Availability and Resiliency to Failures of Networks and Services (Disponibilidade e Robustez a Falhas de Redes e Serviços)

Modelação e Desempenho de Redes e Serviços

Prof. Amaro de Sousa (asou@ua.pt)

DETI-UA, 2025/2026

Availability

- The availability of an element is the probability of the element being operational on any instant of time.
- For a given element *i*, consider:
  - $MTBF_i$ is the *Mean Time Between Failures* of element *i*
  - $MTTR_i$ is the *Mean Time To Repair* of element *i*
- then, the availability a_i of element *i* is given by:

$$a_i = \frac{MTBF_i}{MTBF_i + MTTR_i}$$

- Examples:
  - If a link fails on average after 1 year of operation (*MTBF* = 365.25 $\times$ 24 = 8766 hours) and takes on average 2 days of being repaired and become operational (*MTTR* = 2 $\times$ 24 = 48 hours), then its availability is 0.99455 (= 99.455%);
  - If a router fails on average after 90 days of operation (*MTBF* = 90 $\times$ 24 = 2160 hours) and takes on average 3 hours to be repaired (or replaced) and reconfigured to become operational (*MTTR* = 3 hours), then its availability is 0.99862 (= 99.862%).

Availability

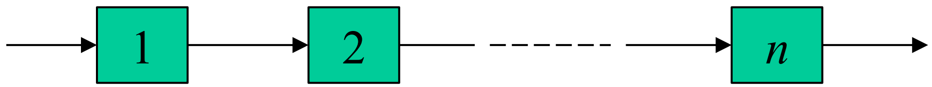
- The measurement of the availability of an element can be defined over different periods of time.

Availability	Downtime per year	Downtime par 30 days
90% (one nine)	36.53 days	73.05 hours
99% (two nines)	3.65 days	7.31 hours
99.9% (three nines)	8.77 hours	43.83 minutes
99.99% (four nines)	52.6 minutes	4.38 minutes
99.999% (five nines)	5.26 minutes	26.3 seconds
99.9999% (six nines)	31.56 seconds	2.63 seconds

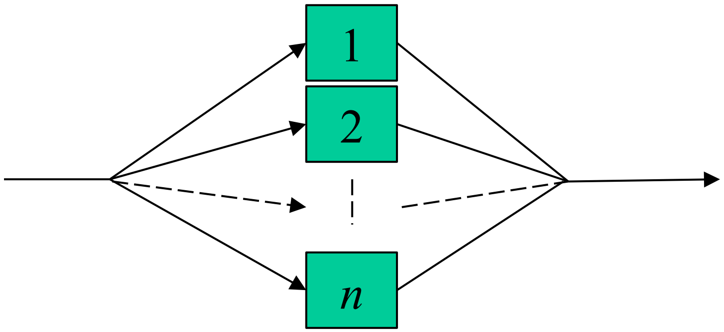
Downtime: total time such that the element is not available

Availability of a system with multiple elements

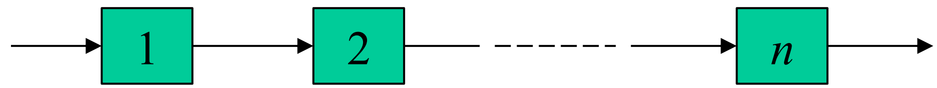
- The availability of a system with multiple elements is determined by modelling the system has a directed graph of elements in series or in parallel.
- The rules to decide if the elements are in series or in parallel are:
  - A set of elements is in series if the system fails when one of the elements fail:



  - A set of elements is in parallel if a system fails only when all elements fail simultaneously:



Availability of a system
<u>with all elements in series</u>

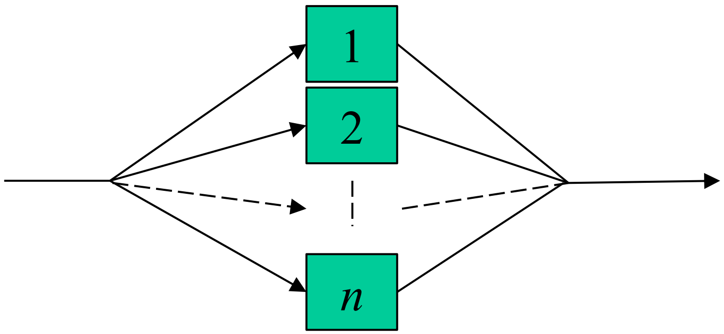


- The availability A of the system is the probability of all elements being available (i.e., working).
- If the availability of element i is a_i (and considering that failures in different elements are statistically independent), the availability of the system is the product of the availability values of all elements:

$$A = a_1 \times a_2 \times \cdots \times a_n$$

- Properties:
  - The availability of the system is lower (i.e., worst) then (or equal to) the availability of the element with the lowest availability value
- Example:
  - A system with 3 elements in series, each one with an availability of 99.9%, has an availability of 99.7%

Availability of a system **with all elements in parallel**



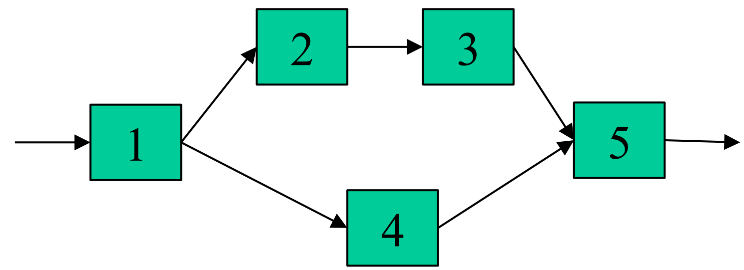
- The availability A of the system is the probability of at least one element being available (i.e., working).
- If the availability of element *i* is a_i (and considering that failures in different elements are statistically independent), the availability of the system is 1 – (the probability of all elements being unavailable):

$$A = 1 - [(1 - a_1) \times (1 - a_2) \times \cdots \times (1 - a_n)]$$

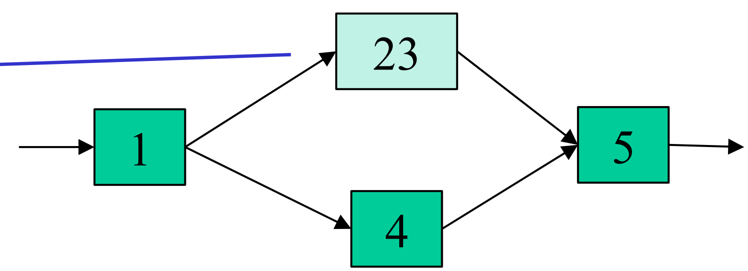
- Properties:
  - The availability of a system is higher (i.e., better) than (or equal to) the availability of the element with the highest availability value
- Example:
  - A system with 3 elements in parallel, each one with an availability of 99.0%, has an availability of 99.9999% (six nines)

Availability of a system with multiple elements

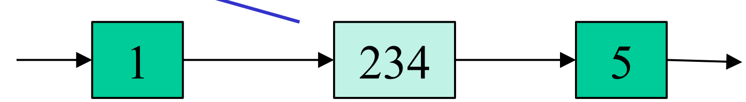
In general, we reduce iteratively the general graph onto a smaller graph (i.e., with a lesser number of elements) replacing a set of elements (in series or in parallel) by a single element with its equivalent availability.



$$A_{23} = a_2 \times a_3$$



$$A_{234} = 1 - [(1 - A_{23}) \times (1 - a_4)]$$



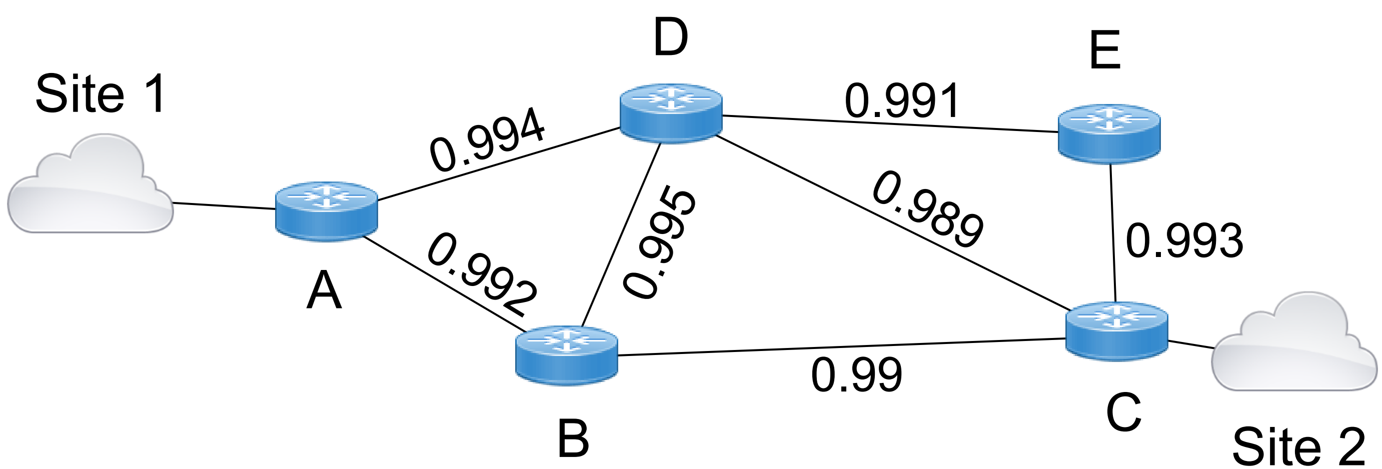
$$A = a_1 \times A_{234} \times a_5$$

Availability of a unicast service in a communications network

The availability of a unicast service depends on the routing path over the network of each flow of the service.

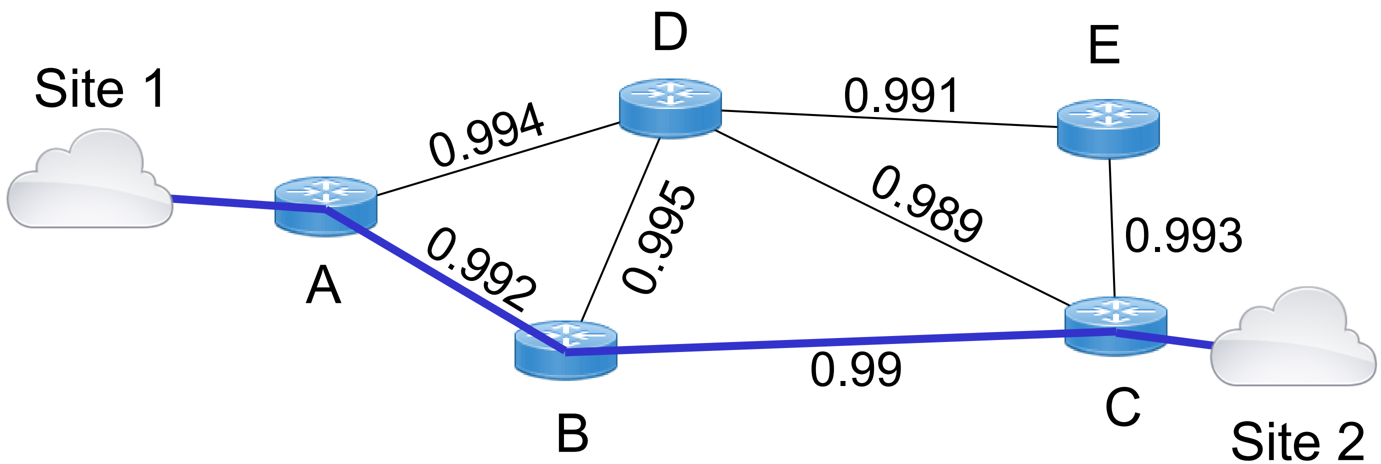
Consider the following VPN (Virtual Private Network) service between two sites of a company (a service with a single flow) provided by the network (in the figure) of a telecommunications operator. Consider that:

- all routers have an availability of 99.99%,
- the figure indicates the availability of each network link.



Availability of the example

If the VPN service flow between the two sites is routed through path A $\rightarrow$ B $\rightarrow$ C (from site 1 to site 2) and through the same path in its opposite direction (from site 2 to site 1),

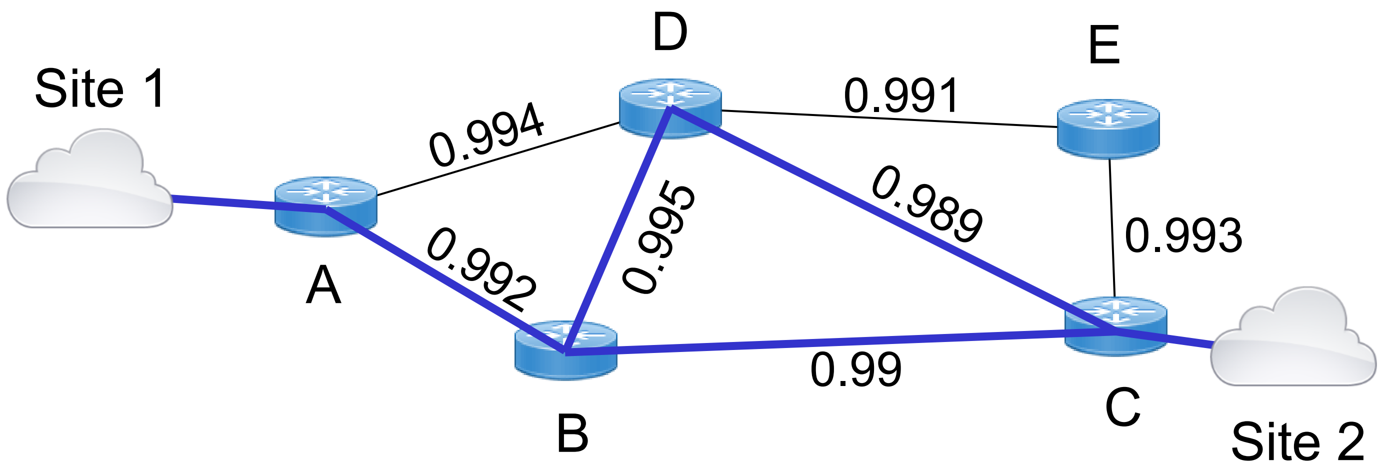


then, the service availability provided by the network is:

$$A = a_A \times a_{AB} \times a_B \times a_{BC} \times a_C$$
$$= 0.9999 \times 0.992 \times 0.9999 \times 0.99 \times 0.9999 = 0.9818$$

Availability of the example

If the VPN service flow is routed through either A → B → C or A → B → D → C (from site 1 to site 2) and through the same paths in the opposite direction (from site 2 to site 1),



then, the service availability provided by the network is:

$$A_{BDC//BC} = 1 - [(1 - a_{BD} \times a_D \times a_{DC}) \times (1 - a_{BC})] =$$

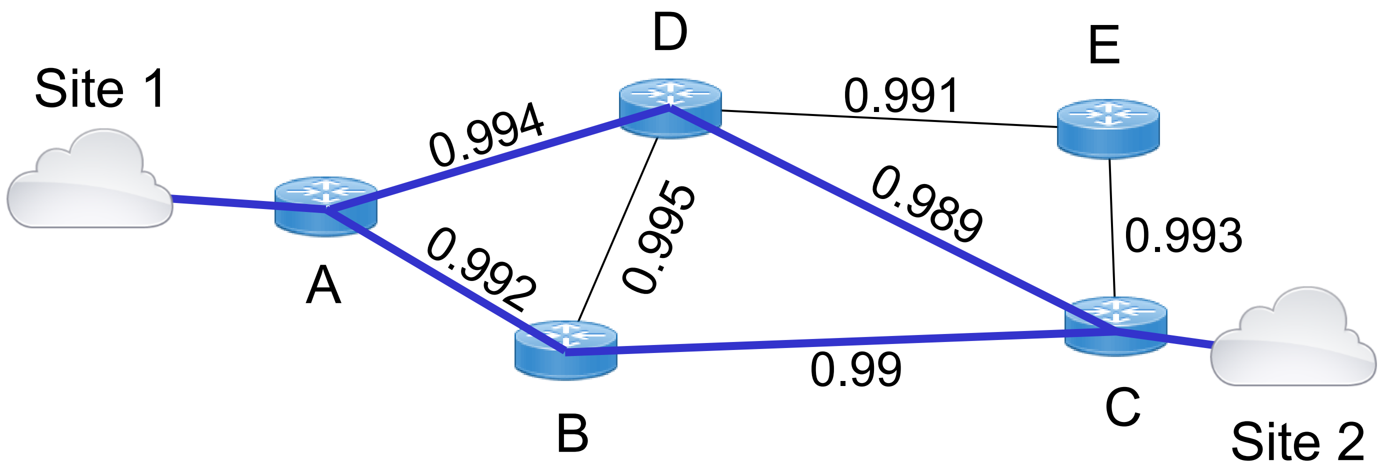
$$= 1 - [(1 - 0.995 \times 0.9999 \times 0.989) \times (1 - 0.99)] = 0.99984$$

$$A = a_A \times a_{AB} \times a_B \times A_{BDC//BC} \times a_C$$

$$= 0.9999 \times 0.992 \times 0.9999 \times 0.99984 \times 0.9999 = 0.9915$$

Availability of the example

If the VPN service flow is routed through either A → B → C or A → D → C (from site 1 to site 2) and through the same paths in the opposite direction (from site 2 to site 1),



then, the service availability provided by the network is:

$$A_{ADC//ABC} = 1 - [(1 - a_{AD} \times a_D \times a_{DC}) \times (1 - a_{AB} \times a_B \times a_{BC})] =$$

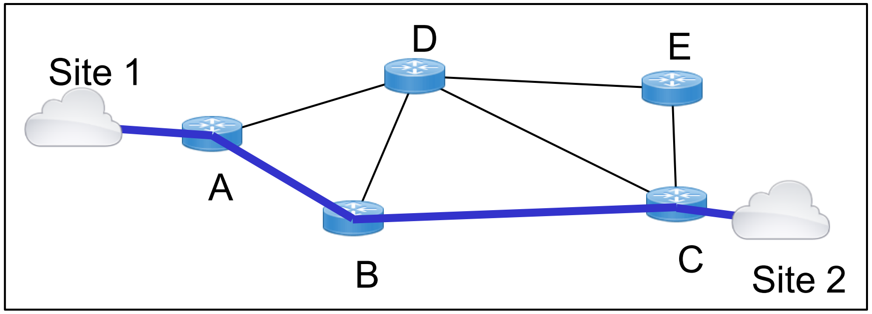
$$= 1 - [(1 - 0.994 \times 0.9999 \times 0.989) \times (1 - 0.992 \times 0.9999 \times 0.99)] = 0.9997$$

$$A = a_A \times A_{ADC//ABC} \times a_C$$

$$= 0.9999 \times 0.9997 \times 0.9999 = 0.995$$

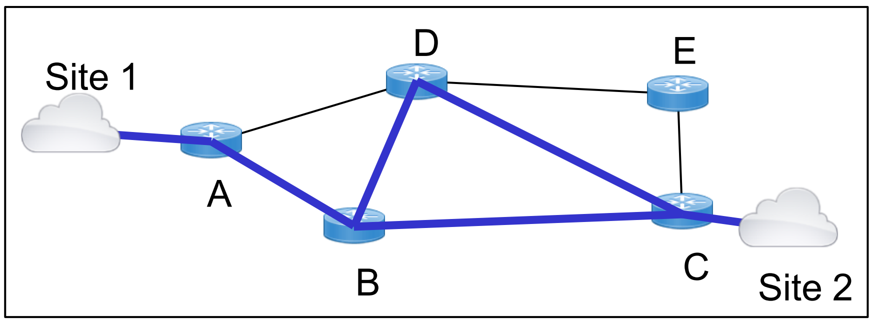
Availability of the example

Comparing the 3 previous solutions:



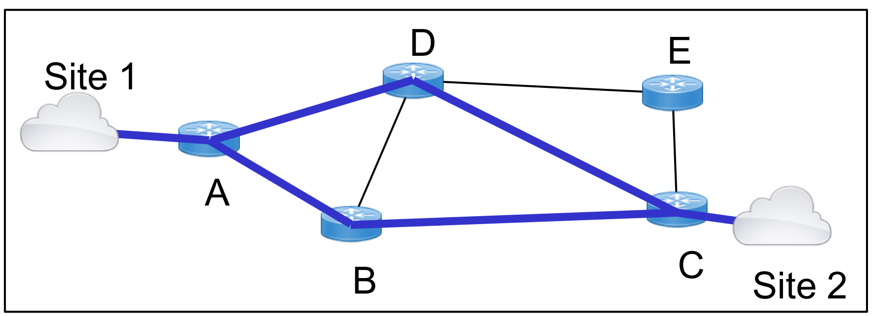
$$A = 0.9818$$

The service fails on average
(1 – 0.9818) × 365.25
= 6.65 days/year



$$A = 0.9915$$

The service fails on average
(1 – 0.9915) × 365.25
= 3.1 days/year



$$A = 0.995$$

The service fails on average
(1 – 0.995) × 365.25
= 1.8 days/year

Availability model for link availability in telecommunication networks

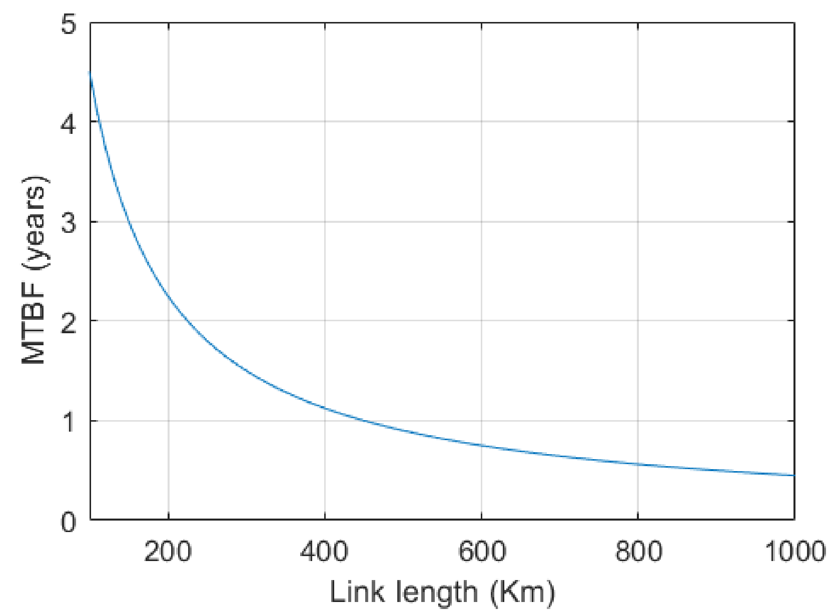
Following [1], an availability model of the optical links in the USA networks was at the beginning of this century given approximately by:

$$\frac{MTBF}{MTBF + MTTR}$$

$MTTR$ = 24 hours

$$MTBF = \frac{CC \times 365 \times 24}{\text{link length [Km]}} \text{ [hours]}$$

CC ($Cable\ Cut\ metric$) = 450 Km



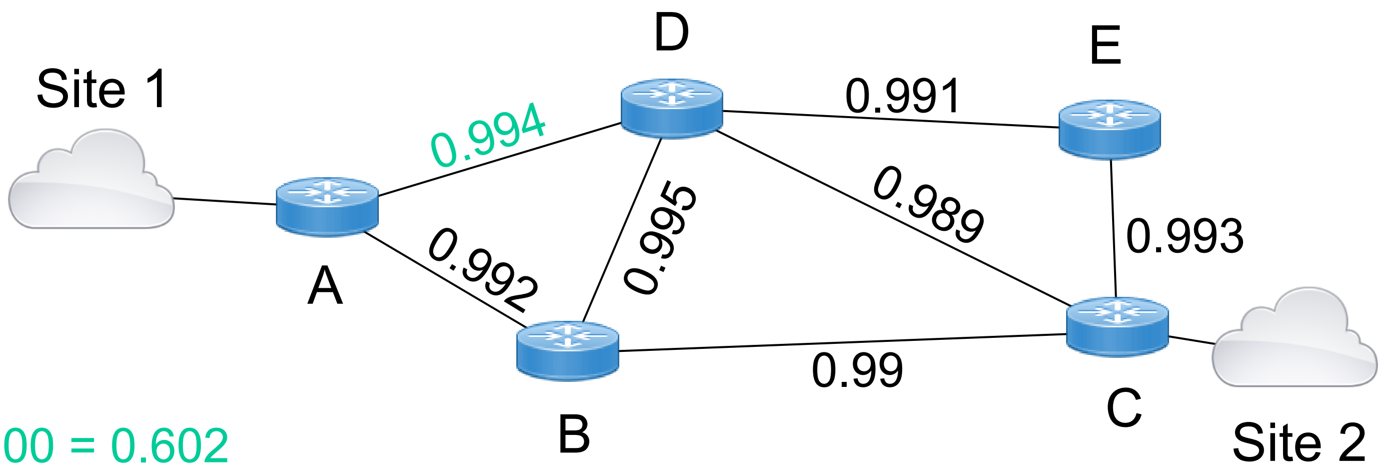
[1] J.-P. Vasseur, M. Pickavet and P. Demeester, “Network Recovery: Protection and Restoration of Optical, SONET-SDH, IP, and MPLS”, Elsevier (2004)

Determination of the *k*-most available routing paths

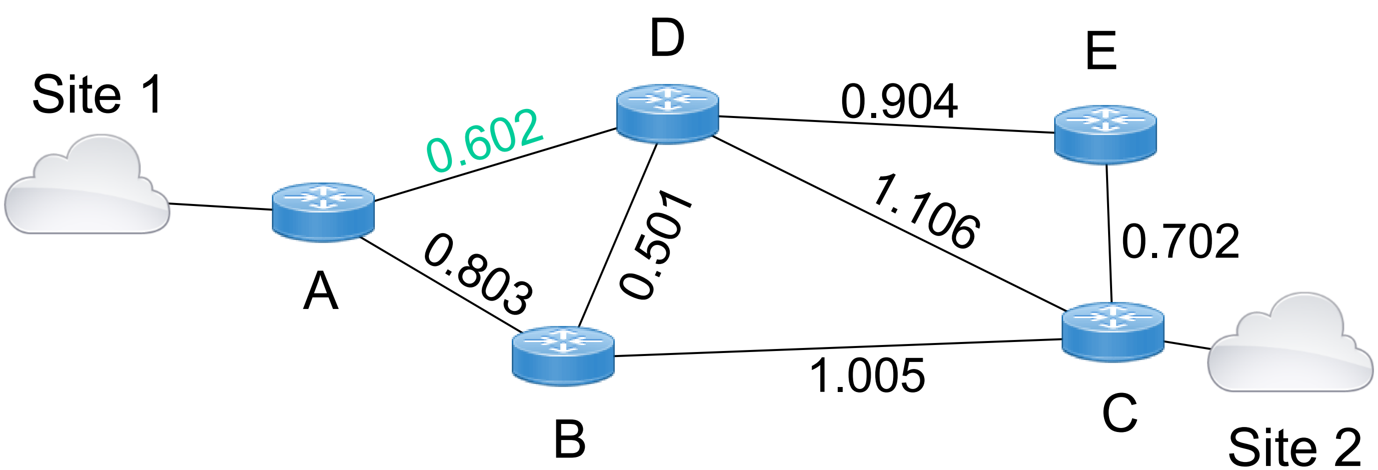
- Consider:
  - a telecommunications network in which the availability of nodes (*routers*, *switches*, *optical cross-connects*) is 1.0.
  - the set P of all routing paths in the network from a given origin node to a given destination node.
- The availability a_p of each routing path $p \in P$ is, therefore, the product of the availabilities of all links belonging to the path.
- The logarithm of the path availability $log(a_p)$ is then the sum of the logarithms of the availabilities of each link.
- The logarithm function is monotonically increasing. So, the k^{th} most available routing path is also the k^{th} routing path with the highest logarithm value of its availability.
- Note that $log(a_p)$ is negative. So, if we consider the "length" of each link as $-log(a_p)$, the k-most available routing paths can be determined by any *k*-shortest path algorithm.

Illustration of the “lengths” of each link in the *k*-most available routing paths

Consider again the example of a VPN service between two sites of a company. The top figure indicates the availability of each link. The bottom figure indicates the resulting link “lengths”.



$-\,log(0.994) \times 100 = 0.602$



Availability of a telecommunications netowrk

Consider a network with *n* links and all links with the same availability *a.* In this case, the number *i* of unavailable links is a binomial random variable with probability *p* = 1 – *a*.

So, the probability of *i* unavailable links is:

$$f(i) = \binom{n}{i} p^i (1-p)^{n-i}, i = 0, 1, 2, \ldots, n$$

The probability *P* of 2 or more links being unavailable is:

CONCLUSION:
In the vast majority of the cases such that there are link failures, only one link is unavailable.

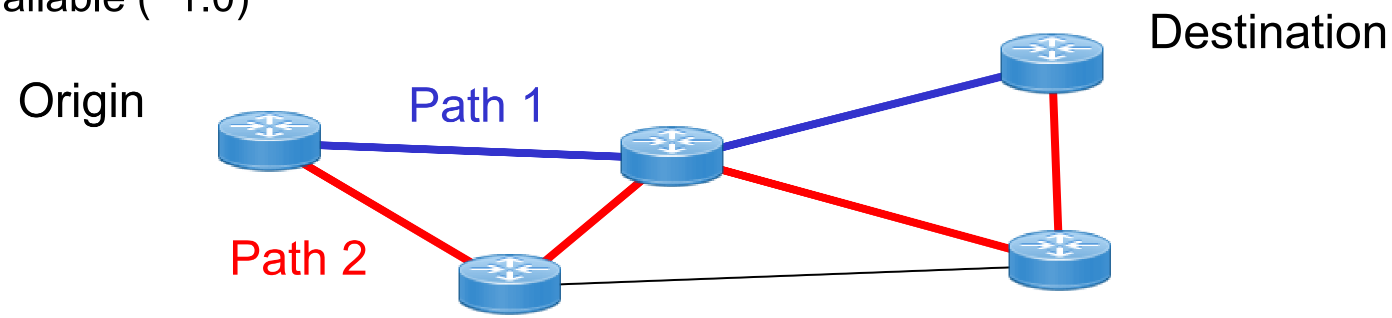
a	*n*	*P*
0.995	10	0.110%
0.995	20	0.447%
0.995	40	1.719%
0.999	10	0.004%
0.999	20	0.019%
0.999	40	0.076%

Resilience of services to network failures

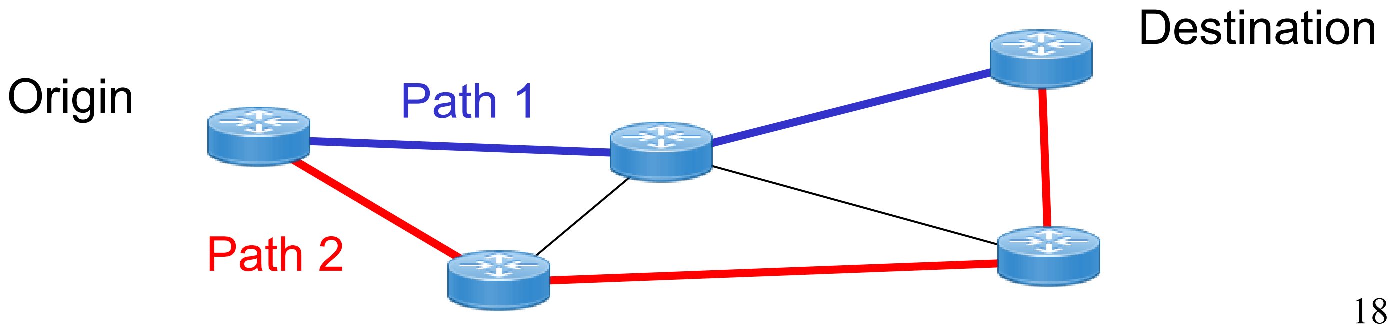
- The resilience of a telecommunications network is generically defined as its capacity to maintain its supported services when one or more of its elements (nodes and/or links) fail.
- The resilience of a network can be enhanced by two types of schemes.
- **Restauration scheme**: the services are routed assuming no failures; when a failure occurs, the network tries to reroute as much as possible the services through the surviving resources (i.e., through the nodes and links that did not fail).
  – Examples: IP networks with traditional routing protocols (such as RIP or OSPF)
- **Protection scheme**: the resources of the network are assigned (to the different flows of the different services) considering not only the no failure case but also a subset of possible failure cases; if one of such cases occurs, the resource assignment guarantees that the services do not fail.
  – The typical cases of interest are the cases with a single element failure (the most likely cases as illustrated in the previous slide)

Protection schemes based on disjoint pairs of routing paths

- Each flow (from its origin node to its destination node) is supported by two disjoint routing paths (both starting on the origin node and finishing on the destination node):
  - disjoint on links (i.e., without common links) used when nodes are highly available (~1.0)



  - disjoint on nodes and links (i.e., without common link nor intermediate nodes) used when the availability of nodes is not close to 1.0



Protection schemes based on disjoint pairs of routing paths

- A pair of link disjoint routing paths protects the flow to any individual link failure.
- A pair of node and link disjoint routing paths protects the flow to any individual element failure (node or link) except the failure of the origin node or the destination node.

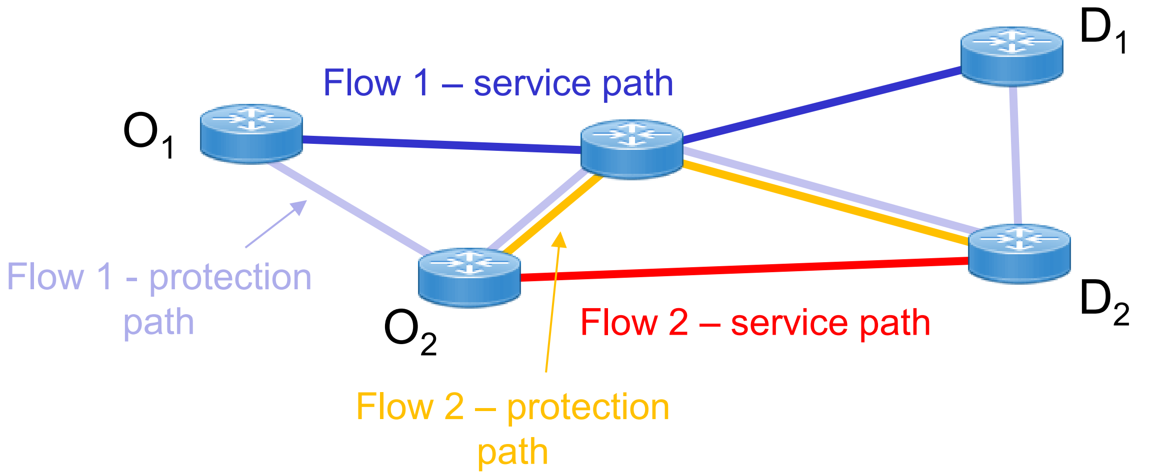
- In both cases, there are two possible protection schemes:
- **Protection 1+1** (one-plus-one): the flow is sent duplicated through the 2 routing paths.
  - The failure recovering time is negligible (i.e., very short).
  - It requires the duplication of network resources used for each flow.
- **Protection 1:1** (one-to-one): the flow is routed by one of the routing paths (named *service path*) and the other routing path (named *protection path*) is used only when the first path fails.
  - The failure recovery time is not negligible (the origin node must be notified of the service path failure to start transmitting through the protection path).
  - The resources of the protection path can be shared by different flows (or used by other lower priority services).

Protection schemes based on disjoint pairs of routing paths - EXAMPLE

Consider the bottom network supporting two flows of a service:

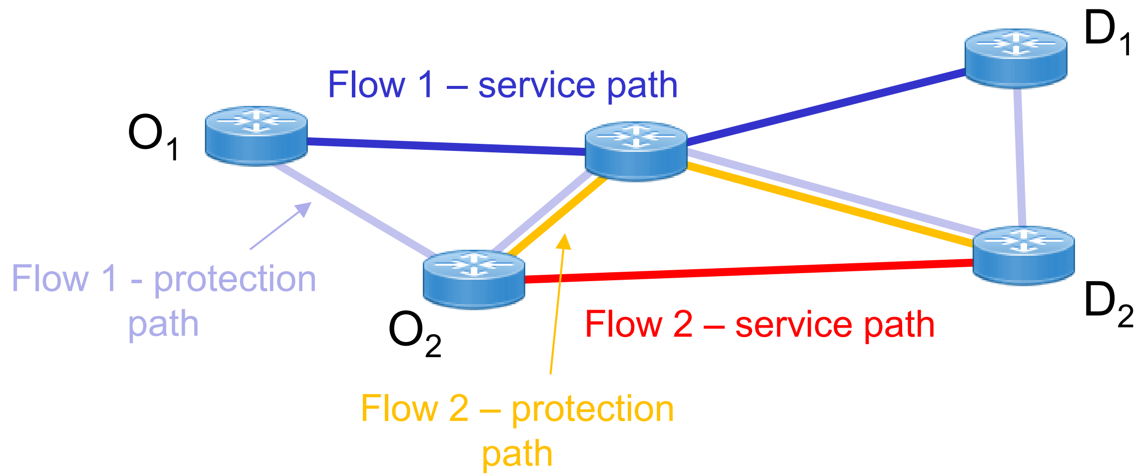
- flow 1 of 10 Gbps from router O_1 to router D_1
- flow 2 of 20 Gbps from router O_2 to router D_2

The figure shows the pair of routing paths used to reoute each flow.



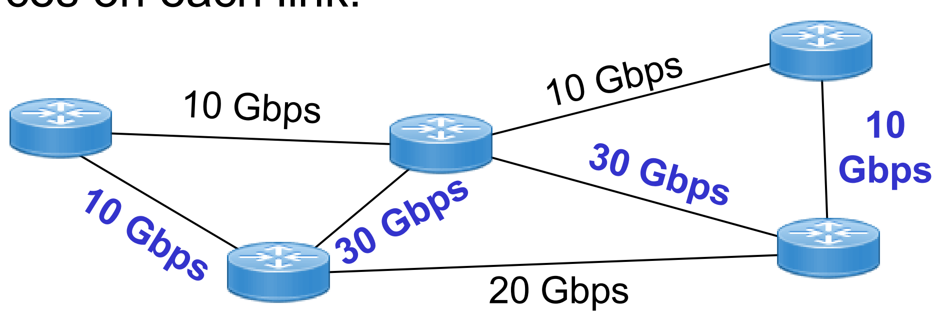
Question: How much resources are needed on each link to support the two flows when the protection schemes is 1+1 and when it is 1:1?

Protection schemes based on disjoint pairs of routing paths - EXAMPLE

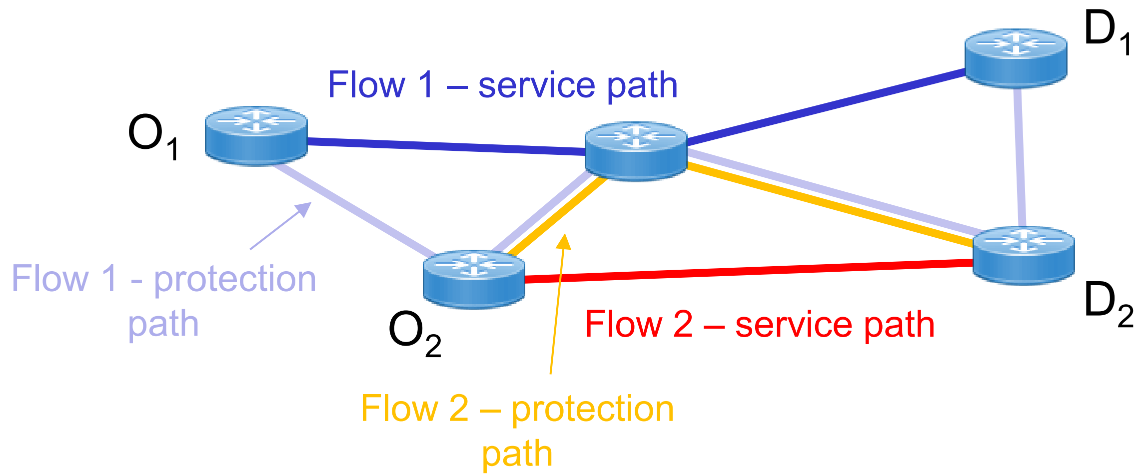


Protection 1+1 - the 10 Gbps of flow 1 and the 20 Gbps of flow 2 are transmitted by the two routing paths of each flow.

Required resources on each link:



Protection schemes based on disjoint pairs of routing paths - EXAMPLE



Protection 1:1 – since the service paths are disjoint, a single element failure can disrupt only one of them. So, in the links of the protection paths, the resources required are only the highest throughput among both flows.

Required resources on each link: