ε.δε.μ²
ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ & ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΕΡΓΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Κατανεμημένη Υλοποίηση Ενισχυτικής Μάθησης:

1. Αλγόριθμος Bellman-Ford

2. Δρομολόγηση BGP στο Internet

καθ. Βασίλης Μάγκλαρης

maglaris@netmode.ntua.gr

www.netmode.ntua.gr

Video Conference μέσω Cisco Webex

Πέμπτη 14/5/2020

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Παράδειγμα Δυναμικού Προγραμματισμού: Βελτιστοποίηση Δρομολόγησης (*Επανάληψη*)

Εύρεση Δρόμων Ελάχιστου Κόστους από Κόμβο A *σε* Κόμβο J μέσω του μονο-κατευθυντικού γράφου όπως στο σχήμα με κατεύθυνση γραμμών ***A → Δ***

Ενδεικτικό κόστος γραμμών: $A \to B: 2,\ B \to A: \infty$

$B \to F: 4,\ F \to B: \infty$

Ενδεικτικό κόστος δρόμου: Δρόμος $\{A, B, F, I, J, Q\}$: $2 + 4 + 3 + 4 = 13$

Κατάσταση Περιβάλλοντος: Κόμβος σε παρούσα διερεύνηση $\{A, B, \dots, J\}$

Αποφάσεις Agent: Επόμενος κόμβος για διερεύνηση $\{up, down, staight\}$

Αναδρομικός Υπολογισμός *Q-Factors*:

$Q(H, down) = 3, \qquad Q(I, up) = 4$

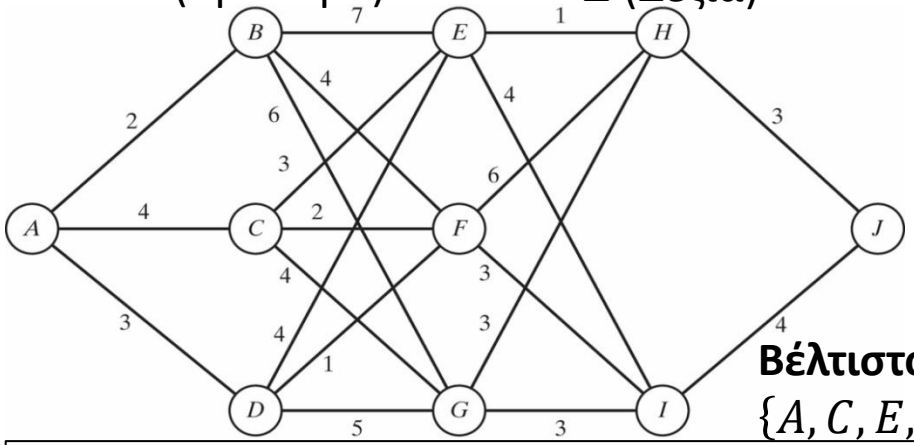
$Q(E, staight) = 1 + 3 = 4, \qquad Q(E, down) = 4 + 4 = 8$

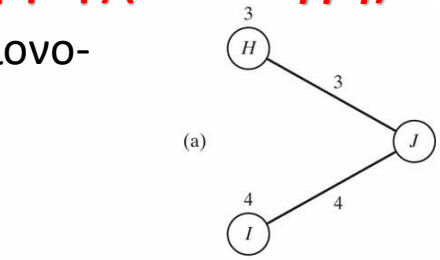
$Q(F, up) = 6 + 3 = 9, \qquad Q(F, down) = 3 + 4 = 7$

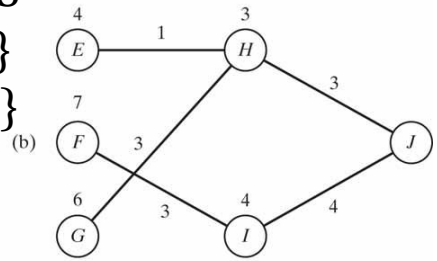
.....

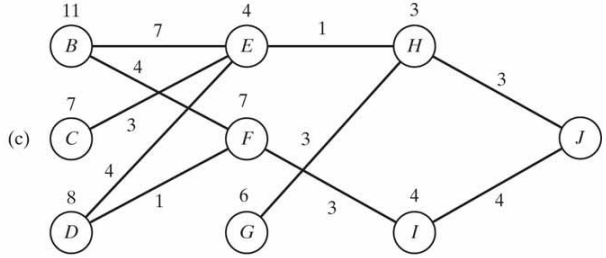
Κατεύθυνση Γραμμών

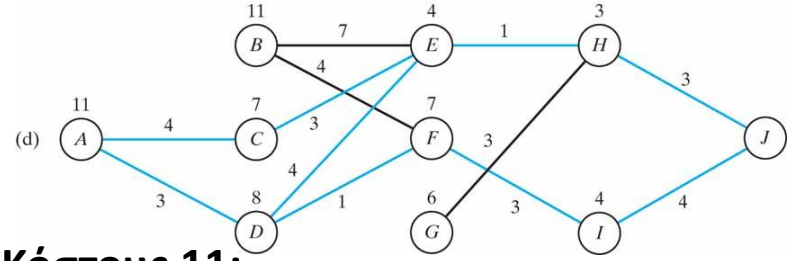
A (Αριστερά) ⟶ ***Δ*** (Δεξιά)











Βέλτιστοι Δρόμοι Κόστους 11:

$\{A, C, E, H, J\}, \{A, D, E, H, J\}, \{A, D, F, I, J\}$

Αλγόριθμοι Δυναμικού Προγραμματισμού ***Bellman-Ford*** στηρίζουν την δρομολόγηση ***Border Gateway Protocols*** (***BGP***) ανάμεσα στα ~66,000 Αυτόνομα Συστήματα (***Autonomous Systems, AS***) στο ***Internet*** (~830,000 γνωστά δίκτυα)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Απευθείας Προσεγγιστικές Μέθοδοι Δυναμικού Προγραμματισμού (1/2) (*Επανάληψη*)

Οι δύο αλγόριθμοι Δυναμικού Προγραμματισμού (***Value Iteration*** & ***Policy Iteration***) προαπαιτούν γνώση των **πιθανοτήτων μεταβάσεων** $p_{ij}(a)$ και του ***άμεσα αναμενόμενου κόστους κατάστασης*** $c(i,a) = \sum_{j=1}^{N} p_{ij}(a) g(i,a,j)$ **εκτιμώμενου** με βάση τα **γνωστά** $g(i,\mu(i),j) = g(i,a,j)$ (***observed κόστη μετάβασης*** $i \to j$ με απόφαση a)

Οι απευθείας προσεγγιστικές μέθοδοι (***Direct Approximate Dynamic Programming Methods***) εκτιμούν τις πιθανότητες μετάβασης και τα αναμενόμενα κόστη μεταβάσεων - αποφάσεων μακροπρόθεσμων πολιτικών με προσομοιώσεις ***Monte Carlo***. Ενσωματώνονται στους δύο βασικούς αλγόριθμους Δυναμικού Προγραμματισμού με τις εξής παραλλαγές:

- **Value Iteration** → **Temporal-Difference *TD*(0) Learning**
- **Policy Iteration** → ***Q*-Learning**

Γενική Μεθοδολογία - Απαιτήσεις

- Οι προσομοιώσεις ***Monte Carlo*** δημιουργούν σενάρια πολλαπλών πιθανών τροχιών (***system trajectories***) της εξέλιξης του ***Markov Decision Process*** σε κάθε **επεισόδιο** (***episode***) από μια αρχική κατάσταση i_0 μέχρι κάποια τελική $i_T = TERMINAL$ (T είναι το βήμα n που η κατάσταση $i_n \to i_T$ και τερματίζεται το **επεισόδιο**). Η διαδικασία μάθησης συνήθως περιλαμβάνει πολλά ανεξάρτητα **επεισόδια** με διαφορετικές ***trajectories***
- Οι τιμές συναρτήσεων ***cost-to-go*** $J(i)$ ανανεώνονται σε κάθε προσομοίωση με προσθήκη του (**γνωστού**) ***άμεσου*** (***observed***) ***κόστους μετάβασης*** $g(i,j)$ σε επισκέψεις προσομοιωμένης τροχιάς μεταβάσεων από κατάσταση i προς κατάσταση j
- Οι μέθοδοι ***Monte Carlo*** απαιτούν γνώση της δομής του περιβάλλοντος από εμπειρία (όχι από πρότερη γνώση πιθανοτήτων), διαχειρίσιμο αριθμό παρατηρήσιμων (***observable***) καταστάσεων και σημαντικό αριθμό από ***trajectories*** για καλές εκτιμήσεις

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Απευθείας Προσεγγιστικές Μέθοδοι Δυναμικού Προγραμματισμού (2/2) (*Επανάληψη*)

Σύνοψη Εννοιών Δυναμικού Προγραμματισμού

Ορισμός ***Άμεσου*** (***Observed***) ***Κόστους***: $g(i,a,j)$ για μετάβαση $i \to j$ με απόφαση a

Ορισμός ***Άμεσου Αναμενόμενου Κόστους***: $c(i,a) \triangleq \sum_{j=1}^{N} p_{ij} g(i,a,j)$ για $\forall i$ και $a \in \mathcal{A}_i$

Ορισμός ***Cost-to-Go*** : $J^{\mu}(i) = c(i,\mu(i)) + \gamma \sum_{j=1}^{N} p_{ij}(\mu(i)) J^{\mu}(j)$ για $\forall i$ και πολιτική $\mu(i)$

Βέλτιστα ***Cost-to-Go*** (***Bellman***): $J^{*}(i) = \min_{a \in \mathcal{A}_i} \left(c(i,a) + \gamma \sum_{j=1}^{N} p_{ij} J^{*}(j) \right), i = 1,2,\dots,N$

Ορισμός ***Q-Factors***: $Q^{\mu}(i,a) \triangleq c(i,a) + \gamma \sum_{j=1}^{N} p_{ij}(a) J^{\mu}(j)$ για $\forall i$ και $a \in \mathcal{A}_i$

Ορισμός ***Βέλτιστων Q-Factors***: $Q^{*}(i,a) = \sum_{j=1}^{N} p_{ij}(a) \left(g(i,a,j) + \gamma \min_{b \in \mathcal{A}_j} Q^{*}(j,b) \right)$

Ορισμοί *on-policy*, *off-policy*

- Η ***on-policy*** σε κάθε βήμα εκτιμά με προσομοιώσεις ***Monte Carlo*** το κόστος $J^{\mu}(i)$ των καταστάσεων i μιας τροχιάς (***trajectory***) όταν ακολουθείται η υπό αξιολόγηση **συνολική** πολιτική μ. Με επαναλήψεις που περιλαμβάνουν διορθωτικές αποφάσεις $i \to \alpha$ οδηγούνται τα $J^{\mu}(i)$ σε διαδοχικές μειώσεις (**Value Iteration → *TD*(0)-Learning**)
- Η ***off-policy*** συγκρίνει εναλλακτικές αποφάσεις σε καταστάσεις του περιβάλλοντος i μιας τροχιάς (***trajectory***) και σε κάθε βήμα **επιλέγει** με απληστία αποφάσεις a με το ελάχιστο $Q(i,a)$ στην παρούσα κατάσταση i. Τα ***Cost-to-Go*** $J^{\mu}(j)$ μιας προσωρινής πολιτικής μ εκτιμώνται από προσομοιώσεις ***Monte Carlo*** των ***trajectories*** χωρίς να συμπεριλαμβάνουν βελτιώσεις $i \to \alpha$ που ίσως προκύψουν από τα ***Q-Factors*** (**Policy Iteration → *Q*-Learning**)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Παράδειγμα Δυναμικού Προγραμματισμού: Δρομολόγηση BGP στο Internet - RFC 4271 (1/7)

Υλικό από Παρουσιάσεις Μαθήματος Διαχείριση Δικτύων – Ευφυή Δίκτυα ΣΗΜΜΥ Ε.Μ.Π.

https://webvm.netmode.ntua.gr/courses/wp-content/uploads/2019/10/NetMan_2019_10_14.pdf

Το παγκόσμιο ***Internet*** αποτελείται (*6/2019*) από $\sim \mathbf{830,000}$ γνωστά δίκτυα τελικούς προορισμούς (π.χ. Δίκτυο ΕΜΠ, IP: 147.102.0.0/16), οργανωμένα σε $\sim \mathbf{66,000}$ Αυτόνομα Συστήματα (***Autonomous Systems, AS***) με διαχειριστική αυτονομία (π.χ. GRNET/ΕΔΕΤ, Autonomous System Number - ***ASN 5408***)

Η δρομολόγηση εντός Αυτόνομης Κοινότητας γίνεται με βάση κεντρικά ρυθμιζόμενα πρωτόκολλα (***Interior Gateway Protocols*** – ***IGP***, π.χ. OSPF) ενώ μεταξύ των $\mathbf{66,000}$ AS's μέσω γενικών πινάκων δρομολόγησης σε συνοριακούς δρομολογητές (***Border Gateways***, ***Border Routers***) με καταχωρήσεις για όλα τα $\sim \mathbf{830,000}$ γνωστά δίκτυα του ***Internet***

Η δημιουργία – ανανέωση των γενικών πινάκων δρομολόγησης (***σε ηλεκτρονική μνήμη των Border Gateways***) ***γίνεται με το Border Gateway Protocol – BGP*** **(RFC 4271)**

- Οι ***Border Routers*** (***Gateways***) των ***AS*** ανακοινώνουν (μέσω ***BGP signaling***) στα $\mathbf{66,000}$ ***AS***'s του ***Internet***, τα $\mathbf{830,000}$ δίκτυα – τελικούς προορισμούς τα οποία είτε ανήκουν σε αυτά ή είναι προσπελάσιμα (***reachable***) διαμέσου αυτών, με εκτιμήσεις κόστους (βάρους) βέλτιστων ***inter-AS*** δρόμων προς κάθε δίκτυο - προορισμό
- Οι ***Border Gateways*** υπολογίζουν αυτόνομα βέλτιστες διαδρομές προς όλους τους τελικούς προορισμούς με βάση τις προτιμήσεις (πολιτικές) των διαχειριστών τους, όποτε κρίνουν πως αλλαγές τοπολογίας ή πολιτικής ή επίδοσης επιβάλλουν ανανέωση δρόμων
- Ο κατανεμημένος προσδιορισμός βέλτιστης δρομολόγησης ορίζει κόστη προς τους $\mathbf{830,000}$ τελικούς προορισμούς βάση πληροφοριών ***reachability*** και μετρήσεων κόστους διασύνδεσης προς τα γειτονικά ***AS***. Βασίζεται στον Αλγόριθμο ***Bellman – Ford***

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Παράδειγμα Δυναμικού Προγραμματισμού: Δρομολόγηση BGP στο Internet - RFC 4271 (2/7)

Αλγόριθμος Distance Vector (Bellman – Ford)
BGP (Bellman – Ford)

- Οι συνοριακοί δρομολογητές (***Border Gateways***) κάθε Αυτόνομης Περιοχής (***AS***) εντοπίζουν τους βέλτιστους δρόμους (***shortest paths***) ενδιάμεσων και τελικού ***AS*** προς όλα τα γνωστά δίκτυα προορισμούς εκτελώντας αλγόριθμο βασισμένο στον δυναμικό προγραμματισμό (***dynamic programming***) που εισήγαγε ο ***Bellman***

- Χρειάζεται γνώση διανυσμάτων κόστους (βαρών) των άμεσων συνδέσεων (***Inter AS Interfaces***) και εκτιμήσεις κόστους (αποστάσεις, ***distance vectors***) προς όλα τα γνωστά δίκτυα προορισμούς στο ***Internet*** (**830,000+**, *6/2019*)

- Η βελτιστοποίηση βασίζεται σε κατανεμημένο αλγόριθμο ***Bellman - Ford*** που υλοποιείται μέσω σηματοδοσίας ανακοινώσεων (***BGP Announcements***) μεταξύ όλων των (**66,000+**, *6/2019*) Αυτόνομων Περιοχών (***AS***) του ***Internet*** με πληροφορίες δρομολόγησης και εκτιμήσεις κόστους

- Από τη σκοπιά του ***Reinforcement Learning*** το ***BGP*** μπορεί να θεωρηθεί κατανεμημένη επέκταση του Δυναμικού Προγραμματισμού με συνεργασία (***cooperative game***) **66,000** αυτόνομων ***Agents*** . Η συνεργατική βελτιστοποίηση κωδικοποιήθηκε σαν ***Multi-Agent Reinforcement Learning – MARL*** από τον ***Michael Littman*** το 1994 https://www2.cs.duke.edu/courses/spring07/cps296.3/littman94markov.pdf

Το BGP αποτελεί κύριο παράγοντα επιτυχίας της παγκόσμιας επανάστασης του *Internet*

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Παράδειγμα Δυναμικού Προγραμματισμού: Δρομολόγηση BGP στο Internet - RFC 4271 (3/7)

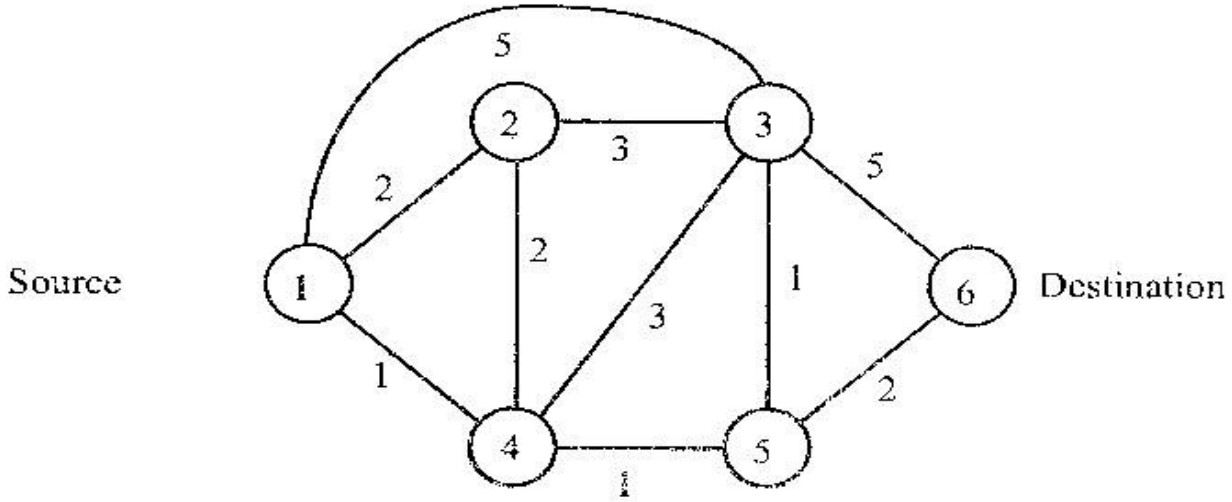
Δίκτυο (Γράφος) Αναφοράς Παραδείγματος, $N = 6$ Κόμβων

Οι κόμβοι του γράφου παριστούν τα διάφορα ***AS*** του ***Internet***

Τα δίκτυα πηγής και προορισμού των χρηστών είναι ενσωματωμένα στους κόμβους (***AS***) ***Source*** – ***Destination*** του γράφου

Τα κόστη των γραμμών του γράφου αφορούν και στις 2 κατευθύνσεις και εκτιμώνται από τους άμεσα συνδεόμενους κόμβους (***Border Gateways***) με βάση προτιμήσεις των διαχειριστών

- Στο παράδειγμα που ακολουθεί υπολογίζονται δένδρα ελαχίστων δρόμων (***shortest path trees***) από όλους τους κόμβους (***AS***) προς την ***ρίζα*** $\{6\}$
- Η επιλογή του ρόλου της ρίζας του δένδρου (πηγή ή προορισμός) έγινε αυθαίρετα. Οι αλγόριθμοι ισχύουν κατ' αναλογία για αντίστροφους ρόλους ρίζας



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Παράδειγμα Δυναμικού Προγραμματισμού: Δρομολόγηση BGP στο Internet - RFC 4271 (4/7)

Υπολογισμός Δένδρου Ελάχιστων Δρόμων (Shortest Path Tree) προς $\{6\}$ από $\{1, 2, 3, 4, 5\}$

Εφαρμογή Αλγορίθμου *Q*-Learning (*Off-policy*) με *Asynchronous Updates*

$\{i\}$ Κατάσταση (***State***) του γράφου, κόμβος (***AS***) $i = 1,2, \dots, N$ (στο παράδειγμα $N = 6$, μέχρι 66,000 στο ***Internet***)

$P^{(n)}(i)$ Απόφαση (***Action***): Επόμενος κόμβος (***AS***) από τον $\{i\}$ προς τον $\{6\}$, ενδιάμεσος ή τελικός στην επανάληψη (***Iteration***) n

d_{ij} Κόστος (βάρος) γραμμής (i, j) στην επανάληψη n (***Transtion Cost***) ρυθμιζόμενο από την πολιτική δρομολόγησης του $\{i\}$ ή/και απευθείας μετρήσεις των αμέσων γειτόνων $\{i, j\}$. Αν $d_{ij} = c, \forall (i, j) \Rightarrow$ ***min hop routing***

$L^{(n)}(i)$ ***Labels***, ***Q-Factors*** $L^{(n)}(i) \triangleq Q(i, P^{(n)}(i))$: Εκτιμήσεις ελάχιστου κόστους (από τον $\{i\}$ προς τον $\{6\}$ στην επανάληψη n (ανανεώνονται ***ασύγχρονα***, σύμφωνα με τις ***πιο πρόσφατες εκτιμήσεις*** ανάλογα με την σειρά εκτέλεσης των ανανεώσεων – ***updates***)

Περιγραφή Αλγορίθμου Bellman – Ford

- Αρχικά έχουμε $L_i^{(0)} = \infty \ \ \forall i \neq 6, \ L_6^{(n)} = 0 \ \ \forall n,$
- Σε κάθε διαδοχική επανάληψη (***iteration***) $n = 1,2, \dots$ και $\forall i$ ανανεώνουμε ***ασύγχρονα*** τις εκτιμήσεις ελαχίστου κόστους από την παρούσα κατάσταση προς τον προορισμό με βάση τις σχέσεις του Δυναμικού Προγραμματισμού σύμφωνα με τις πιο πρόσφατες εκτιμήσεις (***updates***) των $L_j^{(n)}$ για όλους τους άμεσους γείτονες j του i:

$$L_i^{(n+1)} = \min_j \left\{ L_j^{(n)} + d_{ij} \right\} \forall i \neq 6$$

- Αν $L_i^{(n+1)} = L_i^{(n)} \ \forall i$ σταματάμε τον αλγόριθμο και προσδιορίζουμε τους βέλτιστους δρόμους από όλα τα $\{i\}$ προς τον προορισμό $\{6\}$ σύμφωνα με τις αποφάσεις $P^{(n)}(i)$ σαν ***Shortest Path Tree*** με ρίζα τον $\{6\}$
- Πολυπλοκότητα αλγορίθμου: $O(N^3)$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ

Παράδειγμα Δυναμικού Προγραμματισμού: Δρομολόγηση BGP στο Internet - RFC 4271 (5/7)

Εκτέλεση Αλγορίθμου για Προορισμό {6}

Παράδειγμα: INITIAL LABELS: L(1)=L(2)=...=L(5)= ∞, L(6)=0

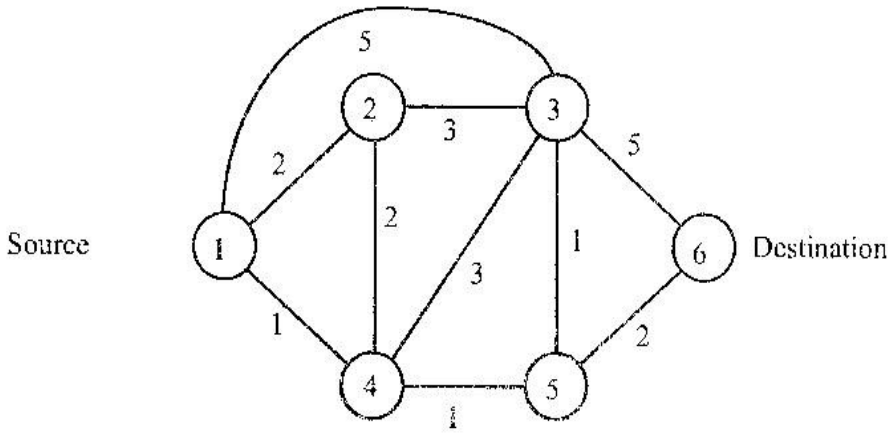
UPDATE ORDER 5,4,3,2,1

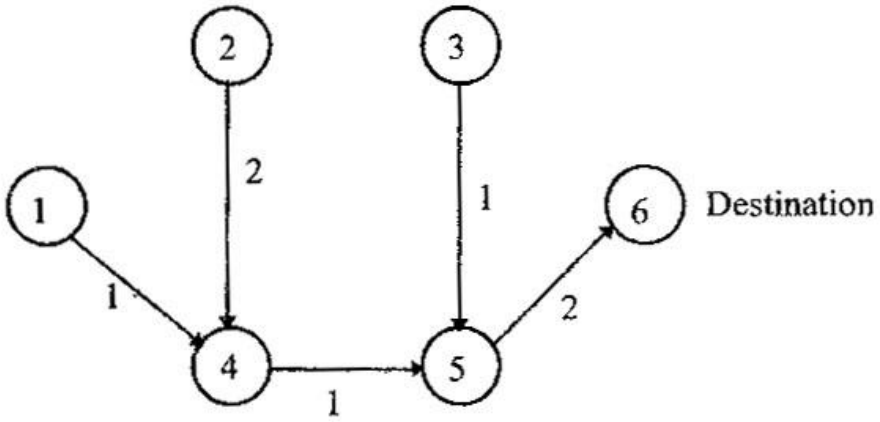
Iteration Number	Labels L(n), Current Predecessor Node P(n)									
	L(5), P(5)		L(4), P(4)		L(3), P(3)		L(2), P(2)		L(1), P(1)	
1	2	6	3	5	3	5	5	4	4	4
2	2	6	3	5	3	5	5	4	4	4

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	**8**	**3**	2	6
2	**9**	**4**	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

Η ταχύτητα σύγκλησης εξαρτάται από την σειρά ανανέωσης των Labels των κόμβων



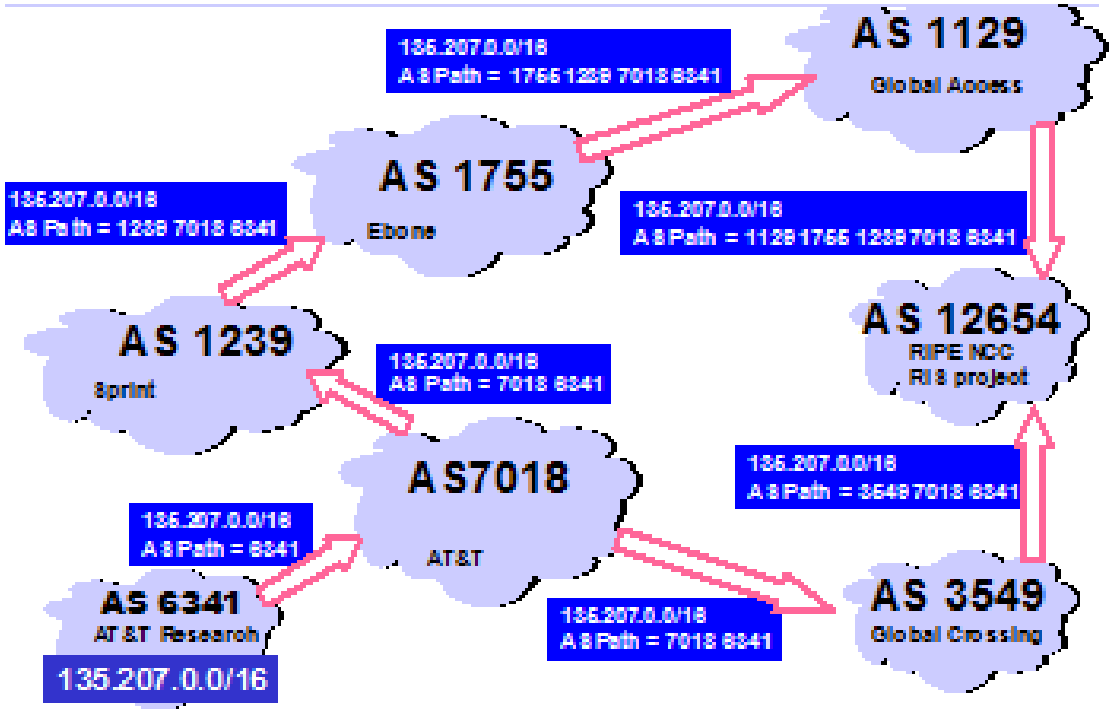


ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ

Παράδειγμα Δυναμικού Προγραμματισμού: Δρομολόγηση BGP στο Internet - RFC 4271 (6/7)

Παράδειγμα Μάθησης - Ανακοίνωσης Δικτύου 135.207.0.0/16

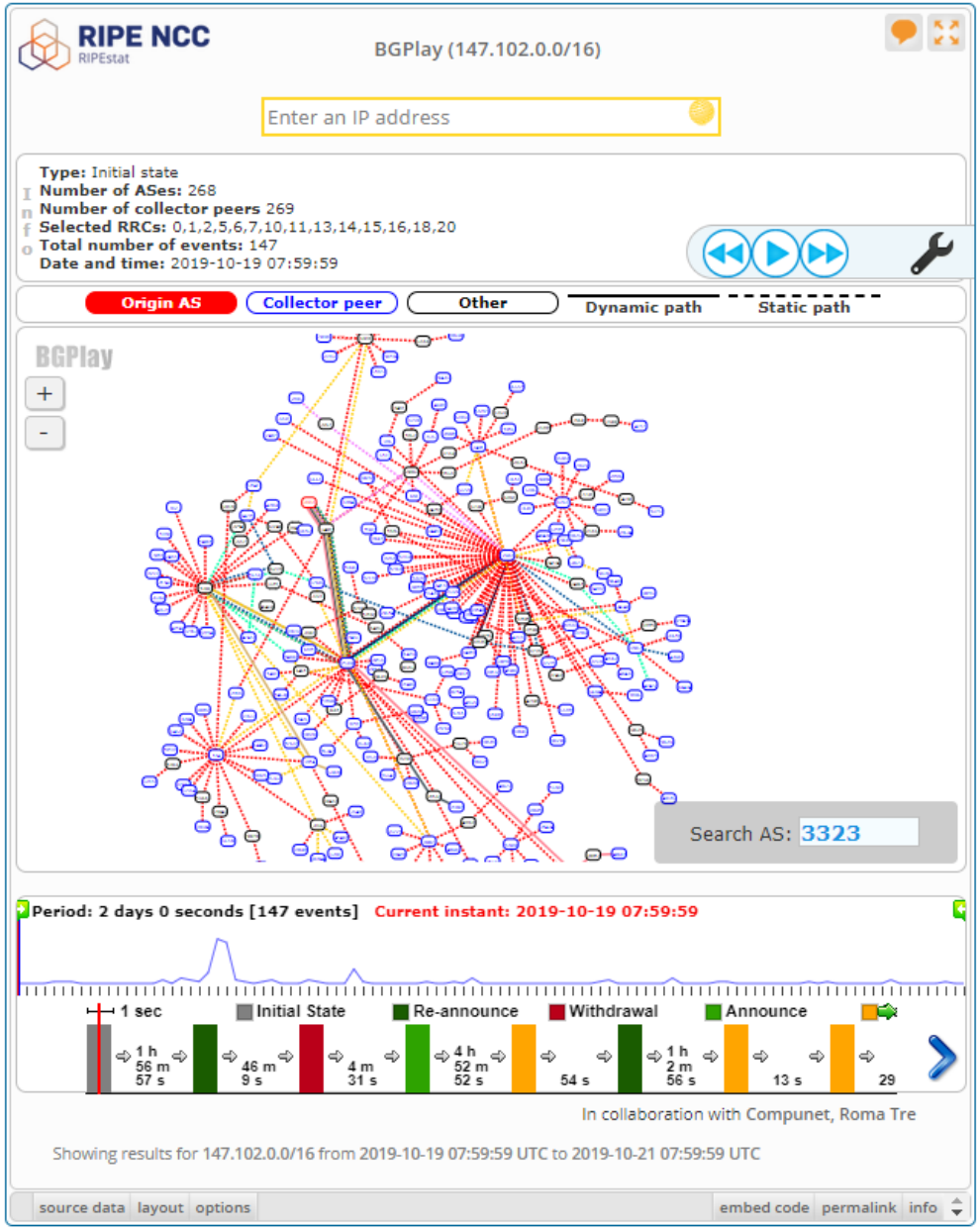
(από παρουσίαση του Timothy G. Griffin, AT&T Research, Paris 2002)



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ

Παράδειγμα Δυναμικού Προγραμματισμού: Δρομολόγηση BGP στο Internet - RFC 4271 (7/7)

Πραγματική Εικόνα των Δρόμων BGP (10-10-2019)



ΠΑΡΟΧΗ INTERNET ΣΤΟ NTUA (ASN: 3323)
GRNET (ASN: 5408), GÉANT (ASN: 20965),
COGENT-174 (ASN: 174), TELIANET (ASN: 1299)
https://stat.ripe.net/special/bgplay

GÉANT Tier 1/2 Providers (Internet feeds)

- **COGENT-174 (174)**
- **TELIANET (1299)**