

ΔΙΑΧΕΙΡΙΣΗ ΔΙΚΤΥΩΝ

Δρομολόγηση Επιπέδου IP στο Internet

Άμεση – Έμμεση Δρομολόγηση

Δρομολόγηση εντός Αυτόνομης Περιοχής (IGP)

Δρομολόγηση μεταξύ Αυτόνομων Περιοχών (BGP)

Αλγόριθμοι Distance Vector (Bellman)

Αλγόριθμοι Link State (Dijkstra)

B. Μάγκλαρης

maglaris@netmode.ntua.gr

www.netmode.ntua.gr

22/10/2018

ΔΡΟΜΟΛΟΓΗΣΗ ΣΤΟ INTERNET

(επανάληψη)

- **Πρώθηση (forwarding):** Μετακίνηση πακέτων από την είσοδο δρομολογητή σε κατάλληλη έξοδο (**λειτουργία data plane**)
- **Δρομολόγηση (routing):** καθορισμός διαδρομής πακέτων από πηγή προς προορισμό, *routing algorithms* (**λειτουργία control plane**)

Σημείωση: Στις επόμενες διαφάνειες χρησιμοποιήθηκε υλικό υποστήριξης του βιβλίου των Kurose & Ross “**Computer Networking: A Top-Down Approach,**” Pearson Publisher, 6th Edition.

ΔΡΟΜΟΛΟΓΗΣΗ ΣΤΟ INTERNET

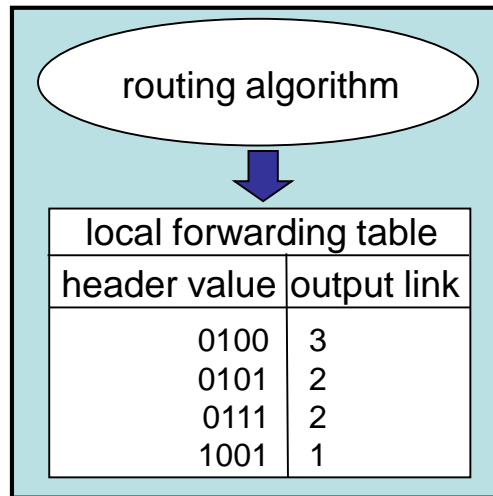
Σε κάθε δρομολογητή (router) τηρούνται σε ηλεκτρονική μνήμη:

1. Πίνακας Δρομολόγησης (**Routing Table**)

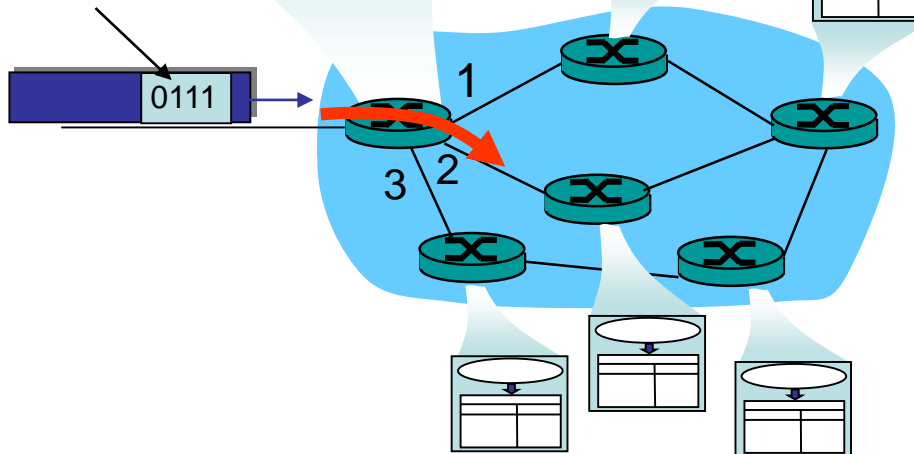
- Επίπεδο Ελέγχου (*Control Plane*)
- Για γνωστά destination IP prefixes (δίκτυα) ορίζει την επόμενη στάση IP (next hop IP)
- Σε περίπτωση διπλών εγγραφών, προτιμάται η επιλογή με βάση τη λεπτομερέστερη διεύθυνση IP (*Longest Prefix Match*, π.χ. η 147.102.13.0 προτιμάται της γενικευμένης 147.102.0.0)

2. Πίνακας Προώθησης (**Forwarding Table**)

- Επίπεδο Δεδομένων (*Data Plane*)
- Τηρείται σε ταχύτατη μνήμη (*TCAM*)
- Αντιστοιχεί ανά πακέτο διευθύνσεις IP του Πίνακα Δρομολόγησης σε φυσικές εξόδους (Interfaces) του δρομολογητή



Τιμή στην επικεφαλίδα του πακέτου



ΔΡΟΜΟΛΟΓΗΣΗ ΕΠΙΠΕΔΟΥ ΔΙΚΤΥΟΥ

Layer 3 Intra-AS Routing (επανάληψη)

- **Intra- AS Routing, Interior Gateway Protocols (IGP):** Μια έξοδος προς επόμενο Interface για κάθε υποδίκτυο (subnet) τελικό προορισμό
 - RIP (Routing Internet Protocol): Πρωτόκολλο παλαιάς γενιάς, βασισμένο σε αλγόριθμους **distance vector (Bellman Ford)**
 - OSPF (Open Shortest Path First): Το πιο διαδομένο σήμερα, βασισμένο σε αλγόριθμους **link state (Dijkstra)**. Για λόγους κλιμάκωσης, μπορεί να υλοποιηθεί ιεραρχικά γύρω από την περιοχή κορμού (**backbone area 0** ή 0.0.0.0) και με συνδεδεμένες περιφερειακές περιοχές (**stub areas**). Για δρομολόγηση πακέτων Pn4, ισχύει η OSPF Version 2: RFC 2328, 1998
 - IS-IS
 - Δυνατότητα πολλαπλών εναλλακτικών δρόμων ίσου κόστους (ECMP): Direct routing μεταξύ γειτονικών δρομολογητών με παράλληλες συνδέσεις και επιλογή εξόδου με proprietary αλγορίθμους (π.χ. Per-Packet Round-Robin, Per-Flow Load Balancing, Per Source -Destination Traffic Engineering...) Για αυτοματισμό του OSPF απαιτείται τροποποίηση του αλγορίθμου **Dijkstra για k-Shortest paths**

ΔΡΟΜΟΛΟΓΗΣΗ ΕΠΙΠΕΔΟΥ ΔΙΚΤΥΟΥ

Layer 3 Inter-AS Routing (επανάληψη)

- **Inter-AS Routing, Exterior (Border) Gateway Protocols (EGP/BGP):**
Πολλές εναλλακτικές διαδρομές με βάρη προς όλα τα γνωστά δίκτυα (περίπου 750.000 σήμερα) μεταξύ ακραίων (border) routers αυτονόμων συστημάτων (Autonomous Systems, AS, περίπου 62.000 σήμερα)
 - Η διαδρομή καταγράφεται στον BGP Table των ακραίων δρομολογητών (border gateways) ενός AS ανά prefix (υποδίκτυο προορισμού) και την σειρά των AS's της προτεινόμενης διαδρομής (μαζί με το βάρος της)
 - Οι πίνακες BGP φυλάσσονται στην ηλεκτρονική μνήμη των border gateways και ανανεώνονται δυναμικά όποτε υπάρχουν αλλαγές στο Internet με ευθύνη των γειτονικών δρομολογητών (border gateways) που ανακοινώνουν τα δίκτυα των αυτονόμων κοινοτήτων (AS's) που γνωρίζουν (*advertising*)
 - Ο υπολογισμός των «βέλτιστων» Inter-AS δρόμων γίνεται κατανεμημένα, βασισμένος σε αλγόριθμους *distance vector (Bellman Ford)* με κόσθη τα βάρη των συνδέσεων μεταξύ border gateways. Οι δρόμοι Inter-AS μπορεί να αλλάξουν στη πορεία ενός πακέτου προς τον τελικό του προορισμό

ΔΡΟΜΟΛΟΓΗΣΗ ΕΠΙΠΕΔΟΥ ΔΙΚΤΥΟΥ

Είδη Intra-AS Routing

- Άμεση δρομολόγηση (**direct routing**)
 - Κάθε κόμβος (PC, router) στέλνει πακέτα IP σε **interface κόμβου** του ίδιου υποδικτύου
- Έμμεση δρομολόγηση (**indirect routing**)
 - Ο τελικός κόμβος στέλνει πακέτα IP σε κόμβο του ίδιου δικτύου – υποδικτύου, χρησιμοποιώντας ενδιάμεσους δρομολογητές (routers). Για την προώθηση πακέτων σε Επίπεδο 2 (L2, Ethernet) μαθαίνει την αντιστοίχιση του **interface δρομολογητή (gateway) & την διεύθυνση L2 (MAC) μέσω ARP (Address Resolution Protocol)**
 - Ο δρομολογητής πρέπει να γνωρίζει τη διαδρομή (επόμενο interface δρομολογητή) προς το δίκτυο – υποδίκτυο προορισμού
 - Μεταξύ δρομολογητών κορμού εντός AS υπάρχουν πίνακες δρομολόγησης (routing tables) προς όλα τα εσωτερικά υποδίκτυα (και τα άμεσα συνδεδεμένα σε γειτονικά AS) που ανανεώνονται με πρωτόκολλα **IGP** (συνήθως **OSPF** με core και stub areas)
 - Οι τελικοί κόμβοι στέλνουν πακέτα με διεύθυνση προορισμού εκτός του δικτύου – υποδικτύου τους σε **default gateway** (π.χ. 147.102.13.200)

ΔΡΟΜΟΛΟΓΗΣΗ ΕΠΙΠΕΔΟΥ ΔΙΚΤΥΟΥ

Intra-AS Routing, Host Routing Table

- Εγγραφές του τύπου (N, R)

- N: Δίκτυο προορισμού
- R: Επόμενο interface δρομολογητή (gateway)

- **Host routing table σε λειτουργικό Windows από το μηχάνημα με IP 147.102.13.32**

> netstat -nr

Routing Table:

Network	Destination	Netmask	Gateway	Interface	Metric
	0.0.0.0	0.0.0.0	147.102.13.200	147.102.13.32	20
	127.0.0.0	255.0.0.0	127.0.0.1	127.0.0.1	1
	147.102.13.0	255.255.255.0	147.102.13.32	147.102.13.32	20
	147.102.13.32	255.255.255.255	127.0.0.1	127.0.0.1	20
	147.102.255.255	255.255.255.255	147.102.13.32	147.102.13.32	20
	224.0.0.0	240.0.0.0	147.102.13.32	147.102.13.32	20

- Προς το ίδιο τοπικό υποδίκτυο **147.102.13.0/24** σαν gateway ορίζεται κατευθείαν (**direct**) το τοπικό interface **147.102.13.32**
 - Προς destination `dolly.netmode.ntua.gr` (**147.102.13.10**) gateway θα είναι το τοπικό interface **147.102.13.32**
- Προς όλα τα άλλα δίκτυα **0.0.0.0** σαν gateway ορίζεται το **147.102.13.200** (**default gateway: router.netmode.ntua.gr**)
- Προς **local host 127.0.0.0/8** (π.χ. για δοκιμή δικτυακών εφαρμογών τοπικά) ή προς το ίδιο το **147.102.13.32/32** «προωθούνται» στο «interface» **127.0.0.1**
- Προς διεύθυνση broadcast **147.102.255.255/32** σαν gateway ορίζεται κατευθείαν (**direct**) το τοπικό interface **147.102.13.32** (η διεύθυνση δεν ισχύει στο δίκτυο του ΕΜΠ)
- Προς διευθύνσεις multicast **224.0.0.0/4** σαν gateway ορίζεται κατευθείαν (**direct**) το τοπικό interface **147.102.13.32**

ΔΡΟΜΟΛΟΓΗΣΗ ΕΠΙΠΕΔΟΥ ΔΙΚΤΥΟΥ

Intra-AS Routing, Links between Routers

- Για ομοιομορφία της δρομολόγησης, κάθε γραμμή ορίζεται (συνήθως) σαν δίκτυο με 4 τουλάχιστον διευθύνσεις (/30)
- Παράδειγμα: Μεταξύ ΕΜΠ **147.102.0.0/16** & Παν. Αθηνών **195.134.64.0/18** ορίζεται το «δίκτυο» **147.102.224.32/30**
 - Υποδίκτυο: **147.102.224.32/30**
 - Άκρο ΕΜΠ: **147.102.224.33/30**
 - Άκρο Παν. Αθηνών: **147.102.224.34/30**
 - Broadcast: **147.102.224.35/30**

ΠΑΡΑΔΕΙΓΜΑ ΕΣΩΤΕΡΙΚΗΣ ΔΡΟΜΟΛΟΓΗΣΗΣ: ΤΟ ΔΙΚΤΥΟ ΤΟΥ Ε.Μ.Π.

ntua.gr (147.102.0.0/16, ASN 3323)

ΠΡΟΣΟΧΗ

Οι πίνακες δρομολόγησης στο Internet για λόγους ομοιομορφίας είναι της μορφής:

- **Prefix Δικτύου Τελικού Προορισμού :: IP Interface Εισόδου Επόμενου Κόμβου**

ΠΑΡΑΔΕΙΓΜΑ:

Ο δρομολογητή του Ε.Μ.Π. 147.102.224.33 βρίσκει τον δρομολογητή του ΕΚΠΑ

147.102.224.34 σαν μέλος του υποδικτύου:

- 147.102.224.32/30 (παροχή διευθύνσεων από Ε.Μ.Π.)

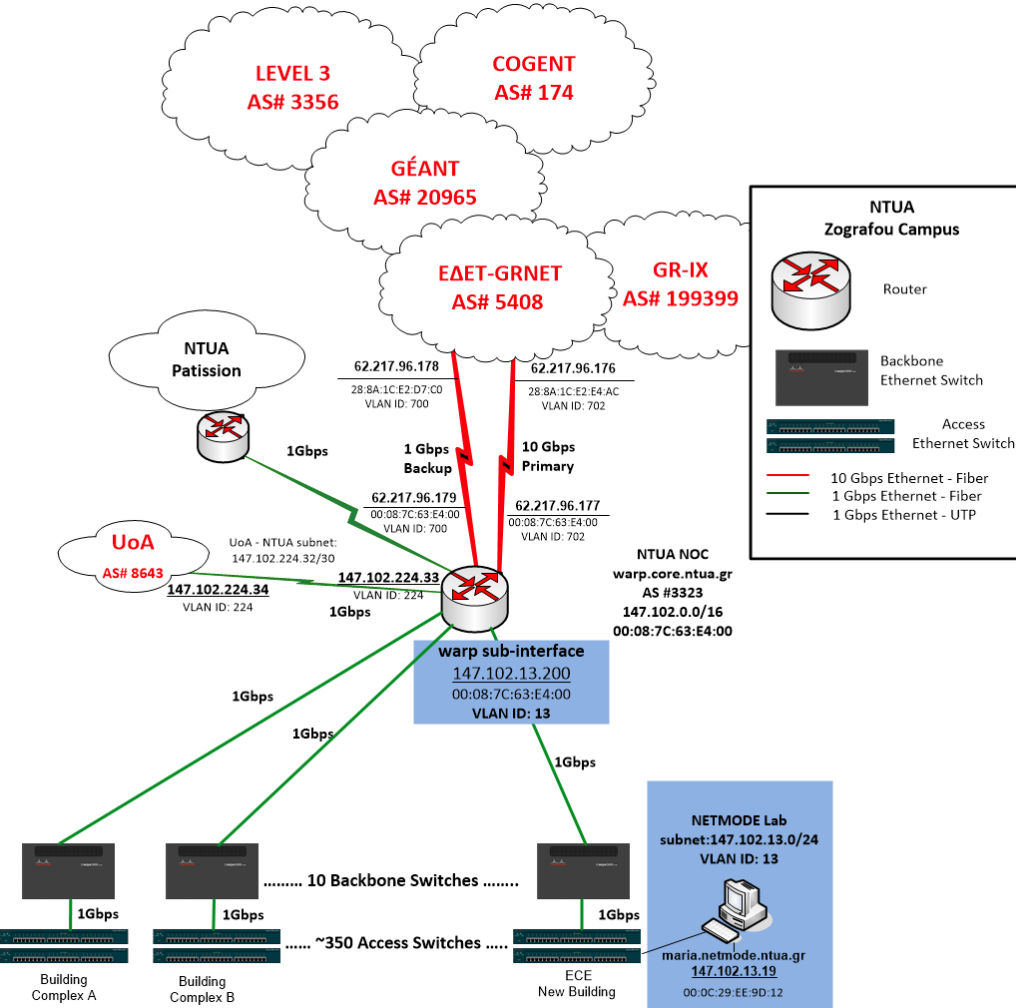
Η γραμμή Ε.Μ.Π. – ΕΚΠΑ (όπως όλες οι γραμμές σε Δίκτυα Internet) ορίζεται σαν υποδίκτυο (prefix) με 4 τουλάχιστον διευθύνσεις IP:

- Υποδίκτυο: 147.102.224.32
- Άκρο Ε.Μ.Π.: 147.102.224.33
- Άκρο ΕΚΠΑ: 147.102.224.34
- Broadcast: 147.102.224.35

ΑΝΤΙ-ΠΑΡΑΔΕΙΓΜΑ:

Ο δρομολογητή του Ε.Μ.Π. 62.217.96.177 βρίσκει τον δρομολογητή του ΕΔΕΤ 62.217.96.176 σαν μέλος του υποδικτύου:

- 62.217.96.176/31 (παροχή διευθύνσεων από ΕΔΕΤ)



ΑΛΓΟΡΙΘΜΟΙ ΕΥΡΕΣΗΣ ΔΡΟΜΩΝ

- **DV: Distance Vector** (αλγόριθμος **Bellman-Ford**)
 - IGP: **RIP** (Routing Information Protocol)
 - EGP: **BGP** (Border Gateway Protocol)
- **LS: Link State** (αλγόριθμος **Dijkstra**)
 - IGP: **OSPF** (Open Shortest Path First): Link State Data Base + αλγόριθμος Dijkstra στον κορμό Αυτόνομου Δικτύου (Core of an Autonomous System - AS)
 - Κόστος γραμμών δικτύου: Ανάλογα με την ταχύτητα ή οριζόμενα από τον Διαχειριστή
 - Ανανέωση κόστους γραμμών μέσω **LSA** (Link State Advertisement) κάθε 30 min (default) ή λόγω μεταβολής κατάστασης. Μετά από 60 min (time out) χωρίς LSA η γραμμή διαγράφεται
 - Σε περιφερειακά υποδίκτυα (stub areas): Default G/W
 - Για δίκτυα εκτός AS: Ανακοινώσεις εντός AS μέσω i-BGP

ΑΛΓΟΡΙΘΜΟΣ Distance Vector

IGP RIP – BGP (Bellman – Ford)

- Κάθε κόμβος υπολογίζει την επόμενη «βέλτιστη» στάση προς όλες τις κατευθύνσεις, σύμφωνα με την εικόνα που έχει τοπικά (**πραγματικά καταναεμημένος αλγόριθμος**)
- Χρειάζεται γνώση του «κόστους» των άμεσων συνδέσεων (interfaces) και το εκτιμώμενο «κόστος» από τους άμεσους γείτονες προς όλους τους προορισμούς (π.χ. όλα τα δίκτυα στο Internet που ανακοινώνουν - announce – σε έναν ακραίο δρομολογητή μιας αυτόνομης κοινότητας - AS - οι γειτονικές του αυτόνομες κοινότητες στο πρωτόκολλο BGP)
- Βασίζεται στον δυναμικό προγραμματισμό (με πιθανές επεκτάσεις για την ανακάλυψη εναλλακτικών δρόμων στο BGP)
- Οι κόμβοι εντοπίζουν τους βέλτιστους δρόμους (shortest paths) προς όλους τους κόμβους εκτελώντας αλγόριθμο βασισμένο στον δυναμικό προγραμματισμό (**dynamic programming**) που εισήγαγε ο Bellman
- Αρχικά παρουσίασε αστάθειες (π.χ. δρόμους με κύκλους - loops) αλλά σήμερα το πρωτόκολλο **EGP BGP** είναι η καρδιά του Internet
- Το πρωτόκολλο **IGP RIP** έχει αντικατασταθεί από το **OPSF** και το **IS-IS**

ΑΛΓΟΡΙΘΜΟΣ Link State

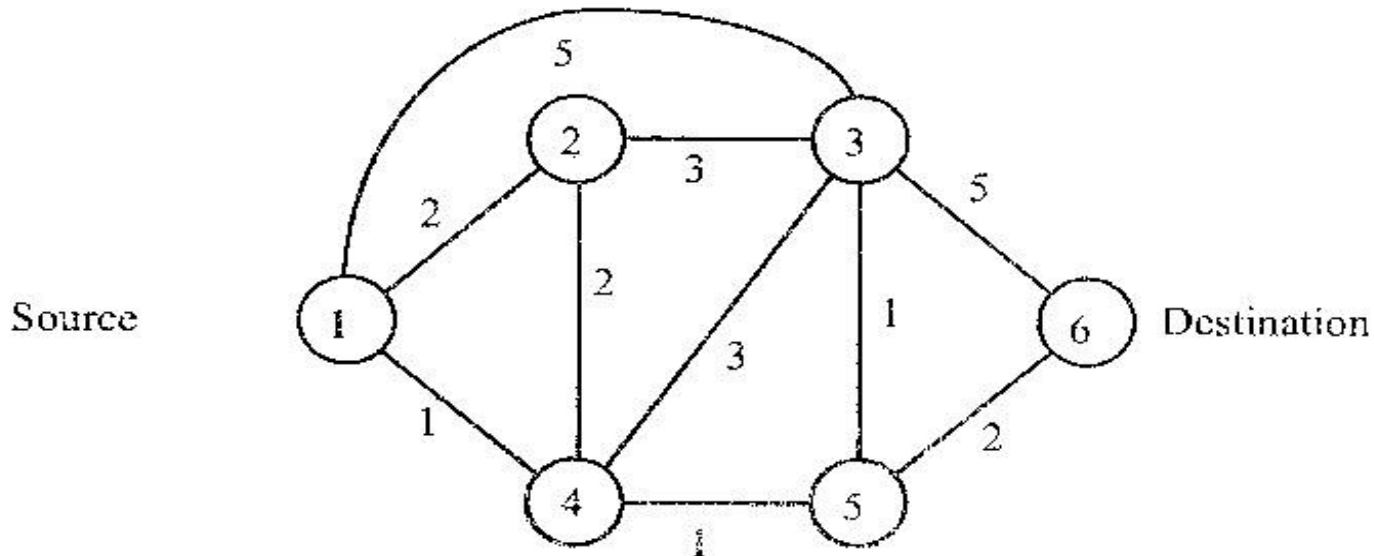
IGP OSPF (Dijkstra)

- Θεωρείται ευσταθής αλγόριθμος, επαρκής για IGP: Μια αυτόνομη κοινότητα ιεραρχείται εσωτερικά σε περιοχές OSPF 0 (μία ή περισσότερες) + περιφερειακές **stub areas** με static routing (*στο δίκτυο του ΕΜΠ, OSPF τρέχουν μόνο 2 δρομολογητές*)
- Κάθε δρομολογητής κορμού έχει πλήρη εικόνα της περιοχής του – τοπολογία, κόστη συνδέσεων
- Όλοι οι δρομολογητές κορμού εκτελούν τον αλγόριθμο Dijkstra για εντοπισμό όλων των δρόμων ελαχίστου κόστους (shortest paths) σε ρόλο κεντρικού συστήματος ελέγχου.
- Πληροφορίες κατάστασης δικτύου (τοπολογία, κόστη) μεταδίδονται με Link State Announcements (LSA) μεταξύ γειτόνων κάθε 1/2 ώρα (default) ή λόγω μεταβολής κατάστασης – θεωρητικά όλοι έχουν την ίδια εικόνα
- **Τα LSA μεταδίδονται σαν αυτόνομα IP datagrams, όχι μέσω πρωτοκόλλων TCP ή UDP**
- Γενίκευση του OSPF: ECMP (Equal-Cost Multi-Path)
- Σε μεγάλα δίκτυα κορμού εφαρμόζεται εναλλακτικά ο Αλγόριθμος **IS-IS** (Intermediate System to Intermediate System)

ΕΝΔΕΙΚΤΙΚΗ ΕΚΤΕΛΕΣΗ ΑΛΓΟΡΙΘΜΩΝ

Bellman Ford & Dijkstra

- Δίκτυο (γράφος) αναφοράς
- Τα κόστη των γραμμών αφορούν και στις 2 κατευθύνσεις
- Στα παραδείγματα που ακολουθούν υπολογίζονται δένδρα ελαχίστων δρόμων (*shortest path trees*) από όλους προς την *ρίζα {6}* (Bellman Ford) και από την *ρίζα {1}* προς όλους (Dijkstra)
- **ΠΡΟΣΟΧΗ:** Η επιλογή του ρόλου της *ρίζας* του δένδρου (πηγή ή προορισμός) έγινε αυθαίρετα. Δεν εξαρτάται από τους αλγόριθμους που ισχύουν κατ' αναλογία για αντίστροφους ρόλους ρίζας



ΑΛΓΟΡΙΘΜΟΣ Bellman-Ford

Distance Vector (1/2)

Υπολογισμός Δένδρου Ελαχίστων Δρόμων (Shortest Path Tree)
προς τον κόμβο {6} από τους κόμβους {1, 2, 3, 4, 5}

$D_i^{(h)}$: Κόστος από κόμβο (j) προς προορισμό (destination) 6 στο βήμα h

$d_{ij}^{(h)}$: Κόστος (βάρος) γραμμής (i,j) στο βήμα h

$L(j) = D_i^{(h)}$ Εκτίμηση ελαχίστου κόστους (label) από τον { j } προς τον {6} στο βήμα h

$P(j)$: Επόμενος κόμβος από τον { j } προς τον {6} στο βήμα h

Αρχικά έχουμε: $D_i^{(h)} = 0 \ \forall h, D_i^{(0)} = \infty \ \forall i \neq 6$ ($D_6^{(0)} = 0, D_j^{(0)} = \infty, j = 1,2,3,4,5$)

Για κάθε διαδοχικό $h \geq 0$: $D_i^{(h+1)} = \min_j [D_j^{(h)} + d_{ij}] \ \forall i \neq 6$

Αν: $\forall i \neq 6 \ D_i^{(n+1)} = D_i^{(n)}$ για δύο διαδοχικά h τότε σταματάμε τον αλγόριθμο, ή διαφορετικά έπειτα από N επαναλήψεις.

Στην χειρότερη περίπτωση ο αλγόριθμος πρέπει να επαναληφθεί $N-1$ φορές για $N-1$ κόμβους και με $N-1$ εναλλακτικές λύσεις. Επομένως, πρόκειται για αλγόριθμο πολυπλοκότητας $O(N^3)$.

ΑΛΓΟΡΙΘΜΟΣ Bellman-Ford

Distance Vector (2/2)

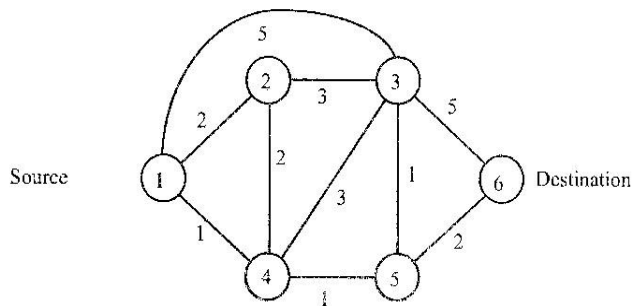
Παράδειγμα: INITIAL LABELS: $L(1)=L(2)=\dots=L(5)=\infty$, $L(6)=0$

UPDATE ORDER 5,4,3,2,1

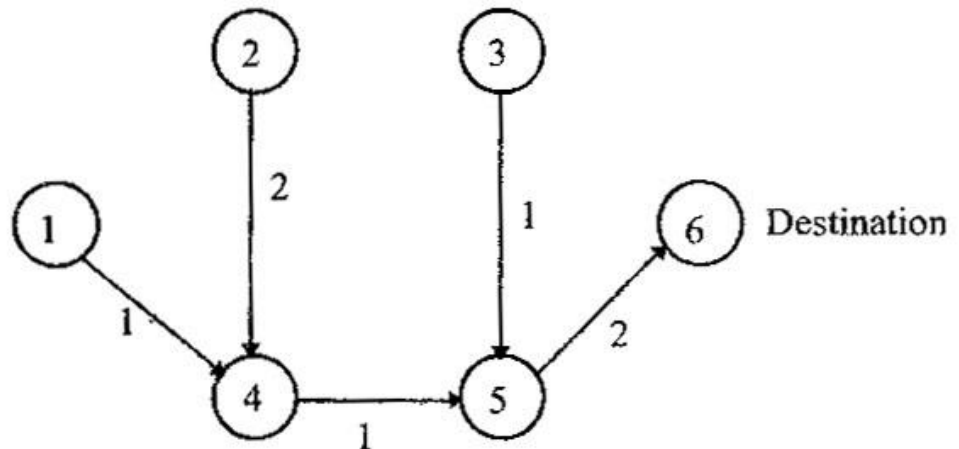
Iteration Number	Labels $L(n)$, Current Predecessor Node $P(n)$				
	$L(5), P(5)$	$L(4), P(4)$	$L(3), P(3)$	$L(2), P(2)$	$L(1), P(1)$
1	2 6	3 5	3 5	5 4	4 4
2	2 6	3 5	3 5	5 4	4 4

UPDATE ORDER 1,2,3,4,5

Iteration Number	Labels $L(n)$, Current Predecessor Node $P(n)$				
	$L(1), P(1)$	$L(2), P(2)$	$L(3), P(3)$	$L(4), P(4)$	$L(5), P(5)$
1	∞ -	∞ -	5 6	∞ -	2 6
2	10 3	8 3	3 5	3 5	2 6
3	4 4	5 4	3 5	3 5	2 6
4	4 4	5 4	3 5	3 5	2 6



SHORTEST PATH TREE



ΑΛΓΟΡΙΘΜΟΣ Dijkstra

Link State (1/2)

Υπολογισμός Δένδρου Ελαχίστων Δρόμων (Shortest Path Tree)
από τον κόμβο $\{1\}$ από τους κόμβους $\{2, 3, 4, 5, 6\}$

P : Σύνολο από *permanent labels*

D_j : Κόστος από πηγή (*source*) $\{1\}$ προς κόμβο $\{j\}$

d_{ij} : Κόστος (βάρος) γραμμής (i,j)

$L(j) = D_j^{(h)}$ Εκτίμηση ελαχίστου κόστους (*label*) από τον $\{1\}$ προς τον $\{j\}$ στο βήμα h

$P(j)$: Προηγούμενος (*Predecessor*) κόμβος από τον $\{1\}$ προς τον $\{j\}$ στο βήμα h

Αρχικά έχουμε: $P = \{1\}$, $D_1 = 0$, $D_j = d_{1j} \forall j \neq 1$

Βήμα 1^ο: Βρες $i \notin P$ τέτοιο ώστε: $D_i = \min_{j \in P} D_j$ και κάνε: $P = P \cup \{i\}$ Εάν το P περιλαμβάνει όλους τους κόμβους τότε ο αλγόριθμος σταματά, αλλιώς:

Βήμα 2^ο: Για όλα $j \notin P$ $D_j = \min[D_p, D_i + d_{ij}]$ και ξανά στο 1^ο Βήμα

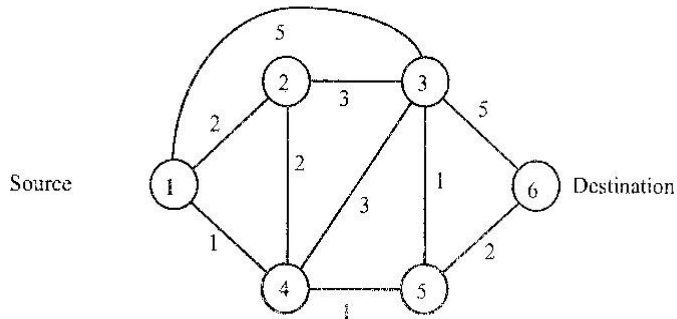
Σε κάθε βήμα ο αλγόριθμος απαιτεί έναν αριθμό πράξεων ανάλογο του N και έχουμε $N-1$ βήματα. Δηλ. είναι αλγόριθμος πολυπλοκότητας $O(N^2)$.

ΑΛΓΟΡΙΘΜΟΣ Dijkstra

Link State (2/2)

INITIAL LABELS: $L(1)=0, L(2)=L(3)=\dots=L(6)=\infty$

Iteration Number	Permanently Labeled Nodes	Labels $L(n)$, Current Predecessor Node $P(n)$				
		$L(2), P(2)$	$L(3), P(3)$	$L(4), P(4)$	$L(5), P(5)$	$L(6), P(6)$
1	1	2 1	5 1	1 1	∞ -	∞ -
2	1,4	2 1	4 4	-	2 4	∞ -
3	1,4,2	-	4 4	-	2 4	∞ -
4	1,4,2,5	-	3 5	-	-	4 5
5	1,4,2,5,3	-	-	-	-	4 5



SHORTEST PATH TREE

