Sisteme Tolerante la Defecte
Sisteme Distribuite

Lect. Dr. Ing. Cristian Chilipirea – cristian.chilipirea@mta.ro

Când discutăm de sisteme distribuite?

Când discutăm de sisteme distribuite?

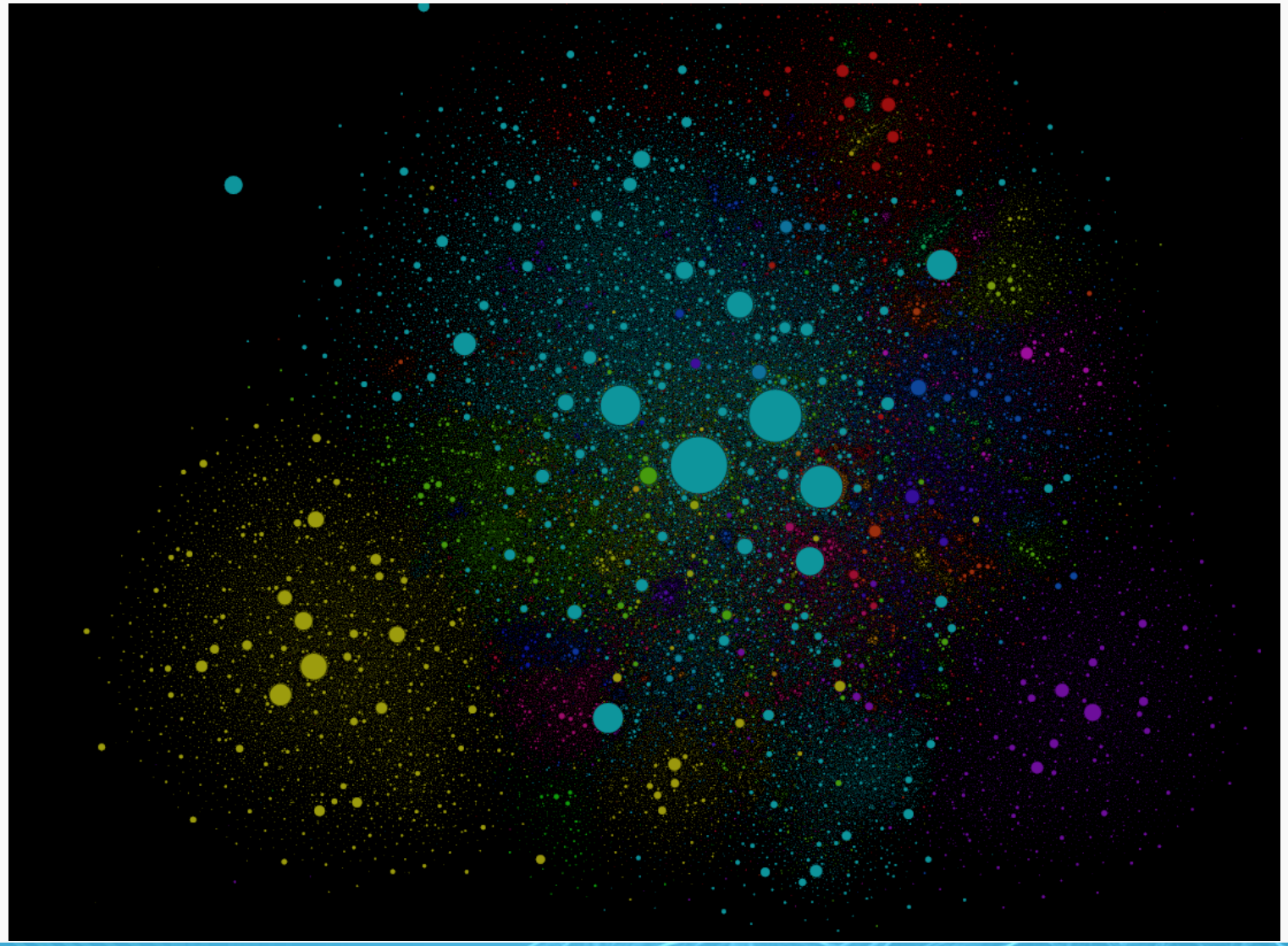
- P2P
- Server-Client
- Clustere
- Cloud
- Sisteme de calcul intensiv (supercomputer)
- Rețele ad-hoc (ex. rețele formate **doar** din senzori WiFi sau telefoane mobile)

Când discutăm de sisteme distribuite?

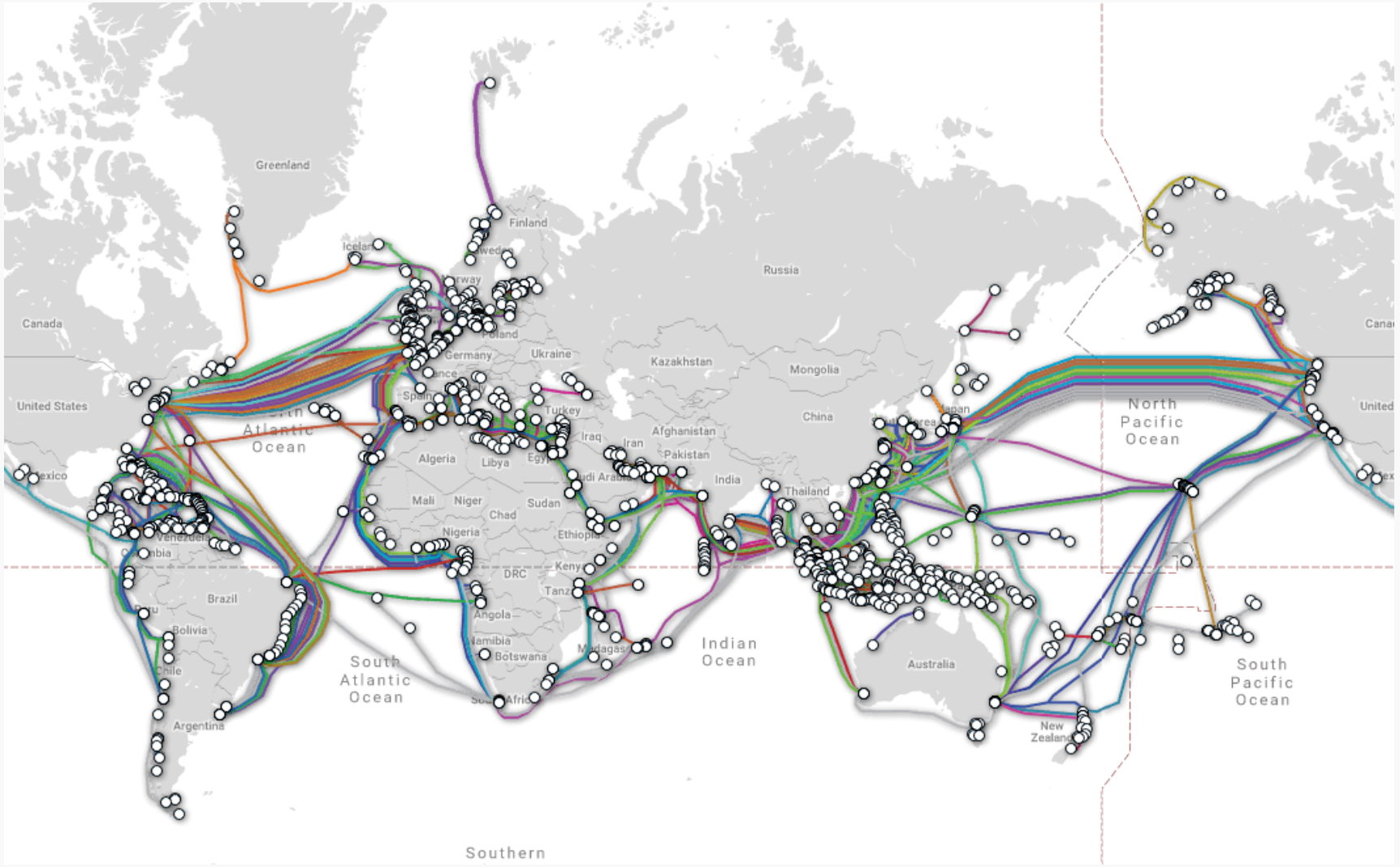
Oricând avem mai mult de un sistem

Care este cel mai mare sistem distribuit?

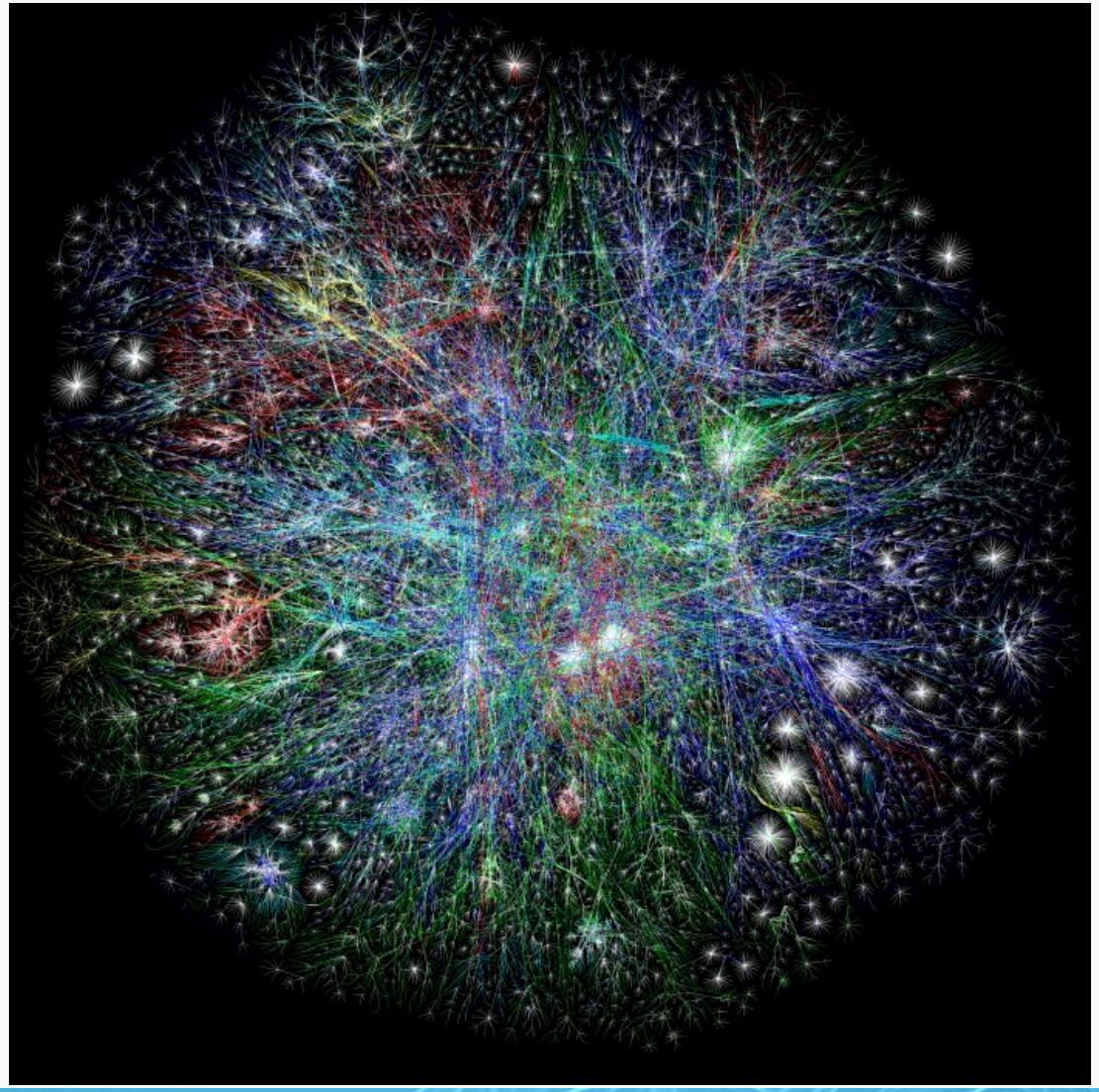
Care este cel mai mare sistem distribuit?



Care este cel mai mare sistem distribuit?



Care este cel mai mare sistem distribuit?



Care este cel mai mare sistem distribuit?

Care este cel mai mare sistem distribuit?

Care este cel mai mare sistem distribuit?

The INTERNET

Distributed Systems – The INTERNET

Distributed Web-based systems

The World Wide Web (**WWW**) can be viewed as a huge distributed system
ng of millions of clients and servers for accessing linked documents.
maintain collections of documents, while clients provide users an
use interface for presenting and accessing those documents.
Web started as a project at CERN, the European Particle Physics
ory in Geneva, to let its large and geographically dispersed group

Cine a construit Internetul?

Cine a construit Internetul? - Vint Cerf

A Protocol for Packet Network Intercommunication

VINTON G. CERF AND ROBERT E. KAHN,
MEMBER, IEEE

Abstract — A protocol that supports the sharing of resources that exist in different packet switching networks is presented. The protocol provides for variation in individual network packet sizes, transmission failures, sequencing, flow control, end-to-end error checking, and the creation and destruction of logical process-to-process connections. Some implementation issues are considered, and problems such as internetwork routing, accounting, and timeouts are exposed.

INTRODUCTION

IN THE LAST few years considerable effort has been expended on the design and implementation of packet switching networks [1]-[7],[14],[17]. A principle reason for developing such networks has been to facilitate the sharing of computer resources. A packet communication network includes a transportation mechanism for delivering data between computers or between computers and terminals. To make the data meaningful, computer and terminals

of one or more *packet sw*
communication media th
switches. Within each H
exist *processes* which
processes in their own o
definition of a process
purposes [13]. These pr
ultimate source and de
network. Typically, witl
there exists a protocol fo
any source and destinatio
and destination processes
convention for commu
Processes in two distinct
use different protocols
ensemble of packet swi
media is called the *pack*

Stiva OSI?

Stiva OSI?

OSI Model				
Layer			Protocol data unit (PDU)	Function[3]
Host layers	7	Application	Data	High-level APIs, including resource sharing, remote file access
	6	Presentation		Translation of data between a networking service and an application; including character encoding, data compression and encryption/decryption
	5	Session		Managing communication sessions, i.e. continuous exchange of information in the form of multiple back-and-forth transmissions between two nodes
	4	Transport	Segment, Datagram	Reliable transmission of data segments between points on a network, including segmentation, acknowledgement and multiplexing
Media layers	3	Network	Packet	Structuring and managing a multi-node network, including addressing, routing and traffic control
	2	Data link	Frame	Reliable transmission of data frames between two nodes connected by a physical layer
	1	Physical	Symbol	Transmission and reception of raw bit streams over a physical medium

Stiva TCP/IP?

Stiva TCP/IP?

OSI	TCP/IP
Application	
Presentation	
Session	Data
Transport	TCP
Network	IP
Data link	Ethernet
Physical	Cupru

Granții stiva TCP/IP?

OSI	TCP/IP	Garanții
Application		
Presentation		
Session	Data (HTTP)	
Transport	TCP	
Network	IP	
Data link	Ethernet	
Physical	Cupru	Mesajul e transmis

Granții stiva TCP/IP?

OSI	TCP/IP	Garanții
Application		
Presentation		
Session	Data (HTTP)	
Transport	TCP	
Network	IP	
Data link	Ethernet	Best efort ca mesajul să ajungă unde trebuie în rețea
Physical	Cupru	Mesajul e transmis

Granții stiva TCP/IP?

OSI	TCP/IP	Garanții
Application		
Presentation		
Session	Data (HTTP)	
Transport	TCP	
Network	IP	Best efort ca mesajul să ajungă unde trebuie în Internet
Data link	Ethernet	Best efort ca mesajul să ajungă unde trebuie în rețea
Physical	Cupru	Mesajul e transmis

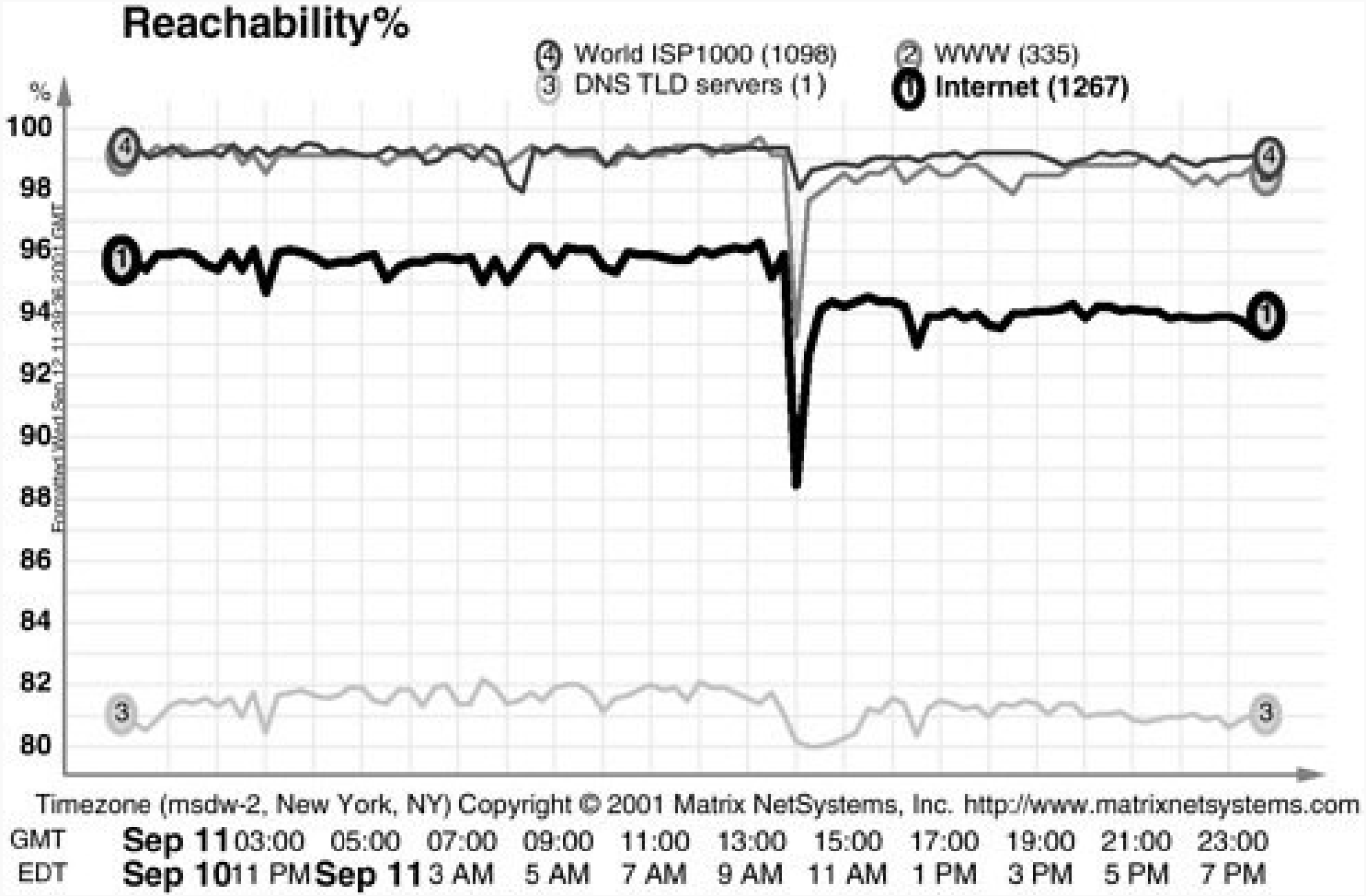
Granții stiva TCP/IP?

OSI	TCP/IP	Garanții
Application		
Presentation		
Session	Data (HTTP)	
Transport	TCP	Mesajul ajunge
Network	IP	Best efort ca mesajul să ajungă unde trebuie în Internet
Data link	Ethernet	Best efort ca mesajul să ajungă unde trebuie în rețea
Physical	Cupru	Mesajul e transmis

Granții stiva TCP/IP?

OSI	TCP/IP	Garanții
Application		
Presentation		
Session	Data (HTTP)	Mesajul va fi înțeles de server/browser
Transport	TCP	Mesajul ajunge
Network	IP	Best efort ca mesajul să ajungă unde trebuie în Internet
Data link	Ethernet	Best efort ca mesajul să ajungă unde trebuie în rețea
Physical	Cupru	Mesajul e transmis

Deci cât de bun e internetul ca sistem distribuit?



Cine a inventat Ethernet?

Cine a inventat Ethernet? - Robert Metcalfe

Computer Systems — G. Bell, S. Fuller and D. Siewiorek, Editors

Ethernet: Distributed Packet Switching for Local Computer Networks

Robert M. Metcalfe and David R. Boggs
Xerox Palo Alto Research Center

Ethernet is a branching broadcast communication system for carrying digital data packets among locally distributed computing stations. The packet transport mechanism provided by Ethernet has been used to build systems which can be viewed as either local computer networks or loosely coupled multiprocessors. An Ethernet's shared communication facility, its Ether, is a pas-

1. Background

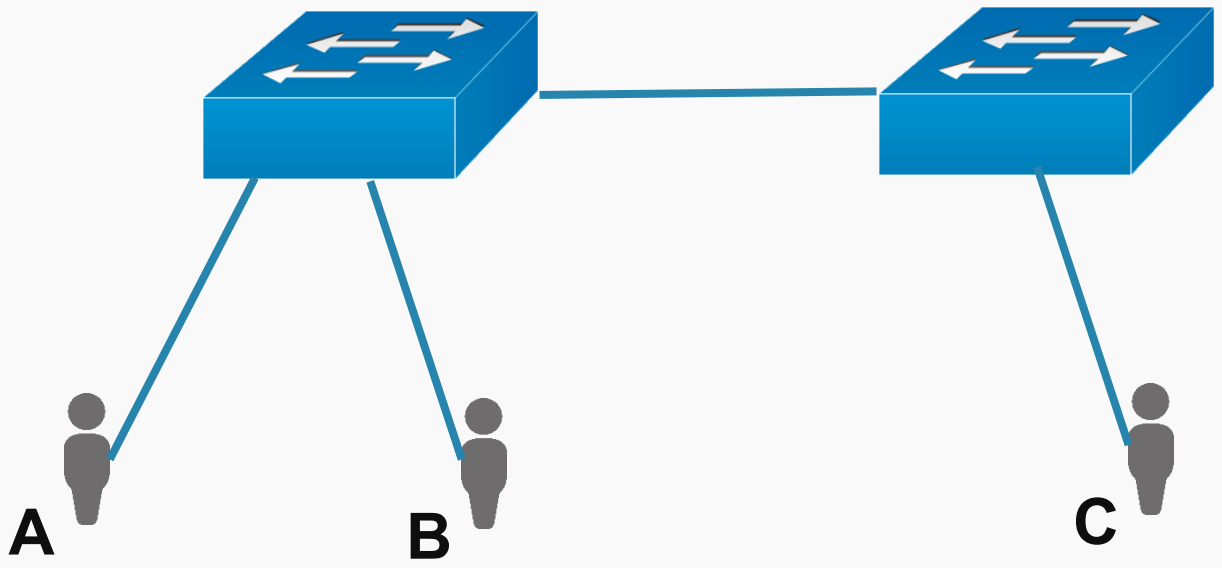
One can characterize distributed computing as a spectrum of activities varying in their degree of decentralization, with one extreme being remote computer networking and the other extreme being multiprocessing. Remote computer networking is the loose interconnection of previously isolated, widely separated, and rather large computing systems. Multiprocessing is the construction of previously monolithic and serial computing systems from increasingly numerous and smaller pieces computing in parallel. Near the middle of this spectrum is local networking, the interconnection of computers to gain the resource sharing of computer networking and the parallelism of multiprocessing.

The separation between computers and the associated bit rate of their communication can be used to divide the distributed computing spectrum into broad activities. The product of separation and bit rate, now about 1 gigabit-meter per second (1 Gbmps), is an indication of the limit of current communication technology and can be expected to increase with time:

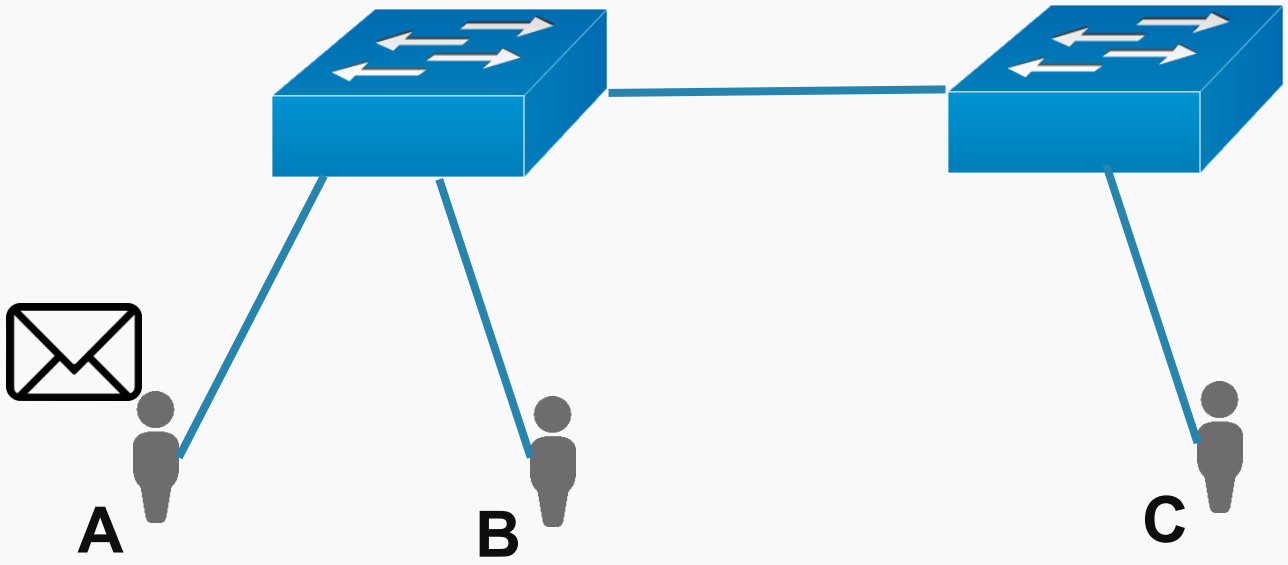
Activity	Separation	Bit rate
Remote networks	> 10 km	< .1 Mbps
Local networks	10–.1 km	.1–10 Mbps
Multiprocessors	< .1 km	> 10 Mbps

1.1 Remote Computer Networking

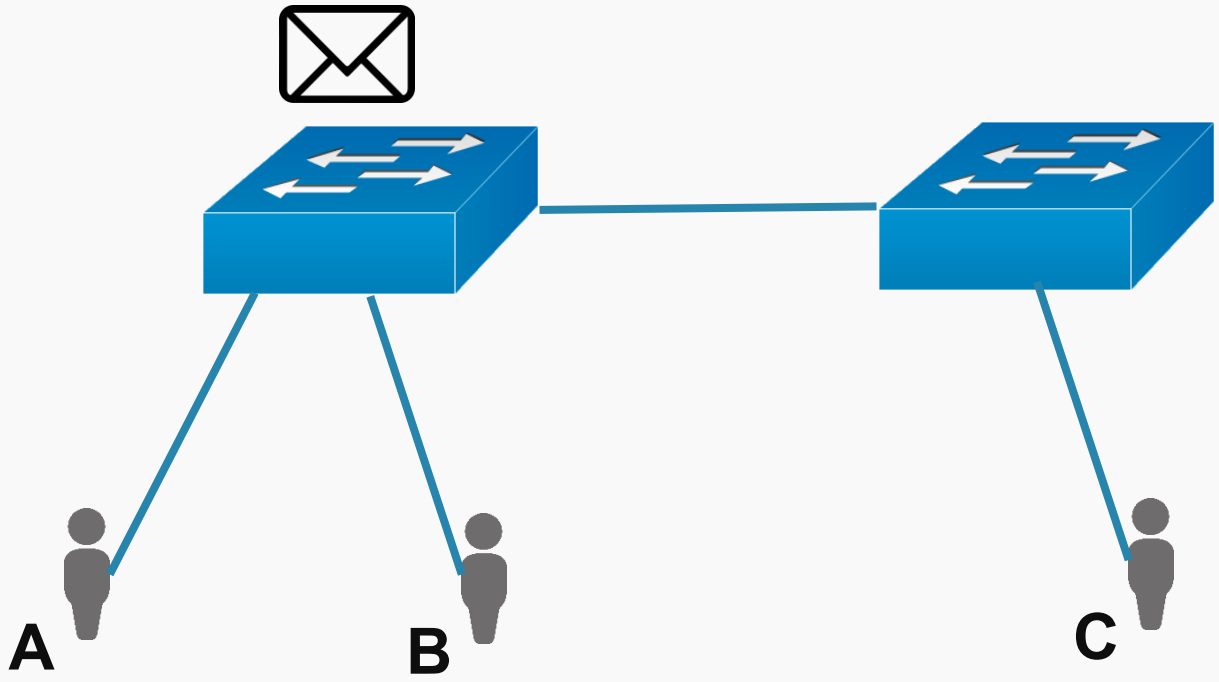
Ethernet – CAM table



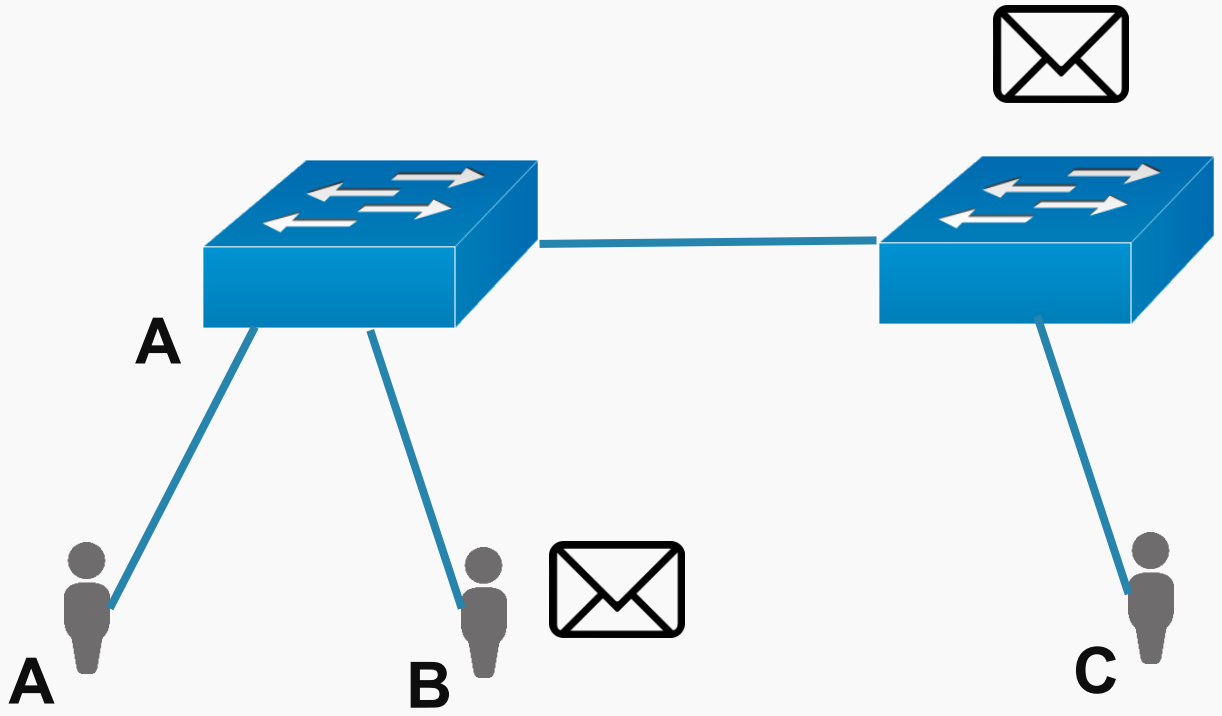
Ethernet – CAM table



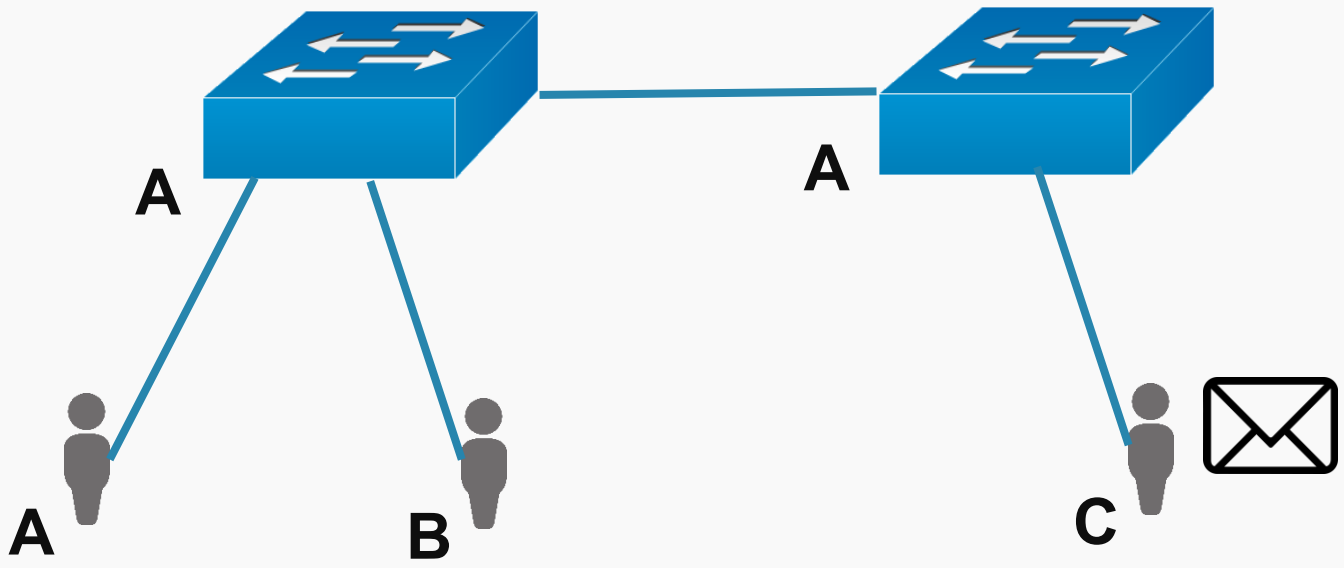
Ethernet – CAM table



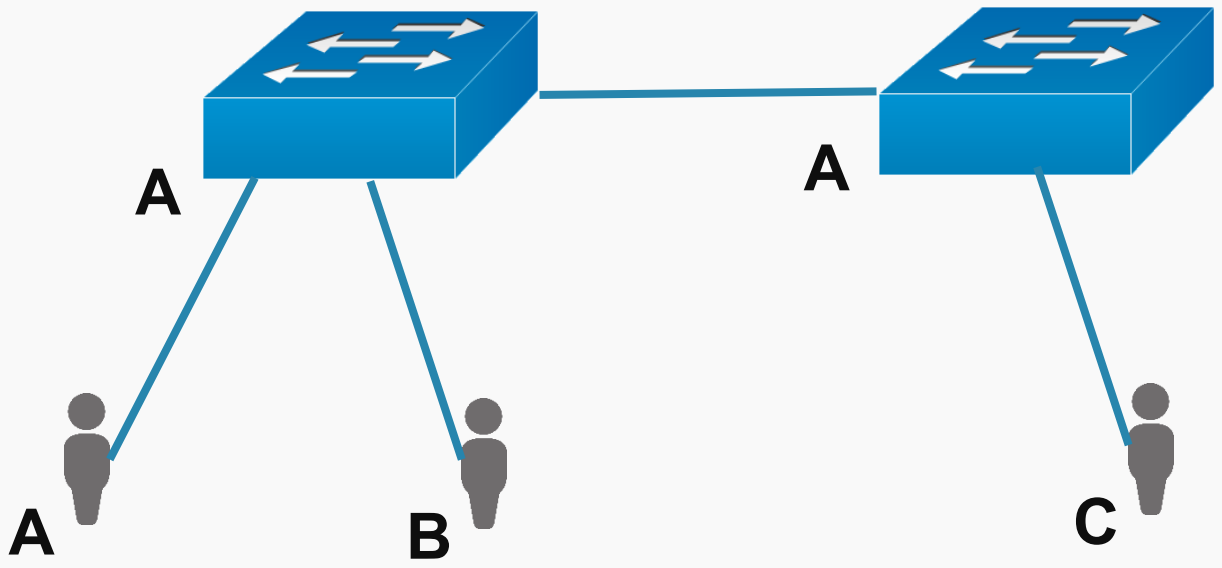
Ethernet – CAM table



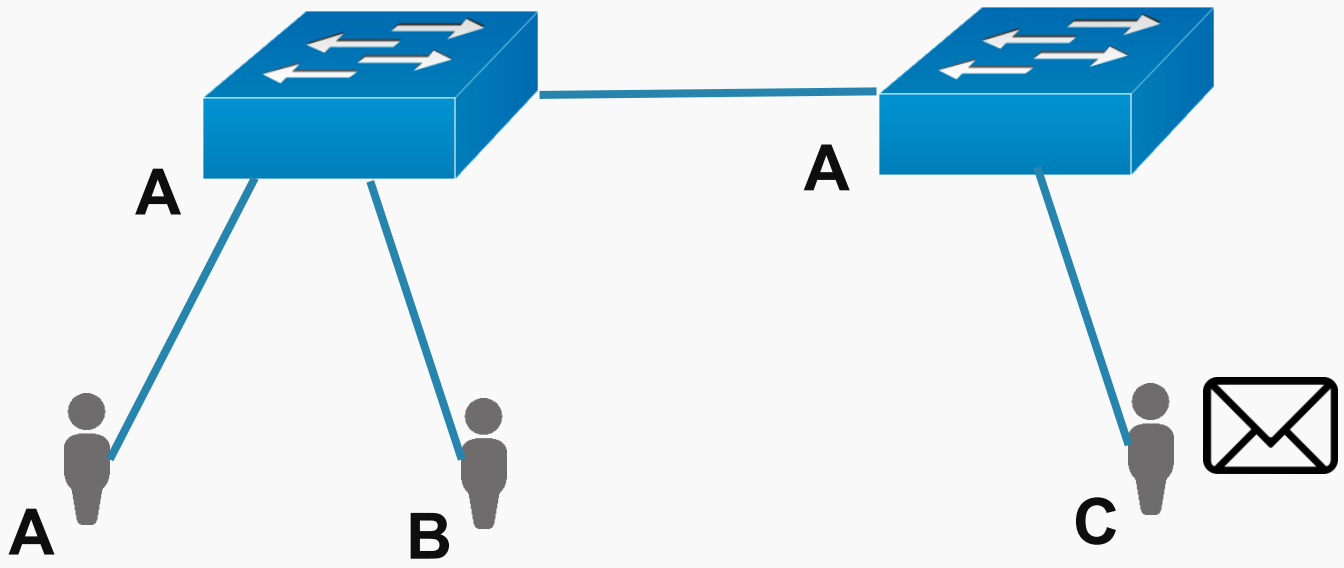
Ethernet – CAM table



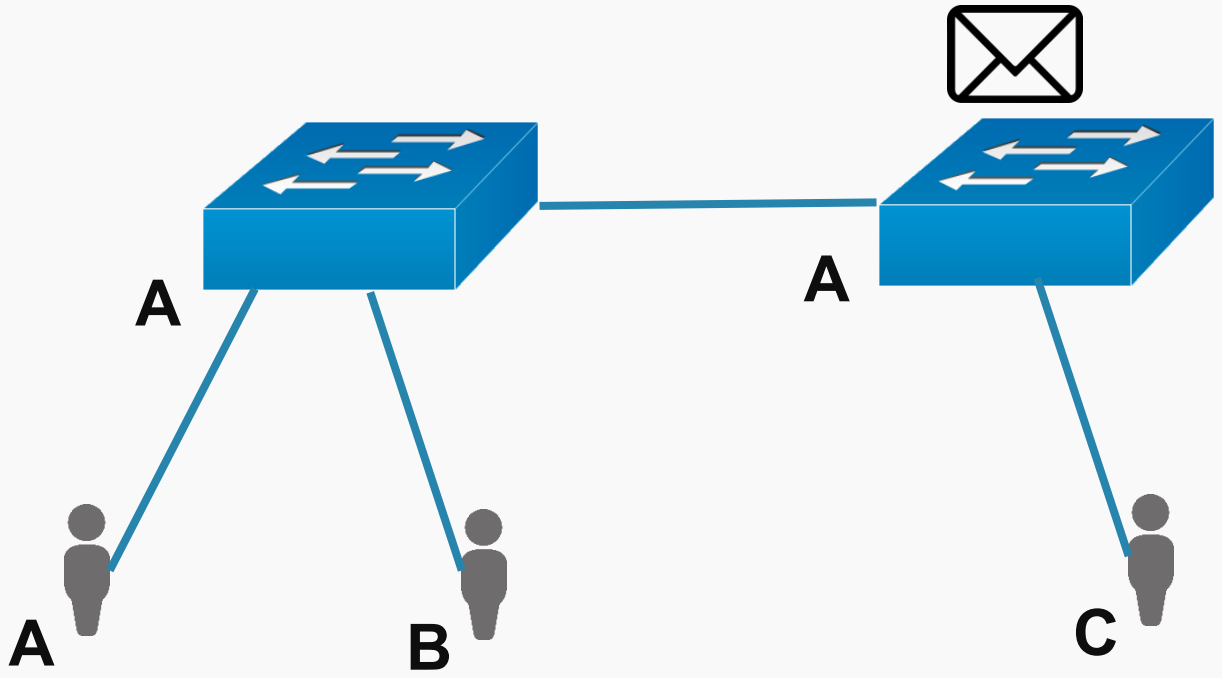
Ethernet – CAM table



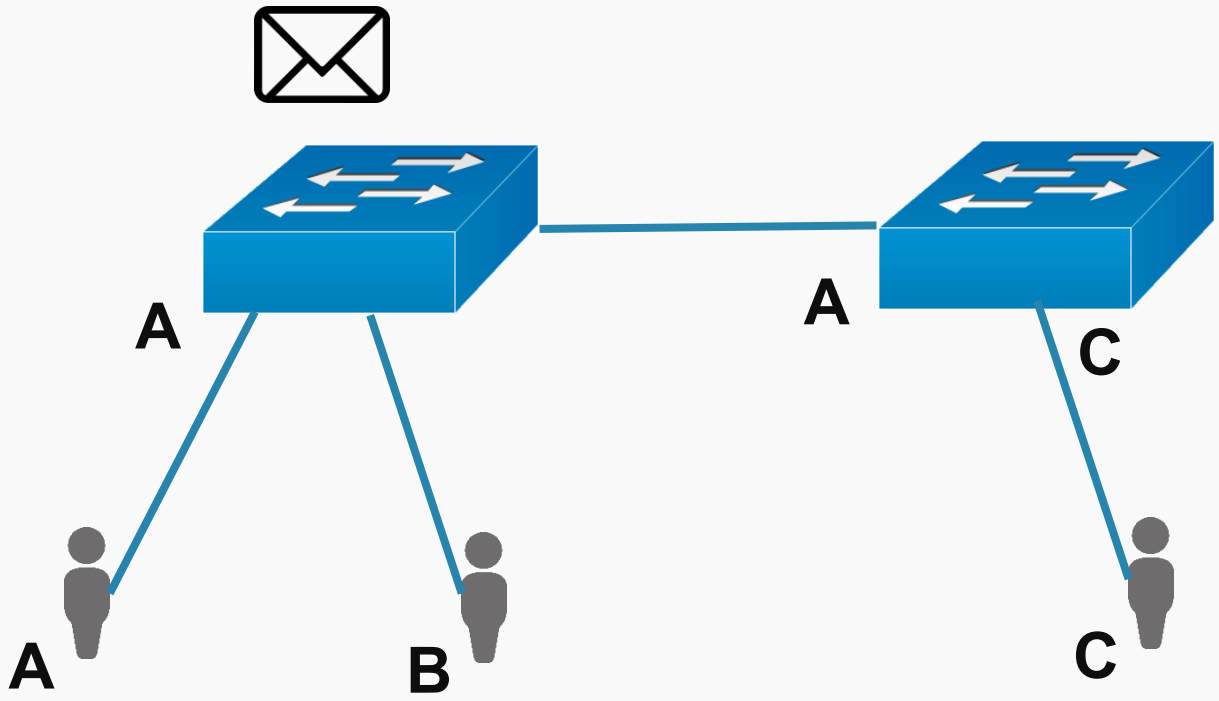
Ethernet – CAM table



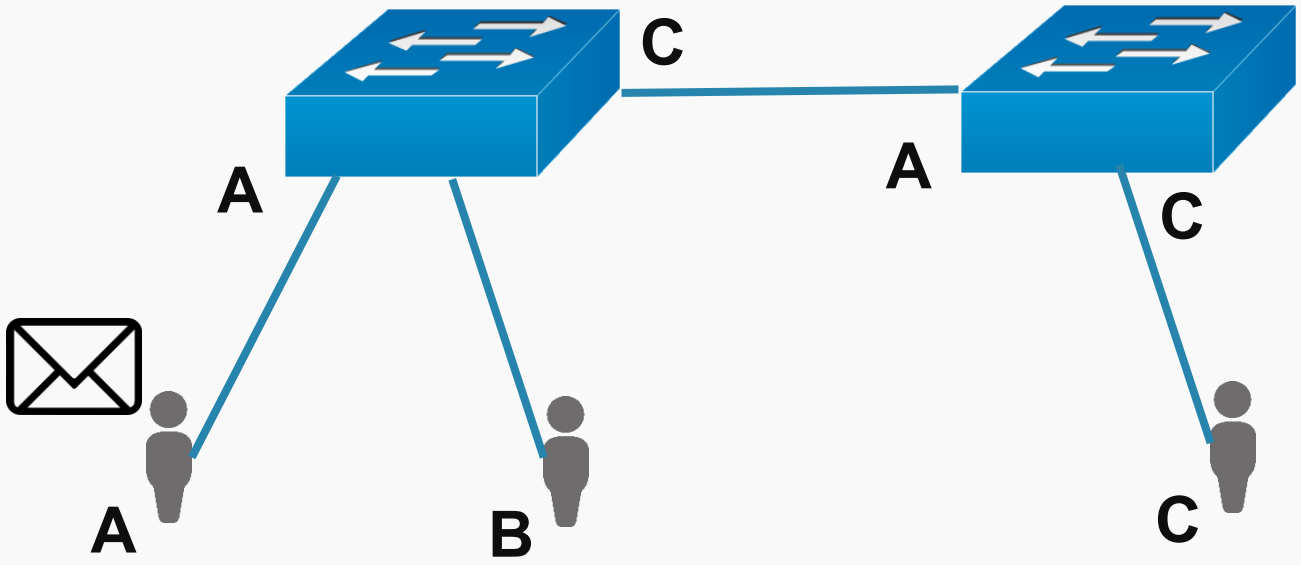
Ethernet – CAM table



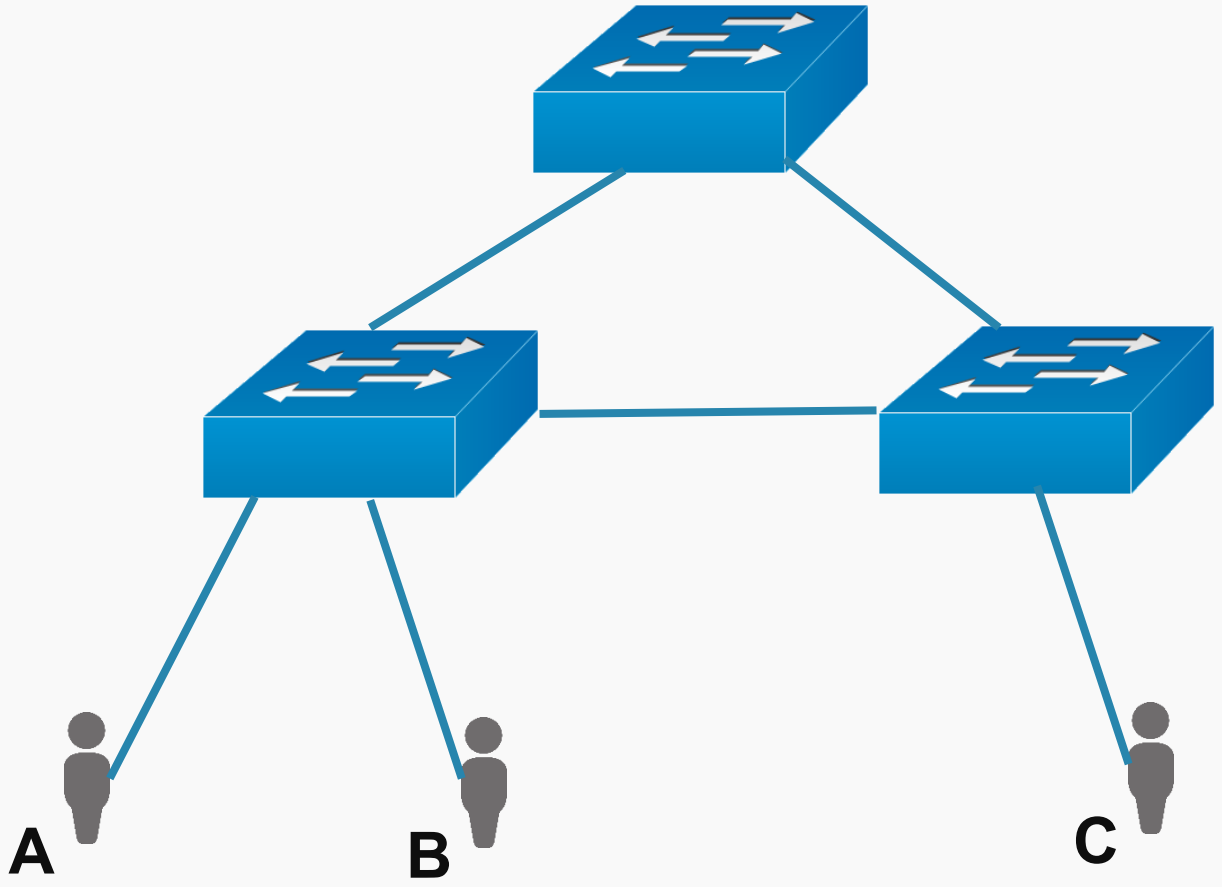
Ethernet – CAM table



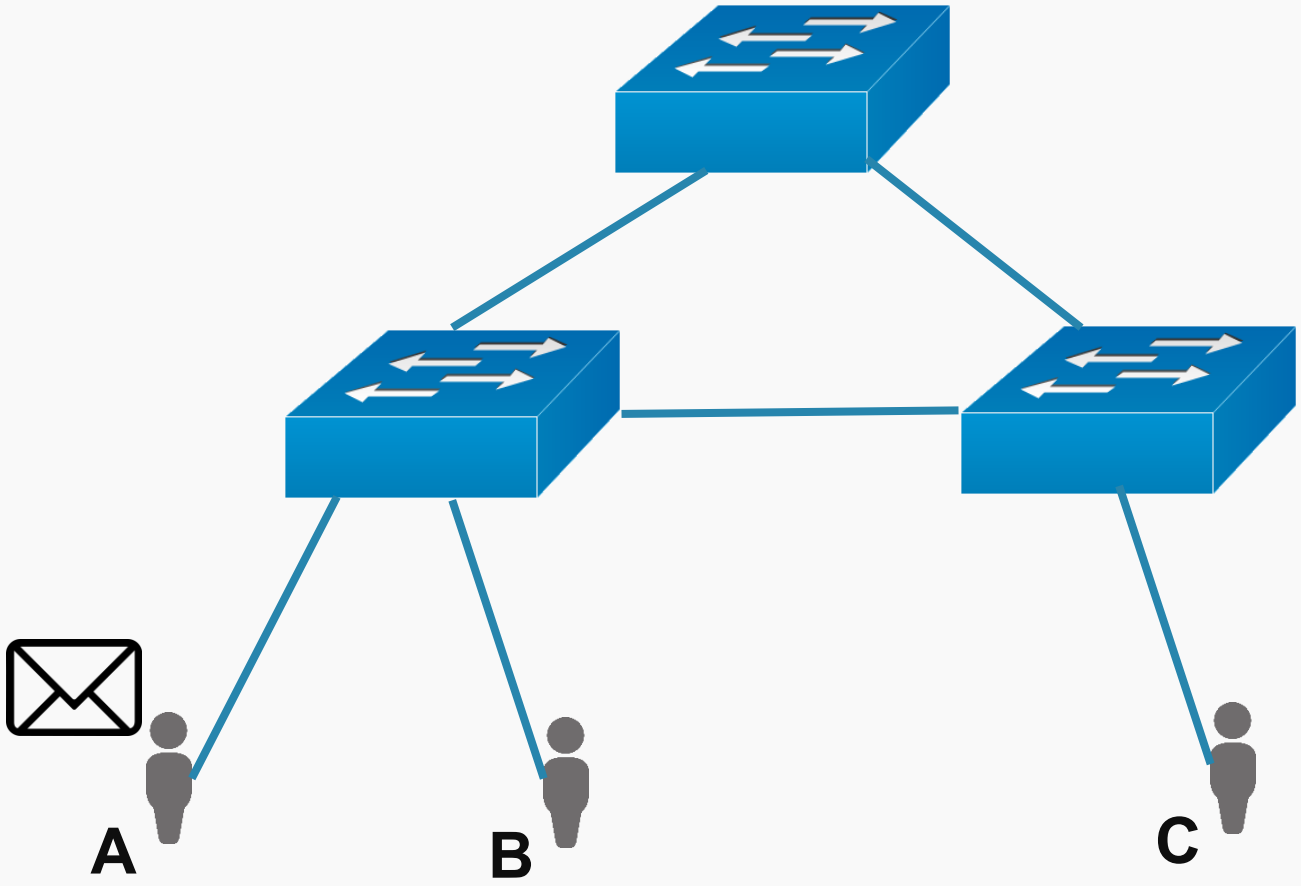
Ethernet – CAM table



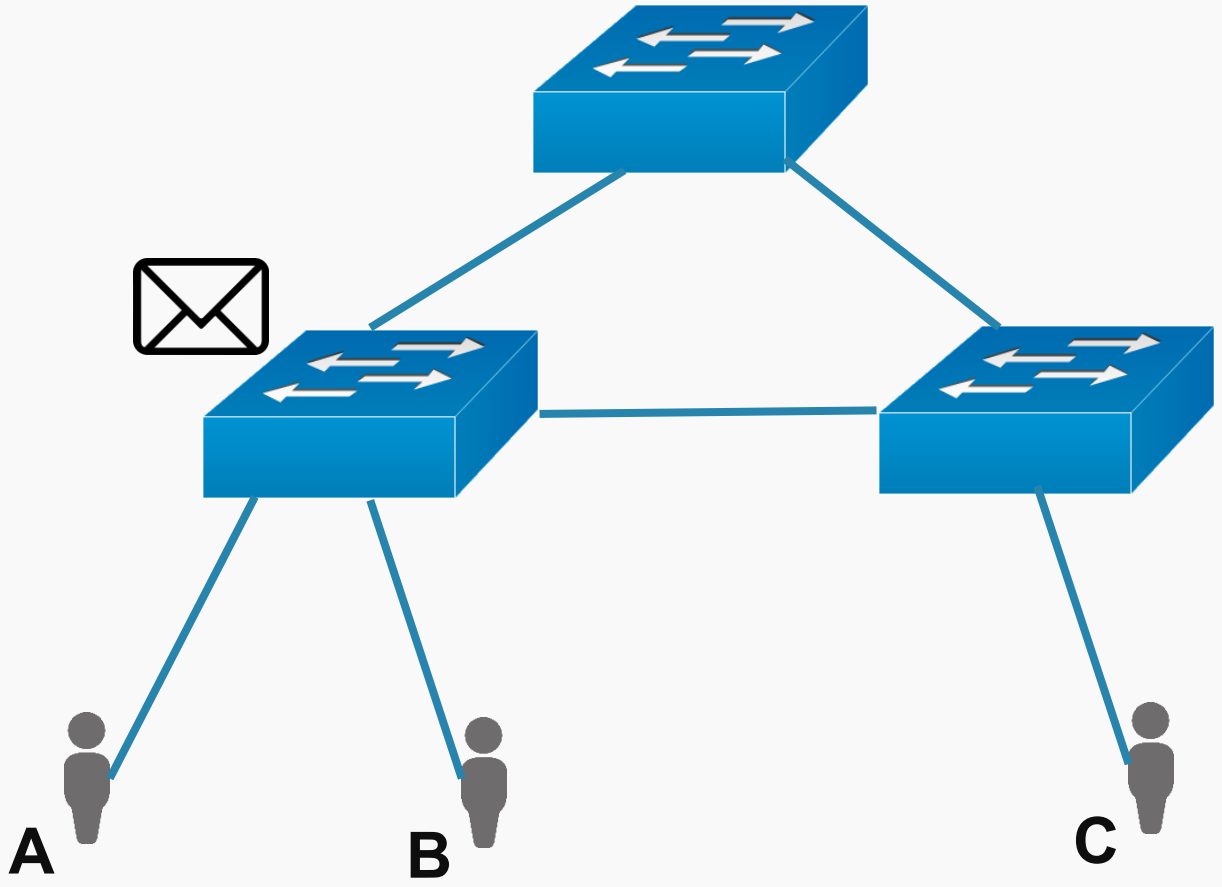
Ethernet – CAM table – Probleme cu cicluri



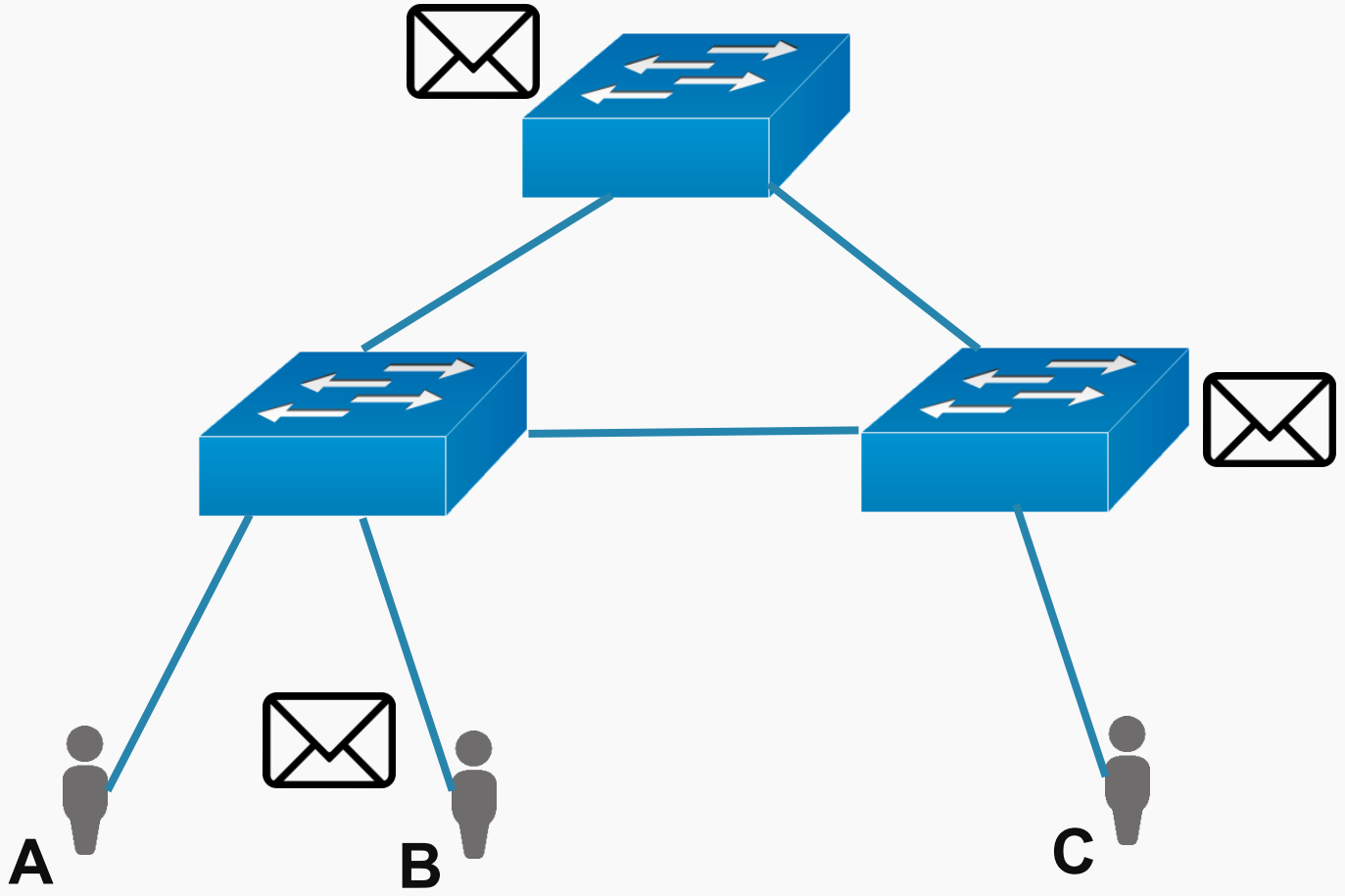
Ethernet – CAM table – Probleme cu cicluri



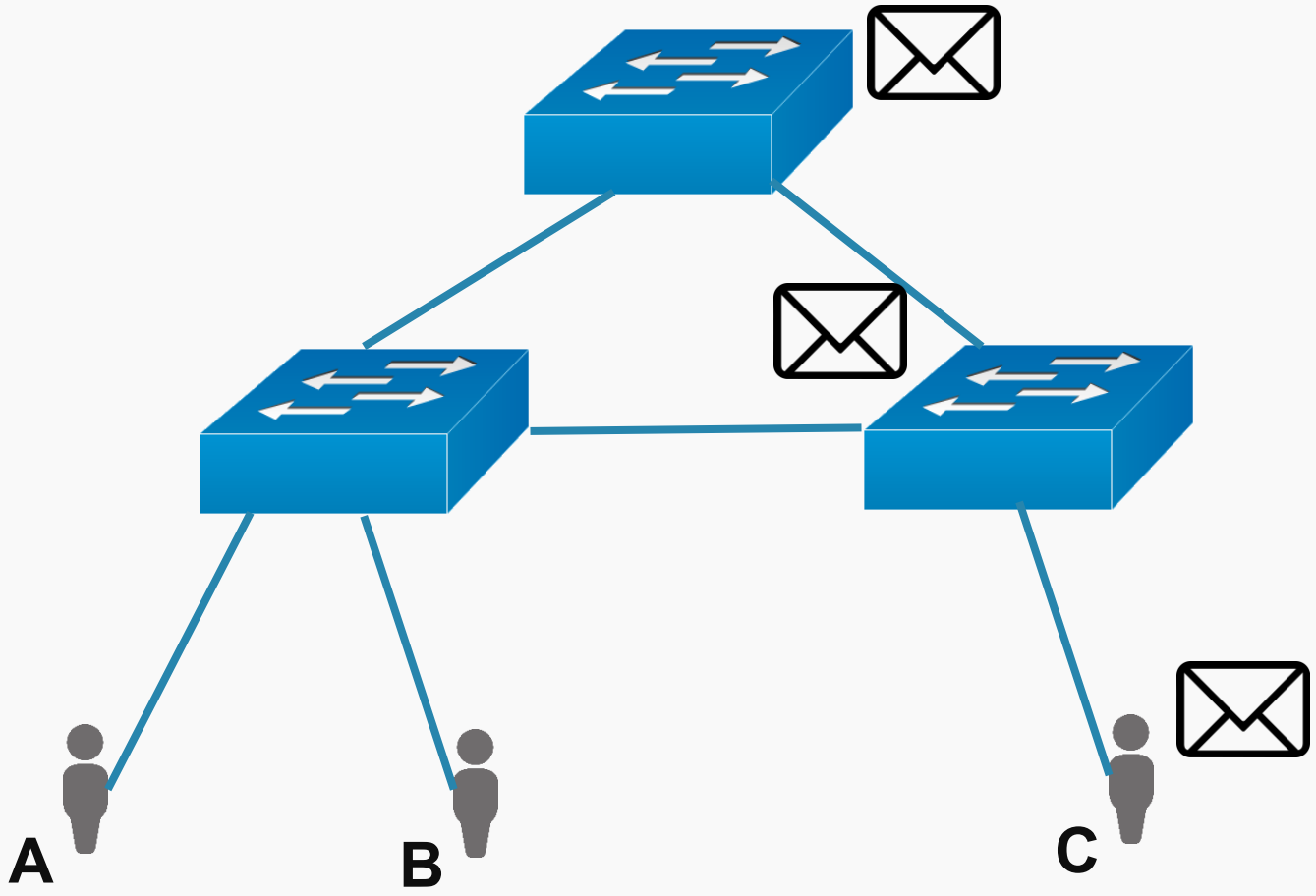
Ethernet – CAM table – Probleme cu cicluri



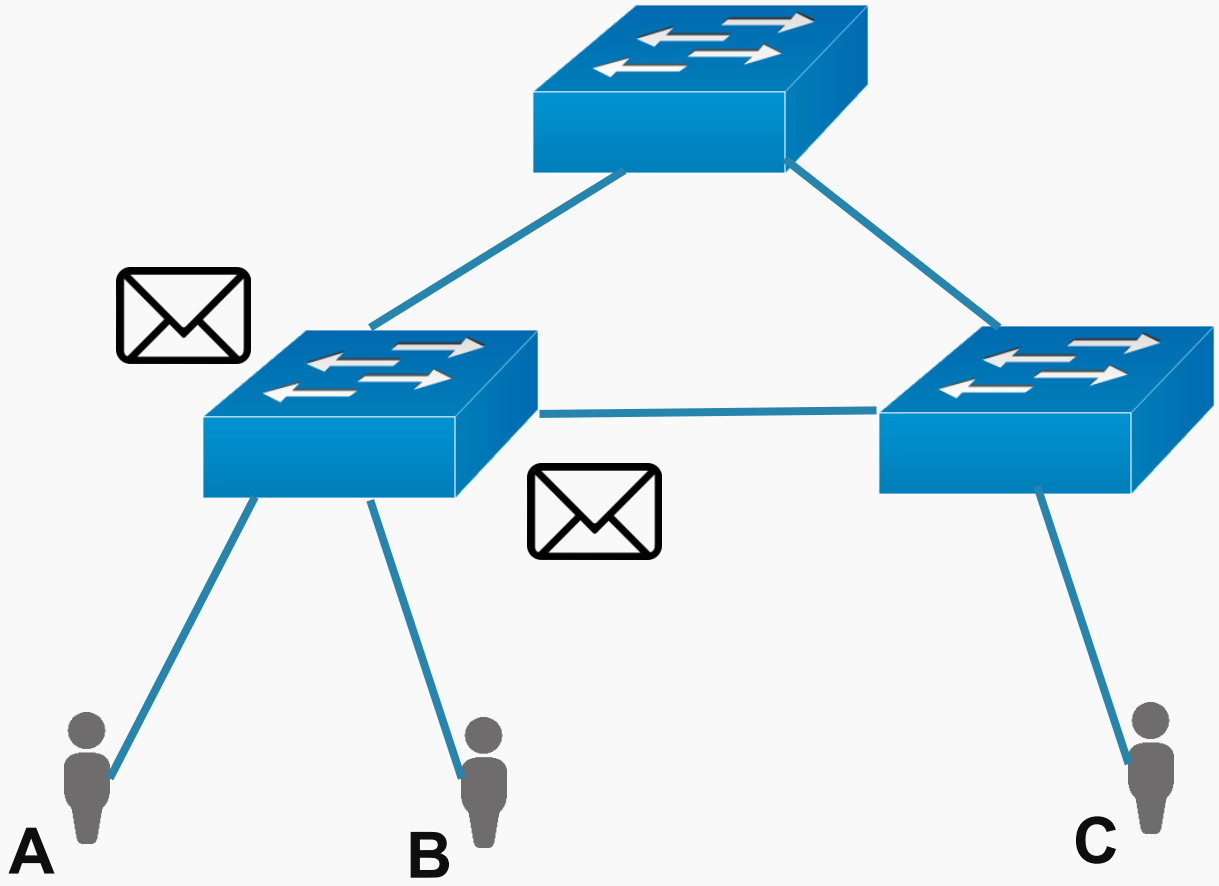
Ethernet – CAM table – Probleme cu cicluri



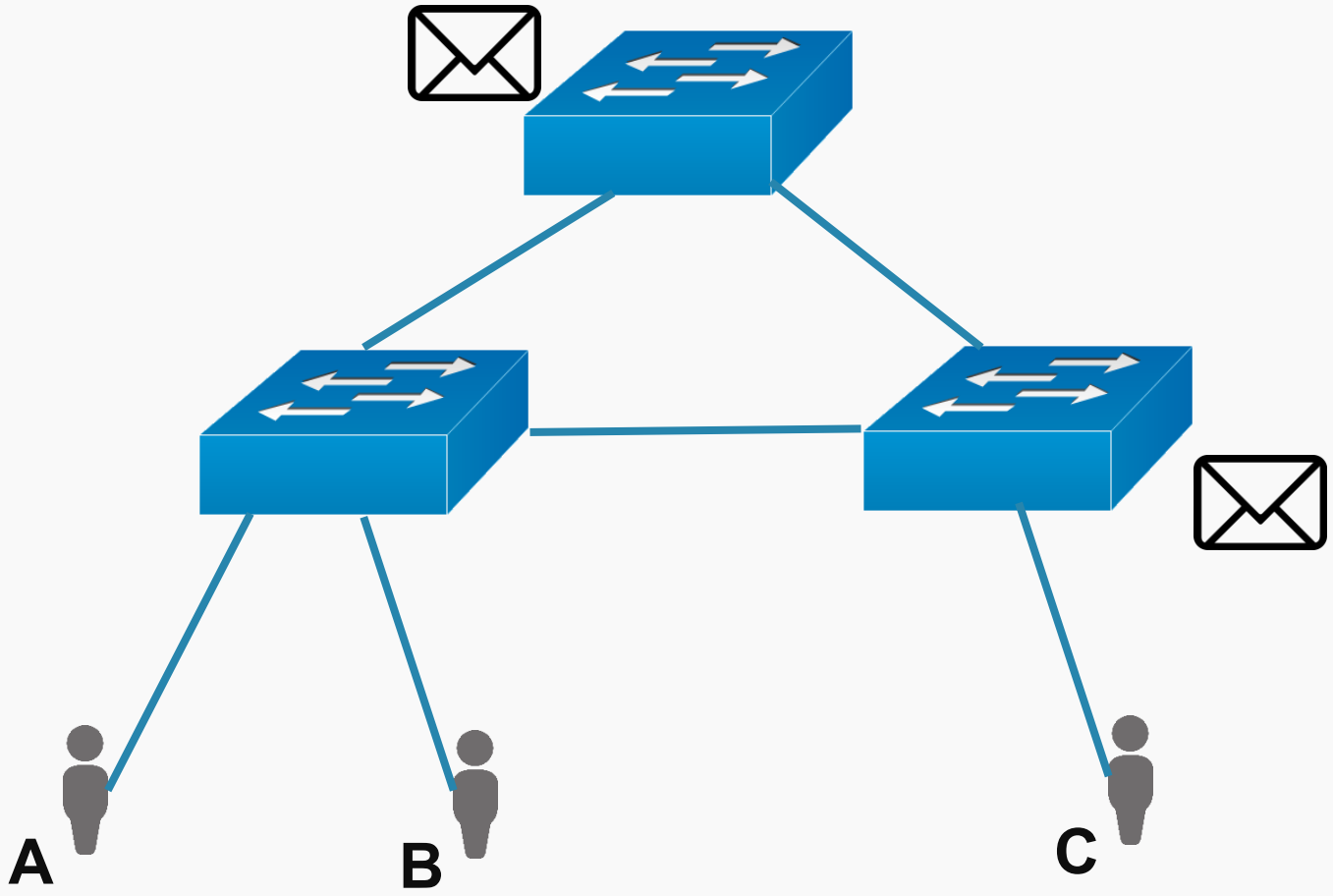
Ethernet – CAM table – Probleme cu cicluri



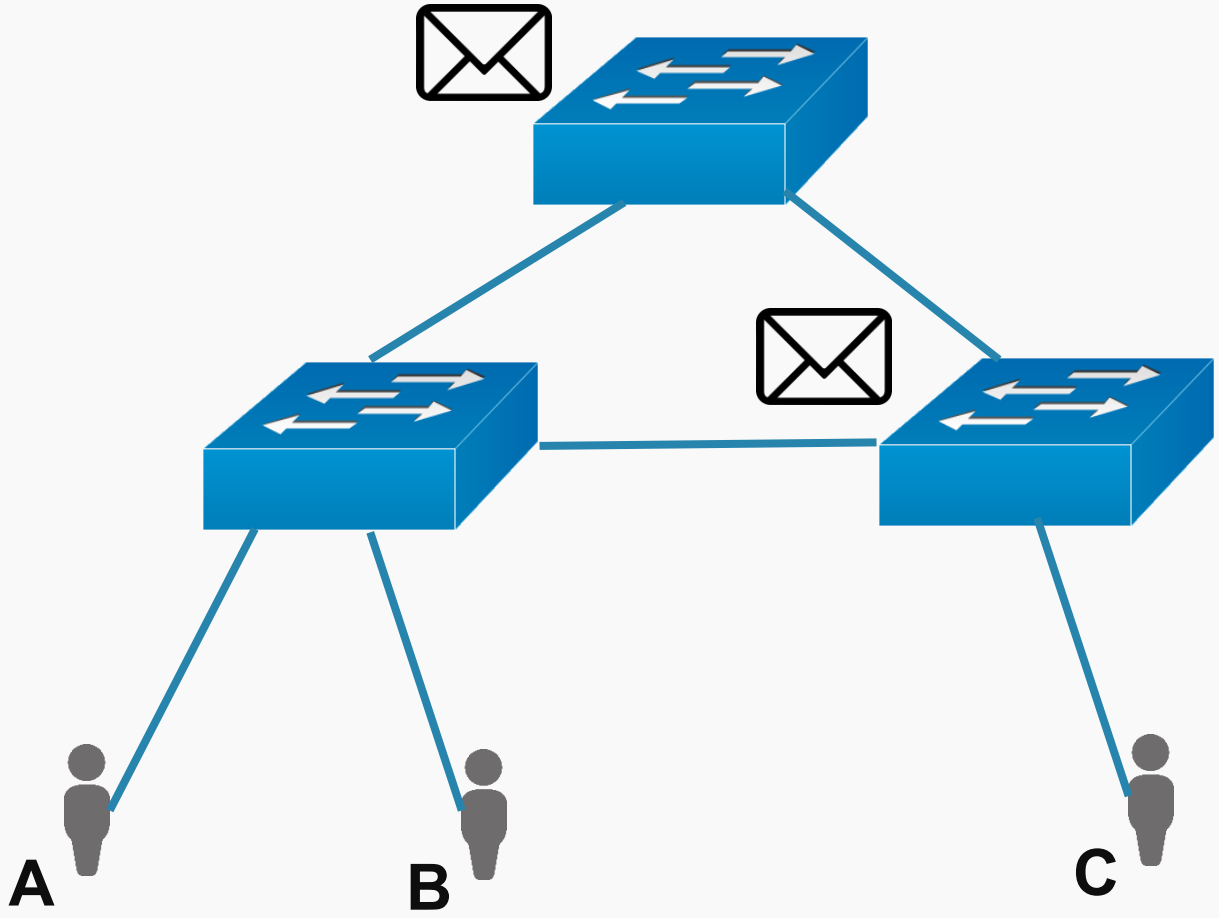
Ethernet – CAM table – Probleme cu cicluri



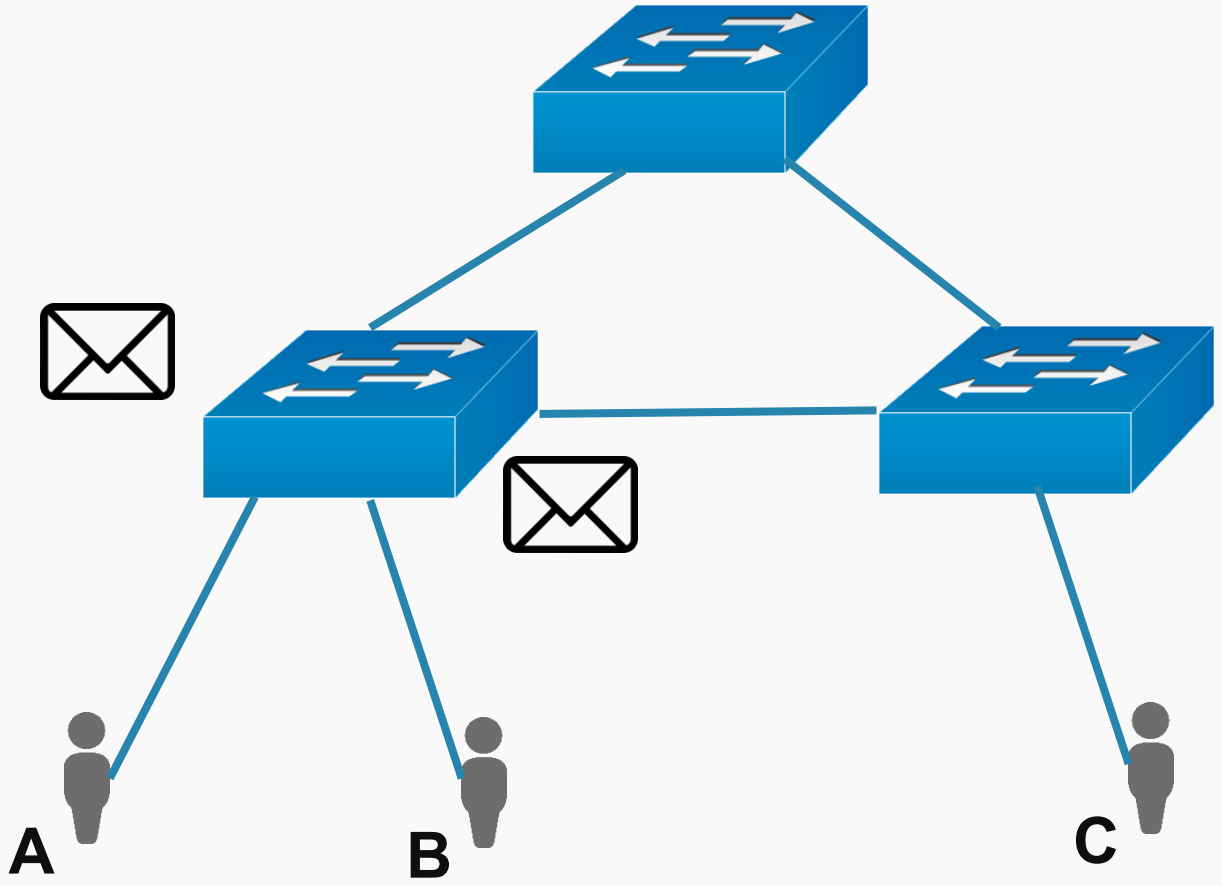
Ethernet – CAM table – Probleme cu cicluri



Ethernet – CAM table – Probleme cu cicluri



Ethernet – CAM table – Probleme cu cicluri

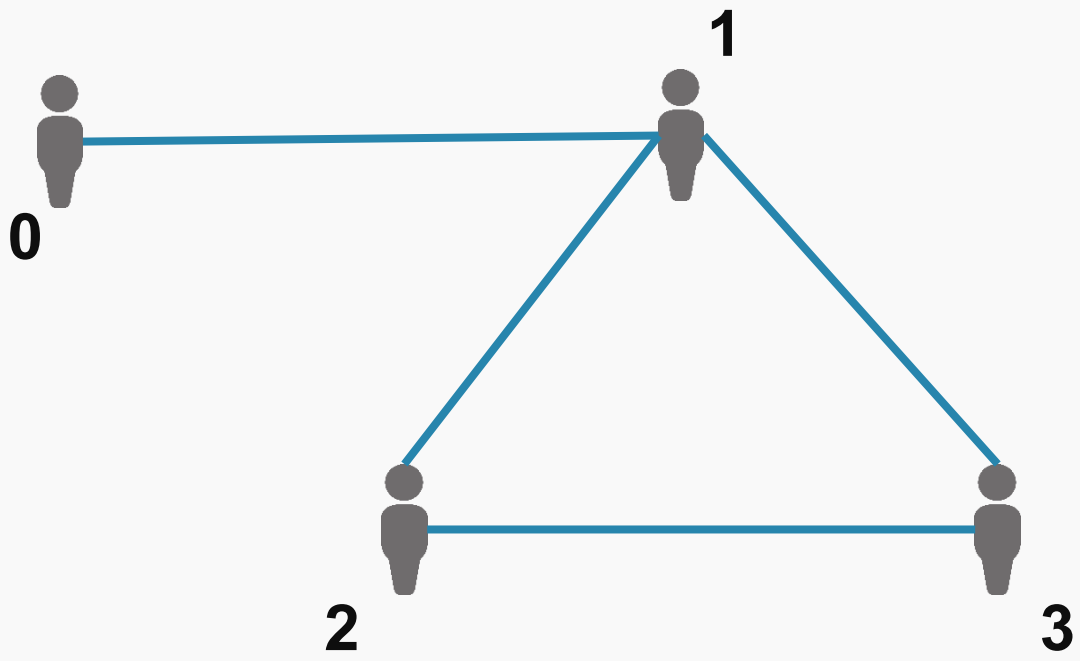


Cum eliminăm cicluri dintr-un graf?

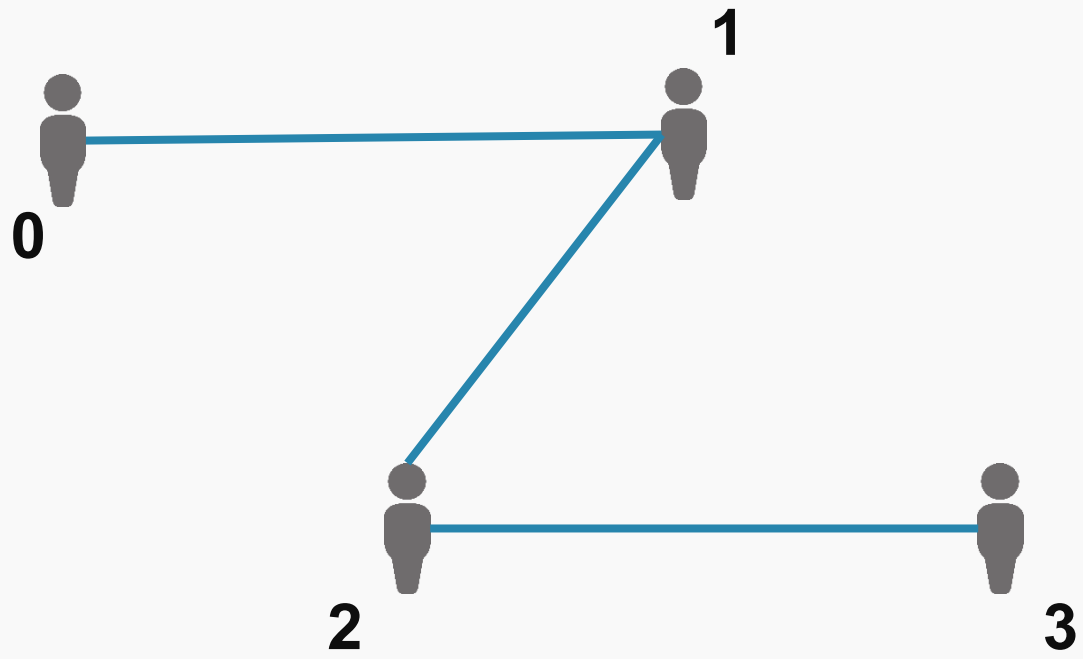
Cum eliminăm cicluri dintr-un graf?

Spanning Tree Protocol

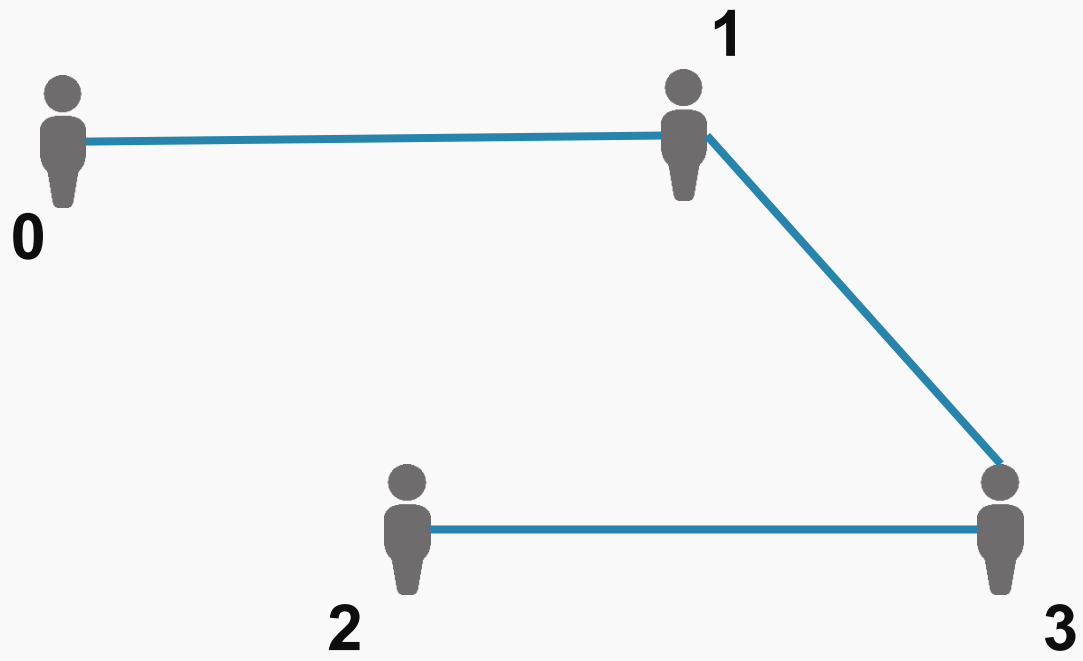
Distributed Spanning Tree



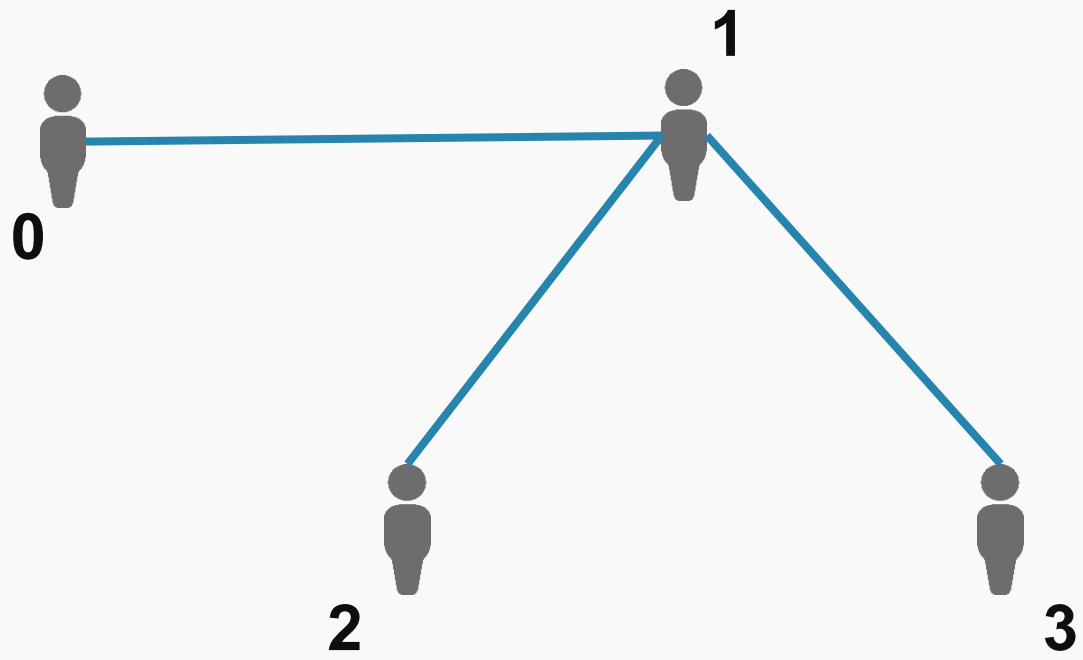
Distributed Spanning Tree – soluția



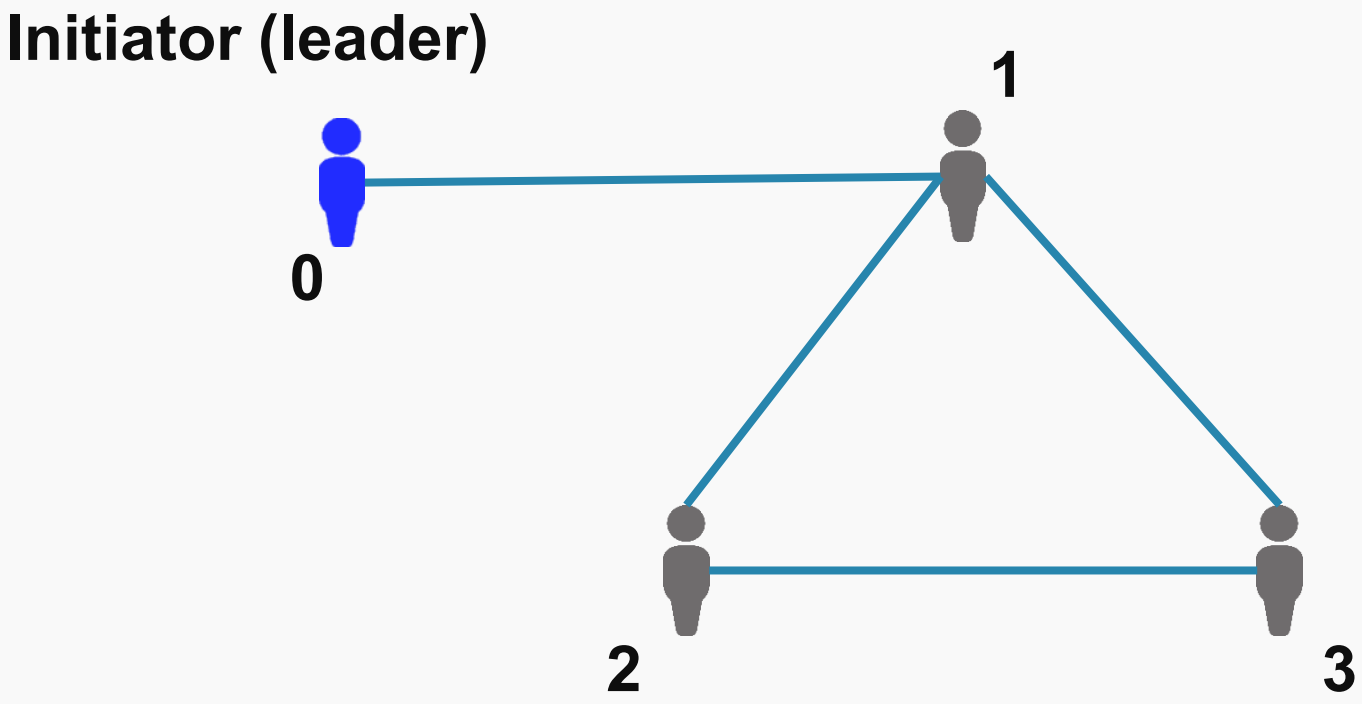
Distributed Spanning Tree - soluția



Distributed Spanning Tree - soluția



Distributed Spanning Tree – Inițiator



Distributed Spanning Tree

Initiator (leader)

- Send Probe to all neighbors
- Receive response from all neighbors
- Compute the entire graph

- Send graph to everyone

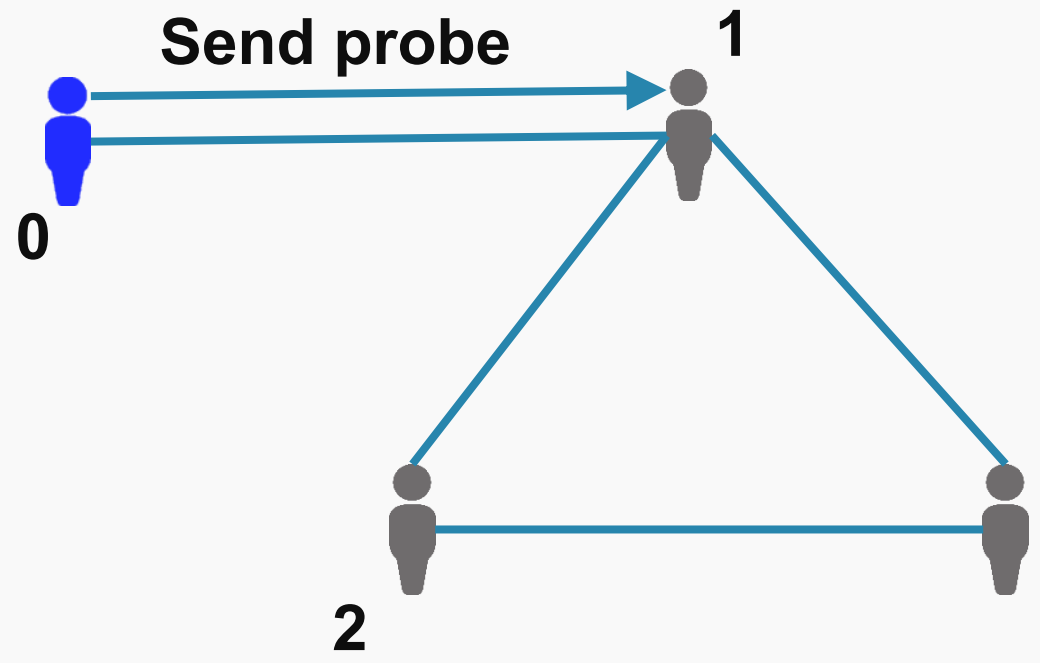
Distributed Spanning Tree

Everyone else

- Receive probe from someone
- That someone is marked as parent
- Forward probe to all neighbors except parent
- Receive response from all neighbors
- Merge responses
- Send response to parent

Distributed Spanning Tree

Node	Parent
0	-
1	
2	
3	

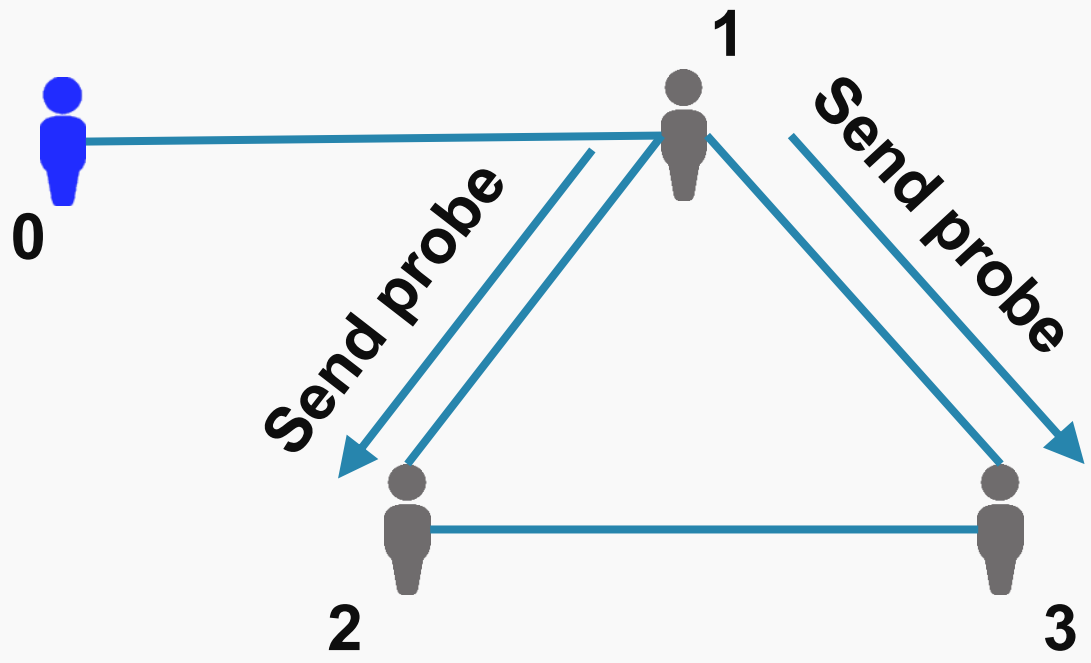


Node 2
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Node 3
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Distributed Spanning Tree

Node	Parent
0	-
1	0
2	
3	

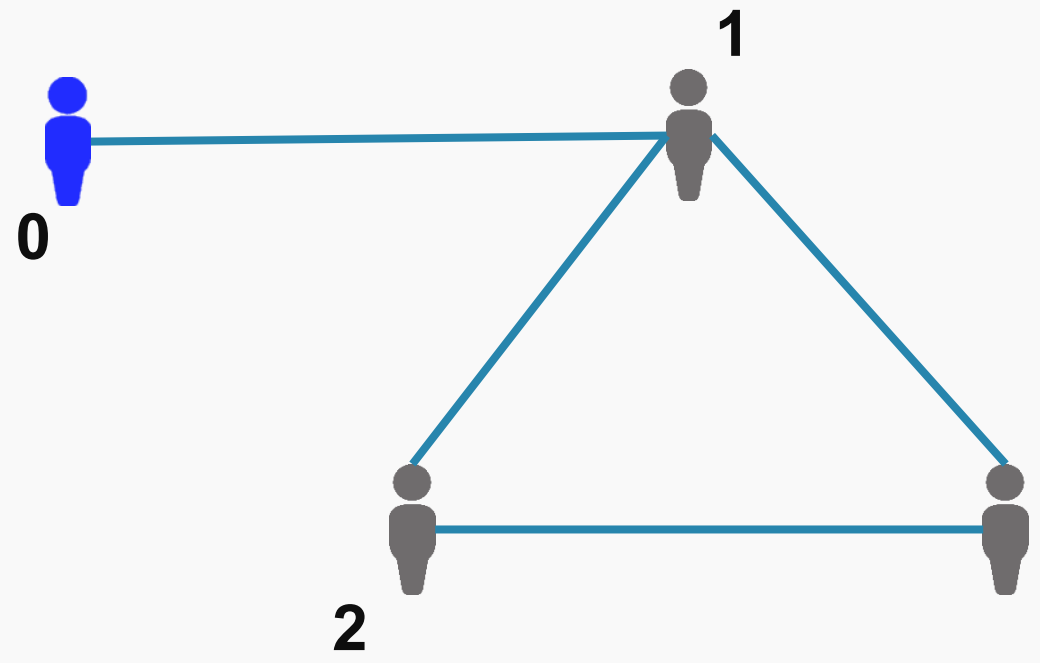


Node 2
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Node 3
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Distributed Spanning Tree

Node	Parent
0	-
1	0
2	1
3	1



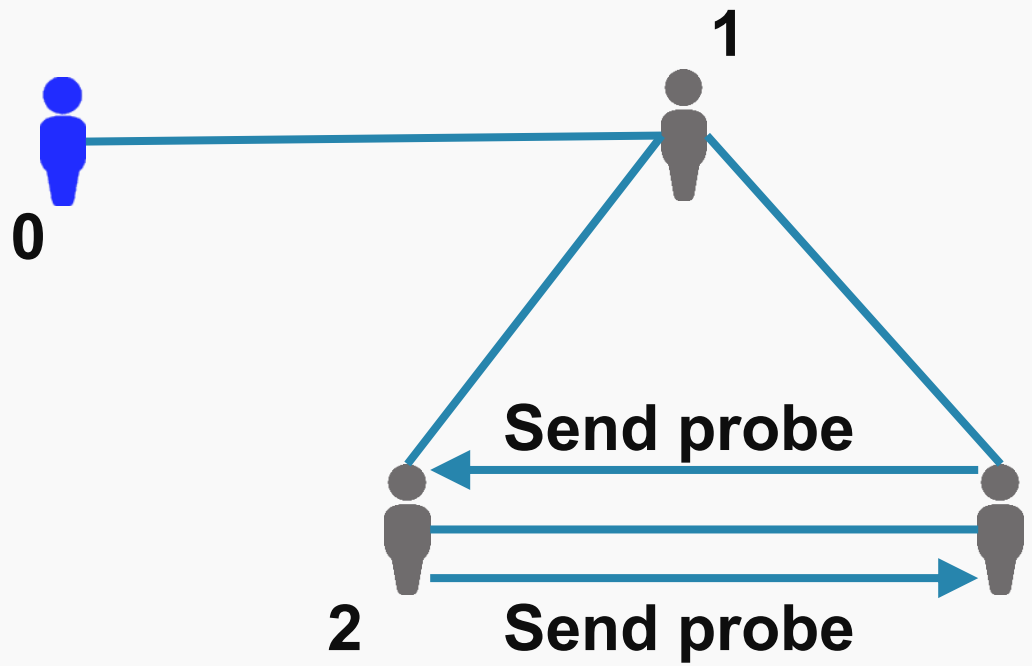
Node 2
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Node 3
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Distributed Spanning Tree – Inițiator

2 și 3 tratează probele unul altuia ca răspunsuri

Node	Parent
0	-
1	0
2	1
3	1

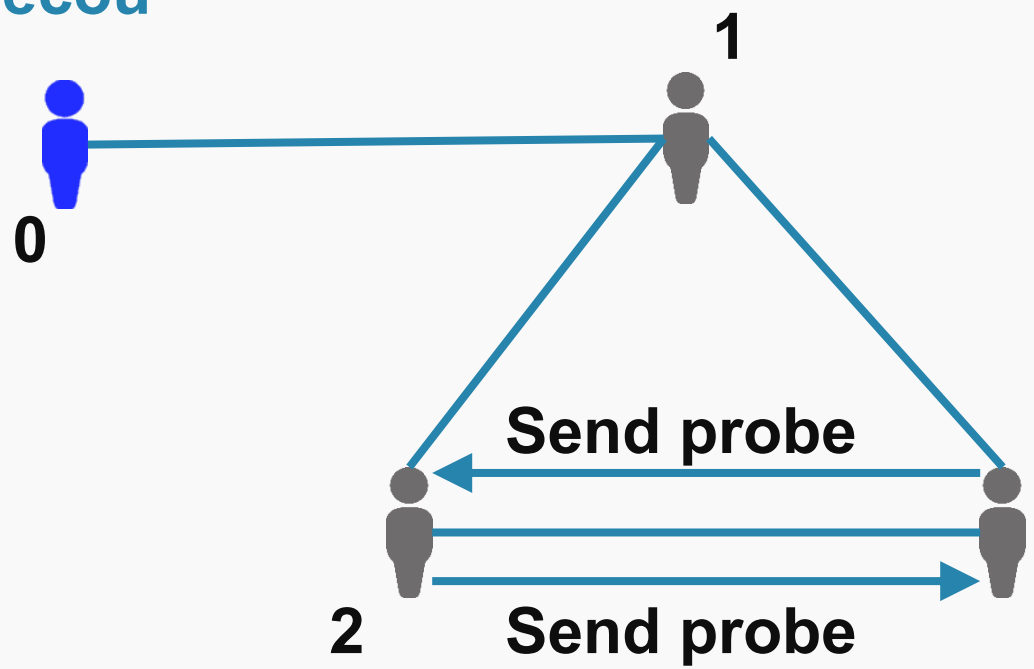


Node 2
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Node 3
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Distributed Spanning Tree – Initiator

Funcționează doar dacă proba are același format ca ecou



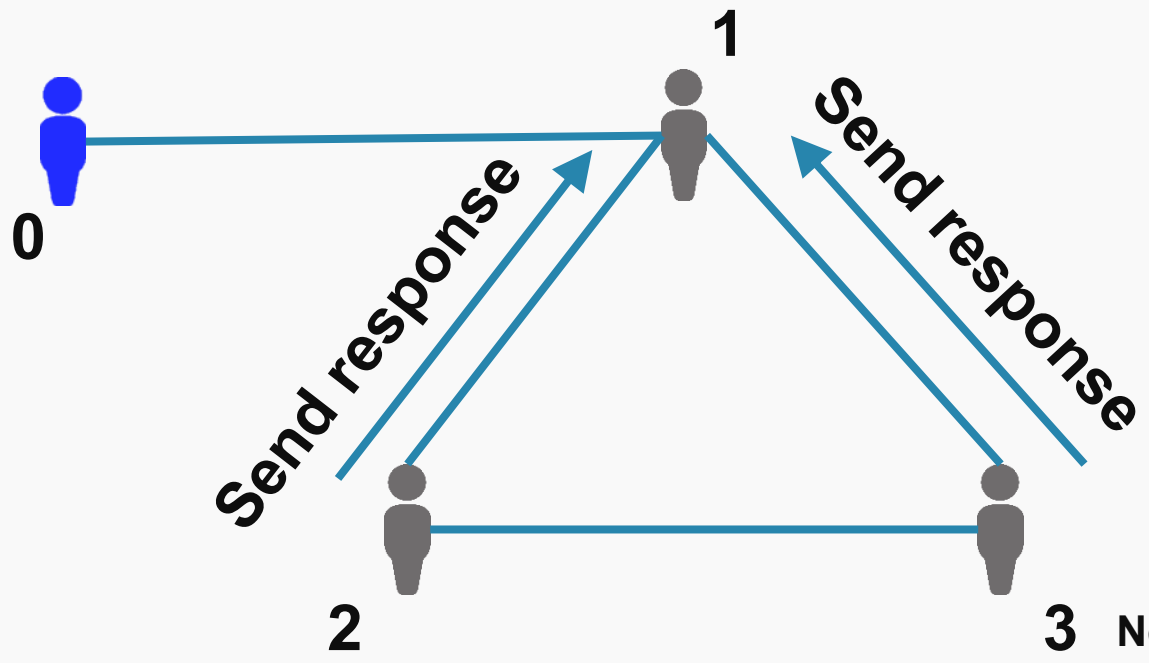
Node	Parent
0	-
1	0
2	1
3	1

Node 2
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Node 3
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Distributed Spanning Tree

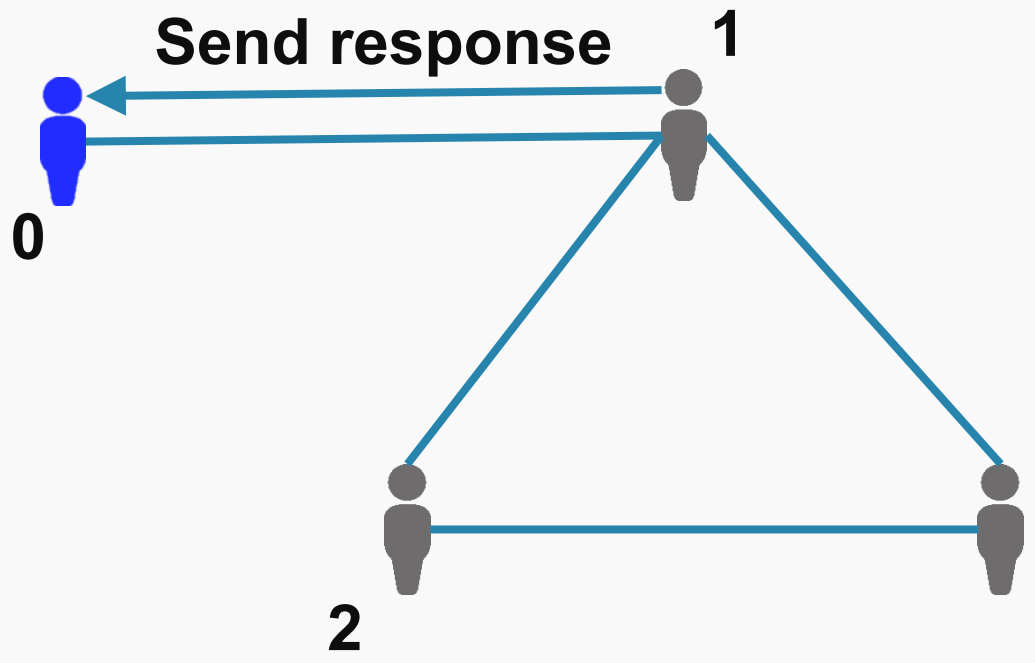
Node	Parent
0	-
1	0
2	1
3	1



Node 2
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Node 3
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Distributed Spanning Tree



Node	Parent
0	-
1	0
2	1
3	1

Node 2
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Node 3
Recv probe
Mark parent
Send probe children
Recv response children
Merge responses
Send response parent

Distributed Spanning Tree

Comunicație full sincronă

Distributed Spanning Tree

Cum alegem inițiatorul?

Alegere lider

Alegere lider

Scopul este transformarea automată a unui sistem distribuit, decentralizat, într-un sistem cu topologie client-server;

Mai mult, dacă serverul moare, poate fi ales un nou ”lider”.

Alegere lider Le Lann

DISTRIBUTED SYSTEMS—TOWARDS A FORMAL APPROACH

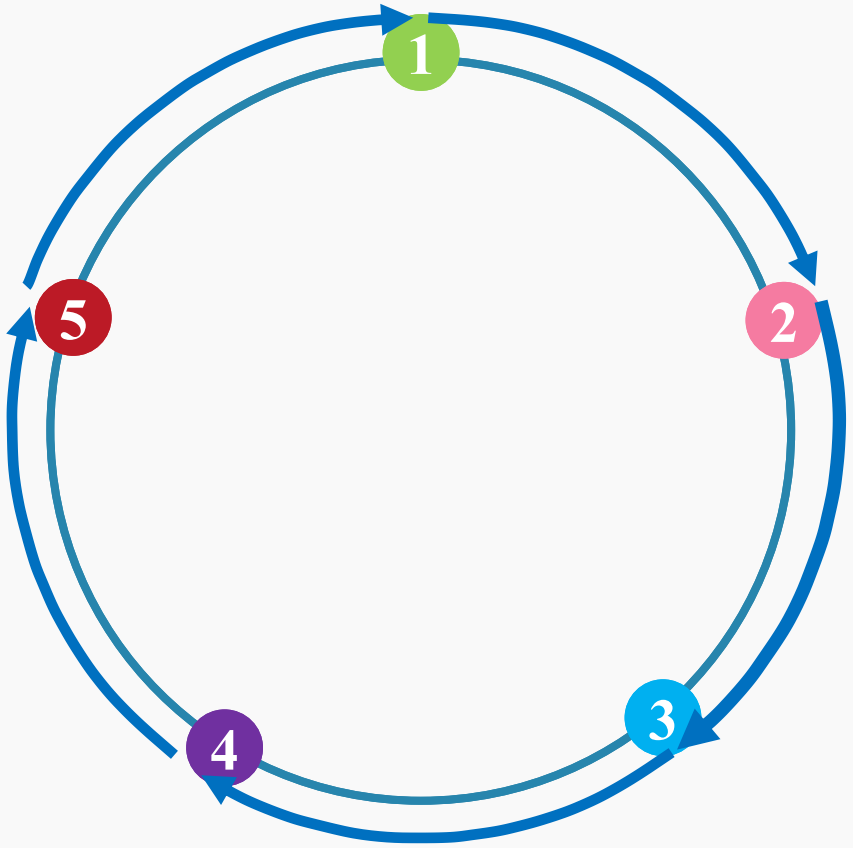
GÉRARD LE LANN
IRISA—Université de Rennes—BP 25 A
35 031 Rennes Cedex, France

Packet-switching computer communication networks are examples of distributed systems. With the large scale emergence of mini and micro-computers, it is now possible to design special or general purpose distributed systems. However, as new problems hav
devised to operate such distributed systems in a
tics of distributed systems are analysed and fund
shown that distributed systems are not just simpl
techniques used in some planned or existing syste
problems is illustrated by the study of a mutual
ronment.

1. INTRODUCTION

Computer communication networks using packet-swit-
ching technology provide for the interconnection of
data-processing equipments of any kind. Such systems,

Algoritmul Le Lann



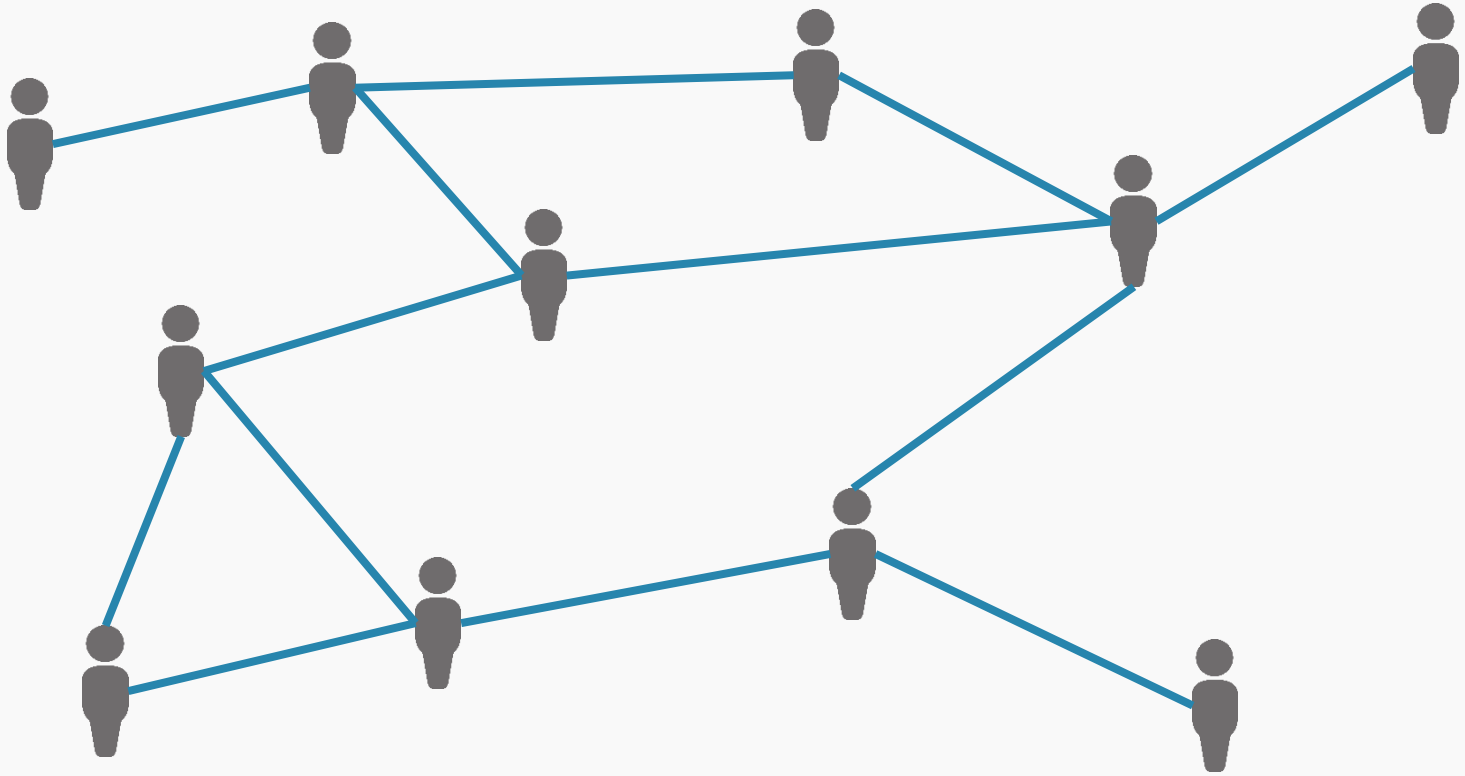
Fiecare transmite ID-ul său în dreapta

Până își primește ID-ul său
Primește ID din stânga
Trimite ID primit în dreapta

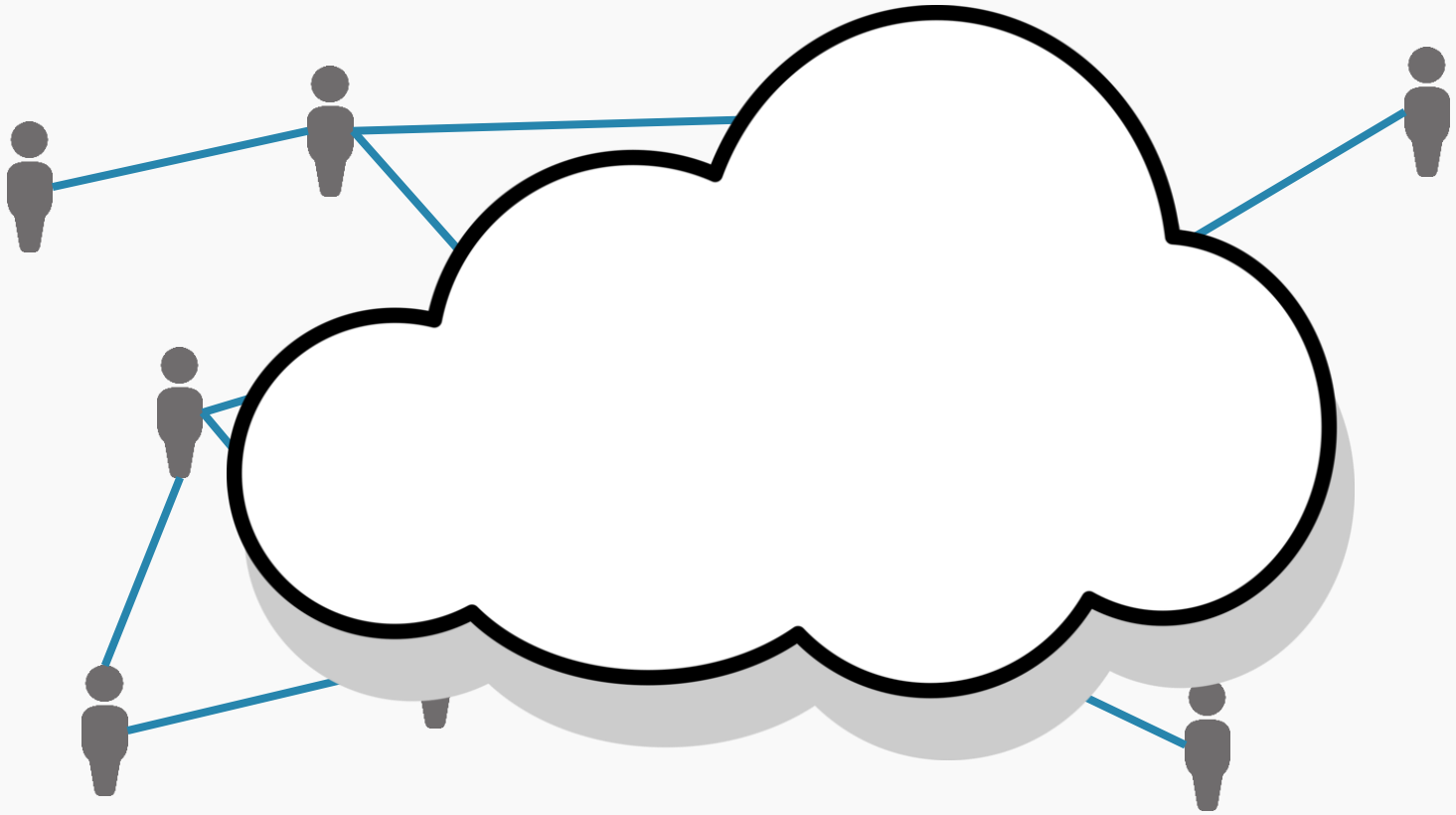
Cel mai mare/mic/etc. este ales lider

De ce mai discutăm despre topologii inel?

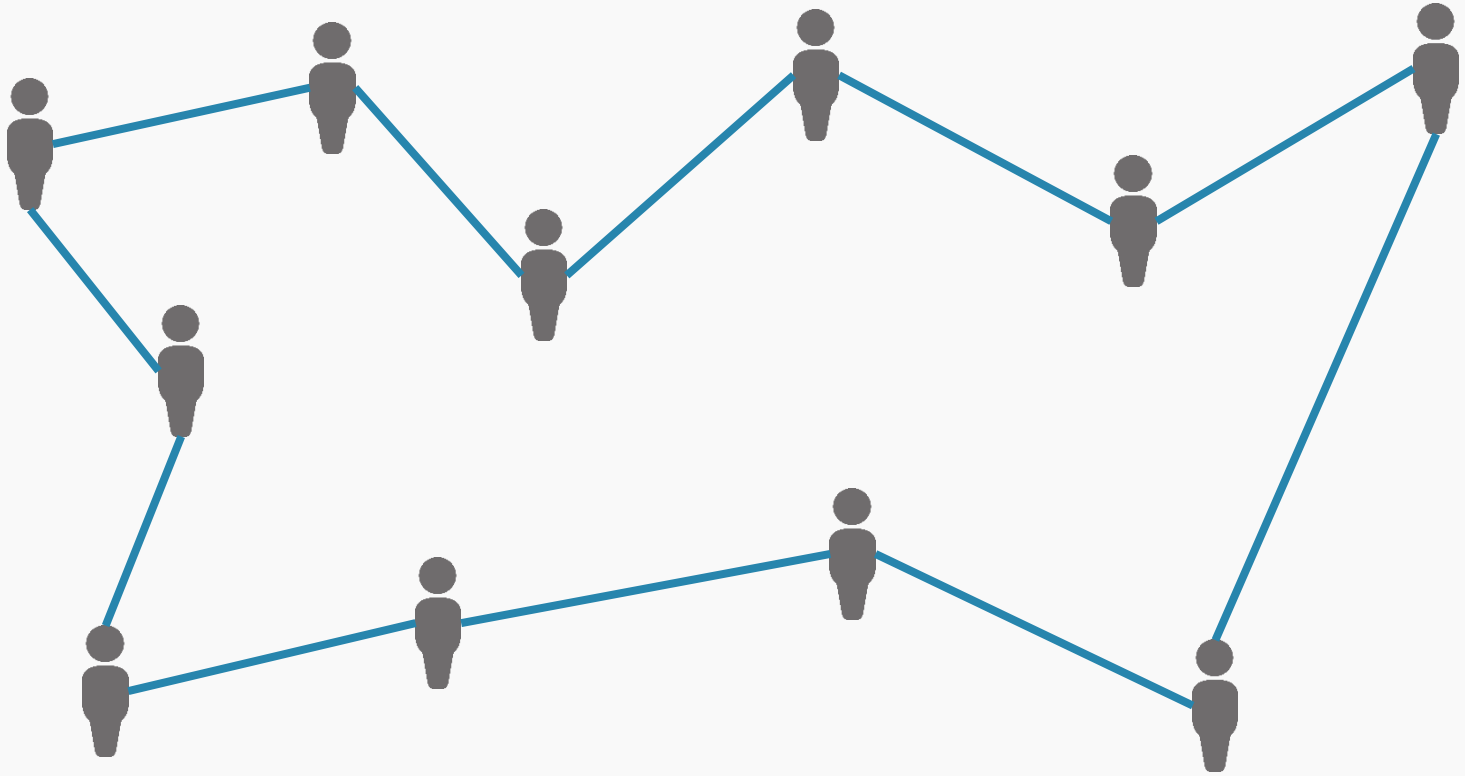
Una e topologia fizică



Una e topologia fizică



Alta e topologia software



Dar dacă nu știm deja ID-urile?

UUID sau GUID (Global Unique Identifier)

Network Working Group
Request for Comments: 4122
Category: Standards Track

P. Leach
Microsoft
M. Mealling
Refactored Networks, LLC
R. Salz
DataPower Technology, Inc.
July 2005

A Universally Unique IDentifier (UUID) URN Namespace

Status of This Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Abstract

This specification defines a Uniform Resource Name namespace for UUIDs (Universally Unique IDentifier), also known as GUIDs (Globally Unique IDentifier). A UUID is 128 bits long, and can guarantee uniqueness across space and time. UUIDs were originally used in the Apollo Network Computing System and later in the Open Software

Stabilirea Topologiei IP

În alte cuvinte: stabilirea rutelor și tabelelor de rutare.

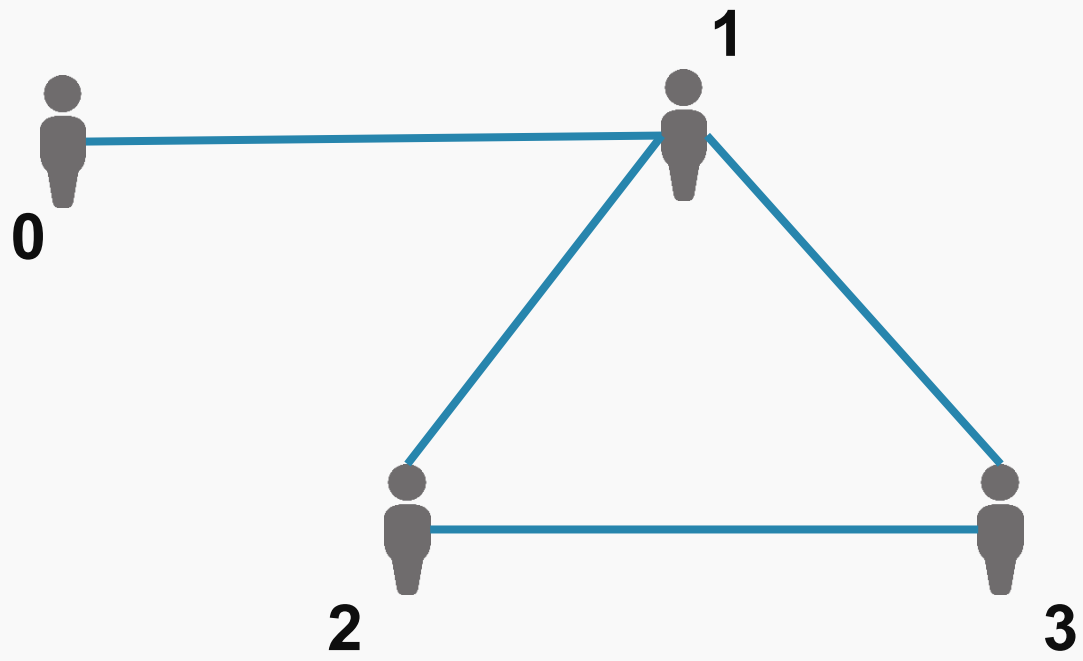
Algoritmi de stabilire rute în IP?

Algoritmi de stabilire rute în IP?

- BGP
- RIP
- IGRP
- EIGRP
- OSPF

…

Algoritm stabilire topologie



Algoritm stabilire topologie

Trimite topologie cunoscută tuturor vecinilor
Primește de la toți vecinii topologiile cunoscute de ei
Adaugă informații la topologia cunoscută.

Repetă la infinit!!

Algoritm stabilire topologie

Trimite topologie cunoscută tuturor vecinilor
Primește de la toți vecinii topologiile cunoscute de ei
Adaugă informații la topologia cunoscută.

Repetă la infinit!!

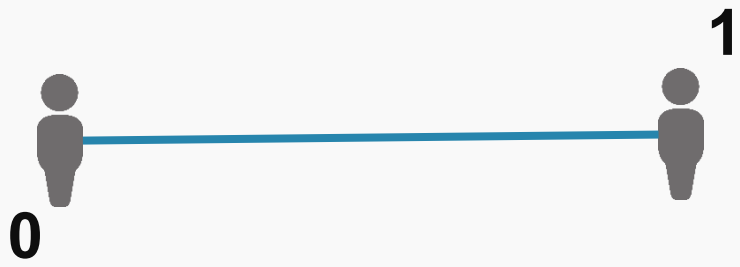
Atenție în forma actuală nu se pot șterge conexiuni.

Algoritm stabilire topologie

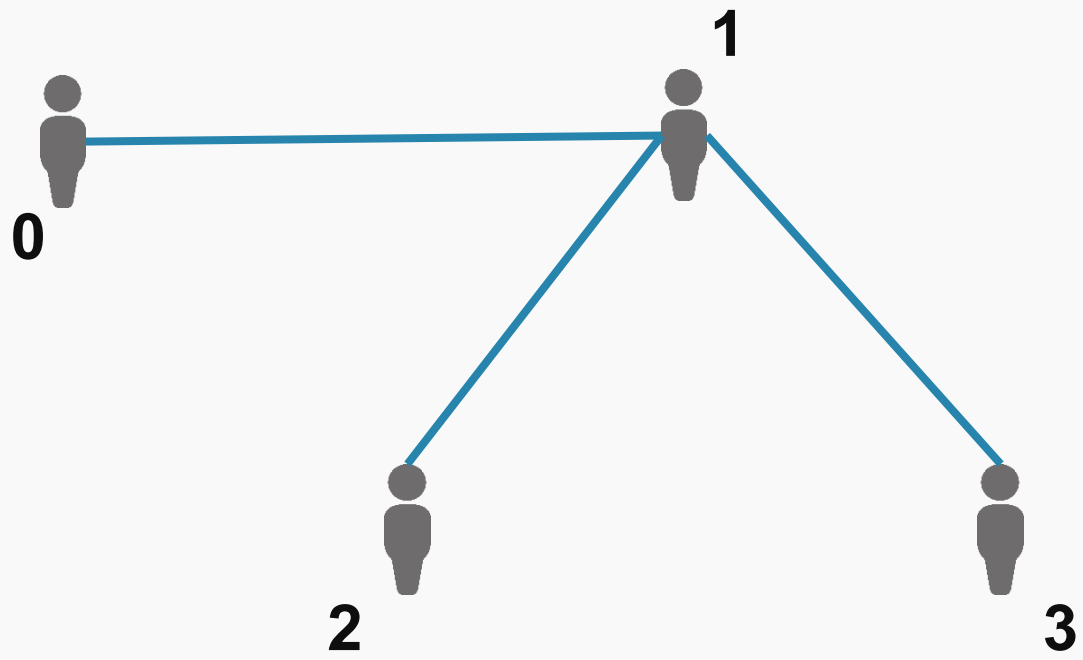
Trimite topologie cunoscută tuturor vecinilor
Primește de la toți vecinii topologiile cunoscute de ei
Adaugă informații la topologia cunoscută.

Dacă topologia s-a modificat, repetă.

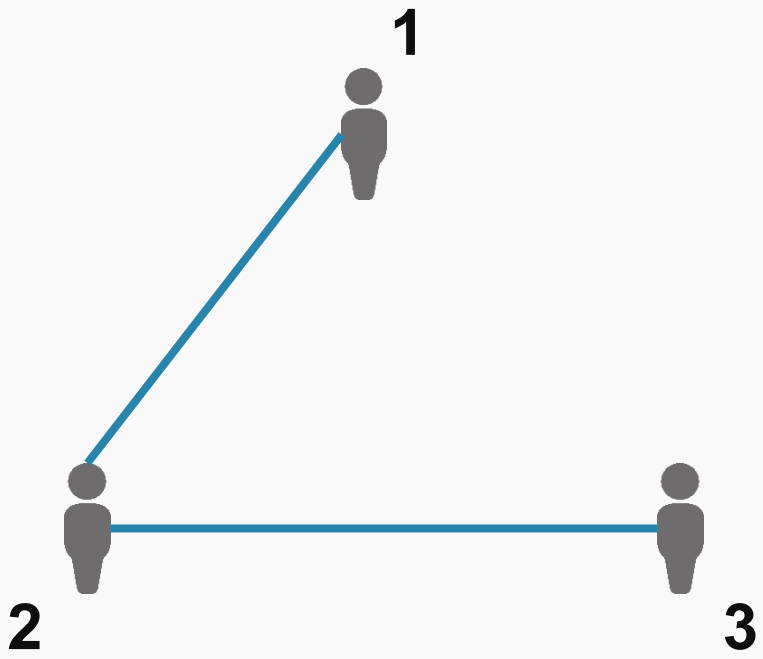
Topologii inițiale - 0



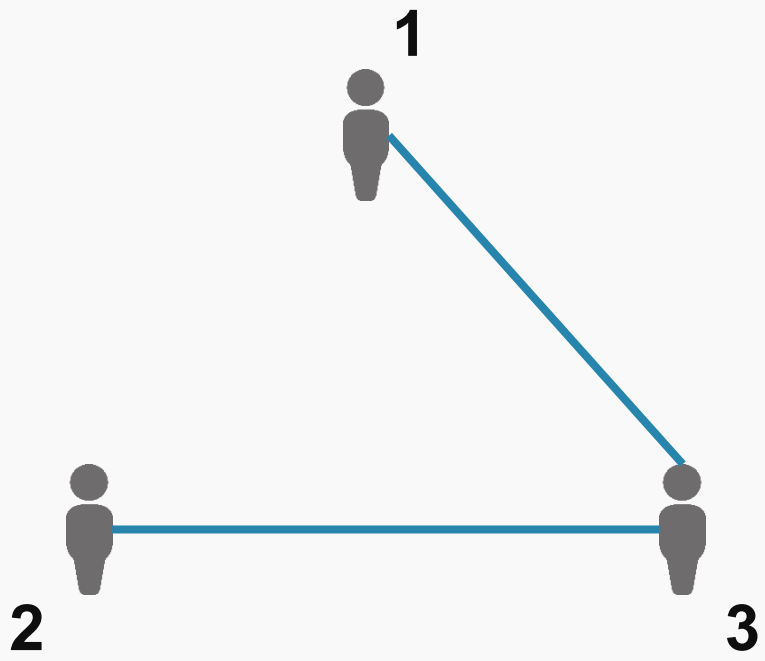
Topologii inițiale - 1



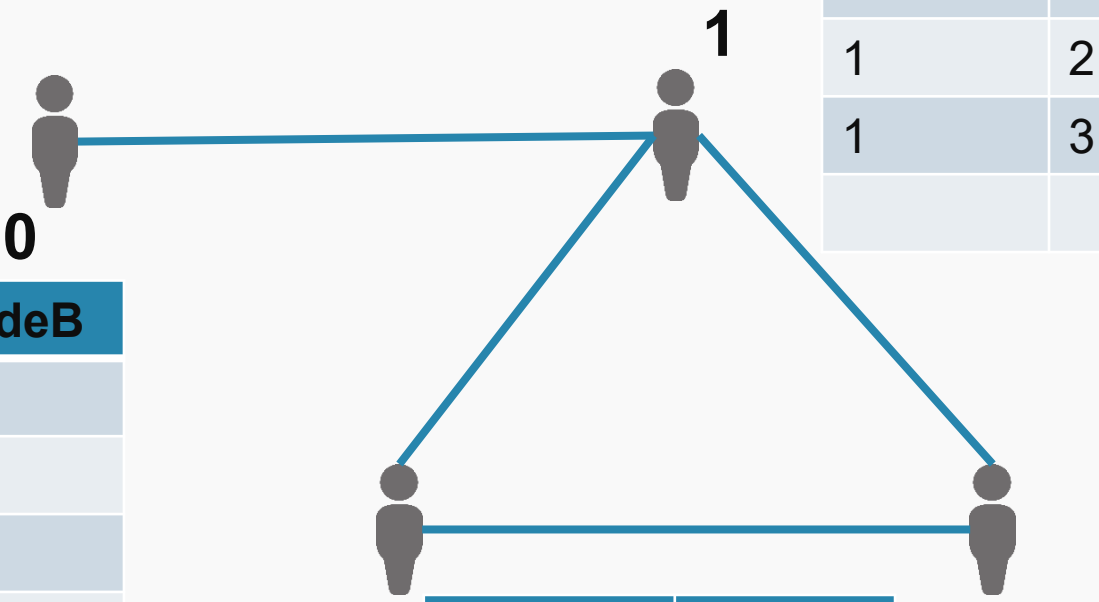
Topologii inițiale - 2



Topologii inițiale - 3



Algoritm stabilire topologie



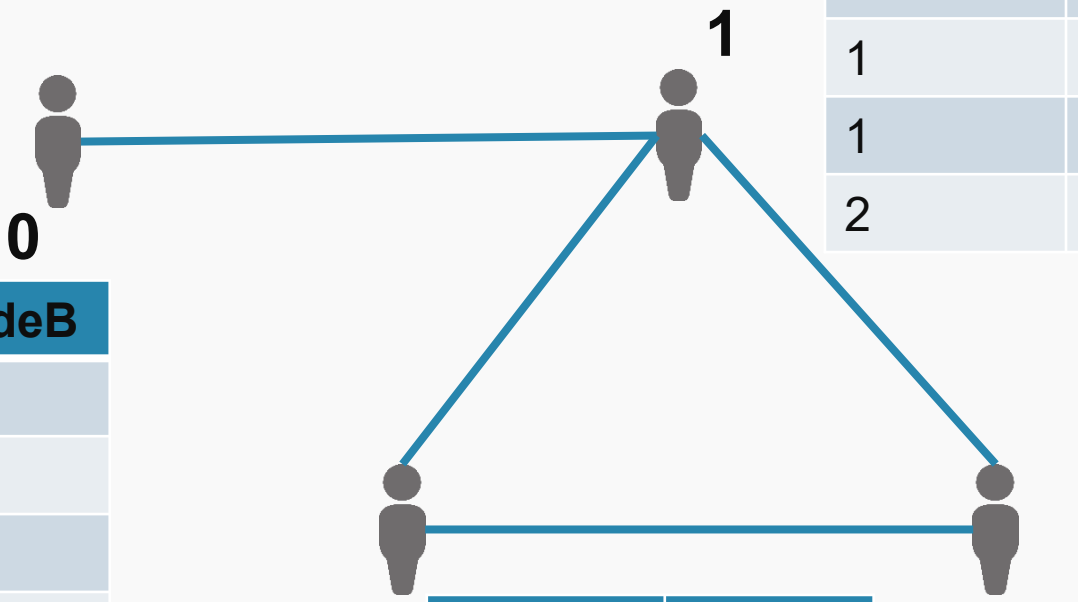
NodeA	NodeB
1	0
1	2
1	3

NodeA	NodeB
0	1

NodeA	NodeB
2	1
2	3

NodeA	NodeB
3	1
3	2

Algoritm stabilire topologie



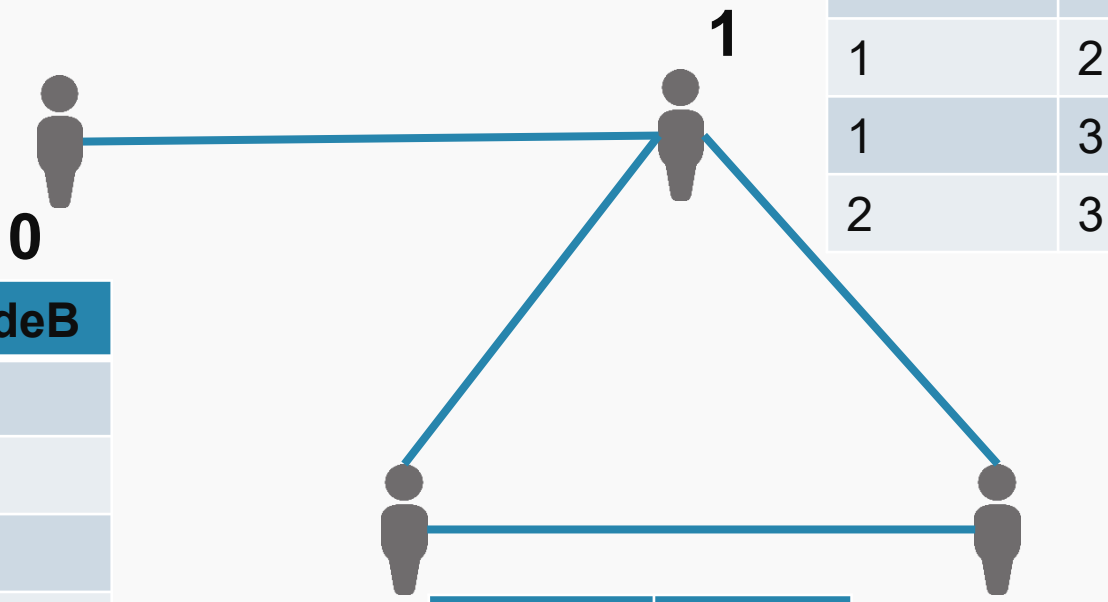
NodeA	NodeB
0	1
1	2
1	3

NodeA	NodeB
1	0
1	2
1	3
2	3

NodeA	NodeB
2	1
2	3
1	0
1	3

NodeA	NodeB
3	1
3	2
1	0
1	2

Algoritm stabilire topologie



NodeA	NodeB
0	1
1	2
1	3
2	3

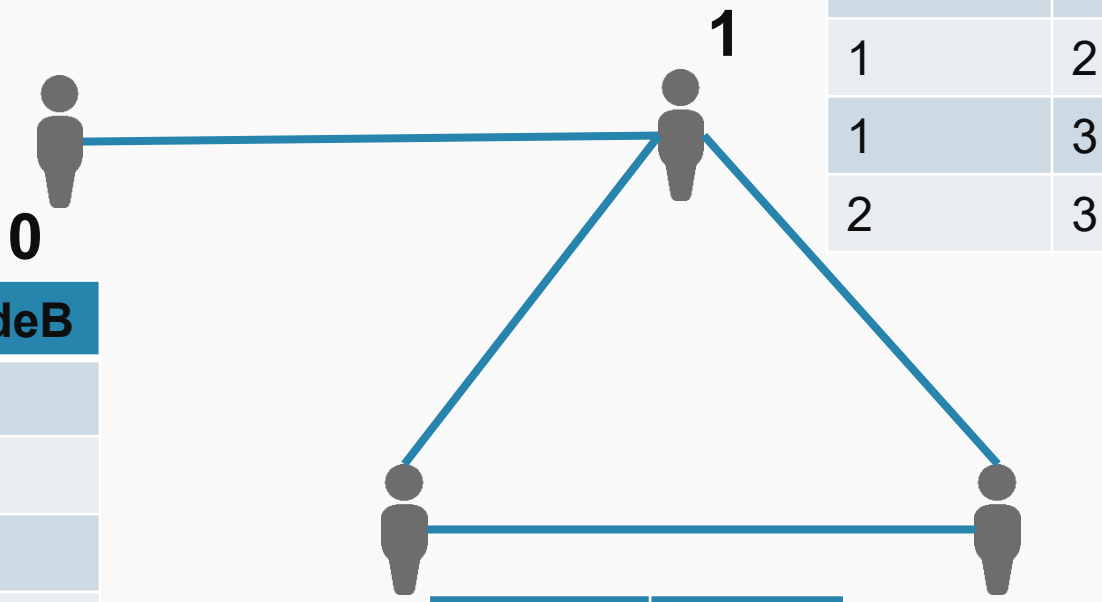
NodeA	NodeB
1	0
1	2
1	3
2	3

NodeA	NodeB
2	1
2	3
1	0
1	3

NodeA	NodeB
3	1
3	2
1	0
1	2

Alegere lider în graf

Alegere lider în graf



NodeA	NodeB
0	1
1	2
1	3
2	3

NodeA	NodeB
1	0
1	2
1	3
2	3

NodeA	NodeB
2	1
2	3
1	0
1	3

NodeA	NodeB
3	1
3	2
1	0
1	2

Cine a construit The World-Wide Web?

Cine a construit The World-Wide Web?
Tim Berners-Lee

Network Working Group
Request for Comments: 2068
Category: Standards Track

R. Fielding
UC Irvine
J. Gettys
J. Mogul
DEC
H. Frystyk
T. Berners-Lee
MIT/LCS
January 1997

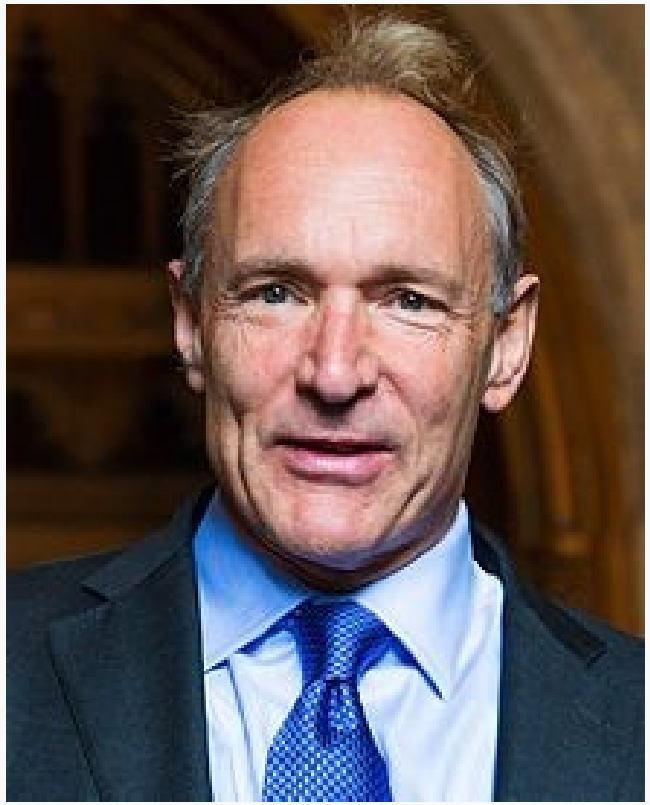
Hypertext Transfer Protocol -- HTTP/1.1

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Abstract

The Hypertext Transfer Protocol (HTTP) is an application-level protocol for distributed, collaborative, hypermedia information systems. It is a generic, stateless, object-oriented protocol which can be used for many tasks, such as name servers and distributed object management systems, through extension of its request methods. A feature of HTTP is the typing and negotiation of data representation, allowing systems to be built independently of the data being transferred.



Cine a construit The World-Wide Web?
Tim Berners-Lee

Network Working Group
Request for Comments: 1738
Category: Standards Track

T. Berners-Lee
CERN
L. Masinter
Xerox Corporation
M. McCahill
University of Minnesota
Editors
December 1994

Uniform Resource Locators (URL)

Status of this Memo

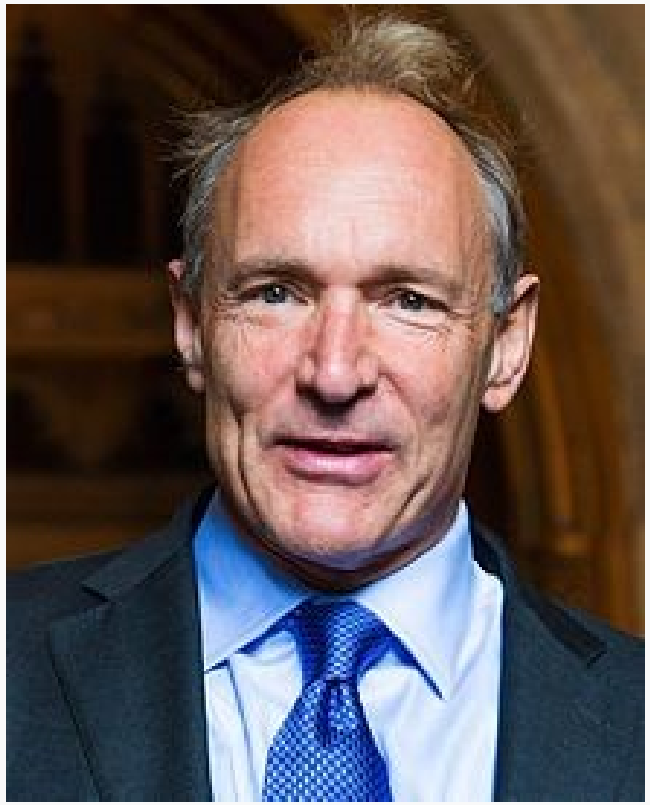
This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Abstract

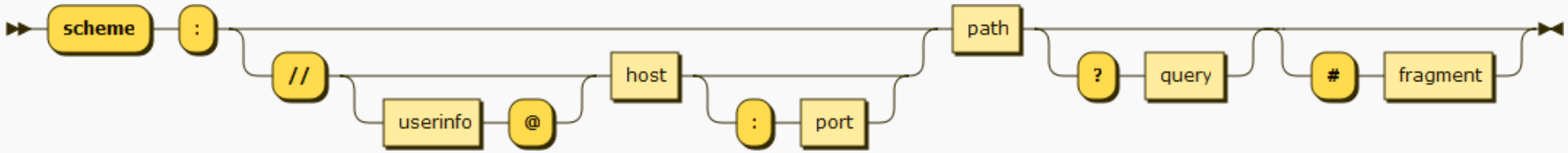
This document specifies a Uniform Resource Locator (URL), the syntax and semantics of formalized information for location and access of resources via the Internet.

1. Introduction

This document describes the syntax and semantics for a compact string representation for a resource available via the Internet. These



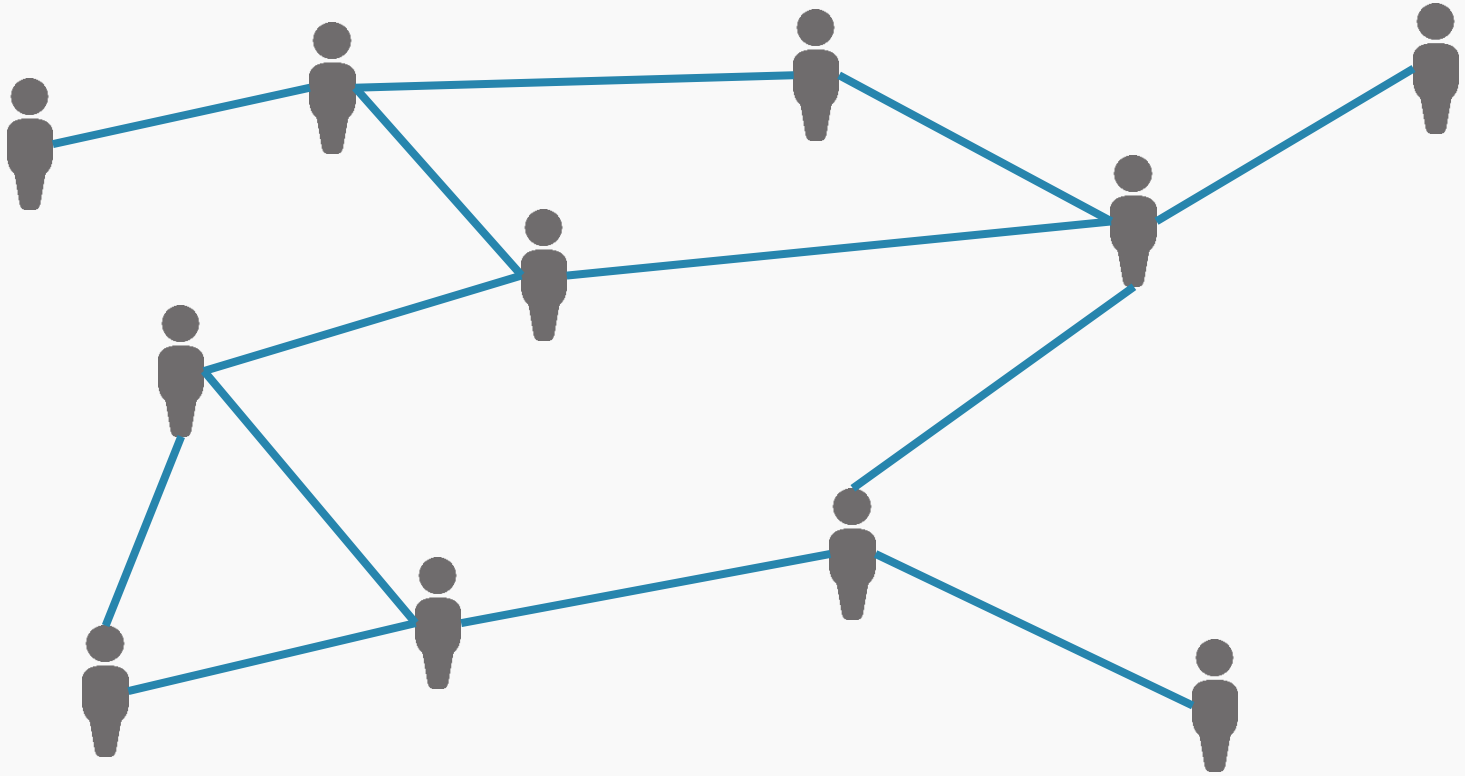
URL



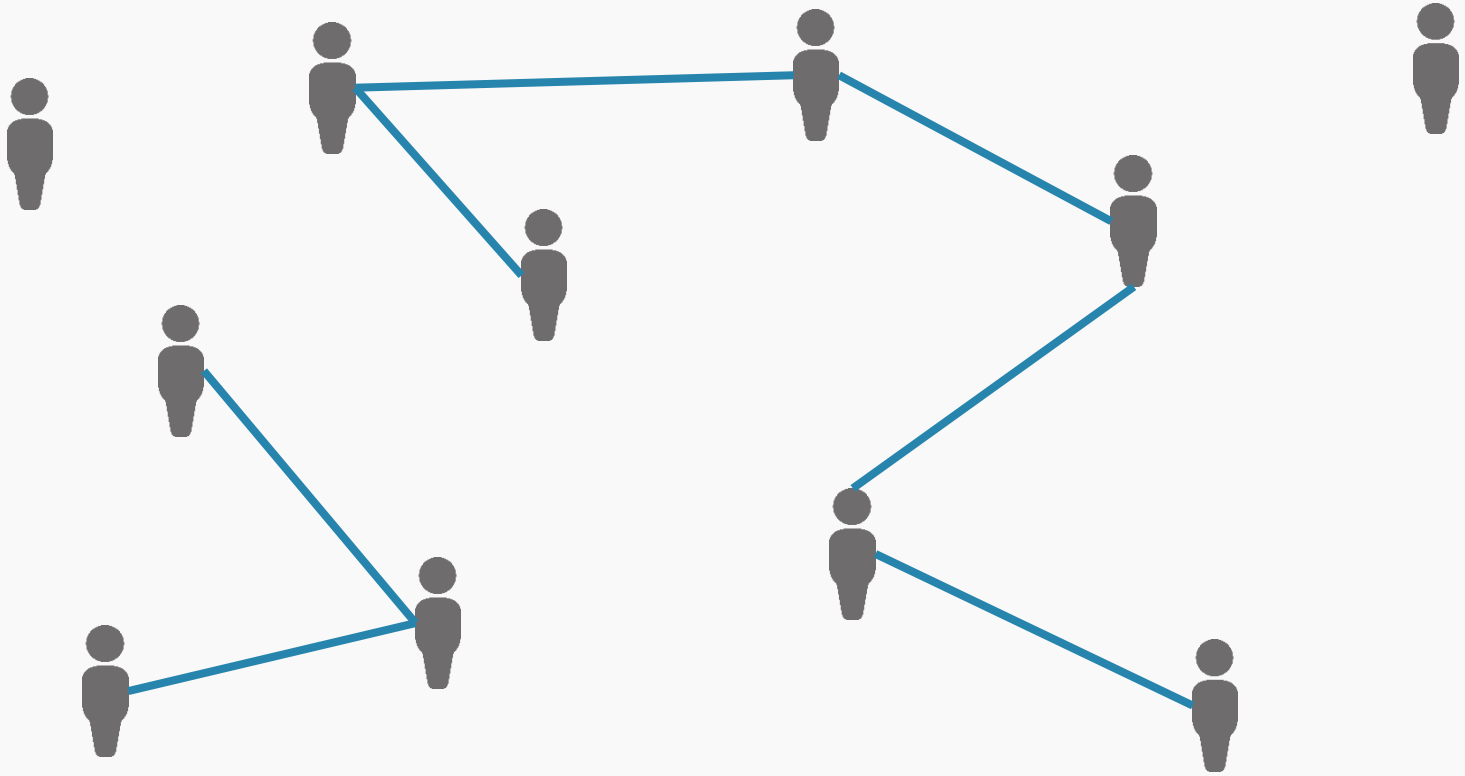
Sisteme distribuite extreme

Comunicare Epidemic

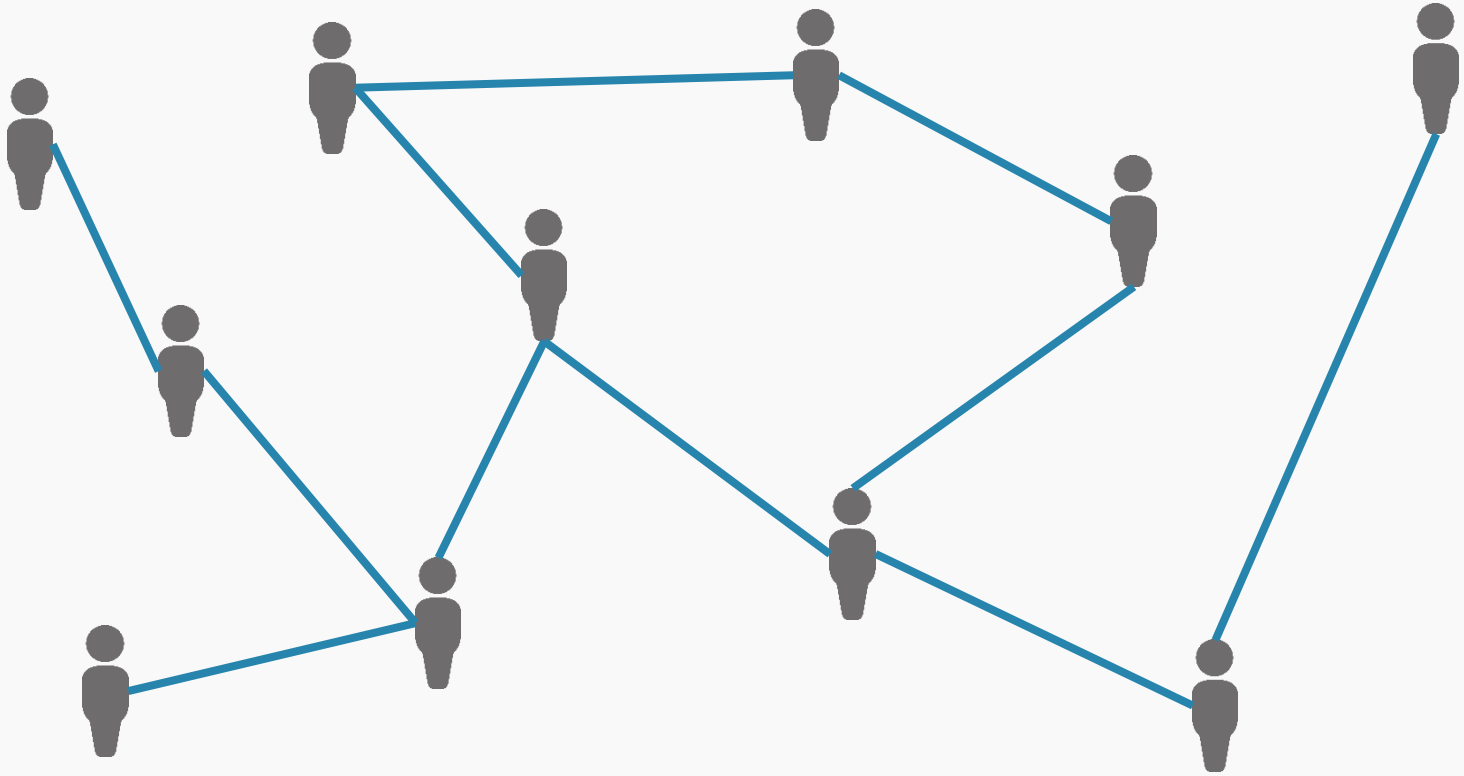
Epidemic



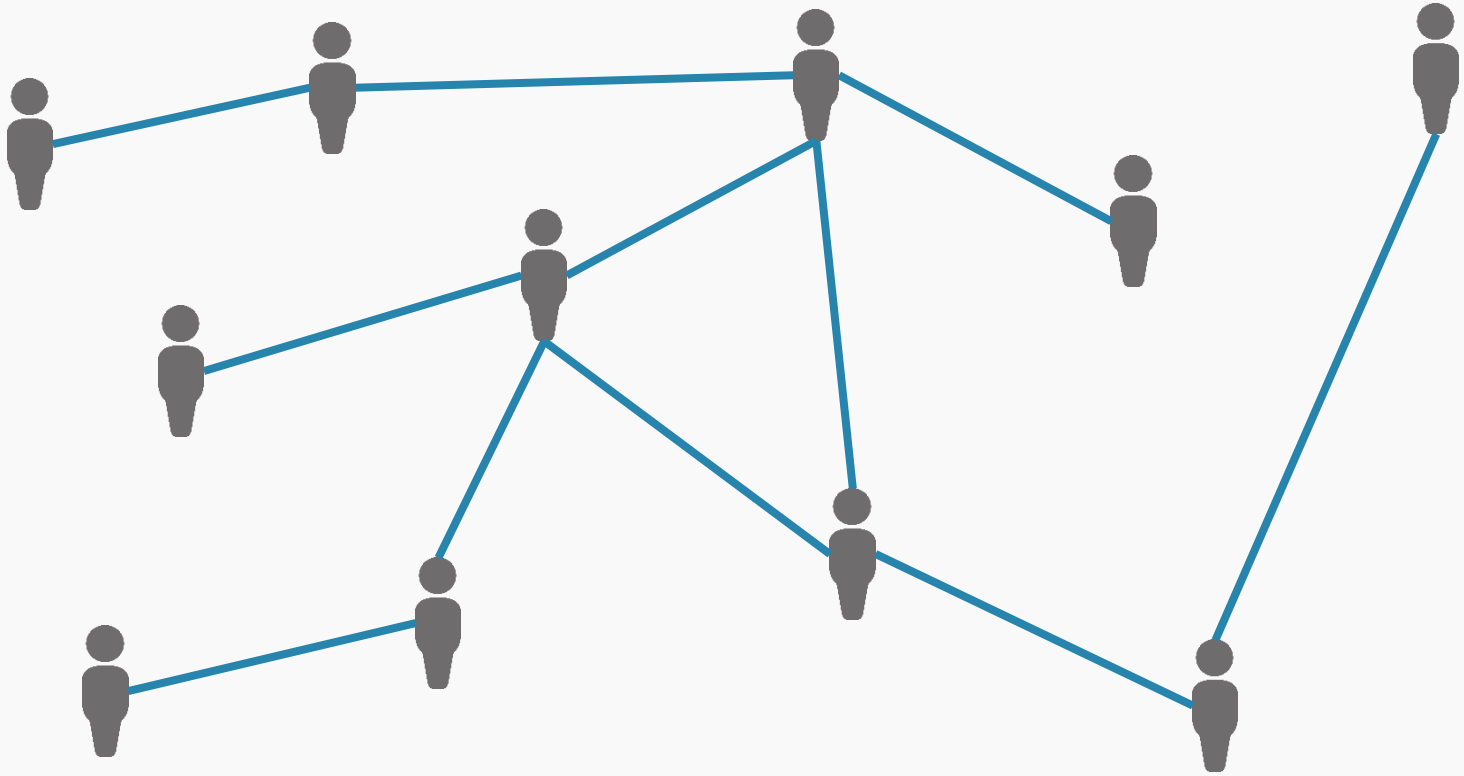
Comunicare Epidemic
Topologia se poate schimba în orice moment



Comunicare Epidemic
Topologia se poate schimba în orice moment

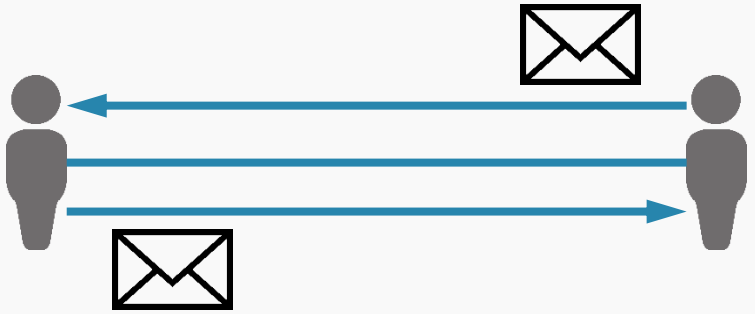


Comunicare Epidemic
Topologia se poate schimba în orice moment

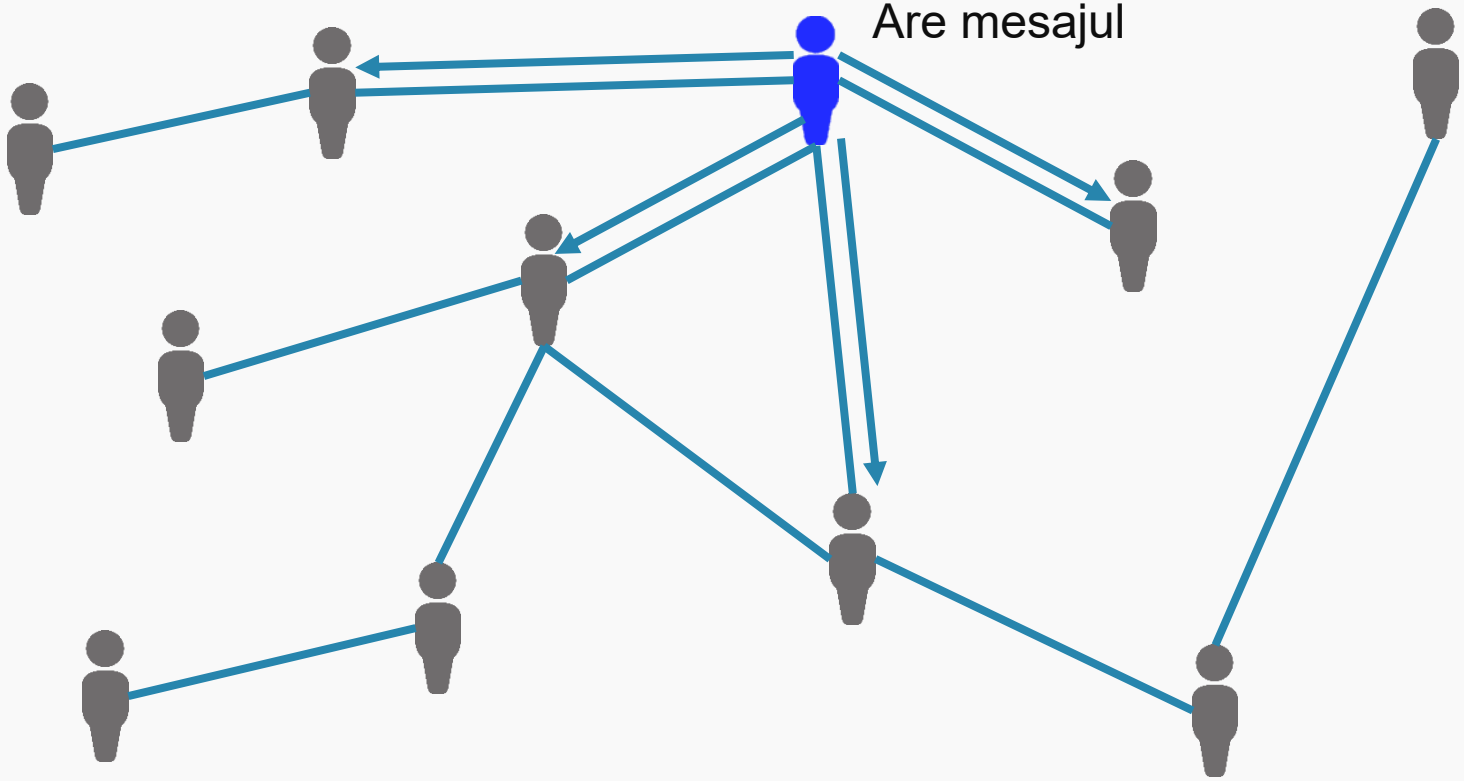


Epidemic

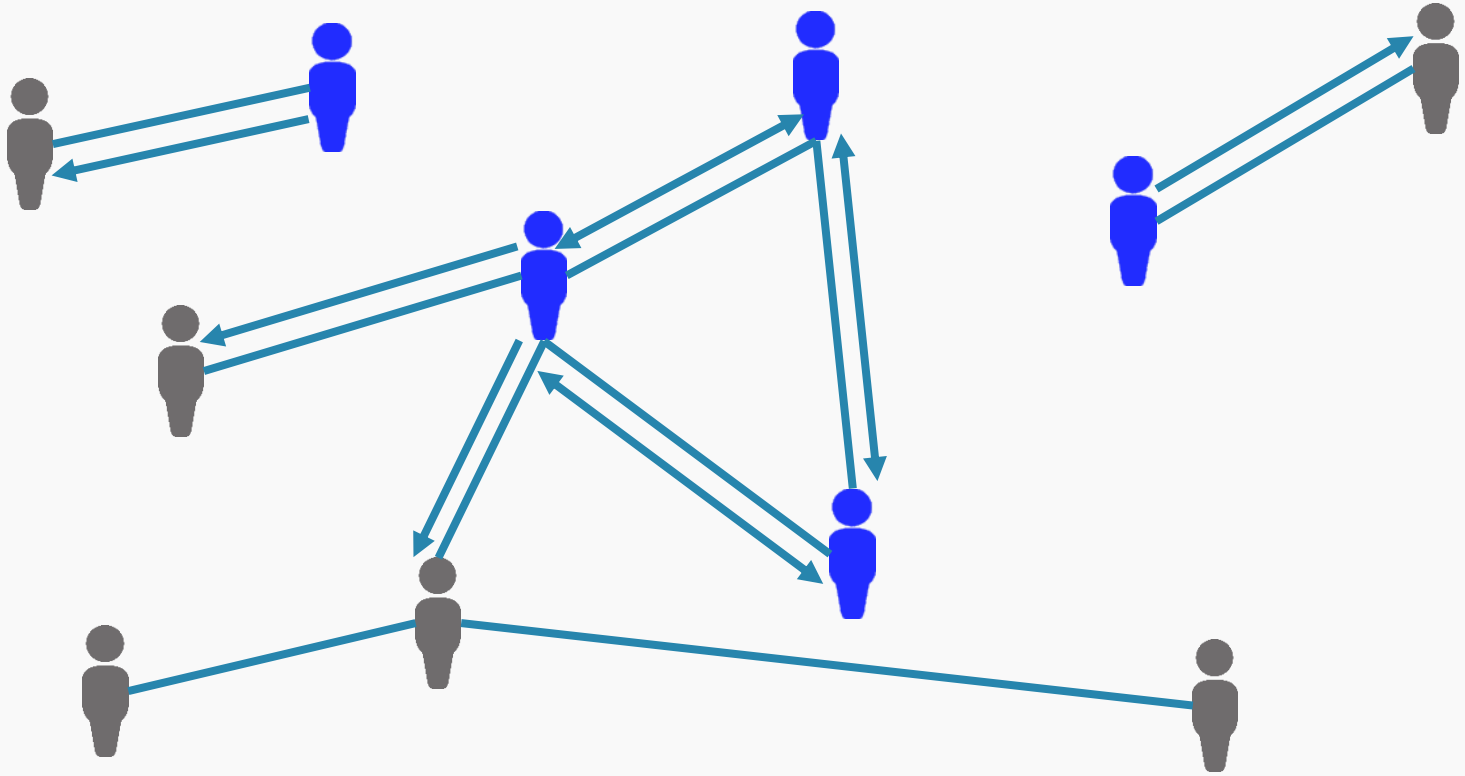
Cât timp avem conexiune nodurile schimbă mesaje.



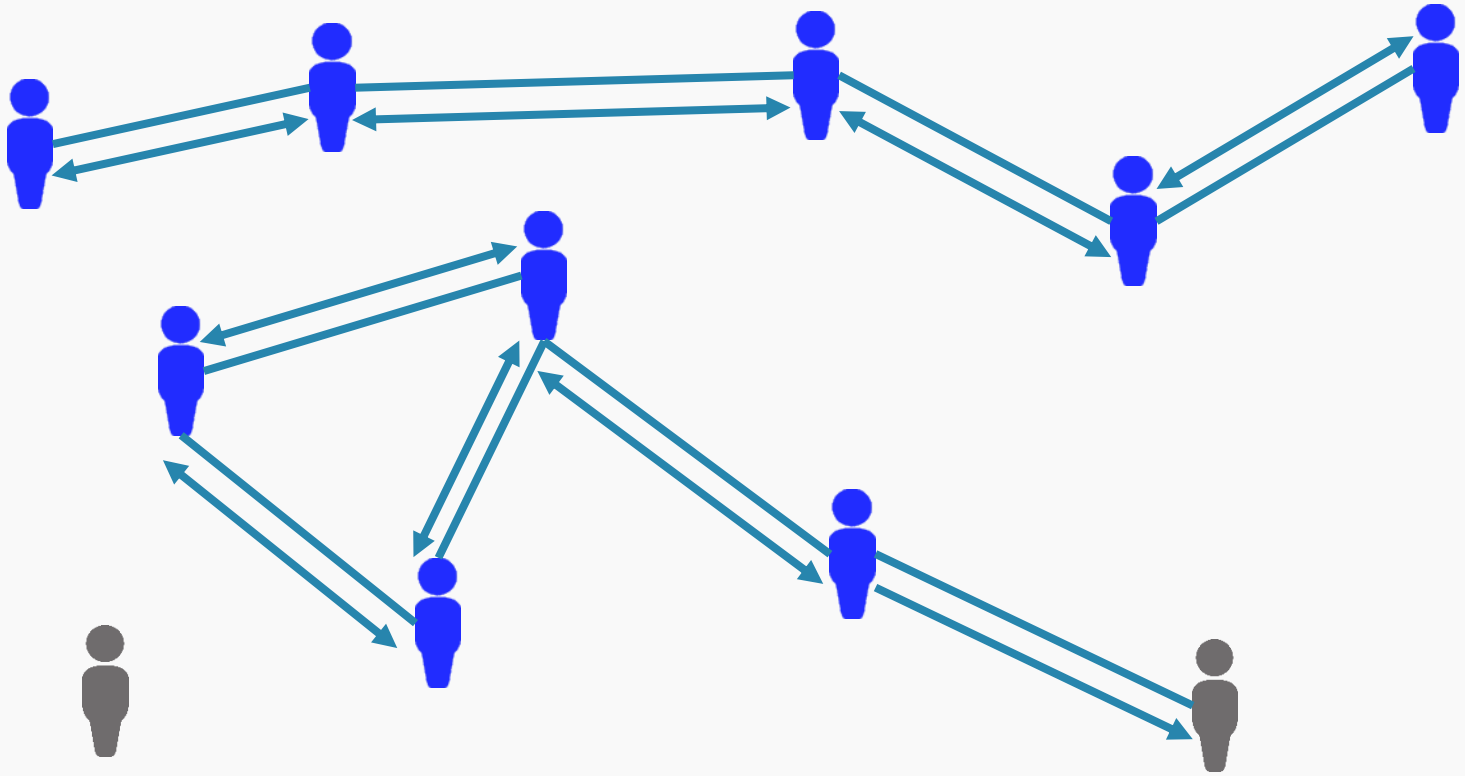
Comunicare Epidemic
Topologia se poate schimba în orice moment



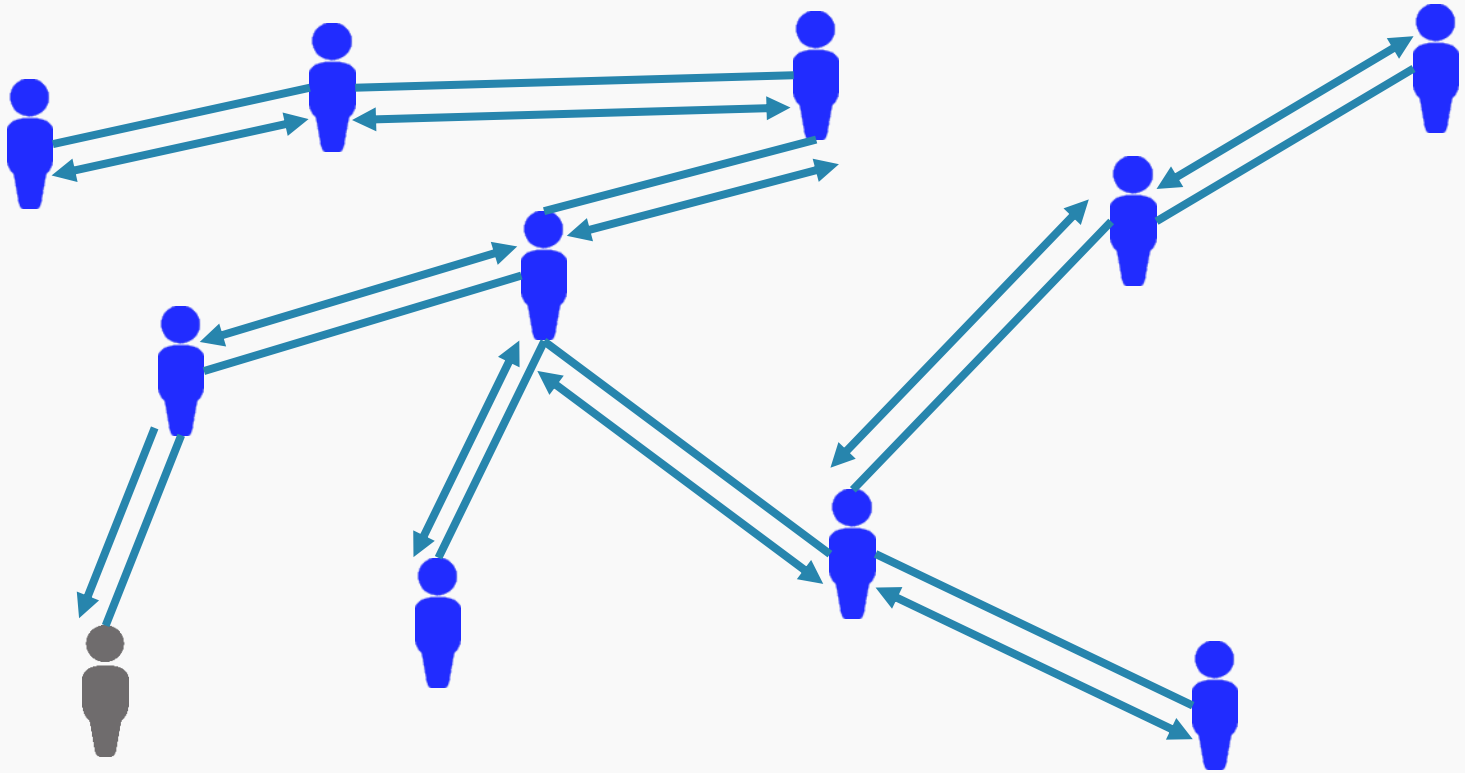
Comunicare Epidemic
Topologia se poate schimba în orice moment



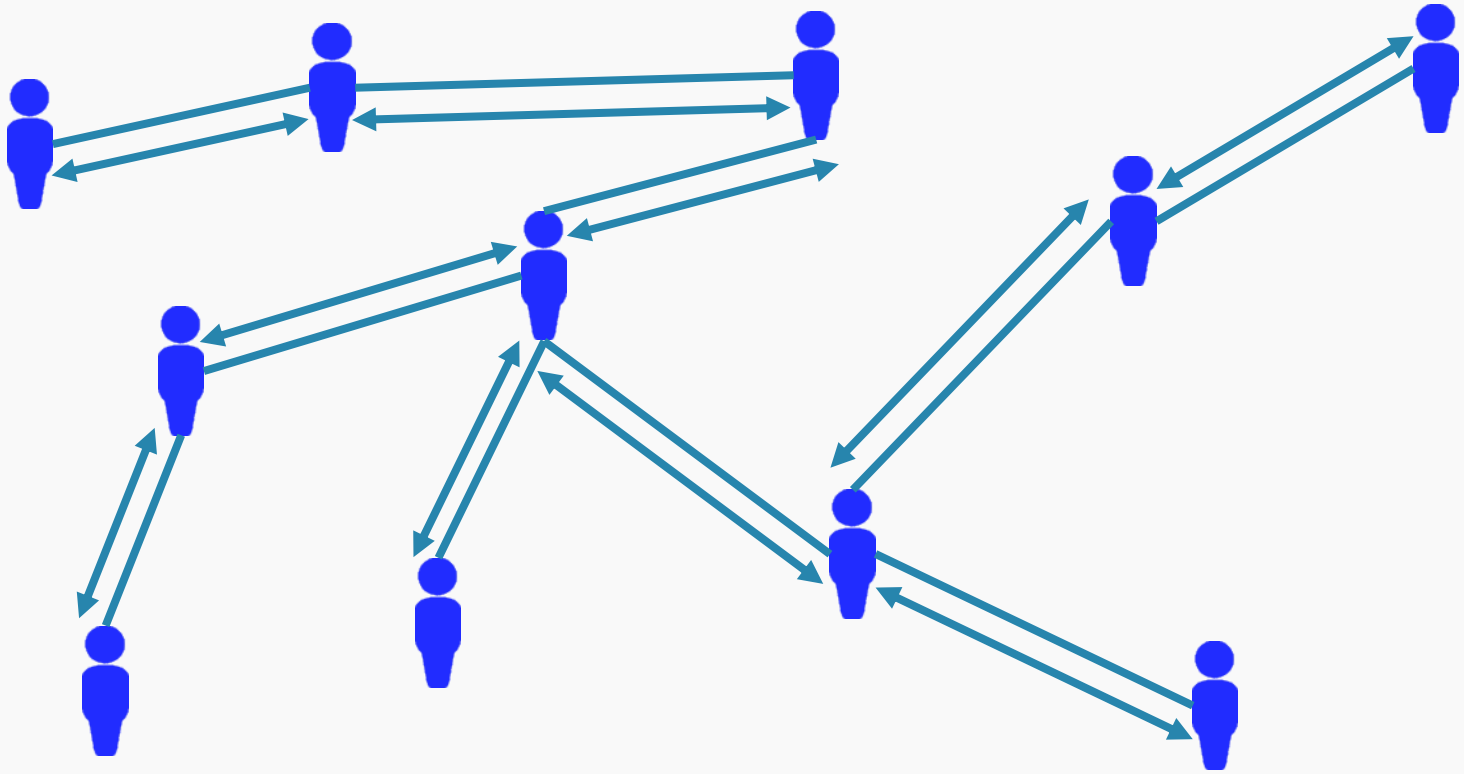
Comunicare Epidemic
Topologia se poate schimba în orice moment



Comunicare Epidemic
Topologia se poate schimba în orice moment



Comunicare Epidemic
Topologia se poate schimba în orice moment



Epidemic – Avantaje

Nu mai există entități centralizate gen ISP

Funcționează în cazuri extreme
(cutremur/potop) când cablurile ar putea fi rupte

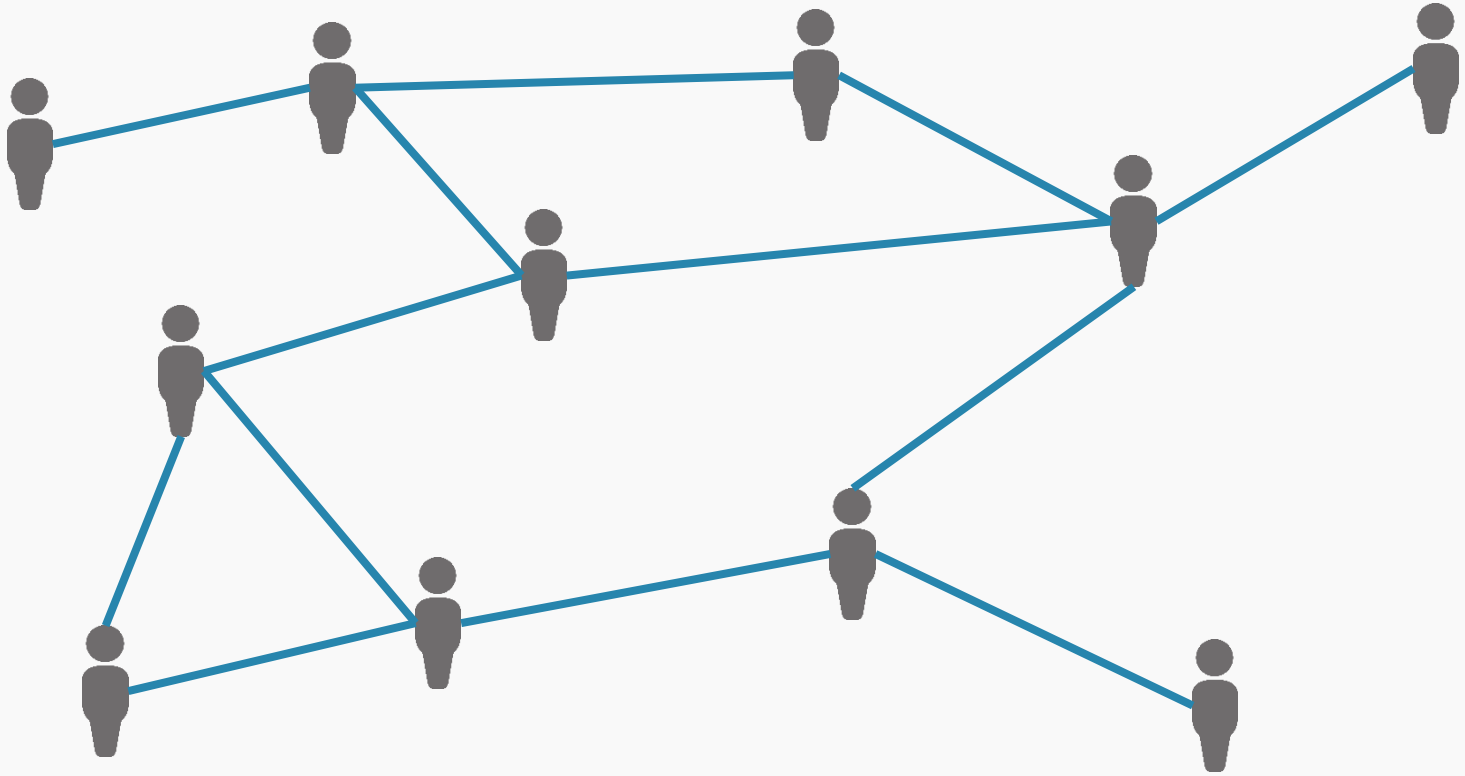
Epidemic – De ce nu e folosit în loc de IP?

Epidemic – De ce nu e folosit în loc de IP?

Necesită multe transmisii inutile pentru orice mesaj

Necesită spațiu de stocare mare pe fiecare dispozitiv pentru a evita cicluri

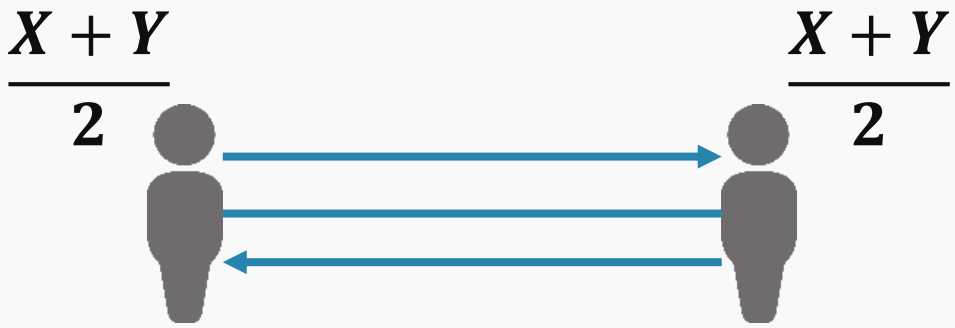
Epidemic – numărare nodurilor



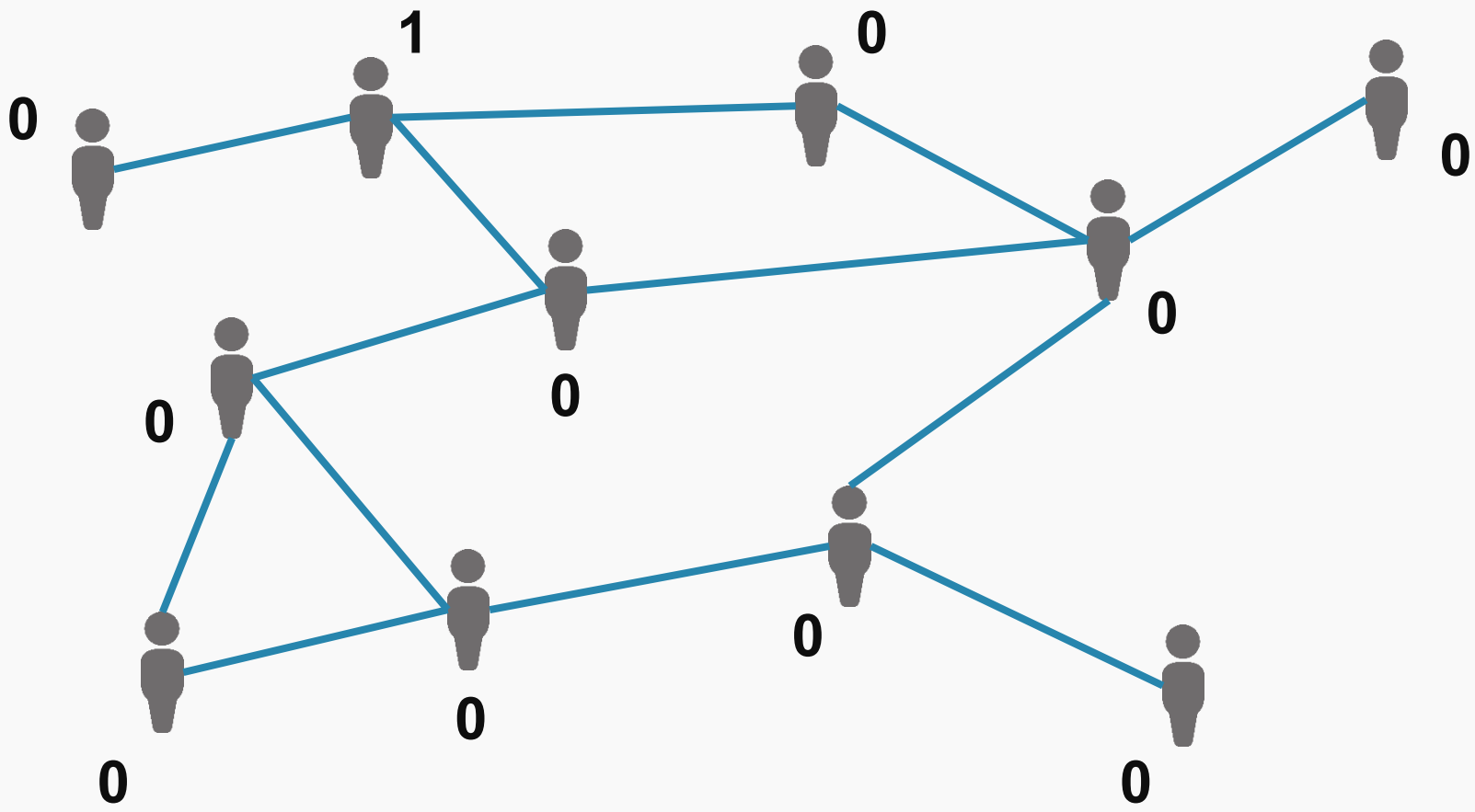
Epidemic – numărare nodurilor

Epidemic – numărare nodurilor

Epidemic – numărare nodurilor

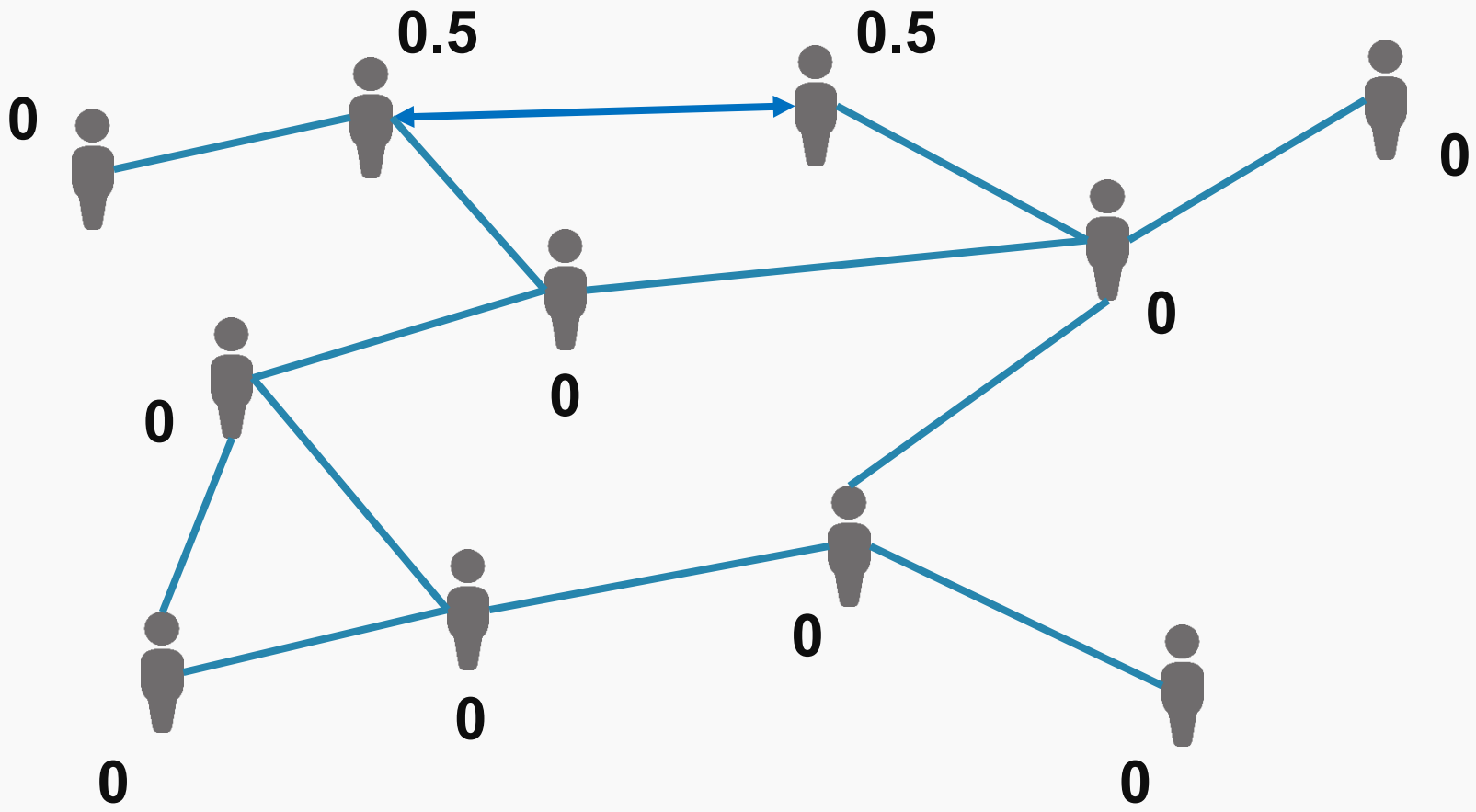


Epidemic – numărare nodurilor



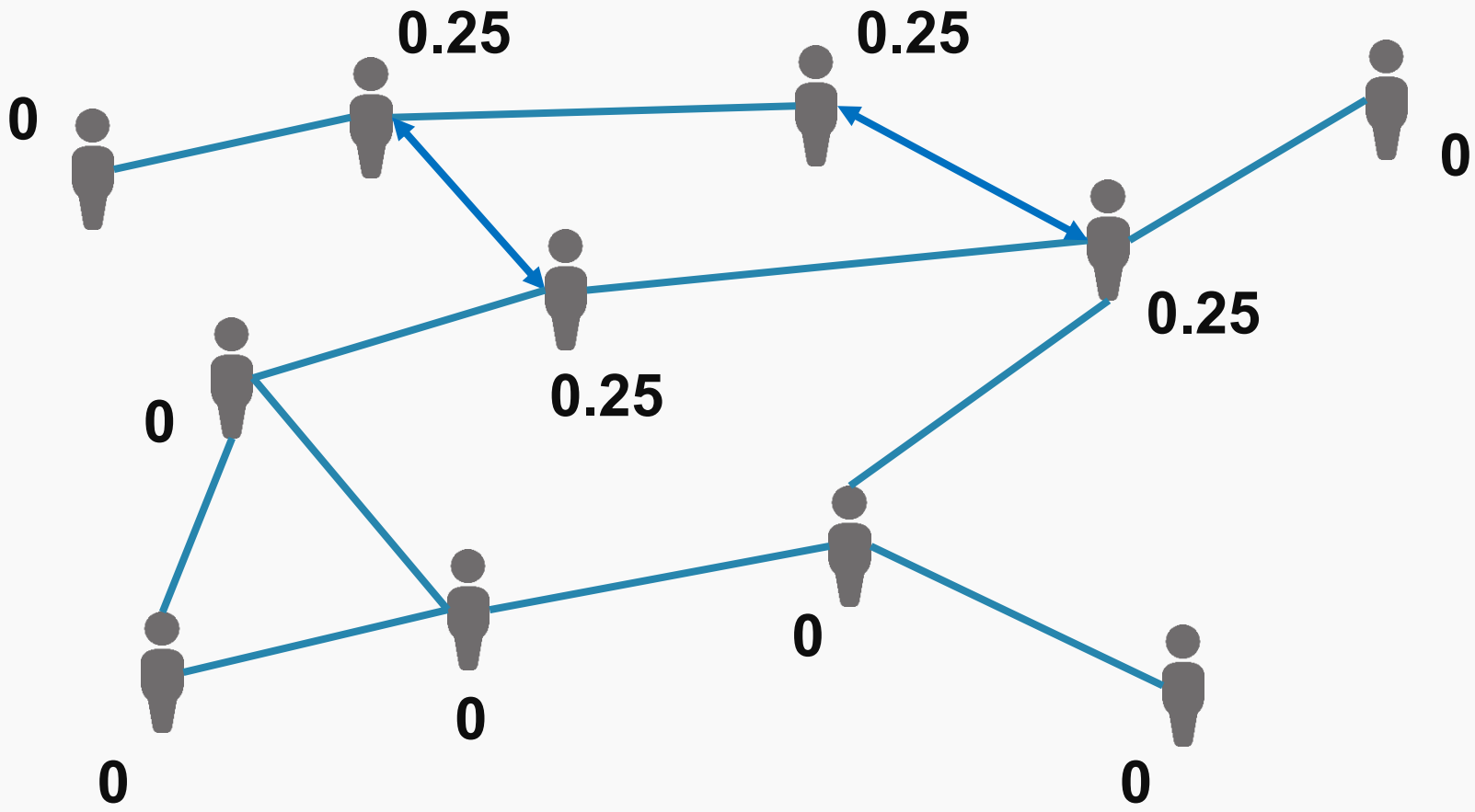
Sum of all = 1

Epidemic – numărare nodurilor



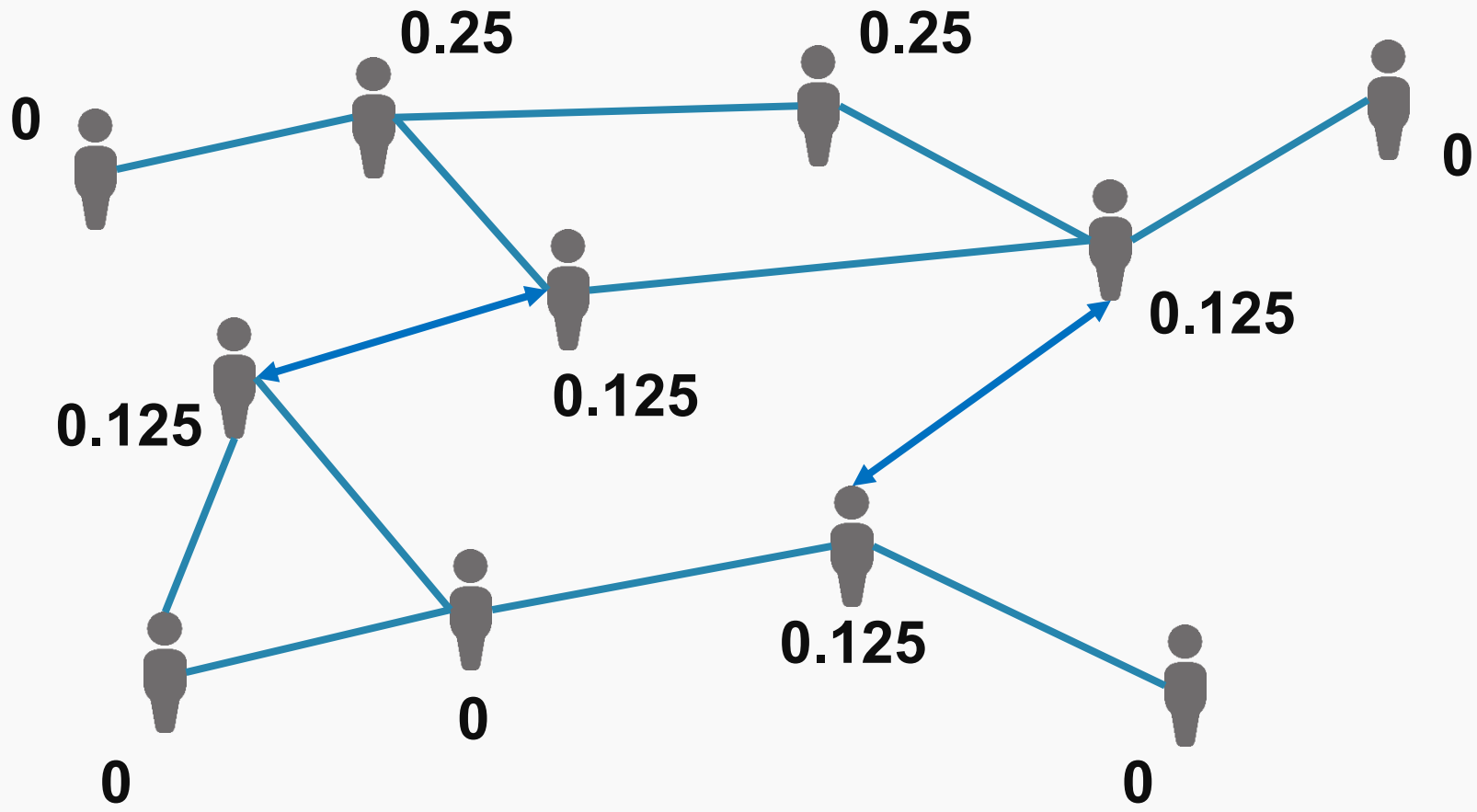
Sum of all = 1

Epidemic – numărare nodurilor



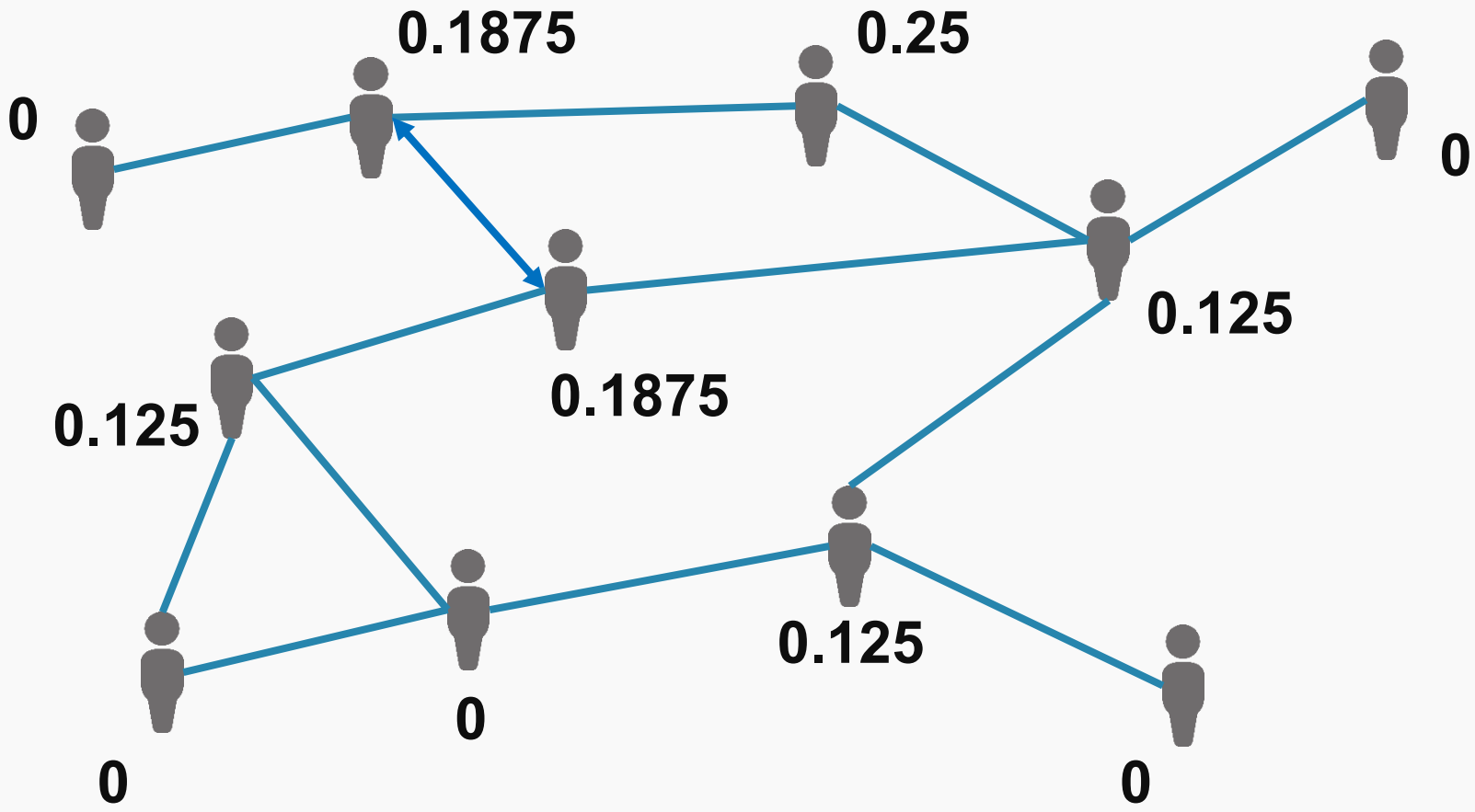
Sum of all = 1

Epidemic – numărare nodurilor



Sum of all = 1

Epidemic – numărare nodurilor

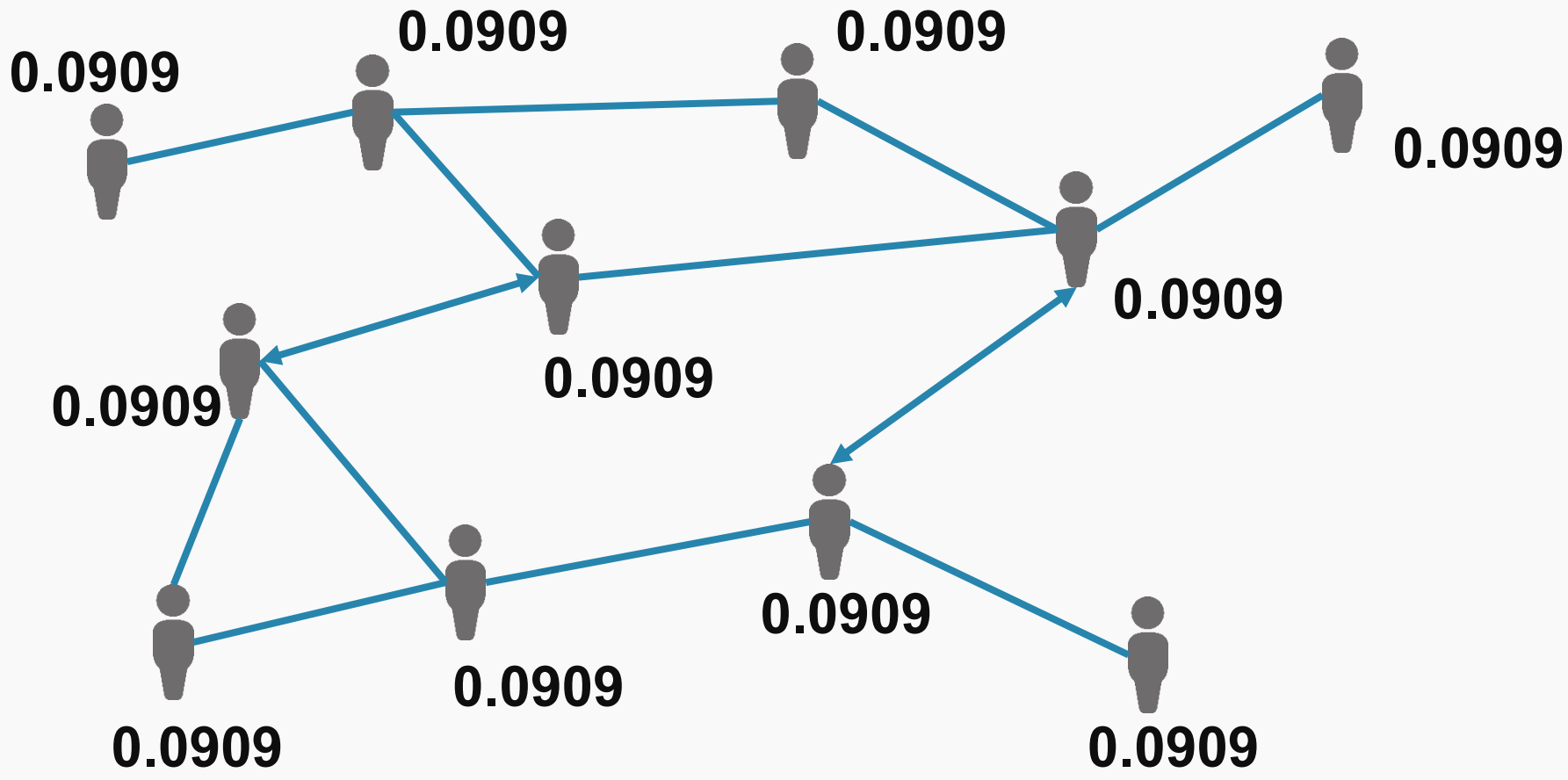


Sum of all = 1

Epidemic – numărare nodurilor

Și tot așa…

Epidemic – numărare nodurilor



Sum of all = 1

Epidemic – numărare nodurilor

$$\frac{1}{0.0909} = 11(\text{numărul nodurilor})$$