

Background and the evolution of the internet

[Science](#), [Biology](#)



The Internet has undergone explosive growth since the first connections were established in 1969. This growth has necessitated an extremely large system scale-up that has required new developments in the technology of information transfer. These new developments allow simplified solutions to the problem of how to reliably get information from point A to point B. Unfortunately, the rapid pace of the required technological advancement has not allowed for optimal solutions to the scale-up problem.

Rather, these solutions appear to have been the most convenient and practical at the time. Thus, the information transfer technology of today's internet does not guarantee the 'best path' for data transmission. The definition of the best path may mean the most cost effective or the fastest path or some path based on optimization of multiple protocols, but the current technology used in the internet cannot guarantee that the best path for data transmission will be chosen. The result is a reduction in economic and system resource efficiency.

The Evolution of the Internet

The Internet has become integrated into the economic, technological and security infrastructure of virtually every country in the world. However, the internet had quite a humble beginning. It was originally designed as a back-up military communications network (MILNET) and as a university research communications network (National Science Foundation Network, NFSNET / Advanced Research Projects Agency Network, ARPANET). The original technology developed for these limited systems was not designed for the massive scale-up that has occurred since inception.

Moreover, the original design of the internet system was based on the sharing of resources. The recent applications of the internet for commerce

<https://assignbuster.com/background-the-evolution-of-the-internet/>

and proprietary information transfer processes make resource sharing an undesirable aspect. A more recent development is resource usage based on policies limiting what part of the internet can use a specific service or data transmission line. An Introduction to Networks and Routing What is a network? A network is a group of computers linked together by transmission lines that allow communication between the computers.

Some of these computers are the equipment used by people on their desktop. Other computers in the network are computers that are designed only to direct traffic on the network or between different networks. Computer scientists often think of networks as large graphs with lines used to connect dots. The dots are called nodes and correspond to computers and the lines correspond to the transmission lines that connect the computers. The Internet is a giant network of smaller networks, called autonomous systems, that allows computers to be connected around the globe.

What is routing? The process of transmitting information from a source computer to a destination computer is called routing. The way this is done can greatly effect how quickly the information is transmitted between the two computers. What is a router? A router is a computer with more than one connection to the rest of the network that is programmed to choose which transmission lines to send information. Some routers are designed to route information between networks, as on the Internet, while other routers work to route information between computers on the same network.

How do routers route? In order for routers to choose the best route (or path) from the source computer to the destination computer, it is necessary that the routers communicate with each other about what computers and

<https://assignbuster.com/background-the-evolution-of-the-internet/>

networks they are connected to and the routes that can be used to reach these computers and networks. Often these routes must go through other routers. What are advertisements? Advertisements are the messages sent between routers to communicate information about routes to reach each destination. What is convergence?

Convergence occurs on the network or internet when all the routers know all the routes to all the destinations. The time required for all the routers to agree on the state of the network, the network topology, is known as the convergence time. When convergence does not occur, then data can be transmitted to a router which does not know how to get to a destination and this data is then lost. This is called a black hole. It is also possible that the data can be passed around a set of routers continuously without getting to the destination. This is called a routing loop. What is a data packet?

When a large message is being transmitted, the message will probably be broken up into smaller messages called data packets, and these data packets may not all be sent by the same path across the Internet, although they will hopefully all reach the same destination. What is a metric? A routing metric is a measure associated with a particular path between a source and a destination used by the router to decide what path is the best path. Typical metrics used by routing algorithms include path length, bandwidth, load, reliability, delay (or latency) and communication cost.

Path length is a geometric measure of how long the transmission lines are. Bandwidth is used to describe the available transmission rate (bps) of a given section the possible transmission path. The load is the data packet transmission per unit time. The reliability of a data transmission path is

<https://assignbuster.com/background-the-evolution-of-the-internet/>

essentially the number of errors per unit time. The delay in data transmission along a certain path is due to a combination of the metrics that have already been discussed, including geometric length of the transmission lines, bandwidth, and data traffic congestion.

The communication cost is essentially the commercial cost of data transmission along a certain transmission line. What is a router protocol? A router protocol is the way the router is programmed to choose the best path for data transmission and communicate with other routers. This algorithm will consider path metrics associated with each path in a way defined by the manager of each AS. What is an internet address? In order for routers to identify the destination of a data transmission, every destination must have an address.

The internet protocol (IP) method of addressing destinations uses a series of digits separated by dots. An example of an Internet address is 227. 130. 107. 5. Each of the 4 numbers separated by a dot has a value between 0 and 255. This range of values is set from the amount of computer memory designated for addressing at the beginning of the internet. The internet addressing scheme is similar to a scheme for international telephone calls. There is a 'country code' which is a fixed number for each country, and then there are other numbers which change on the phone number to refer to specific locations within the country.

The numbers on the IP address for a network on the internet correspond to what would be the country code on an international phone number are referred to as 'prefix'. The other numbers on the IP address change to refer to individual computers on that particular network. A 'netmask' can also be

<https://assignbuster.com/background-the-evolution-of-the-internet/>

used to specify which numbers on the IP address for a given network are fixed and which ones can be changed. A netmask is a series of ones and zeroes that can be put over the IP address. The part of the IP address under the ones is fixed as a network address.

The part of the IP address under the zeros can be changed to indicate specific computers on the network. What is a Domain Name System (DNS), the domain name and the Uniform Resource Locator (URL)? The DNS is a combination of computer hardware and software that can rapidly match the text specification of an IP address, like `www.helpmegetoutofthis.com`, to an IP address. The part, `helpmegetoutofthis.com`, is called the domain name. The whole text, `www.helpmegetoutofthis.com`, is called the Uniform Resource Locator (URL).

When you send an e-mail or use the Internet, you use the domain name and the URL to locate specific sites. This allows people to type in the text name, or domain name, of an internet site into the Netscape browser instead of trying to remember the numerical IP address. The DNS automatically matches the text name to the IP address for the user when the transmission request is submitted. What are servers and clients? All of the computers on the Internet are classified as either servers or clients. The computers that provide services to other computers are called servers.

The computers that connect to servers to use the services are called clients. Examples of servers are Web servers, e-mail servers, DNS servers and FTP servers. The computers used at the desktop are generally clients. How the internet works. Although the details of routing and software are complex, the operation of the internet from the users' perspective is fairly straight

<https://assignbuster.com/background-the-evolution-of-the-internet/>

forward. As an example of what happens when the Internet is used, consider that you type the URL `www. helpmegetoutofthis. com` into the Netscape browser.

The browser contacts a DNS server to get the IP address. A DNS server would start its search for an IP address. If it finds the IP address for the site, then it returns the IP address to the browser, which then contacts the server for `www. helpmegetoutofthis. com`, which then transmits the web page to your computer and browser so you can view it. The user is not aware that of the operation of an infrastructure of routers and transmission lines behind this action of retrieving a web page and transmitting the data from one computer to another.

The infrastructure of the internet can be seen as a massive array of data relay nodes (routers) interconnected by data transmission lines, where each node can service multiple transmission lines. In the general case where information must be sent across several nodes before being received, there will be many possible pathways over which this transmission might occur. The routers serve to find a path for the data transmission to occur. The routing of a file or data packets of a file is either be done by the technique of source routing or the technique of destination routing.

In source routing, the path the data transmission will follow is specified at the source of the transmission, while destination routing is controlled by the routers along the path. In the modern internet, almost all routing is done by destination routing because of security issues associated with source routing. Thus, the routers must be programmed with protocols that allow a reasonable, perhaps optimum, path choice for each data packet. For the <https://assignbuster.com/background-the-evolution-of-the-internet/>

routers to choose an optimum path also requires that the interconnected routers communicate information concerning local transmission line metrics.

Router communication is thus itself a massive information transfer process, given that there is more than 100, 000 networks and millions of hosts on the Internet. When viewing the enormity of the problem, it is perhaps easier to understand why engineers have accepted a sub-optimal solution to the problem of efficiency in data transfer on the Internet. When initially confronting a problem, the practical engineering approach is to simplify the problem to the point where a working solution can be obtained and then refine that solution once the system is functional.

Some of the simplifying assumptions used by engineers for the current internet data transmission system include 1) A transmission line is never over capacity and is always available as a path choice. 2) The performance of the router and transmission line does not depend on the amount of traffic. These two assumptions do simplify the problem of path choice considerably because now all the transmission lines and nodes may be considered equal in capacity and performance completely independent of traffic. As such, it is a much simpler optimization problem consisting of finding the route with the shortest path length.

To simplify the problem even further, another assumption is made: 3) Consider that an " Autonomous System" (AS), is a small internet inside the Internet. An AS is generally considered to be a sub-network of an Internet with a common administrative authority and is regulated by a specific set of administrative guidelines. It is assumed that every AS is the same and provides the same performance. The problem of Internet routing can now be

<https://assignbuster.com/background-the-evolution-of-the-internet/>

broken down into the simpler problem of selecting optimum paths inside the AS and then considering the optimum paths between the AS.

Since there are 'only' around 15,000 active AS's on the Internet, the overall problem is reduced to finding the best route over 15,000 AS nodes, and then the much simpler problem of finding the best route through each AS. There is an important (to this thesis) set of protocols which control the exchange of routing information between the AS's. The sort of routers in an AS which communicates with the rest of the internet and other AS's are called border routers. Border routers are controlled by a set of programming instructions known as Border Gateway Protocol, BGP.

A more detailed discussion of computer networking principals and the Internet facts can be found in e. g. [7]. An Introduction to Router Protocols. Routers are computers connected to multiple networks and programmed to control the data transmission between the networks. Usually, there are multiple paths that are possible for transmission of data between two points on the Internet. The routers involved in the transmission between two points can be programmed to choose the 'best path' based on some metric. The 'protocols' used to determine the path for data transmission are routing algorithms.

Typical metrics used by routing algorithms include path length, bandwidth, load, reliability, delay (or latency) and communication cost. Path length. Path length is a geometric measure of how long the transmission lines are. The routers can be programmed to assign weights to each transmission line proportional to the length of the line or each network node. The path length is then the sum of the weights of the nodes, lines or lines plus nodes along

the possible transmission path. Bandwidth. Bandwidth is used to describe the available transmission rate (bps) of a given section the possible transmission path.

An open 64 kbps line would not generally be chosen as the pathway for data transmission if an open 10 Mbps Ethernet link is also open, assuming everything else is equal. However, sometimes the higher bandwidth path is very busy and the time required for transmission on a busy, high bandwidth line is actually longer than on a path with a lower bandwidth. Load. This data packet transmission per unit time or the percent of CPU utilization of a router on a given path is referred to as the load on this path. Reliability.

The reliability of a data transmission path can be quantitatively described as the bit error rate and results in the assignment of numeric reliability metrics for the possible data transmission pathways. Delay. The delay in data transmission along a certain path is due to a combination of the metrics that have already been discussed, including geometric length of the transmission lines, bandwidth, and data traffic congestion. Because of the hybrid nature of the communications delay metric, it is commonly used in routing algorithms. Communication Cost.

In some cases, the commercial cost of data transmission may be more important the time cost. Commercial organisations often prefer to transmit data over low capacity lines which they own as opposed to using public, high capacity lines that have usage charges. The routing algorithms do not have to use just one metric to determine the optimum route; rather it is possible to choose the optimum route based on multiple metrics. In order for the optimum path to be chosen by the routers between the data source and the

data destination, the routers must communicate information about the relevant metrics with other routers.

This nature of this communication process is also defined by the routing algorithm and the transmission time is linked to the time required for the routers to have the necessary information about the states of the surrounding routers. The time required for all the routers to agree on the state of the network, the network topology, is known as the convergence time and when all routers are aware of the network topology, the network is said to have converged. Some of the common routing algorithm types can indeed affect the convergence of the network.

Some of the different algorithms characteristics that must be chosen when designing are static or dynamic routing, single path or multi-path routing and link state or distance vector routing. Static Routing. Static routing is done by use of a static list of attributes describing the network topology at the initiation of the network. This list, called a routing table, is used by the routers to decide the optimum routes for each type of data transmission and can only be changed manually. Therefore, if anything changes in the network, such as a cable breaking or a router crashing, the viability of the network is likely to be compromised.

The advantage is that there is no communication required between routers, thus the network is always converged. Dynamic Routing. In contrast to static routing, dynamic routing continually updates the routing tables according to changes that might occur in the network topology. This type of real time information processing allows the network to adjust to variations in data traffic and component reliability, but does require communication between

the routers and thus there is a convergence time cost associated with this solution.

Single Path vs Multi-path Routing. Single path and multi-path routing are accurate descriptive terms regarding the use of either a single line to send multiple packets of data from a given source to a given destination as opposed to using multiple paths to send all the data packets from the source to the destination. Multiple path algorithms achieve a much higher transmission rate because of a more efficient utilization of available resources. Link State vs Dynamic Routing Protocols.

Link-state algorithms are dynamic routing algorithms which require routers to send routing table information to all the routers in the network, but only that information which describes its own operational state. Distance-vector algorithms, however, require each router to send the whole of its router table, but only to the neighbouring routers. Because the link-state algorithms require small amounts of information to be sent to a large number of routers and the distance vector algorithm requires large amounts of information sent to a small number of routers, the link state algorithm will converge faster.

However, link state algorithms require more system resources (CPU time and memory). There is a new type of algorithm developed by CISCO which is a hybrid of the link-state algorithm and the distance vector algorithm [8].. This proprietary algorithm converges faster than the typical distance-vector algorithm but provides more information to the routers than the typical link-state algorithm. This is because the routers are allowed to actively query one another to obtain the necessary information missing from the partial tables communicated by the link-state algorithms.

<https://assignbuster.com/background-the-evolution-of-the-internet/>

At the same time, this hybrid algorithm avoids communication of any superfluous information exhibited in the router communications of the full tables associated with distance-vector algorithm. Switching. The distance vector, link state or hybrid algorithms all have the same purpose, to insure that all of the routers have an updated table that gives information on all the data transmission paths to a specific destination. Each of these protocols requires that when data is transmitted from a source to a destination, the routers have the ability to 'switch' the address on the data transmission.

When a router receives a data packet from a source with the destination address, it examines the address of the destination. If the router has a path to that destination in the routing table, then the router determines the address of the next router the data packet will 'hop' to and changes the physical address of packet to that of the next hop, and then transmits the packet. This process of physical address change is called 'switching'. It will be repeated at each hop until the packet reaches the final destination.

Although the physical address for the forwarding transmission of the data packet changes as the packet moves across the Internet, the final destination address remains associated with the packet and is a constant. The internet is divided up into hierarchical groups that are useful in the description of the switching process. At the bottom of this hierarchy are network devices without the capability to switch and forward packets between sub-networks, where an AS is a sub-network.

These network devices are called end systems (ESs), because if a packet is transmitted there, it cannot be forwarded and has come to the end. At the top of the hierarchy are the network devices that can switch physical

addresses are called intermediate systems (ISs). An IS which can only forward packets within a sub-network are referred to as intra-domain ISs while those which communicate either within or between sub-networks are called inter-domain ISs. Details of Routing Algorithms Link State Algorithms

In a link state algorithm, every router in the network is notified of a topology change at the same time. This avoids some of the problems associated with the nearest neighbour update propagation that occurs in the distance vector algorithms. The 'Open Shortest Path First' (OSPF) protocol uses a graph topology algorithm like Dijkstra's Algorithm to determine the best path for data transmission between a given data source and a data destination. The metric used for route optimisation is specific to the manual configuration of the router.

However, the default metric is the speed of the interface. The OSPF uses a two level, hierarchical network classification. The lower level of hierarchy is groups of routers called areas. All the routers in an area have full knowledge of all the other routers in the area, but reduced knowledge of routers in a different area. The different areas organized within the OSPF algorithm are connected by border routers, which have full knowledge of multiple areas. The upper level of the hierarchy is the backbone network, to which all areas must be connected.

That is, all data traffic going from one area to another must pass through the backbone routers. Distance Vector Algorithms In order for data to be transmitted from a source to a destination on the Internet, the destination must be identified using some mechanism. That is, each possible destination for data transmission must be described with an address. The scheme

<https://assignbuster.com/background-the-evolution-of-the-internet/>

currently used to address the internet space is the Internet Protocol (IP) version 4. The IP version 4 uses an address length limited by 32 bits. An example of an Internet address is 227. 130. 107.

5 with the corresponding bit vector 11100011 10000010 01101011 00000101. An initial difficulty in managing the available address space was the implementation of a class structure, where large blocks of internet address space was reserved for organisations such as universities, leaving commercial applications with limited address space. Routing of data transmission in this address environment was referred to as class-full routing. To alleviate this problem of limited address space, the internet community has slowly evolved to a classless structure, with classless routing.

In distance vector protocols, each router sends adjacent routers information about known paths to specific addresses. The neighbouring routers are sent information giving a distance metric of each one from a destination address. The distance metric could be the number of routers which must be used to reach the destination address, known as the ' hop count', or it could be the actual transmission distance in the network. Although this information is advertised only to the adjacent routers, these routers will then communicate the information with their neighbouring routers, and so on, until the entire network has the same information.

This information is then used to build the routing table which associates the distance metric with a destination address. The distance vector protocol is implemented when a router receives a packet, notes the destination, determines the path with the shortest distance to the destination and then forwards the packet to the next router along the shortest distance path. One

of the first distance vector protocols implemented on the Internet was the Routing Information Protocol (RIP). RIP uses the distance metric of hop count to determine the shortest distance to the destination address.

It also implements several protocols to avoid having data packets pass through the same router more than once (router loops). The path vector protocol is a distance vector protocol that includes information on the routes over which the routing updates have been transmitted. It is this information on path structure which is used to avoid routing loops. Path Vector Protocols are also somewhat more sophisticated than RIP because an attempt is made to 'weight' each path based on a locally defined criteria that may not simply reflect the highest quality of service, but rather the highest profit for an ISP.

The implementation of these types of router algorithms may be different in different parts of the Internet. When the algorithms are implemented inside an autonomous system, they are called Interior Gateway Protocols (IGP). Because the different autonomous systems that make up the Internet are independent from one another, the type of routing algorithm used within the autonomous systems can also be independent of one another.

That is, the managers of each autonomous system are free to choose the type of algorithm which best suits their particular network, whether it is static or dynamic link-state or dynamic distance-vector. When the algorithms are implemented to control data transmission between autonomous systems, they are referred to as Exterior Gateway Protocols (EGP). The EGP connect all autonomous systems together to form the Internet and thus all EGP should use the same algorithm.

The specific algorithm currently used as the EGP on the Internet is the Border Gateway Protocol (BGP), which is a type of distance vector algorithm called a path vector algorithm [9]. A path vector algorithm uses information about the final destination of the data transmission in addition to the attributes of the neighbouring links. It should be noted that the BGP algorithm can also be used as a router protocol within an autonomous system and is called an interior BGP (IBGP) in that instance. This necessitates calling the BGP an EBGp when it is implemented as an EGP.