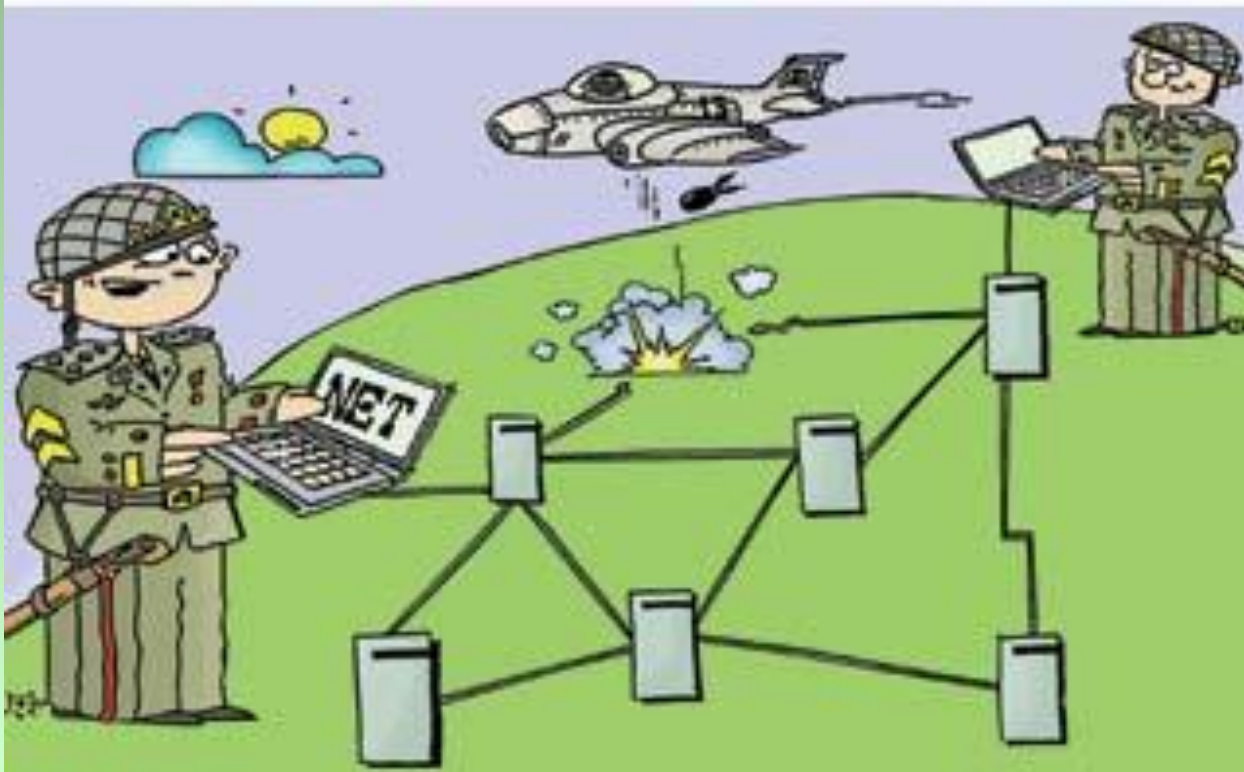


# Интернет (Internet)





**Интернет – это глобальная  
компьютерная сеть,  
объединяющая локальные,  
территориальные и  
корпоративные сети**

**WWW**

**(World Wide Web – всемирная  
паутина)**

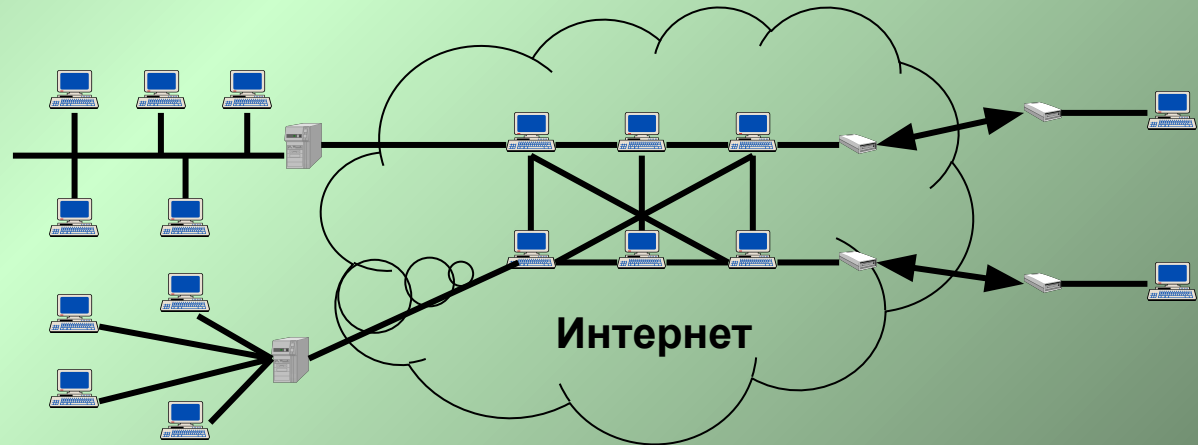
**или просто Web (Веб).**



# Глобальная компьютерная сеть Интернет

**Интернет** - это глобальная компьютерная сеть, объединяющая многие локальные, региональные и корпоративные сети и включающая десятки миллионов компьютеров.

Интернет фактически является **сетевой базой данных**. Гиперссылки связывают между собой сотни миллионов документов в единую сетевую базу данных.



- Слово *Интернет* (англ. *Internet*), обозначающее глобальную компьютерную сеть, возникло как сокращение *Interconnected Networks* — объединённые сети или «сеть сетей». В отличие от локальных сетей, ее «элементы» - не отдельные компьютеры, а сети.
- Информация в Интернете хранится на *серверах*, связанных скоростными линиями связи (оптоволоконными, спутниковыми). Практически все услуги Интернета основаны на использовании технологии «клиент-сервер»: программа-клиент на компьютере пользователя запрашивает информацию, сервер возвращает ответ.



# Всемирная паутина

- Всемирная паутина или «*веб*» (англ. И/И/И/ = *World Wide Web*) - это служба для доступа к гипертекстовым документам (веб-страницам), хранящимся на серверах. Сейчас *WWW* - наиболее популярная служба Интернета.
- **Гипертекст** - это текст, в котором есть активные ссылки (*гиперссылки*) на другие документы.
- Гиперссылки обычно подчеркиваются и выделяются цветом (по умолчанию - синим). Если щелкнуть левой кнопкой мыши на гиперссылке, в окно браузера загружается документ, на который указывает ссылка.
- На современных веб-страницах встречается не только текст, но и графика, звук, видео, причем каждый элемент может быть гиперссылкой. Такие документ называются *гипермедиа*.
- **Сайт (веб-сайт)** - это группа веб-страниц, которые расположены на одном сервере, объединены общей идеей и связаны с помощью гиперссылок.
- Чтобы сайт стал доступен другим компьютерам, на сервере должна быть запущена специальная программа - веб-сервер. Наиболее популярные веб-серверы:

- *Apache* (<http://httpd.apache.org>), свободный веб-сервер для различных операционных систем, включая *Windows*, *Linux*, *Mac OS*;
- *IIS* ([www.iis.net](http://www.iis.net)) - коммерческий веб-сервер для *Windows*;
- *nginx* ([sysoev.ru/nginx](http://sysoev.ru/nginx)) - бесплатный веб-сервер и почтовый сервер для крупных сайтов (есть версии для *Windows* и UNIX-подобных систем).
- Для просмотра веб-страниц на экране используются программы-браузеры (*Internet Explorer*, *Mozilla Firefox*, *Safari*, *Chrome*, *Opera*). Браузер отправляет веб-серверу запрос, содержащий URL-адрес документа (веб-страницы, рисунка, файла и т.п.), а сервер в ответ передает запрошенные данные. Обмен происходит по протоколу HTTP.

# Способы подключения к провайдеру:

- Пользователь получает доступ к глобальной сети через *провайдера* - фирму, локальная сеть которой непосредственно связана с Интернетом.
- с помощью *модема* по обычной телефонной линии; скорость обмена данными не превышает 56 кбит/с, поэтому такой способ уже практически не используется;
- помощью *ADSL-модема*, который также использует телефонную линию, но позволяет одновременно разговаривать по телефону и работать в Интернете; скорость передачи данных из Интернета к пользователю может достигать 25 Мбит/с, однако на телефонной станции необходимо устанавливать дополнительное оборудование (*сплиттер*, отделяющий низкочастотный телефонный сигнал от высокочастотного сигнала, передающего цифровые данные);
- через *локальную сеть* провайдера (если она существует в вашем доме); в этом случае телефонная линия не задействована;
- с помощью *беспроводных модемов* (USB-модемов), которые используют сети сотовых операторов и работают везде, где доступна мобильная связь; скорость передачи данных для сетей 3-го поколения (англ. *3G = 3<sup>rd</sup> generation*) достигает 10 Мбит/с, а в сетях 4-го поколения (4G) - до 1 Гбит/с.



# Новые поколения мобильной

- связи начинали разрабатываться практически через каждые десять лет с момента перехода от разработок первого поколения аналоговых сотовых сетей в 1970-х годах (1G) к сетям с цифровой передачей (2G) в 1980-х годах. От начала разработок до реального внедрения проходило достаточное количество времени (например, сети 1G были внедрены в 1984 году, сети 2G — в 1991 году). В 1990-х годах начал разрабатываться стандарт 3G. Сети поколения 4G, основанные на IP-протоколе, стали разрабатываться в 2000 году и начали внедряться во многих странах с 2010 года.
- **4G** — поколение мобильной связи с повышенными требованиями. К четвёртому поколению принято относить перспективные технологии, позволяющие осуществлять передачу данных со скоростью, превышающей 100 Мбит/с - подвижным и 1 Гбит/с — стационарным абонентам.
- Технологии LTE Advanced (LTE-A) и WiMAX 2 (WMAN-Advanced, IEEE 802.16m) были официально признаны беспроводными стандартами связи четвёртого поколения 4G (IMT-Advanced) Международным союзом электросвязи на конференции в Женеве в 2012 году.
- Скорость 100 Мбит/с должна предоставляться высокоподвижным абонентам (например, поездам и автомобилям), а абонентам с небольшой подвижностью (например пешеходам и фиксированным абонентам) должна предоставляться скорость 1 Гбит/с

# Браузеры



Internet Explorer

Google Chrome

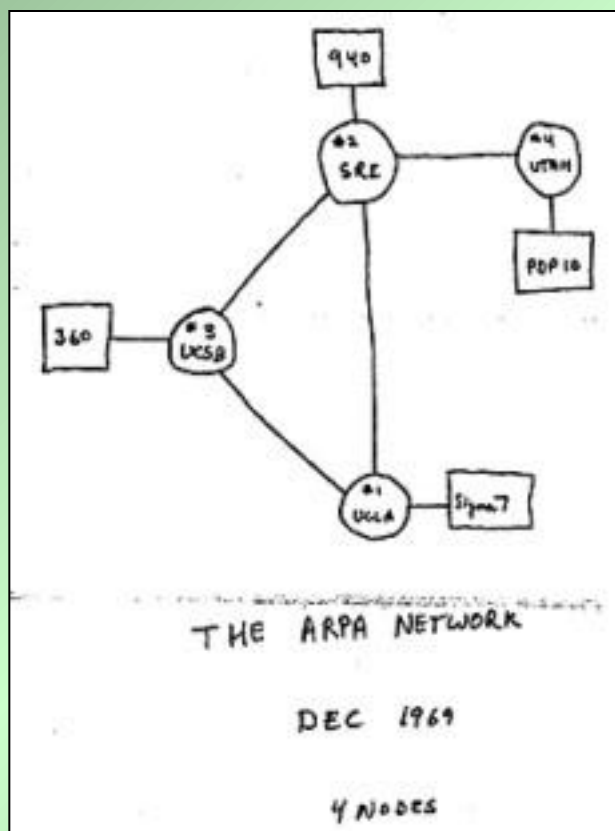
Opera

Mozilla FireFox

Yandex

Safari

# Рабочий набросок схемы первой глобальной сети ARPANET



Первая сеть состояла из 4 компьютеров

# Краткая история

- В 1960-е годы в министерстве обороны США начали разработку компьютерной системы передачи данных, которая получила название *ARPANET* (англ. *Advanced Research Projects Agency Network*, сеть агентства передовых исследований). В основу этого проекта были положены следующие идеи:
- сеть объединяет компьютеры, имеющие разное аппаратное и программное обеспечение;

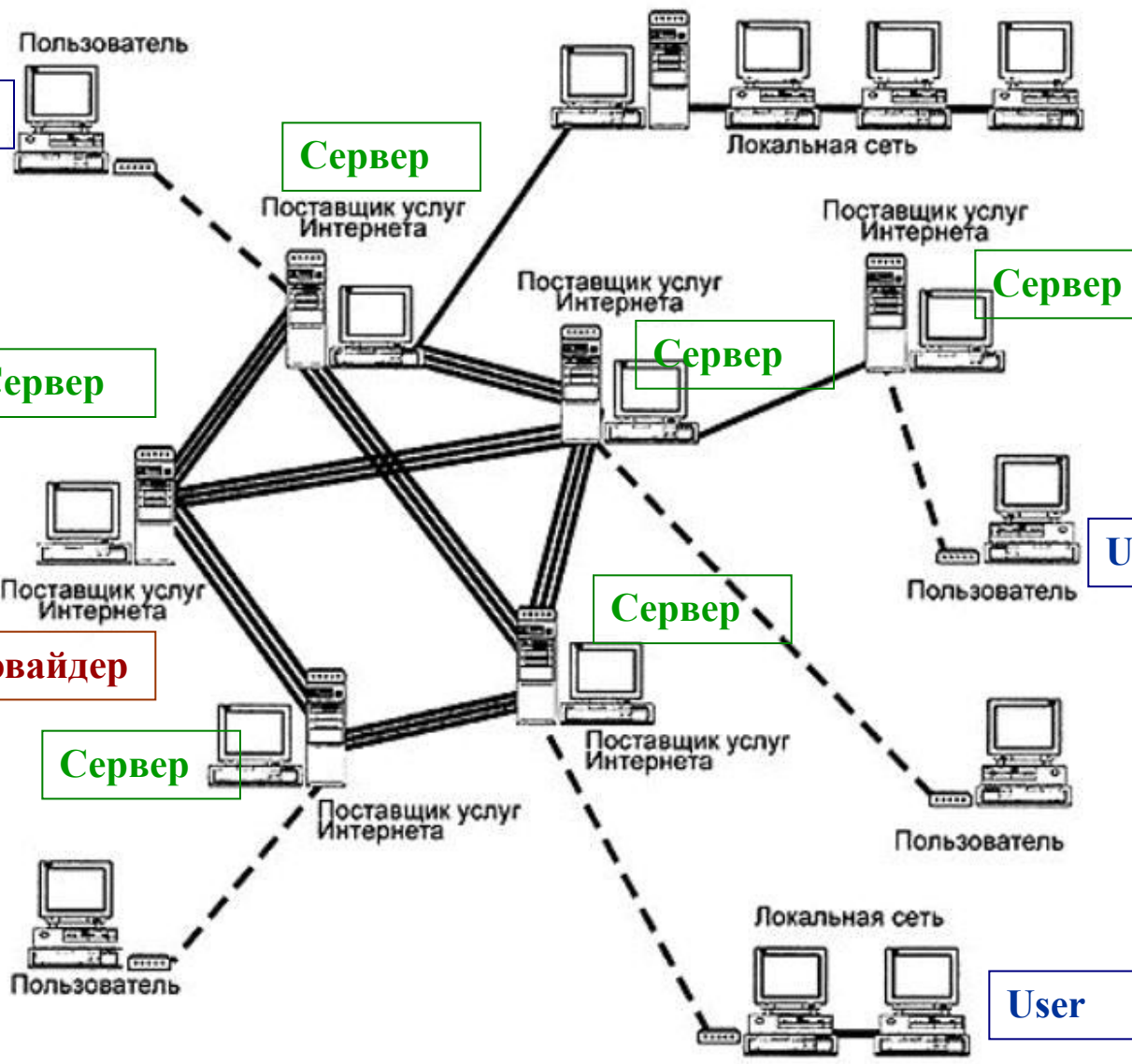
- при подключении новой сети не требуется переделка существующей части;
- нет единого центра (такая сеть называется *распределенной*), это обеспечивает живучесть в случае выхода из строя любого узла;
- пакетная передача данных: передаваемые данные разбиваются на пакеты небольшого размера, одна линия связи используется для одновременной передачи нескольких блоков данных.



- В 1969 году состоялся первый обмен данными по сети между компьютерами, установленными в Калифорнийском университете и Стэнфордском исследовательском центре. В 1971 году была создана программа для работы с электронной почтой, которая сразу стала очень популярной. Начиная с 1973 года, к новой сети подключаются университеты и колледжи не только США, но и Европы. В 1983 году сеть разделяется на две части: военную сеть *MilNet* и общедоступную сеть, которая получила название Интернет.
- История российского Интернета начинается с 1990 года, когда была организована почтовая сеть «Релком» - первый провайдер в Советском Союзе.
- В 1991 году британский ученый Тим Бернес-Ли разработал систему обмена данными в виде *гипертекста* - текста с активными ссылками на другие документы. Сейчас она называется Всемирной паутиной (англ. *WWW = World Wide Web*) и является самой мощной службой Интернета. Многие ошибочно считают, что Интернет и Всемирная паутина - это одно и то же. В самом деле это не так, потому что в Интернете есть и другие службы - электронная почта, обмен файлами, чаты, форумы и т.д.

- Основу «каркас»  
составляют 150 млн.  
серверов. Из них 500  
тыс. в России

User



Сервер

Сервер

Сервер

Сервер





Сервер

Сервер

Провайдер

User

User

-  Высокоскоростные выделенные линии связи
-  Выделенные линии связи
-  Коммутируемые линии связи
-  Модемы для подключения компьютеров к линиям связи





# Магистралы Российского сегмента Интернета

Европа



США

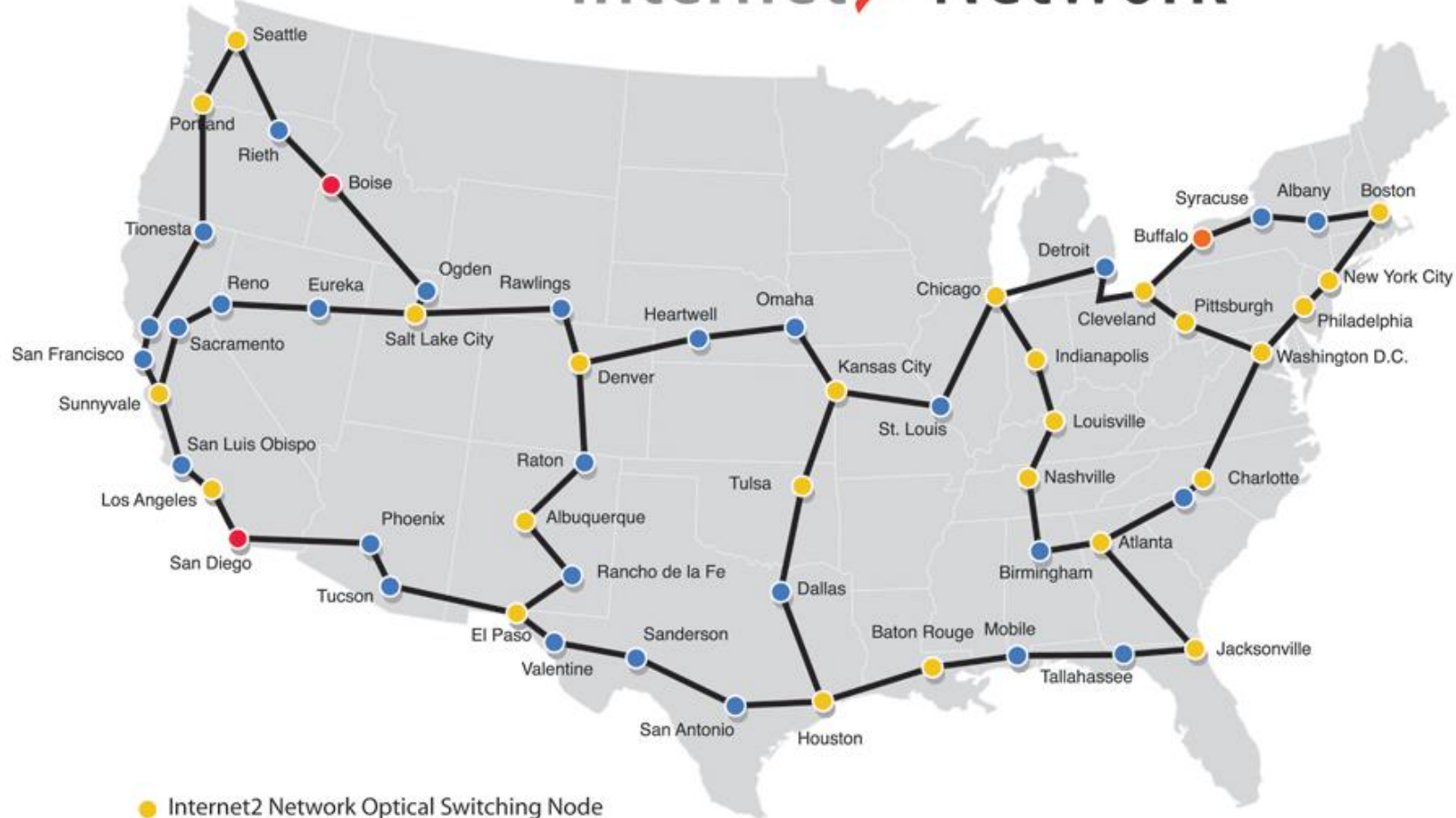
Азия

Азия

Япония

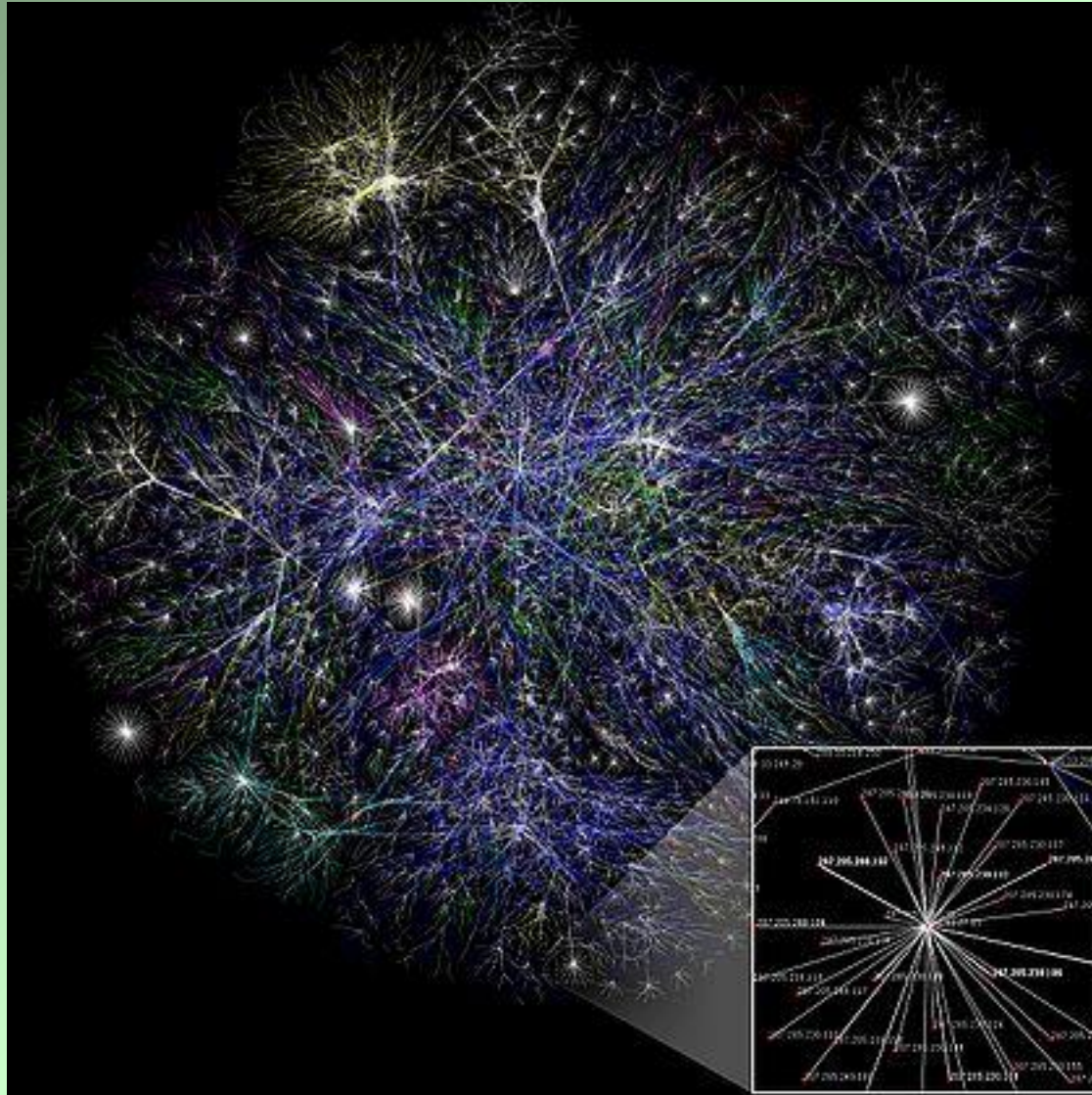


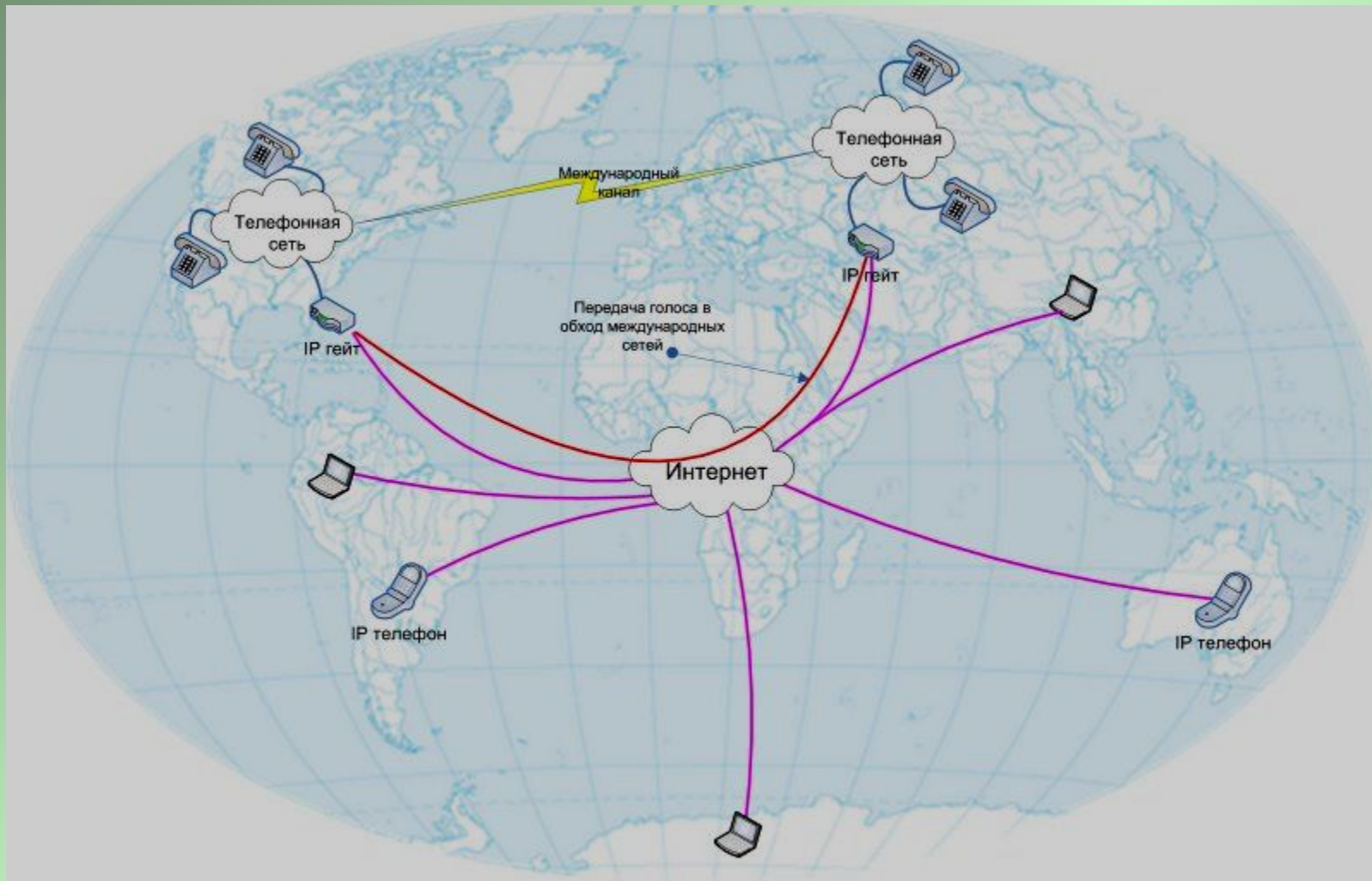
# Internet2 Network



- Internet2 Network Optical Switching Node
- Level3 Regen Site
- Internet2 Redundant Drop/Add Site
- ESnet Drop/Add Site

# Графическое изображение связей между сетями Интернета. Изображены только связи между серверами



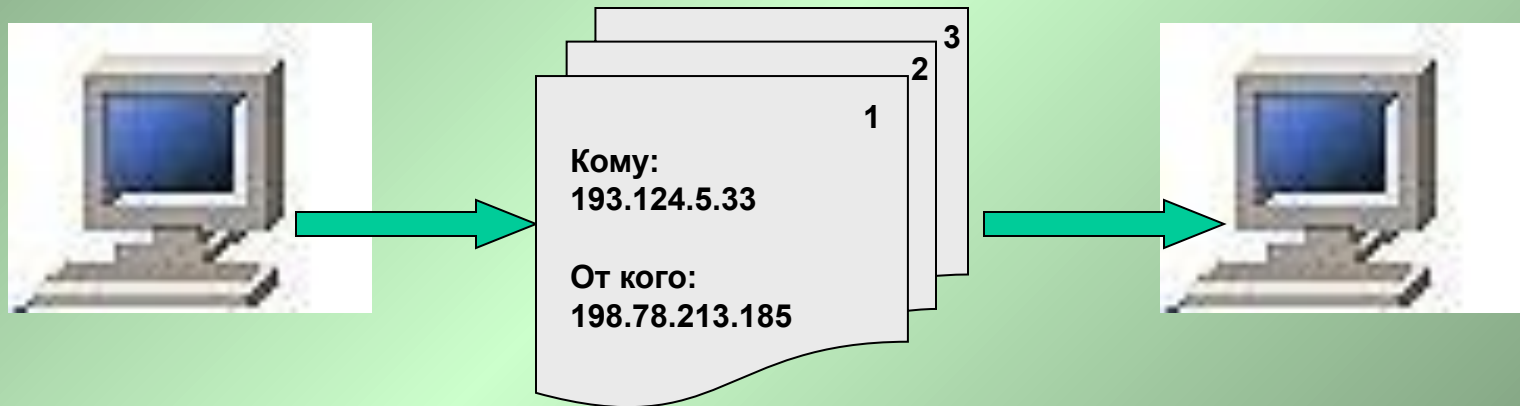




# Протокол передачи данных TCP/IP

Сеть Интернет функционирует и развивается благодаря использованию единого протокола передачи данных *TCP/IP*.

**Internet Protocol (IP)** - протокол маршрутизации - обеспечивает маршрутизацию IP-пакетов, т.е. доставку информации от компьютера-отправителя к компьютеру-получателю.



**Transmission Control Protocol (TCP)** - транспортный протокол - обеспечивает разбиение передаваемых файлов на IP-пакеты в процессе передачи и сборку файлов в процессе получения.

# Адресация в Интернете

Каждый компьютер, подключённый к Интернету, имеет свой уникальный 32-битовый **IP-адрес** (Internet Protocol). Возможно  $2^{32} = 4\,294\,967\,296$  **IP-адресов**, записываемых в виде четырех десятичных чисел от 0 до 255, разделенных точкой: **123.45.67.89**.

Для того чтобы учесть миллионы ПК в сети используют уникальные коды, сетевые протоколы TCP/IP. Этот номер состоит из 4 секций, каждая от 0 до 255

198.168.10.65

Провайдер – лицо или организация, поставщик услуг Интернета



## Протоколы

Вы уже знаете, что для передачи информации источник и приемник должны использовать один и тот же протокол - набор правил и соглашений, определяющих порядок обмена данными в сети. В Интернете в качестве стандарта принят протокол TCP/IP, разработанный в 1974 году. Вообще говоря, это не один протокол, а целое семейство, название которого происходит от двух самых важных протоколов - TCP (англ. *Transfer Control Protocol* - протокол управления передачей) и IP (англ. *Internet Protocol* - межсетевой протокол).

Попробуем разобраться, почему для работы в Интернете нужно использовать несколько протоколов. Предположим, что браузер на компьютере А запрашивает веб-страницу с сервера, который находится на компьютере Б. «Разговор» между браузером и сервером идет с помощью протокола HTTP (англ. *HyperText Transfer Protocol* - протокол передачи гипертекста). Браузер и веб-сервер не могут связаться напрямую. Чтобы послать запрос серверу, браузер передает адрес сервера и текст запроса операционной системе, которая вызывает драйвер протокола TCP.

Задача драйвера ТСР - установить соединение с удаленным компьютером и обеспечить доставку данных. Передаваемый блок данных разбивается на пакеты (размер пакета обычно не превышает 1,5 Кбайта), и каждый пакет передается на следующий уровень - драйверу протокола IP, который посылает его в сеть по указанному адресу

Обычно при работе в Интернете компьютеры А и Б напрямую не связаны, поэтому задача протокола IP - определить *узел-маршрутизатор*, на который нужно отправить пакет, чтобы он дошел до компьютера Б. Когда маршрут определен, пакет (с добавленной служебной информацией) передается на физический уровень (например, в сетевую карту), где передается просто как цепочка байт. Протоколы физического уровня могут быть любыми, они не определены в стандарте.

Протокол IP не гарантирует доставку пакетов, поэтому драйвер ТСР должен (с помощью установленного соединения) проверить, что данные получены, и в случае сбоя передать пакет повторно. На другом конце соединения драйвер ТСР «собирает» пакеты в единый блок данных и передает на уровень приложения (запрос дошел до сервера).

Маршрутизаторы обмениваются информацией друг с другом, сообщая о выходе из строя или подключении каких-то участков сети. Таблицы маршрутизации обновляются автоматически, так что при выборе маршрута пакетов учитывается фактическая структура сети в данный момент.

## ***Классы IP-адресов***

Существует 5 классов IP-адресов – А, В, С, D, Е. Принадлежность IP-адреса к тому или иному классу определяется значением первого октета (W). Ниже показано соответствие значений первого октета и классов адресов.

Класс IP-адреса	А	В	С	Д	Е
Диапазон первого октета	1-127	128-191	192-223	224-239	240-254
		1	3	9	7

# *Диапазоны:*

Класс	С	По
A	0.0.0.0	127.255.255.255
B	128.0.0.0	191.255.255.255
C	192.0.0.0	223.255.255.255
D	224.0.0.0	239.255.255.255
E	240.0.0.0	255.255.255.255

Таким образом, в Интернете используется четырехуровневая система протоколов, каждый из которых «занимается своим делом»:

- 1) *уровень приложений* - формат запросов и ответов, которыми обмениваются программы;
- 2) *транспортный уровень* (TCP) - правила пакетной передачи блоков данных без учета их содержания;
- 3) *сетевой уровень* (IP) - правила выбора маршрута для отдельных пакетов без гарантии их доставки;
- 4) *физический уровень* - правила передачи отдельных байтов по кабельной, оптоволоконной или другой линии связи.



На уровне приложений (который находится «ближе всего» к пользователю) чаще всего применяются протоколы:

HTTP - для передачи веб-страниц;

FTP - для передачи файлов;

SMTP - для передачи на сервер сообщений электронной почты;

POP3 или IMAP - для приема сообщений электронной почты с сервера. Существуют и другие протоколы (для чатов, новостных групп и т.п.), но все они используют TCP и IP соответственно на транспортном и сетевом уровнях

<b>HTTP</b>	Hypertext Transfer Protocol	Протокол передачи гипертекста	Используется службой World Wide Web (WWW) для передачи Web-страниц со всем их содержимым
<b>TCP/IP</b>	Transmission Control protocol / Internet protocol	Протокол управления передачей/ Интернет протокол	Используется для кодирования IP пакетов, передачи и приема пакетов, и декодирования пакетов
<b>FTP</b>	File Transfer Protocol	Протокол передачи файлов	Используется службой FTP для передачи файлов, например с программами
<b>SMTP</b>	Simple Mail Transfer Protocol	Простейший протокол передачи почты	Используется службой E-Mail для пересылки исходящих почтовых отправлений
<b>POP3</b>	Post Office Protocol, v.3	Протокол почтового отделения, версия 3	Используется службой E-Mail для доставки входящих почтовых отправлений
<b>NNTP</b>	Network News Transfer Protocol	Сетевой протокол передачи новостей	Используется службой телеконференций для пересылки сообщений между подписчиками

## Адреса в Интернете

## IP-адреса

В Интернете любые два компьютера могут связаться друг с другом. Для этого каждый из них должен иметь уникальный адрес. С «точки зрения» компьютеров, удобнее работать с числовыми адресами, каждый из которых занимает одинаковое место в памяти. Такие адреса (их называют *IP-адресами*) состоят из четырех чисел в интервале от 0 до 255, например, **192.168.104.115**

212.96.118.82,

$256*256*256*256$

В этих числах закодированы номер сети и номер компьютера в сети. Для того, чтобы выделить эти две части из IP-адреса, используют маски-шаблоны. Маска - это тоже четыре числа в интервале 0-255, но она строится особым образом, по принципу «n единиц, потом - нули» в двоичном коде. Например, маска 255.255.255.0

В связи с бурным развитием Интернета адресов, которые можно использовать при таком кодировании, скоро не будет хватать для всех желающих. Поэтому предполагается, что произойдет постепенный переход на новую (шестую) версию протокола IP, которая обозначается как IPv6. В ней на каждый адрес отводится 128 бит, а не 32.

Уже сейчас существует более 1600 сетей, где применяется IPv6; он поддерживается всеми современными операционными системами и производителями оборудования. Полный переход на IPv6 займет несколько лет, он потребует больших денежных затрат и замены всех устаревших устройств.

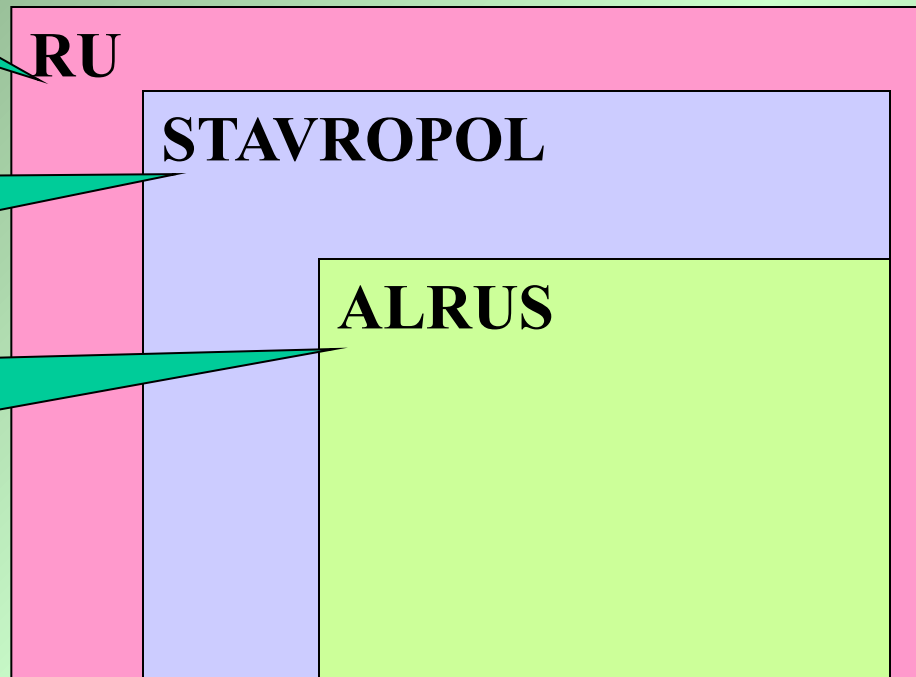
Вообще говоря, IP-адрес присваивается не компьютеру, а *интерфейсу* - каналу передачи данных (сетевой карте, модему). Поэтому один компьютер может иметь несколько IP-адресов, например, если на нем установлено две сетевых карты (или сетевая карта и модем).

# Доменный принцип построения имен

Домен верхнего  
уровня - Страна

Домен нижнего  
уровня - Город

Домен нижнего  
уровня - Узел



# alrus.stavropol.ru

Доменная система имён (DNS - Domain Name System) ставит в соответствие числовому IP-адресу компьютера уникальное доменное имя.



# Доменная система имён

Домены верхнего уровня бывают двух типов: **географические** (двухбуквенные - каждой стране соответствует двухбуквенный код) и **административные** (двух-трёхбуквенные).

Административные	Тип организации	Географическая	Страна
com	Коммерческая	ca	Канада
edu	Общеобразовательная	de	Германия
gov	Правительственная США	jp	Япония
int	Международная	ru	Россия
mil	Военная США	su	бывший СССР
net	Компьютерная сеть	uk	Англия / Ирландия
org	Некоммерческая	us	США

Основной сервер  
компании Microsoft  
*www.microsoft.com*



Основные имена доменов верхнего уровня

Com - коммерческие

Org – некоммерческие

Gov – правительственные

Edu - образовательные

Mil - военные

Net – работа в сети

Biz - бизнес;

Info - информационные сайты;

Name - личные сайты;

Museum - музеи;

Кроме того, каждая страна имеет свой двухбуквенный домен первого уровня

Основные имена доменов верхнего уровня для стран

ru - Россия

au- Австралия

by - Белорусия

ca - Канада

de - Германия

fr - Франция

jp - Япония

Распределением IP-адресов и доменов первого уровня занимается международная организация ICANN (англ. *Internet Corporation for Assigned Names and Numbers*). Российский домен **.ru** был зарегистрирован в 1994 году. Свободный домен второго уровня может зарегистрировать любой желающий за небольшую плату. Такие услуги оказывают специальные организации - *регистраторы доменных имен*, на пример, *RU-Center (nic.ru)*. Домены третьего уровня часто можно получить бесплатно. Например, сайт **narod.yandex.ru** предоставляет всем желающим место под сайт и домен третьего уровня вида **ivanov.narod.ru**.

Раньше в доменных именах было разрешено использовать только латинские буквы, цифры и дефис. Сейчас можно регистрировать домены, содержащие другие знаки, входящие в кодировку *UNICODE*, например, буквы русского алфавита. За Россией закреплен домен **.рф**, в котором все желающие могут зарегистрировать домены второго уровня.

Таким образом, сейчас в Интернете используется две системы адресов: IP-адреса и доменные имена. Чтобы установить соответствие между ними, на специальных серверах, которые называются *DNS-серверами*, хранятся таблицы, состоящие из пар «IP-адрес - доменное имя». Их задача - по запросу компьютера-клиента вернуть IP-адрес для заданного доменного имени (или наоборот).



# При регистрации адреса провайдер предоставляет следующие данные:

- IP-адрес
- Сетевая маска
- Сервер имени
- Шлюз
- Номер телефона модема провайдера
- Инструкция для входа в систему
- и др

Для того, чтобы компьютер смог установить связь с сетью, в настройках сетевой карты (или модема) указывается IP-адрес, маска сети и адрес DNS-сервера. Иногда эти данные определяются автоматически при подключении к сети провайдера.

Когда вы вводите адрес сайта (доменное имя) в адресной строке браузера, сначала отправляется запрос на DNS-сервер, цель которого - определить IP-адрес сервера. Если это удалось, на правляется запрос на получение веб-страницы, причем драйвер протокола IP использует полученный IP-адрес, а не доменное имя. Заметим, что одному доменному имени может соответствовать несколько IP-адресов. такой прием применяется для распределения нагрузки на сайты с большим количеством посетителей (например, [www.yandex.ru](http://www.yandex.ru), [www.google.com](http://www.google.com)). Таким образом, соответствие между доменными именами и IP-адресами можно описать как «многие ко многим»: с одним IP-адресом может быть связано несколько доменных имен и наоборот.

## Адрес ресурса (URL)

Точный адрес имеет не только каждый компьютер в Интернете, но и каждый документ. Для такого адреса чаще всего используется английское сокращение *URL = Uniform Resource Locator* -универсальный указатель ресурса. Типичный URL-адрес состоит из четырех частей: протокола, имени сервера (или его IP-адреса), каталога и имени документа (файла).

Такую систему записи придумал в 1990 году создатель Всемирной паутины Т. Бернес-Ли. Например, адрес

**<http://example.com/doc/new/vasya-new.htm>**

включает

1. протокол HTTP - протокол для обмена гипертекстовыми документами (это веб-страница);
2. доменное имя сервера **example.com**;
3. каталог на сервере **/doc/new**;
4. имя файла **vasya-new.htm**.

Иначе говоря, для обращения к документу **vasya-new.htm**, который находится в каталоге **/doc/new** на сервере **example.com** нужно использовать протокол HTTP.

Иногда каталог и имя файла не указывают, например, **http://example.com**. Это означает, что мы обращаемся к главной странице сайта. Она может иметь разные имена, в зависимости от настроек сервера (чаще всего - **index.htm**, **index.html**, **index.php**).

Для скачивания и загрузки файлов часто используется протокол FTP, тогда адрес документа выглядит примерно так: **ftp://files.example.com/pub/new/vasya-new.zip**

# Примеры интернет-адресов

Протокол  
передачи

Зона Web (Интернета)

<https://www.yandex.ru>

<https://www.mail.ru>

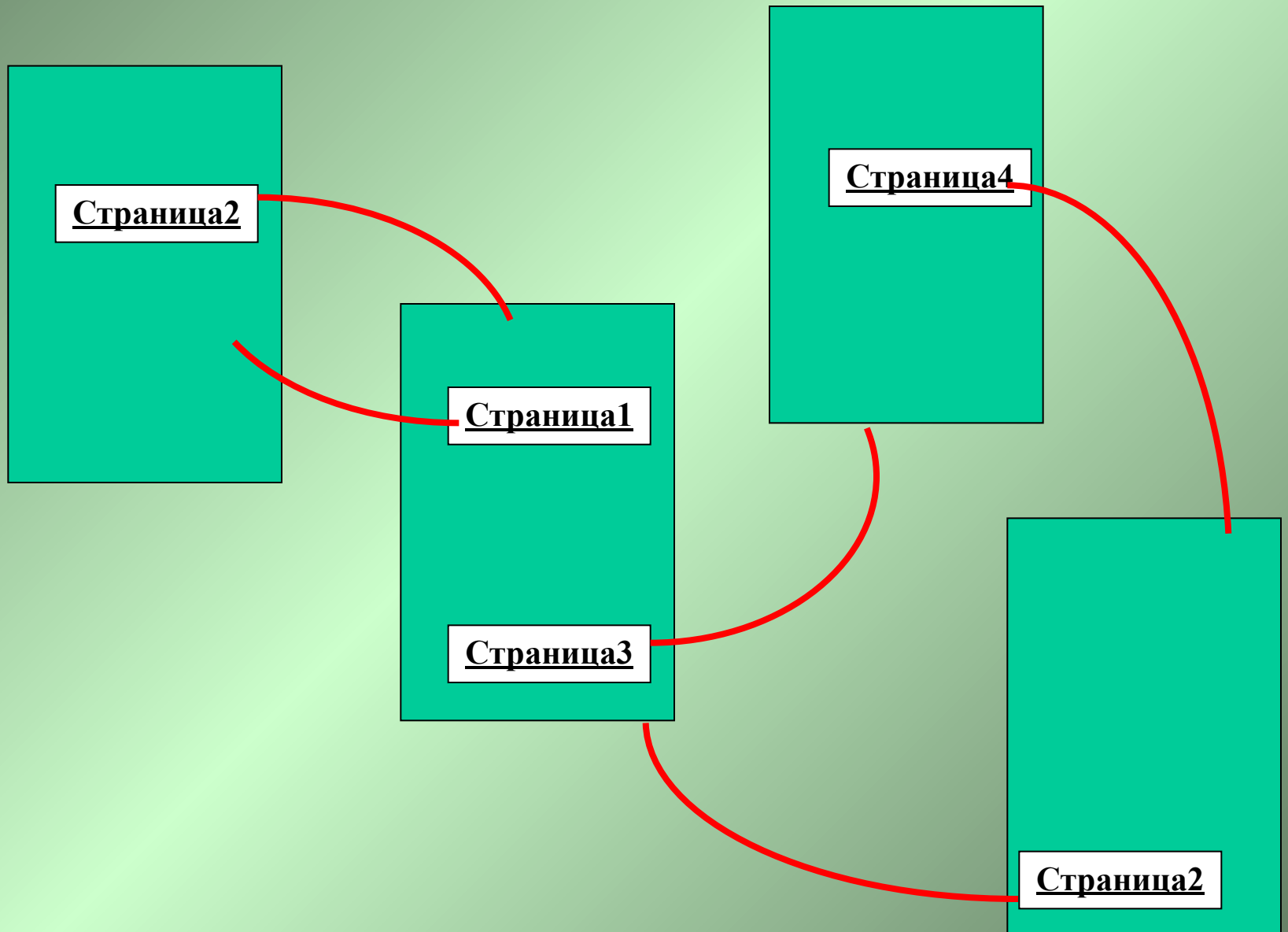
<http://www.rp5.ru/>

<http://acxk.pф>

<http://support.kaspersky.ru/>



# Принцип гипертекста



# Основные услуги Интернета

- ✓ **Хранение и предоставление информации**
- ✓ **Электронная почта**
- ✓ **Поисковые системы**
- ✓ **Голосовая и видео связь**
- ✓ **Передача файлов**
- ✓ **Форумы и блоги**
- ✓ **Социальные сети**
- ✓ **Интернет-магазины**
- ✓ **Электронные платежные системы и др.**

Особенность современной веба - привлечение пользователей к наполнению сайтов информацией и ее корректировке. Это привело к появлению термина «Web 2.0», которым иногда обозначают современный этап развития Всемирной паутины. Сайты, использующие технологии *Web 2.0*, как правило, требуют регистрации пользователей, для этого необходим действующий адрес электронной почты. Любой желающий может создать «личную зону» с собственными настройками и хранить там файлы, фотографии, видео, заметки. Другие могут комментировать эти материалы.

Пользователи объединяются в группы (сообщества) для того, чтобы вместе обсуждать интересующие их вопросы. Часто участники могут оценивать сообщения друг друга, таким образом, изменяется «репутация» (или «карма») участников, появляется некоторое соперничество

*Социальные сети:*

*ВКонтакте ([vk.ru](http://vk.ru)),*

*Одноклассники ([www.odnoklassniki.ru](http://www.odnoklassniki.ru)),*

*Facebook ([www.facebook.com](http://www.facebook.com)) для многих стали местом общения с друзьями и одноклассниками.*

Появились специальные сайты, где пользователи могут вести *блоги* - сетевые дневники ([www.livejournal.com](http://www.livejournal.com), - [www.blogspot.com](http://www.blogspot.com)). Влияние блогов настолько возросло, что их стали приравнивать к средствам массовой информации/

Активно развиваются *вики-системы* (англ. *wiki*) - веб-сайты, структуру и содержимое которых пользователи могут изменять с помощью инструментов, которые есть на самом сайте. Самый известный вики-сайт - это свободная энциклопедия *Википедия* (русская версия размещена на сайте [ru.wikipedia.org](http://ru.wikipedia.org))

## Доменные имена

В отличие от компьютеров, человеку неудобно работать с числовыми адресами. Они плохо запоминаются, при вводе IP-адреса легко сделать ошибку, а заметить ее иногда достаточно сложно. Поэтому в 1984 году была разработана *система доменных имен* (англ. *DNS = Domain Name System*), которая позволила использовать символьные имена сайтов, например, **www.mail.ru**.

Домен (англ. *domain* - область, район) - это группа символьных адресов в Интернете. Домены образуют многоуровневую структуру (*иерархию, дерево*), вкладываются друг в друга, как матрешки

Чем-то такая система напоминает почтовый адрес, в котором указывается страна, город, улица, дом, квартира.

Точка в корне дерева - это *корневой домен*. Домены первого уровня (они называются *доменные зоны*) могут обозначать тип организации, например



С одной стороны, *Web 2.0* расширяет возможности пользователей. С другой стороны, нужно понимать, что размещенные данные хранятся где-то на серверах, куда в принципе может получить доступ злоумышленник. Известны случаи массовых взломов учетных записей в социальных сетях и блогах. Поэтому не следует размещать в Интернете информацию, опубликование которой как-то может вам повредить, даже теоретически.

Фактически на таких сайтах пользователи сами заполняют базу данных о себе, своих друзьях, карьере и даже личной жизни. Изучая и анализируя эти данные, владельцы сайтов и спецслужб получают возможность манипулировать людьми, используя полученную информацию в своих целях, например, для рекламы товаров. Очень часто социальные сети используются для распространения вредоносных программ и рекламных сообщений (*спама*).

В начале XXI века Т. Бернес-Ли (автор Всемирной паутины) предложил развивать веб в направлении создания «семантической паутины» (Web 3.0), в которой все документы связаны по ключевым словам, как в базе данных. Это потребует переделки всех сайтов (добавления специальных смысловых «ярлыков» - *тэгов*), что обеспечит возможность полностью автоматического поиска и обработки информации. Вместо ручного поиска человек будет использовать программу-агент, которая подберет возможные ответы на вопрос и даст ему право окончательного выбора. С другой стороны, поиск нового типа позволит автоматически собирать всю информацию о личности (или организации), так что область его «частного пространства», «личной тайны» значительно уменьшится

## Поиск информации в Интернете

В Интернете сейчас содержится огромное количество данных, при этом найти нужную информацию иногда оказывается достаточно сложно.

**Поисковая система** - это веб-сайт, который предназначен для поиска информации в Интернете.

В начале развития Интернета, когда сайтов было немного, веб-мастера (создатели сайтов) составляли списки ссылок на интересные сайты. Когда ссылок стало много, их начали объединять в группы по темам. В результате развития этой идеи появились *каталоги*.

**Каталог** (англ. *web directory*) - это разбитый по темам список ссылок на сайты с их кратким описанием.

В каталогах обычно используют многоуровневую группировку ссылок (*дерево*): в каждой из крупных тем (*Новости, Наука, Образование* и др.) есть разделы, в разделах - подразделы и т.д.

Первым крупным сайтом-каталогом стал *Yahoo* ([www.yahoo.com](http://www.yahoo.com)), созданный в 1995 году. Самые крупные из российских каталогов - *Яндекс-каталог* ([yasa.yandex.ru](http://yasa.yandex.ru)) и *Каталог@Mail.ru* ([list.mail.ru](http://list.mail.ru)).

Каталоги заполняются вручную людьми-экспертами (редакторами каталога), каждый из которых отвечает за определенный раздел. Кроме того веб-мастера могут предложить редакторам свои сайты для включения в каталог (бесплатно или платно).

**Поисковая машина** - это автоматическая система, которая хранит информацию обо всех известных ей веб-страницах и выдает по запросу адреса тех из них, где встречаются введенные пользователем ключевые слова.

Робот-браузер поисковой машины (его часто называют «паук», англ. *crawler*) выкачивает с сайтов веб-страницы, переходя по всем встречающимся на них ссылкам

**Ключевые слова** - это набор слов и выражений, которые отражают требуемую информацию.

Поисковый робот с помощью индекса находит те страницы, где встречаются эти слова.

Каждая поисковая машина имеет свой язык, который позволяет составлять сложные запросы, например, исключать некоторые ключевые слова из поиска или искать одно из заданного набора слов. Во многих системах для обозначения логической операции «ИЛИ» (нужно одно из указанных слов) используется символ |, а для логической операции «И» (нужны оба слова) - символ &. Если нужно найти словосочетание, в запросе его берут в кавычки. Обычно поисковый робот находит тысячи страниц, соответствующих запросу. Они выдаются пользователю в том порядке, который определяется разработчиками. Чаще всего учитывается *цитируемость* - число ссылок с других сайтов на эту страницу; чем ссылок больше, чем выше «ранг» данной страницы и тем выше она расположена в результатах поиска.

Самая крупная международная поисковая машина - *Google* ([www.google.com](http://www.google.com)). В России лидирующие позиции занимает *Яндекс* ([www.yandex.ru](http://www.yandex.ru)). Эти системы умеют искать не только текст, но также картинки и видео (правда, при поиске изображений используется текстовая информация рядом с ними). Поисковая система *TinEye* ([tinneye.com](http://tinneye.com)) позволяет находить изображения, похожие на образец