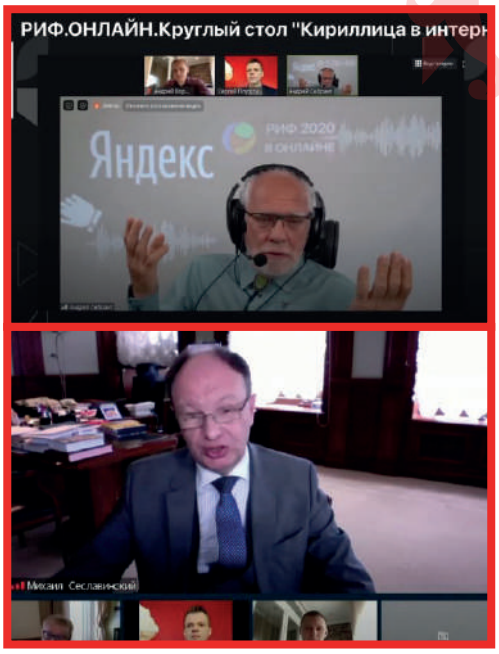
Несмотря на то, что провести праздник, соответствующий торжественности момента, из-за пандемии Covid-19 и введенного карантина оказалось невозможным, юбилей все равно широко отметили – в виртуальном пространстве. 12 мая весь день был посвящен именно 10-летию домена .рф.

О том, как десять лет назад на свет появился домен .рф, на своей странице в социальной сети «ВКонтакте» рассказал один из самых главных его «родителей» - заместитель председателя Совета безопасности Российской Федерации **Дмитрий Медведев**, который в 2008-2012 годах занимал пост президента РФ: «Когда мы планировали создание домена, то прежде всего задумывались о том, что он станет символом значимости русского языка и кириллицы в мире. Однако сегодня его значение выходит далеко за рамки гуманитарного проекта. Домен .рф – одна из важнейших гарантий того, что Рунет будет стабильно работать при любых кризисах и внешних воздействиях. Это особенно важно сейчас, когда из-за пандемии коронавируса значительная часть общения, в том числе и делового, перешла в онлайн. Во многом благодаря этой зоне наш Интернет укрепил свою независимость, стал более безопасным и надежным. Будем и дальше продолжать развитие крупнейшего кириллического домена мира».



В этот же день, 12 мая, состоялся **круглый стол «Кириллица в Интернете»**, который стал фактически кульминацией праздника. Участники круглого стола, которые физически находились в самых разных точках страны, не просто вспомнили, как Россия получала домен .рф, но и поговорили о том, что произошло с кириллическим Интернетом за прошедшие десять лет, какую роль кириллический Интернет играет в объединении, сохранении и развитии русскоязычного сообщества в сетевом мире.

Приветствие участникам круглого стола направил чрезвычайный и полномочный представитель президента РФ в ЦФО, министр связи РФ в 2008-2012 годах Игорь Щёголев. «Десять лет назад нашей стране был передан сертификат на домен верхнего уровня .рф, который стал доступен всем интернет-пользователям. Торжественной церемонии предшествовали длительные международные переговоры и напряженная работа специалистов. Это событие стало знаменательным и на тот момент беспрецедентным и для России, и для Всемирной сети... Сейчас сегмент .рф сети Интернет объединяет сотни тысяч доменных имен и прочно занимает позиции ключевого кириллического домена верхнего уровня, подтверждая статус России в качестве одной из ведущих интернет-держав», – написал **Щёголев**.



А руководитель Федерального агентства по печати и массовым коммуникациям Михаил Сеславинский привел совершенно неожиданные данные: в США, например, работают **833 сайта** в домене .рф, в Германии – **682,** в Финляндии - **182**, в Канаде - **54**, в Перу - **75**, в Мексике - **50**, в Аргентине - **12**. **«Я хочу сказать слова огромной благодарности всем тем, кто поддерживает доменную зону .рф за рубежом»**, - сказал **Сеславинский.**

Директор Координационного центра доменов .RU/.РФ **Андрей Воробьев** рассказал о пути, пройденном доменом .рф с момента первого упоминания в 2007 году до официального делегирования его России в 2010 году. Он представил статистику регистраций и использования доменных имен в домене .рф. «На протяжении девяти лет мы каждый год наблюдали, как домены, которые были зарегистрированы на фоне ажиотажного спроса, частично не продлевались, но к 10-летнему юбилею мы вышли уже на абсолютное плато, уже нет проседания, как это было восемь лет назад», – отметил директор КЦ. Также Воробьев рассказал о проектах, приуроченных к 10-летию домена: сайте 10.кц.рф и проекте поддержки доменных имен и почтовых адресов на национальных языках Поддерживаю.рф.

На круглый стол были приглашены не только те, кто стоял у «колыбели» юбиляра и чьими стараниями домен .рф появился на свет. Представители библиотек, научных и учебных учреждений поделились своим опытом использования доменных имен в домене .рф на практике, а также рассказали и о некоторых проблемах с ним. Одна из проблем – это внедрение концепции «универсального принятия» IDN-доменов и почтовых адресов. Эта тема является сегодня одной из самых горячих в неанглоязычном Интернете. Все участники подчеркивали, что доменные имена должны обслуживаться одинаково, и необходимо активнее работать над техническими решениями для использования EAI-адресов.

Календарь событий: 2020 год

* Глобальная эпидемия коронавируса сделала большинство конференций и физических встреч невозможными. Некоторые события просто не состоятся, но многие регулярные встречи будут проведены в онлайн-режиме. Многие мероприятия перенесли даты проведения. Хотя элемент человеческого общения будет во многом утерян, виртуальные конференции гораздо более доступны для участия – ни транспортных расходов, ни регистрационных взносов!

Международные события

1-3 июня
NANOG 79,
Онлайн*

Североамериканская группа сетевых операторов (The North American Network Operators Group, NANOG) является одной из самых активных профессиональных ассоциаций в области сетевой архитектуры, конфигурации и технического администрирования сетей в Интернете. Основной фокус NANOG на технологиях и системах, обеспечивающих работу Интернета. Поскольку инженерные вопросы NANOG имеют глобальный характер, участие в списке рассылки и конференциях может быть полезно широкому кругу технических специалистов в области сетевых технологий Интернета. Летняя встреча NANOG 79 будет проведена в виртуальном режиме. https://www.nanog.org/meetings/nanog-79/

22-25 июня
ICANN 68,
Онлайн*

Встречи ICANN проводятся три раза в год в различных регионах земного шара для того, чтобы предоставить возможность активным членам сообщества ICANN лично поучаствовать в обсуждении насущных проблем. Общей темой, конечно, является DNS - глобальная система трансляции имен. Здесь обсуждаются как технические вопросы обслуживания услуг DNS, так и юридические и бизнес-аспекты предоставления регистрационных услуг. Участие во встречах ICANN бесплатно.
https://meetings.icann.org/en/remote68

27-31 июля
IETF 108,
Онлайн*

IETF (Internet Engineering Task Force) является одной из основных организаций по разработке стандартов Интернета. В основном работа в IETF проходит в многочисленных списках рассылки, соответствующих различным рабочим группам (этих групп более 100). Три раза в год IETF проводит недельные совещания на которые приезжают разработчики протоколов, инженеры и операторы со всего мира (в среднем около 1200 участников из более 50 стран мира). Совещания IETF – это хорошая возможность познакомиться с новейшими тенденциями в области сетевых технологий и принять участие в их разработке. В этом году встречи IETF проводятся онлайн. https://www.ietf.org/how/meetings/108/

2-10 сентября
APNIC 50,
Дакка, Бангладеш

Встреча APNIC – одна из крупнейших конференций по интернет-технологиям азиатского региона и Океании. Здесь обсуждаются вопросы внедрения и использования интернет-технологий, технического администрирования сетей и инфраструктурных услуг Интернета. Эти встречи проводятся два раза в год – весной совместно с конференцией APRICOT, а осенью – как самостоятельное событие. Поскольку APNIC является региональной интернет-регистратурой, отвечающей за этот регион, помимо технических вопросов здесь обсуждаются политики администрирования адресного пространства.

APNIC и ISPAB, местный хост APNIC 50, наблюдают за развитием ситуации, связанной с эпидемией, и ожидается, что в мае будет принято решение о формате конференции. Регистрация и стипендии приостановлены до этого времени. https://blog.apnic.net/2020/04/24/apnic-50-update/

28 сентября - 2 октября
RIPE 81*

Встреча RIPE проводятся два раза в год и собирают более 800 участников для обсуждения вопросов политики распределения номерных ресурсов (IP-адресов и номеров автономных систем) в зоне обслуживания RIPE NCC, сотрудничества, а также решения технических вопросов, связанных с маршрутизацией, DNS, связностью, измерениями и инструментарием. Встреча длится пять дней и начинается с двухдневной пленарной программы, за которой следуют несколько параллельных сессий заседаний рабочих групп. В этом году майская встреча прошла в онлайн-формате. Как будет организован RIPE 81, пока не известно. https://www.ripe.net/participate/meetings/ripe-meetings

В России

20-23 июля
Владикавказ,
п. Фиагдон

Инфофорум - Северный Кавказ 2020
3-я Межрегиональная конференция по информационной безопасности и информационному взаимодействию «Инфофорум – Северный Кавказ». В тематику конференции вошли следующие направления: Информационная безопасность Северного Кавказа в цифровую эпоху; Развитие защищенных электронных услуг и систем межведомственного электронного взаимодействия; Безопасность кредитно-финансовой системы региона. https://infoforum.ru/conference/vladikavkaz-20

27 октября
Санкт-Петербург,
СПбГУ телекоммуникаций
им. М. А. Бонч-Бруевича

Telecomtrend 2020 (технологии мобильной и беспроводной связи)
Съезд TELECOMTREND по праву считается авторитетной площадкой и профессиональной платформой для компетентного диалога, обмена мнениями и демонстрации передовых решений в сфере телекоммуникаций. Среди постоянных участников такие компании, как МТС, Мегафон, Вымпелком, Радуга 2, Ericsson, Huawei, Nokia, РЖД, ВСІ и др.
http://23vek.ru/telecomtrend

В Москве

3 сентября
отель «Хилтон
Гарден Инн Москва
Красносельская»

«Critical Communications Russia: Цифровые технологии для обеспечения связи и безопасности государства, общества, бизнеса»
Индустрия Critical Communications с каждым годом расширяет поле деятельности и список выполняемых задач, охватив многие области ИКТ. Среди поднимаемых тем конференции: Объединение технологий и инструментов «Безопасного города», системы «112» и «ЭРА-ГЛОНАСС»; Техническая архитектура комплекса «Безопасный город» и т.д. https://www.comnews-conferences.ru/ru/conference/critical2020/

9-10 сентября
Онлайн*

Capacity Russia & CIS 2020
Capacity Russia & CIS 2020, ведущее событие в телекоммуникационном секторе, связывающее российскую и центральноазиатскую отрасль связи со своими международными партнерами.
https://events.capacitymedia.com/event/3147cee9-d7a0-4019-bfc2-78e96ddadd8c/summary

9-11 сентября
Моск. обл.,
п. Горки-10

РИФ+КИБ 2020
Российский Интернет Форум, который соберет ключевых игроков и специалистов рынка высоких технологий. https://2020.rif.ru/
РИФ.онлайн – ежедневные онлайн-секции, предваряющие сентябрьский РИФ+КИБ.
https://2020.rif.ru/online

6 октября
Lotte Hotel Moscow

Телеком 2020: новый уровень цифровой трансформации
XVI ежегодный международный форум операторов связи. Форум «Телеком», сессии в рамках World Mobile Congress и ПМЭФ ежегодно собирают всех знаковых представителей отрасли для обсуждения острых проблем, актуальных вопросов и успешных бизнес-кейсов. Только конструктивный диалог бизнеса и власти и эффективный нетворкинг. https://events.vedomosti.ru/events/telekom20

21-22 октября
отель «Холидей Инн
Лесная»

«TRANSPORT NETWORKS RUSSIA & CIS»
Развитие телекоммуникационных транспортных сетей в эпоху цифровой экономики, распределенных дата-центров и облачных услуг накануне запуска 5G. https://www.comnews-conferences.ru/ru/conference/tn2020/

2-6 ноября
ЦВК Экспоцентр

СВЯЗЬ-2020
32-я международная выставка «Информационные и коммуникационные технологии».
https://www.sviaz-expo.ru/

Осень*
ЦВК Экспоцентр

RIGF 2020
Одиннадцатый российский форум по управлению интернетом (RIGF 2020) состоится осенью 2020 г. в Москве. Во время RIGF 2020 эксперты вспомнят о становлении экосистемы управления интернетом в России, расскажут о его главных характерных особенностях и попробуют представить, как будет выглядеть управление глобальной сетью в будущем. https://rigf.ru/

MSK IX

10
ГОРОДОВ

500+
УЧАСТНИКОВ

42
ПЛОЩАДКИ

21
УЗЛЕЛ DNS-СЕТИ

ПОДКЛЮЧЕНИЯ ДО
100G

ТРАФИК
3,3Тбит/с

MSK-IX ускоряет коммуникации между интернет-компаниями, предоставляя нейтральную платформу Internet eXchange для обмена IP-трафиком между сетями и глобальную распределенную сеть DNS-серверов для поддержки корневых доменных зон.

Более 500 организаций используют сервисы MSK-IX для развития сетевого присутствия в 10 городах. К MSK-IX подключены операторы связи, социальные сети, поисковые системы, видеопорталы, провайдеры облачных сервисов, корпоративные и научно-образовательные сети.

127083, г. Москва, ул. 8 Марта, д. 1, стр. 12
тел.: +7 495 737-92-95
www.msk-ix.ru

+7 495 737-92-95

WWW.MSK-IX.RU

Интернет изнутри

2020