Bringing Order To BGP: Decreasing Time and Message Complexity

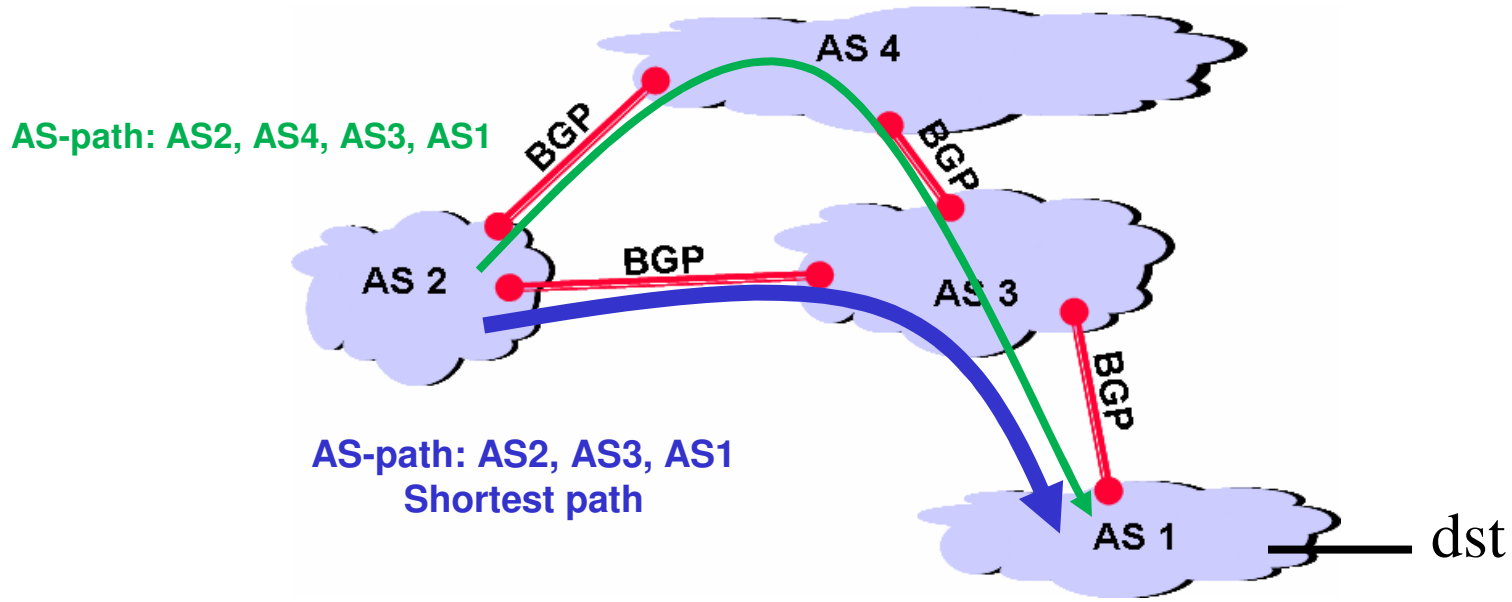
Nir Chen
Interdisciplinary Center Herzliya (IDC)

Joint work with

Anat Bremler-Barr	Jussi Kangasharju	Osnat Mokryn	Yuval Shavitt
IDC	Darmstadt University	IDC	Tel-Aviv University

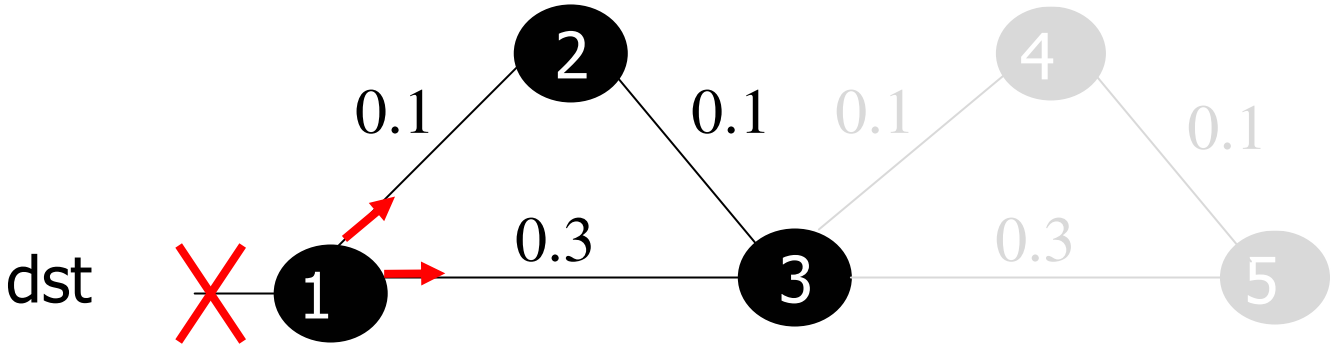
BGP protocol

- BGP – the routing protocol of the internet
- Distance (Path) vector protocol between ASes (AS~ISP)
- Router receives AS-path from the neighbors for a destination
- Chooses the best AS-path (**shortest, policy**)
  - Assume for now BGP is shortest path



Path Exploration Example

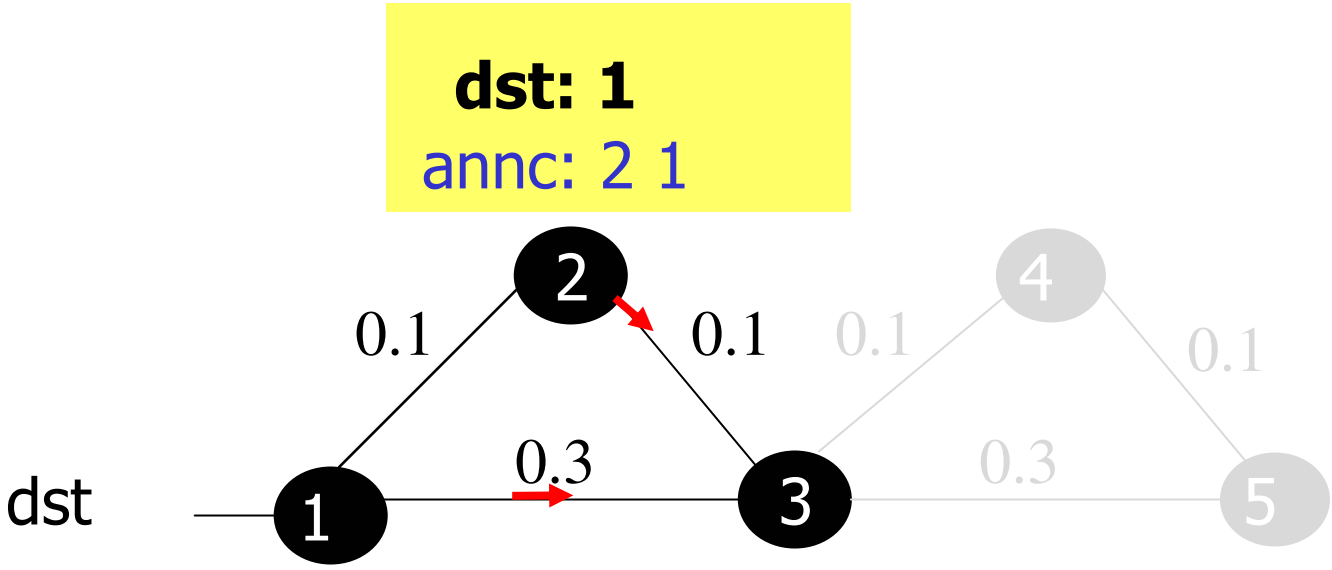
Path exploration: **One event** → **The router sends messages more than once**



t=0

Path Exploration Example

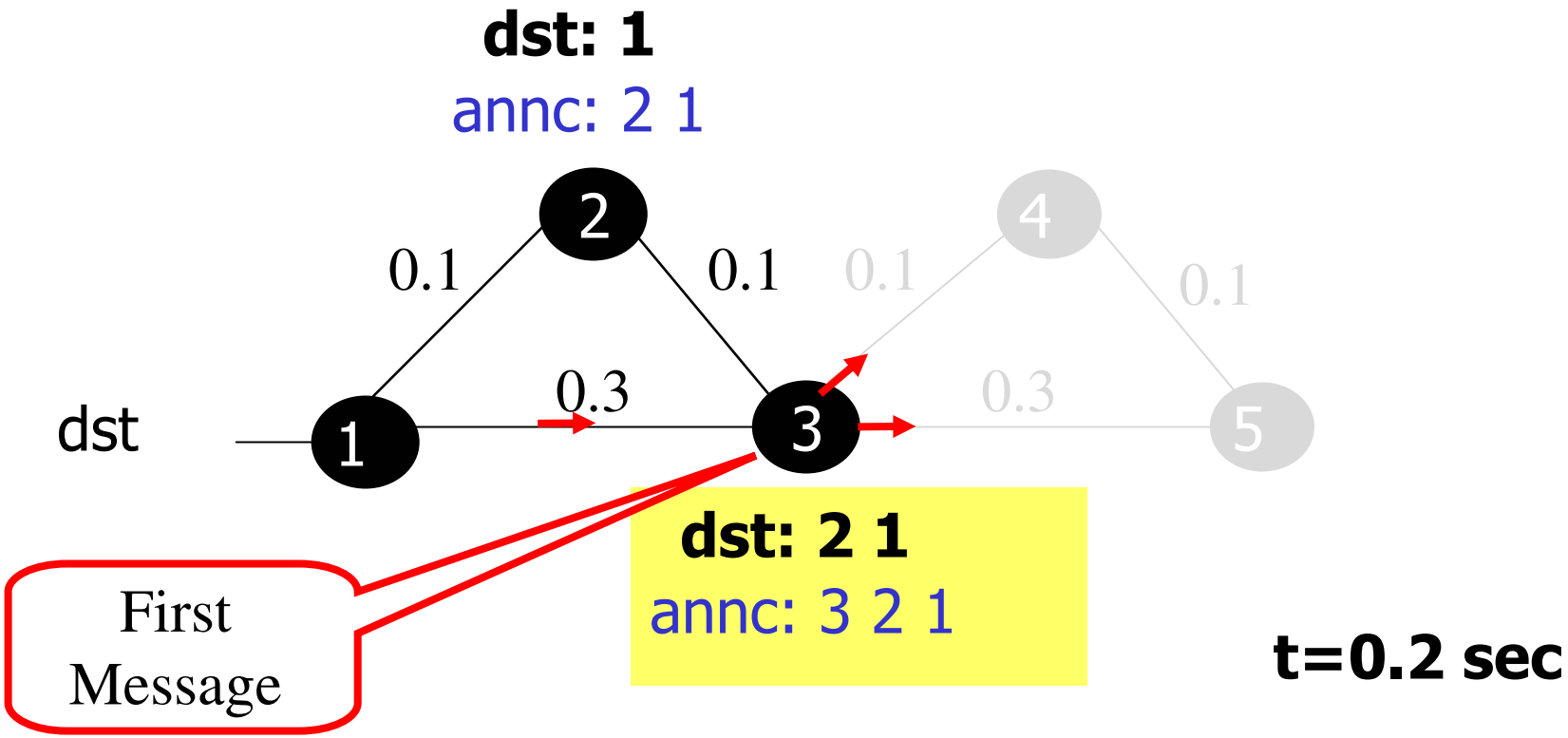
Path exploration: **One event** → **The router sends messages more than once**



t=0.1 sec

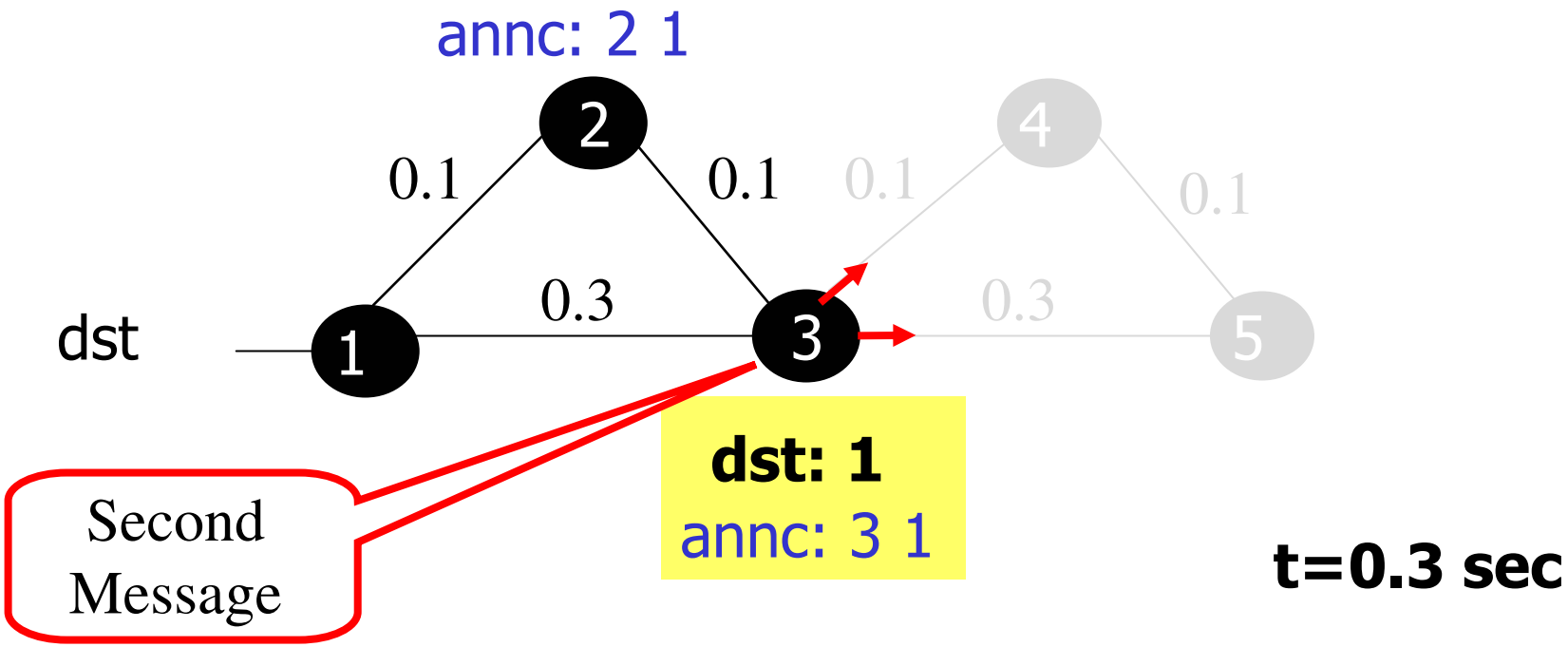
Path Exploration Example

Path exploration: **One event** → **The router sends messages more than once**



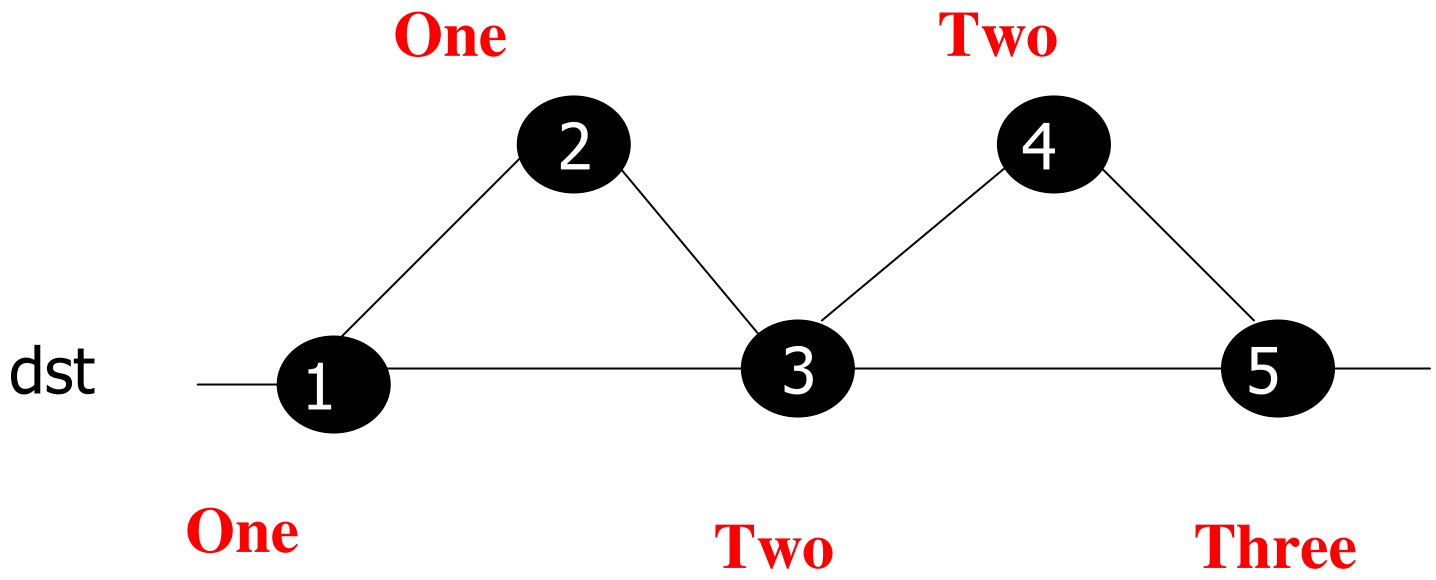
Path Exploration Example

Path exploration: **One event** → **The router sends messages more than once**

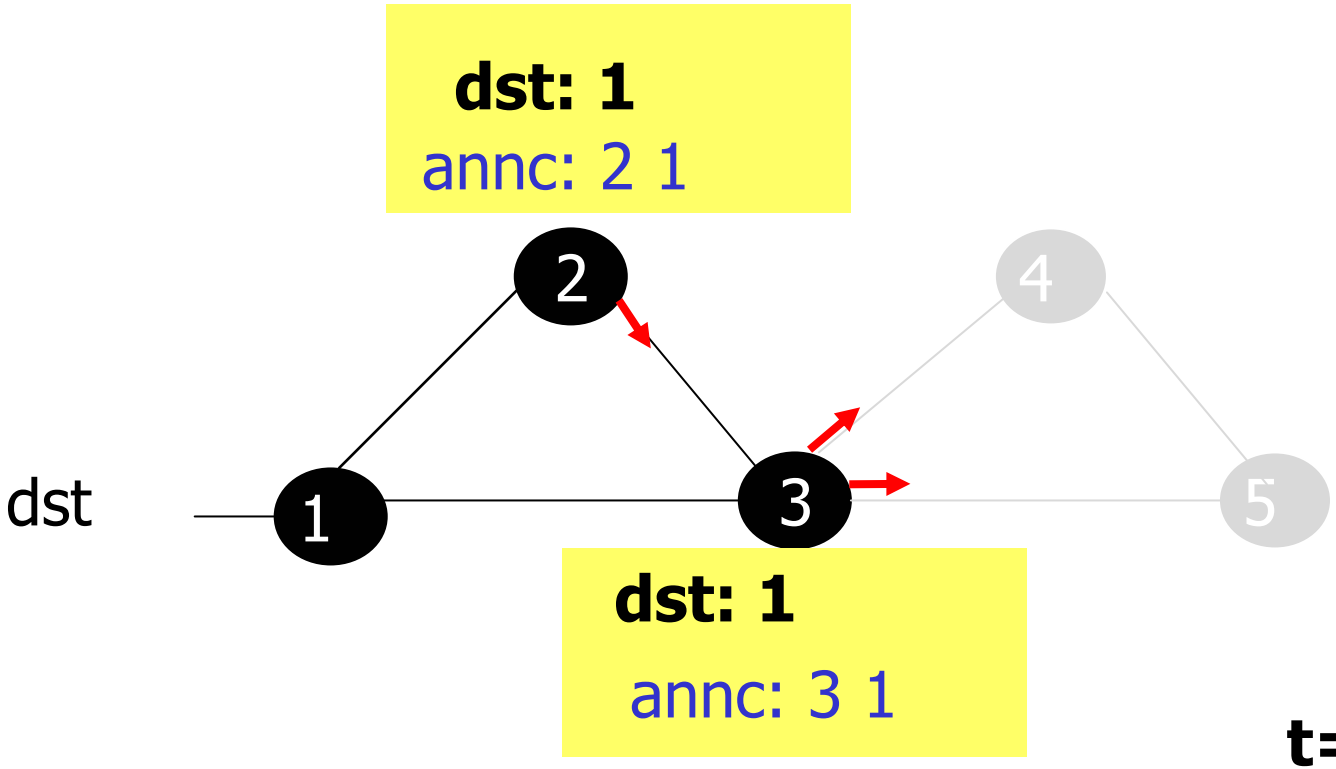


Path Exploration Example

	Time	Message
BGP without MinRouteAdver	Dh h~link delay D- Internet Diameter	DE (shortest)



Synchronous Network – no Path Exploration



t=1 round

	Time	Message
Synchronous BGP	hD	E

Our Goal

- Reduce message (and time) complexity in case of reattachment of a network in BGP – by eliminating the *path exploration* phenomenon
  - 25% of the sent messages can be eliminated by our modification (simulation and real-life data)
- Understanding of distributed algorithm can improve BGP protocol

BGP: MinRouteAdver Timer

- MinRouteAdver rule:

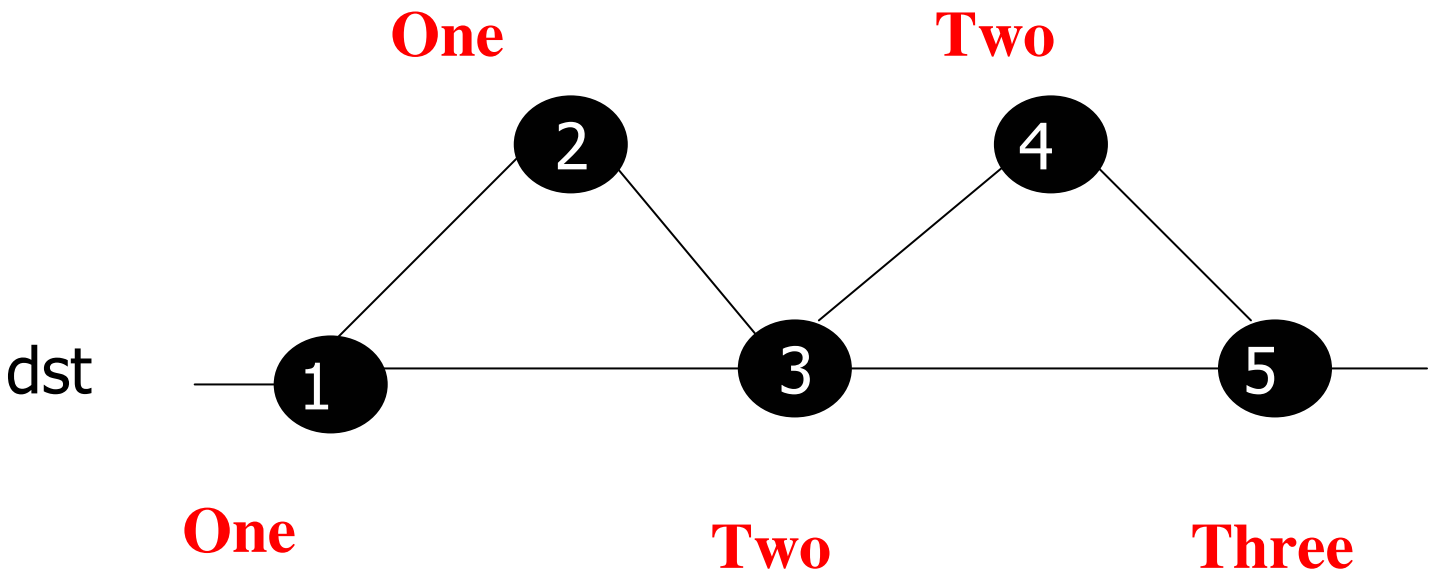
  A router announces the preferred new ASpath for a route

  iff at least 30 sec has passed from the last announcement

- Motivation to reduce messages

BGP: MinRouteAdver Effect

Still suffers from the same ***Path Exploration*** phenomenon



	Time	Message
BGP with MinRouteAdver	30D	1E

Path Exploration: Negative Effects

- Load on Router
- Increase convergence time
- False trigger of route flap dumping mechanism (punishment of oscillated destination)
- **There is a simple alternative …**

Solution #1: Pseudo Ordering rule

- **Change MinRouteAdver rule:**

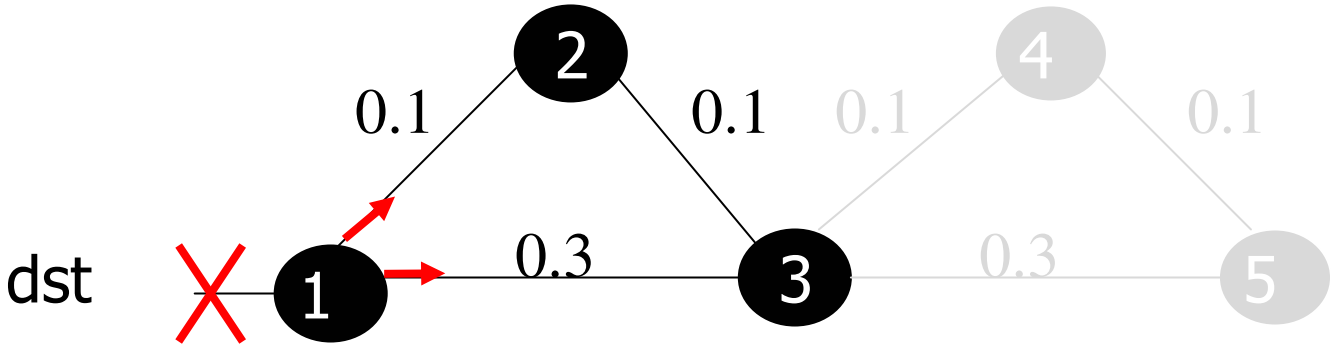
  A router announces the preferred new ASpath for a route

  iff it received the announcement about the new ASpath at least Dh time ago

- **In today internet:**
  - **h - maximum link delay <1**
  - **D – Diameter <15**

Method	Delay	From
BGP MinRouteAdver Rule	**30**	the last time a router sent a message
BGP with Pseudo ordering rule	**Dh~15**	the time the router received the message about the new ASpath

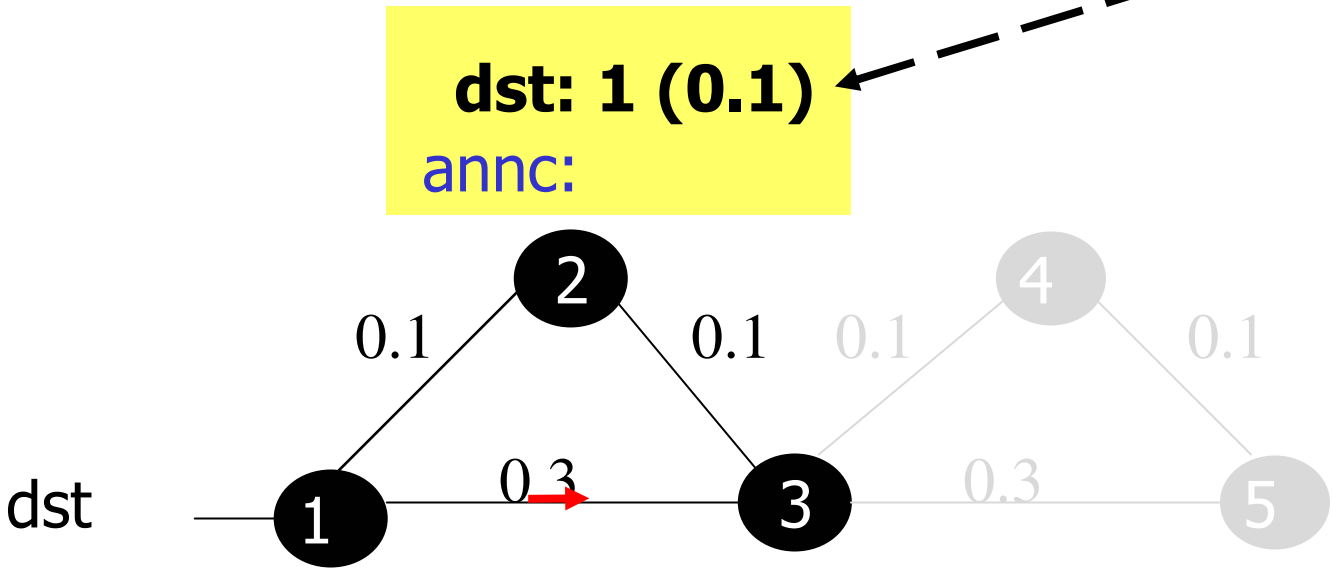
Pseudo Ordering Example



t=0

Pseudo Ordering Example

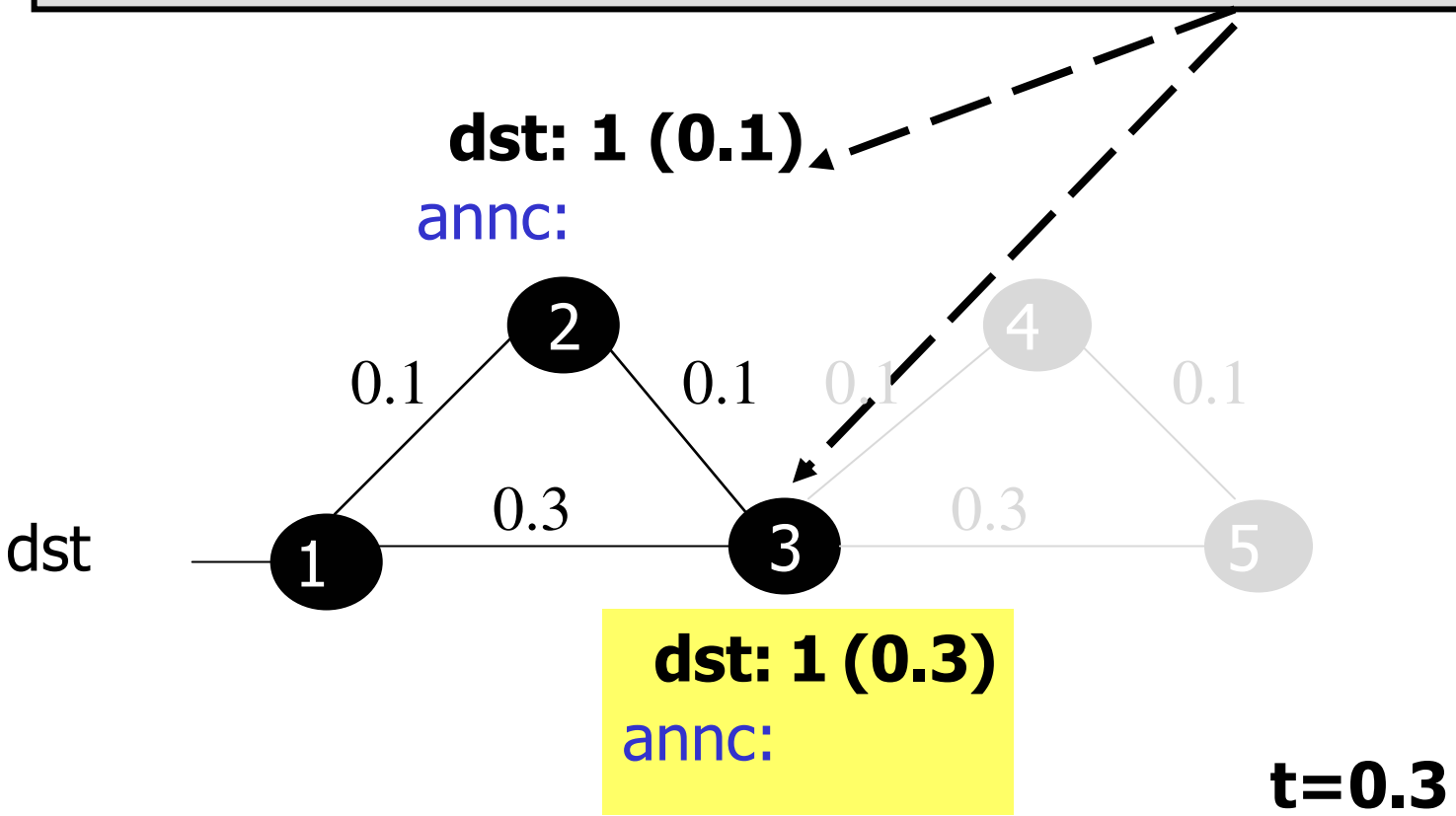
Pseudo Ordering: Wait Dh=15 sec's before sending the announcement



t=0.1

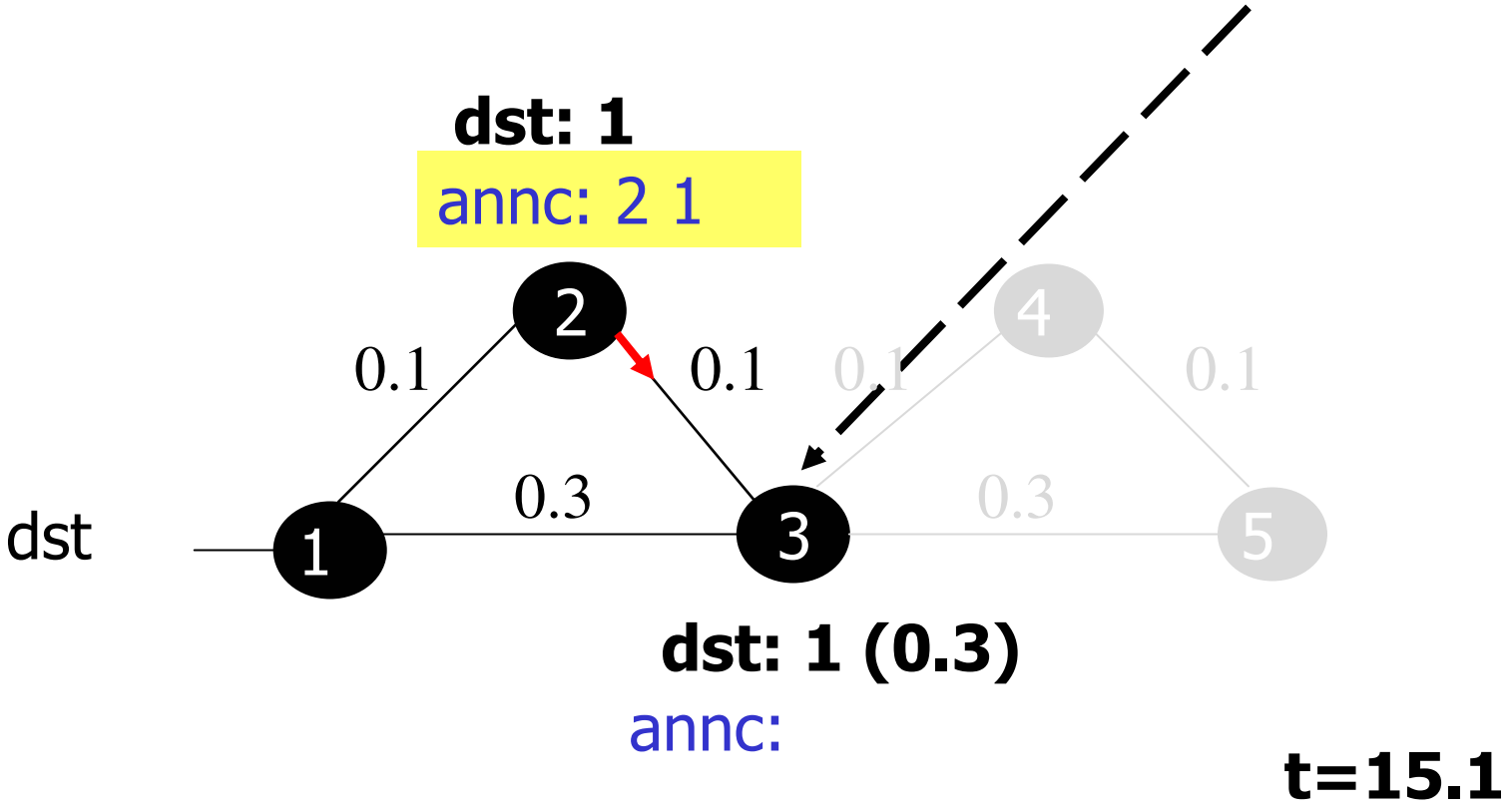
Pseudo Ordering Example

Pseudo Ordering: Wait Dh=15 sec's before sending the announcement

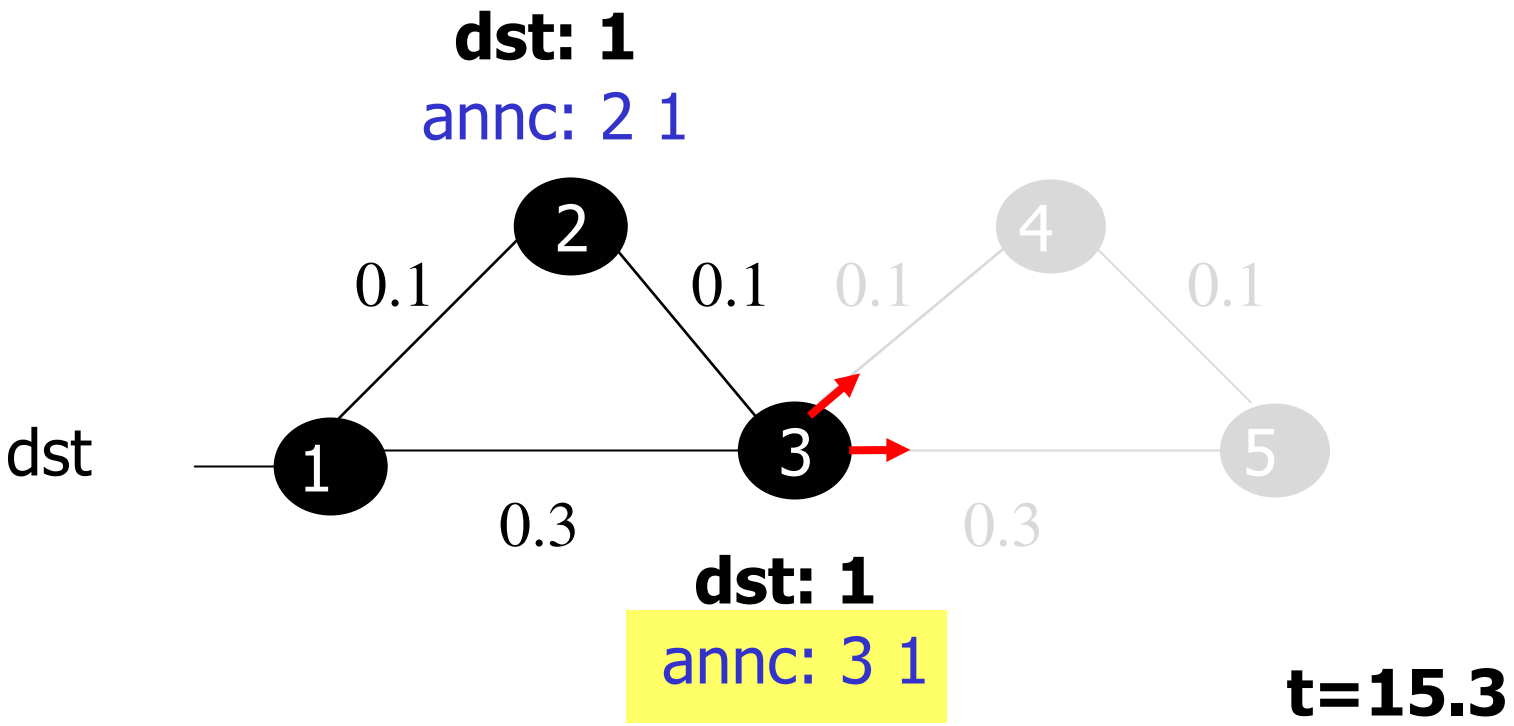


Pseudo Ordering Example

Pseudo Ordering: Wait Dh=15 sec's before sending the announcement

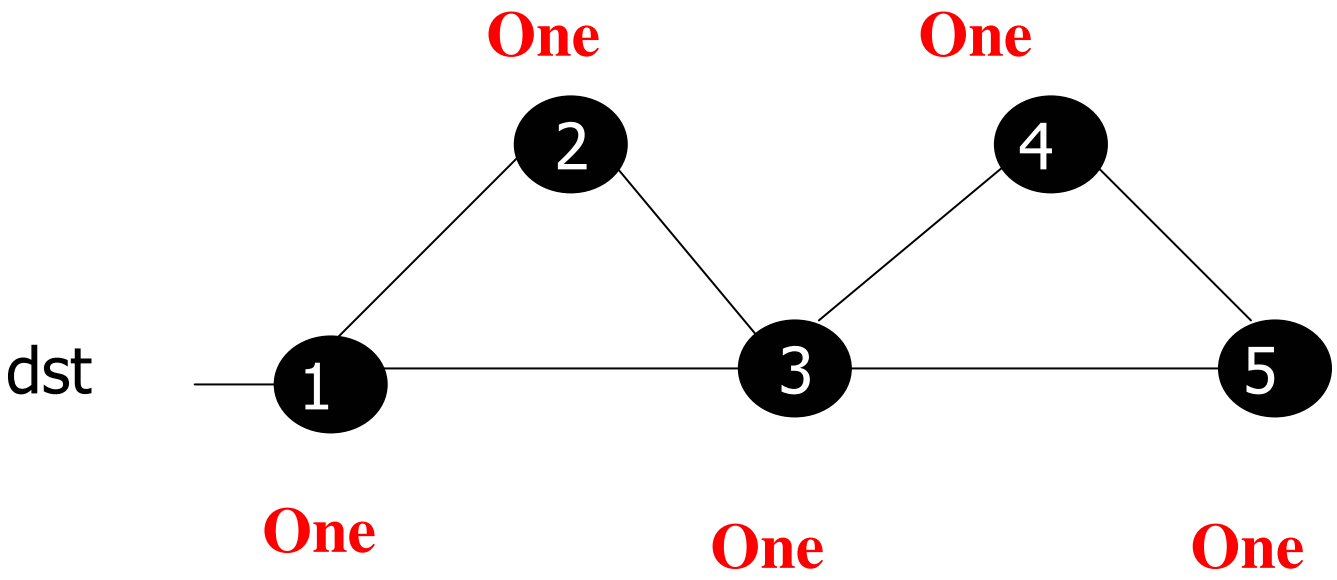


Pseudo Ordering Example



Pseudo Ordering Example

- The number of times each router sends messages to its neighbours



	Time	Message
Pseudo ordering	(Dh + h)D Worst case of BGP is 30D > Pseudo Ordering Average case of BGP is 15D ~ Pseudo Ordering	E

Solution #2: Adaptive Pseudo Ordering rule

- **Change MinRouteAdver rule:**

  A router announces the preferred new ASpath for a route of length L

  iff it received the announcement about the new ASpath at least Lh time ago

Method	Delay	From
BGP MinRouteAdver Rule	**30**	from the last time a router sent a message
BGP with Pseudo ordering rule	**Dh**	the time the router received the message about the new ASpath
BGP with Adaptive Pseudo ordering rule	**Lh**	the time the router received the message about the new ASpath with ASpath of length L

Adaptive Pseudo Ordering rule (cont')

- **Outcome :** We reduce the convergence time almost by half

	Time	Message
Adaptive Pseudo Ordering	$\sum_{l=1}^{D} (lh + h)$	E

BGP shortest path vs. policy

- BGP policy – captures economic relationship between ASes

- **In the full paper we prove using the common policy rules :** BGP chooses the shortest path from the announced routes.
  - I.e.

    BGP chooses the local shortest path -

    BGP does not choose the global shortest path

    ✓**Our algorithm works under the local shortest path trait**

Simulation and Empirical Investigation

- In order to evaluate the **average case** we use simulation with SSFNET
- The pseudo-ordering algorithm reduces the number of messages by 23% (peer) and the **convergence time by 80%** (peer)

- Data: Ripe – feed of BGP announcements
- We analyze reattachment events during 14 days of June 2005
- Estimation: **the pseudo-ordering algorithm reduces the number of messages by 25%**

Related Work

- Main focus of BGP convergence problem – the **fail-down** event ~ Counting to Infinity like problem
  - [Labovitz,Ahuja,Bose,Jahanian],
    [Labovitz,Wattenhofer,Venkatachary,Ahuja]
    [Pei,Zhao,Wang,Massey,Mankin,Wu,Zhang]
    [Bremler-Barr,Afek,Schwartz]
    [Jaideep, Chandrashekar Zhenhai] etc …

- Path Exploration in BGP
  [zhang,Arora,Lie][Labovitz,Wattenhofer,Venkatachary,Ahuja, Jahanian,Bose] – **no solution**

Questions ?

chen.nir@idc.ac.il