BGP Route Flap Damping Algorithms

Steve Shen and Ljiljana Trajković
{wshen, ljilja}@cs.sfu.ca

Communication Networks Laboratory
http://www.ensc.sfu.ca/cnl
Simon Fraser University

Roadmap

- Introduction to Route Flap Damping (RFD)
- ns-2 implementations of RFD
- Simulation scenarios
- Performance analysis
- Improvements to RFD algorithms
- Conclusions and references

BGP

- Border Gateway Protocol:
  - inter-AS (Autonomous System) routing protocol
  - used to exchange network reachability information among BGP systems
  - BGP-4 is the current de facto inter-domain routing protocol
- Path vector protocol:
  - distributes route path information to peers
- Incremental:
  - sends updates as routing tables change

Introduction to RFD

- A route flaps when the route oscillates between being available and unavailable
- Routing oscillations can be caused by:
  - router configuration errors
  - transient data link failures
  - software defects
- BGP employs RFD mechanism to prevent persistent routing oscillations:
  - reduce the number of BGP update messages sent within the network
  - decrease the processing load imposed on BGP speakers

Common approaches to RFD

- Assign a penalty to a route and increment the penalty value when the route flaps
- The route is suppressed and not advertised further when the penalty exceeds the suppress limit
- Penalty of a route decays exponentially based on half life
- If the penalty decreases below the reuse limit, the route is reused and may be advertised again

RFD algorithms

- Existing RFD algorithms that identify and penalize route flaps:
  - Original RFD
  - Selective RFD
  - RFD+

Original RFD:

C. Villamizar, R. Chandra, and R. Govindan, "BGP route flap damping," *IETF RFC 2439*, Nov. 1998.

Selective RFD:
Z. Mao, R. Govindan, G. Varghese, and R. Katz, "Route flap damping exacerbates Internet routing convergence," in *Proc. SIGCOMM 2002*, Pittsburgh, PA, Aug. 2002, pp. 221–233.

RFD+:
Z. Duan, J. Chandrashekar, J. Krasky, K. Xu, and Z. Zhang, "Damping BGP route flaps," in *Proc. IPCCC 2004*, Phoenix, AZ, Apr. 2004, pp. 131–138.

Original RFD

- Defined in RFC 2439
- Each route withdrawal or route attribute change is considered as a flap and penalized accordingly
- It may significantly delay the convergence of relatively well-behaved routes (routes that flap only occasionally):
  - BGP searches for alternatives if a route is withdrawn
  - path exploration leads to increase of penalty due to interim updates
  - BGP may suppress a route due to a single withdrawal

Selective RFD

- Distinguishes path explorations from genuine route flaps:
  - routes are selected in order of non-increasing preference during path exploration after withdrawal
- How to identify flaps:
  - sender attaches its local preference to each route advertisement
  - receiver compares the current route with previous route in terms of route preference
  - a flap is identified if a change of direction in route preference is detected (an increase following a decrease)

Selective RFD

- Simulations in small networks indicated that selective RFD identifies genuine flaps better than original RFD
- Assumes incorrectly that route preference changes are monotonic during path exploration:
  - currently feasible paths at the router may change with time
  - a better path may become available afterwards during path exploration

RFD+

- Overcomes the problem of the selective RFD algorithm
- A flap is identified when:
  - current route has a higher degree of preference than the previous route
  - BGP speaker has received the current route more than once since its previous flap
- Simulations in small networks indicated that RFD+ correctly identified route flaps in the case of a single flap

RFD+

- Algorithm:

  when receiving a route r with prefix d from peer j

  if (r $\notin$ R(d, j))

  insert r into R(d, j)

  else if (r $\in$ R(d, j) and dop(r) > dop(p))

  a flap is identified

  clear R(d, j)

R(d, j): set of all routes with prefix d from peer j

dop(p): degree of preference for the previous route p

Key difference: how to identify flaps?

- Original RFD:
  - any route withdrawal or route attribute change is considered a flap
- Selective RFD:
  - a flap is identified if a change of direction in route preference is detected (an increase following a decrease)
- RFD+:
  - a flap is identified when the current route has a higher preference than the previous one and the BGP speaker has received the current route more than once since the previous flap

Roadmap

- Introduction to Route Flap Damping (RFD)
- ns-2 implementations of RFD
- Simulation scenarios
- Performance analysis
- Improvements to RFD algorithms
- Conclusions and references

ns-2 implementation of RFD

- Based on a BGP model developed for ns-2: ns-BGP 2.0
- Used relevant source code from the SSFNet BGP-4 module:
  - two algorithms implemented in SSFNet BGP-4 v1.5.0: original RFD and selective RFD
- Added implementation of RFD+ in ns-2

ns-BGP 2.0: http://www.ensc.sfu.ca/~ljilja/cnl/projects/BGP/
SSFNet: http://www.ssfnet.org/

ns-2 implementation of RFD

- New C++ classes:
  - DampInfo: stores the damping structure for a prefix advertised from a peer of a BGP speaker and implements all damping algorithms
  - ReuseTimer: keeps track of the reuse timer associated with a flapping route
- Modified C++ files and TCL files:
  - deal with route flap damping when receiving update messages and making routing decisions
  - set default global variables used in route flap damping

Roadmap

- Introduction to Route Flap Damping (RFD)
- ns-2 implementations of RFD
- Simulation scenarios
- Performance analysis
- Improvements to RFD algorithms
- Conclusions and references

ns-2 simulations of RFD

- Four elements of a simulation scenario:
  - network topology
  - inter-arrival time between updates
  - simulation time
  - nature of flaps

Network topology

- Generated by BRITE:
  - AS-level topology
  - Generalized Linear Preference (GLP) model
  - network size: ranging from 100 to 500 nodes
- Built from genuine BGP routing tables:
  - network size: 29 and 110 nodes

BRITE:
http://www.cs.bu.edu/brite

Multi-AS topologies from routing tables:
http://www.ssfnet.org/Exchange/gallery/asgraph

Inter-arrival time between updates

- Use at least three values for the inter-arrival time between updates:
  - one value smaller than the default MRAI value of 30 s: 10 s
  - one intermediate value: 100 s
  - one value large enough for BGP to converge: 1,000 s

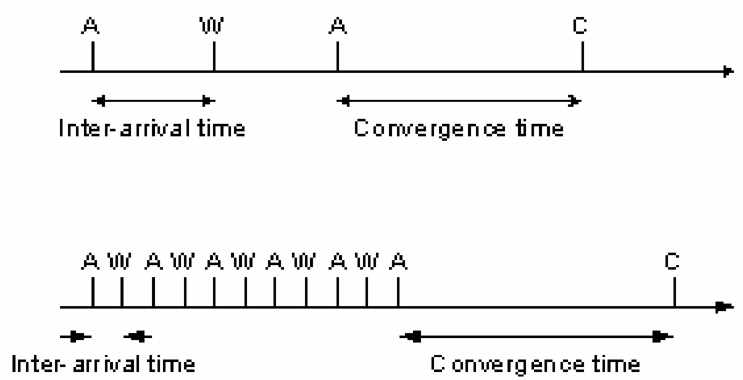
MRAI: Minimum Route Advertisement Interval

Simulation time

- Depends on the route suppression period:
  - with suppression period
  - without suppression period
- Comparison shows the impact of route suppression on individual BGP speakers and on the network

Nature of flaps

- Occasional flaps:
  - one flap
  - inter-arrival time between updates: 1,000 s
- Persistent flaps:
  - five flaps
  - inter-arrival time between updates: 300 s



A: Advertise

W: Withdraw

C: Converge

Roadmap

- Introduction to Route Flap Damping (RFD)
- ns-2 implementations of RFD
- Simulation scenarios
- Performance analysis
- Improvements to RFD algorithms
- Conclusions and references

RFD performance analysis

- Use default MRAI value (30 s) and apply jitter
- Use default Cisco settings:

Suppress limit	2000
Reuse limit	750
Half life (s)	900
Withdrawal penalty	1000
Attribute change penalty	500
Re-advertisement penalty	0
Maximum suppression time (s)	3600

- Compare RFD disabled, original RFD, selective RFD, and RFD+ in cases of occasional and persistent flaps in various networks

RFD performance analysis

- Examine advertisement and withdrawal phases, effect of inter-arrival time between updates, and effect of the origin router location
- Compare:
  - overall number of updates
  - overall number of reported flaps
  - number of flaps reported by each BGP speaker
  - maximum number of flaps associated with a single peer of each BGP speaker
  - overall number of suppressions caused by all the flaps
  - convergence time

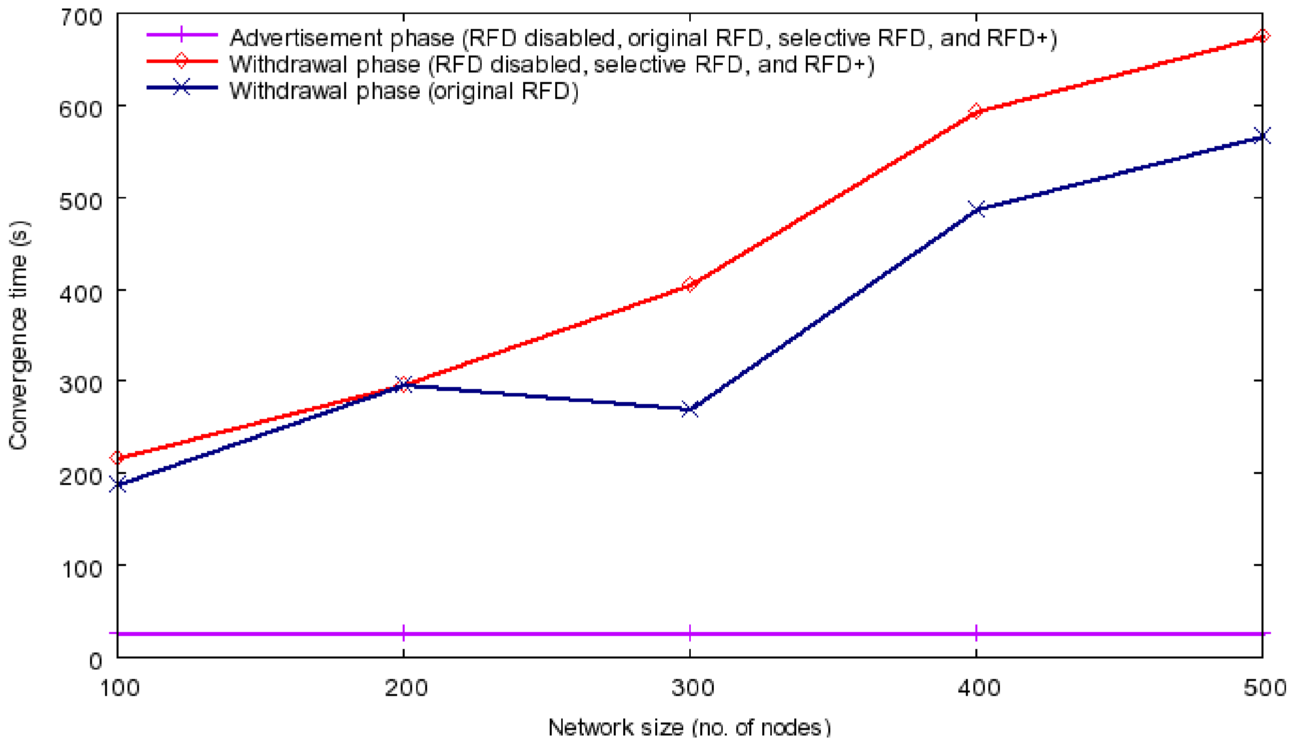
Advertisement vs. withdrawal

- A withdrawal message causes BGP to converge significantly slower than in the case of an advertisement message:

RFD algorithm	Phase	Network size (no. of nodes)				
		100	200	300	400	500
RFD disabled	Advertisement	27.017	27.017	27.017	27.017	27.018
	Withdrawal	216.21	297.31	405.3	594.21	675.3
Original RFD	Advertisement	27.017	27.017	27.017	27.017	27.018
	Withdrawal	189.21	297.31	270.31	486.21	567.21
Selective RFD	Advertisement	27.017	27.017	27.017	27.017	27.018
	Withdrawal	216.21	297.31	405.3	594.21	675.3
RFD+	Advertisement	27.017	27.017	27.017	27.017	27.018
	Withdrawal	216.21	297.31	405.3	594.21	675.3

Advertisement vs. withdrawal: convergence time

- A withdrawal message causes BGP to converge significantly slower than in the case of an advertisement message:



Advertisement vs. withdrawal

- Withdrawal phase: original RFD has the fastest convergence
- Withdrawal phase depends heavily on network size and network topology (dense or sparse)
- Damping algorithms have little effect on advertisement phase, but play an important role during the withdrawal phase

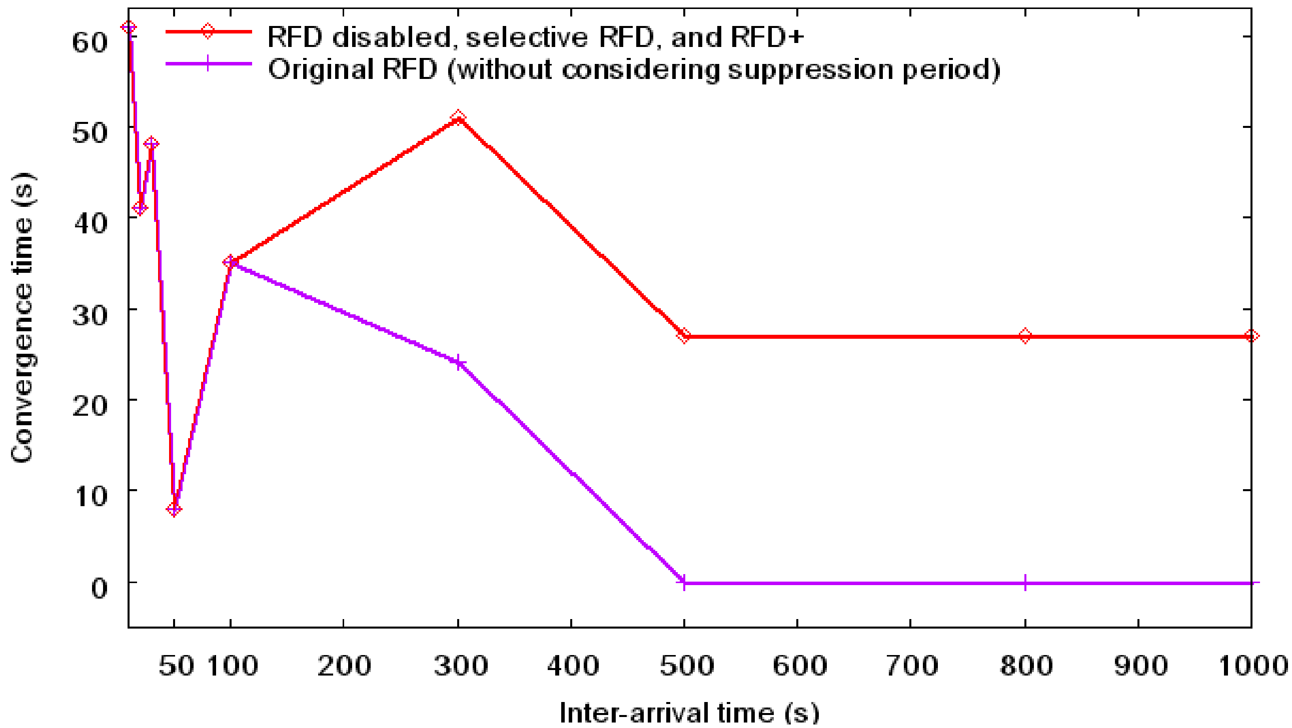
Effect of inter-arrival time

- No visible trend in the relationship between inter-arrival time and convergence time:

RFD algorithm	Evaluation parameters	Inter-arrival time (s)								
		10	20	30	50	100	300	500	800	1000
RFD disabled	Convergence time	61.017	41.017	48.017	8.017	35.018	51.016	27.02	27.02	27.02
	No. of updates	1832	1930	2434	2434	3725	8052	8056	8056	8056
	No. of flaps	/	/	/	/	/	/	/	/	/
Original RFD	Convergence time	61.017	41.017	48.017	8.017	35.016	24.116	0.02	0.02	0.02
	No. of updates	1832	1930	2434	2434	3678	7169	7202	7202	7202
	No. of flaps	961	1001	1430	1430	2229	3800	3829	3829	3829
Selective RFD	Convergence time	61.017	41.017	48.017	8.017	35.018	51.016	27.02	27.02	27.02
	No. of updates	1832	1930	2434	2434	3725	8051	8055	8056	8056
	No. of flaps	487	530	563	563	624	800	800	802	802
RFD+	Convergence time	61.017	41.017	48.017	8.017	35.018	51.016	27.02	27.02	27.02
	No. of updates	1832	1930	2434	2434	3725	8052	8056	8056	8056
	No. of flaps	435	454	491	491	493	496	497	497	497

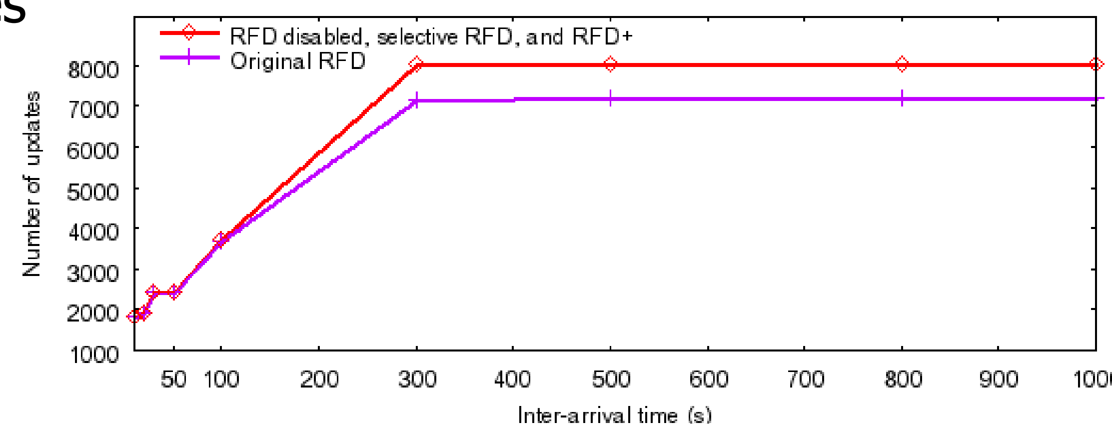
Effect of inter-arrival time: convergence time

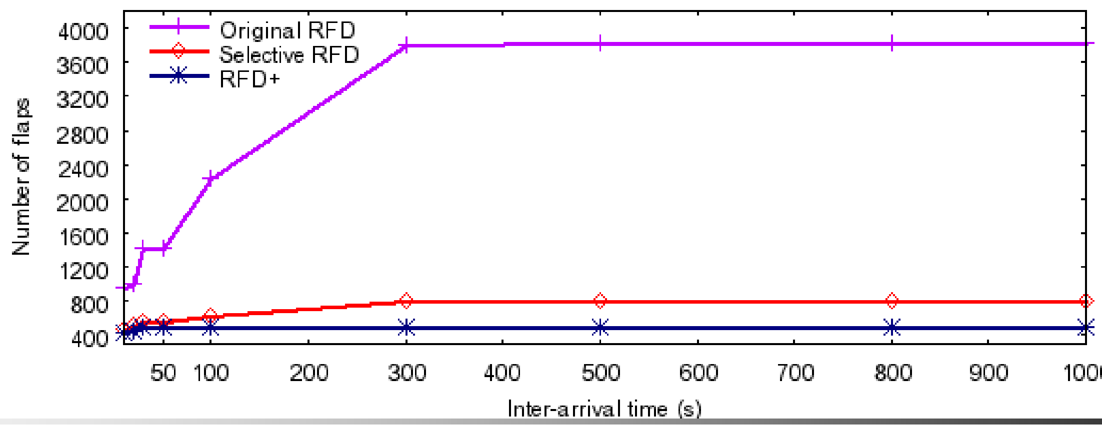
- No visible trend in the relationship between inter-arrival time and convergence time:



Effect of inter-arrival time: number of updates/flaps

- Number of updates and flaps tends to grow as inter-arrival time increases





Effect of inter-arrival time: summary

- Convergence time and number of updates: not affected by increase of inter-arrival time beyond a certain threshold
- Convergence time: affected by the differences between the instances when an update is ready to be sent and when the MRAI timer expires
- Convergence time and number of updates: not affected by damping algorithms when inter-arrival time is short
- Number of updates and number of flaps: no decrease as inter-arrival time increases
- Number of flaps and route suppressions:
  - RFD+: least sensitive to changes in inter-arrival time
  - original RFD: most sensitive

Location of the origin router

- Location:
  - core of the network
  - edge of the network:
    - often takes up to ~ 20% longer for BGP to converge
    - usually increases the number of updates by up to ~ 25%
- Effect on BGP performance depends on:
  - network topology
  - phase: advertisement or withdrawal
  - damping algorithm

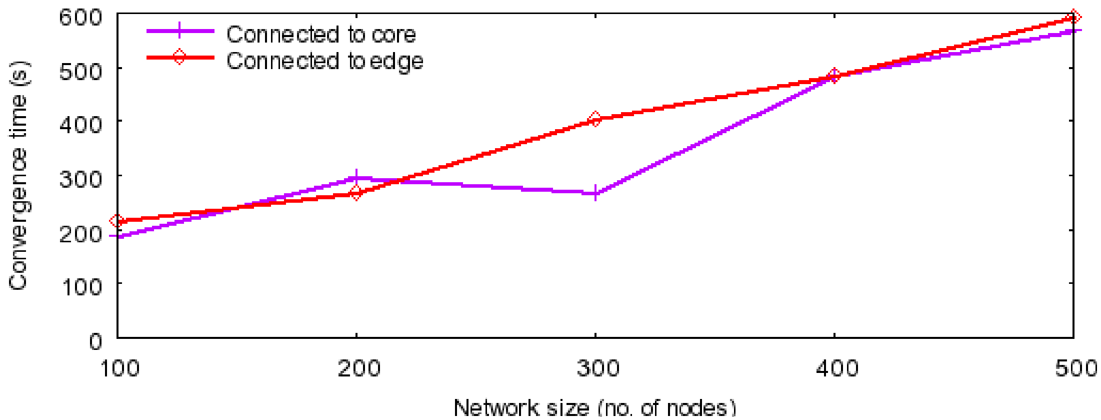
Location of the origin router

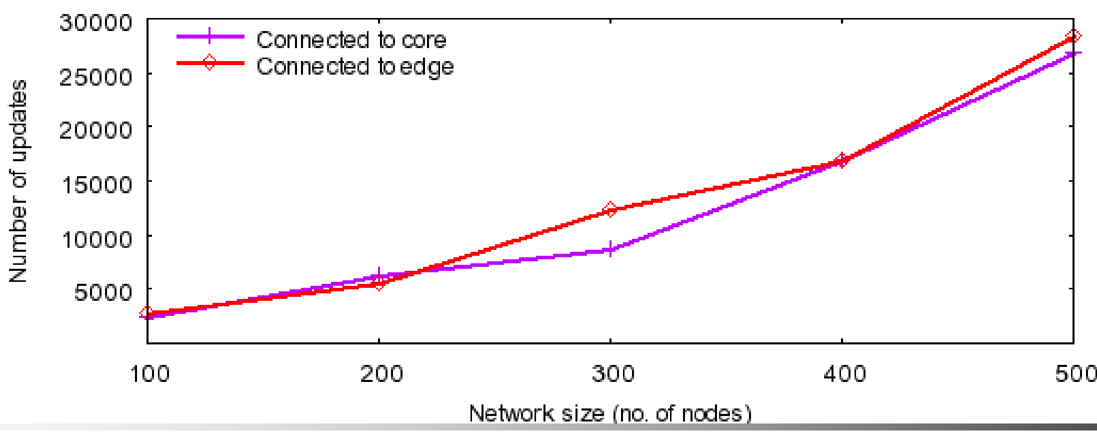
- Effect of the origin router's location on network performance during the withdrawal phase under original RFD:

Location of origin router	Evaluation parameters	Network size (no. of nodes)				
		100	200	300	400	500
Connected to core	Convergence time (s)	189.21	297.31	270.31	486.21	567.21
	No. of updates	2450	6237	8695	16896	26892
Connected to edge	Convergence time (s)	216.21	270.3	405.2	486.21	594.31
	No. of updates	2720	5531	12321	16896	28488

Location of the origin router: convergence time/number of updates

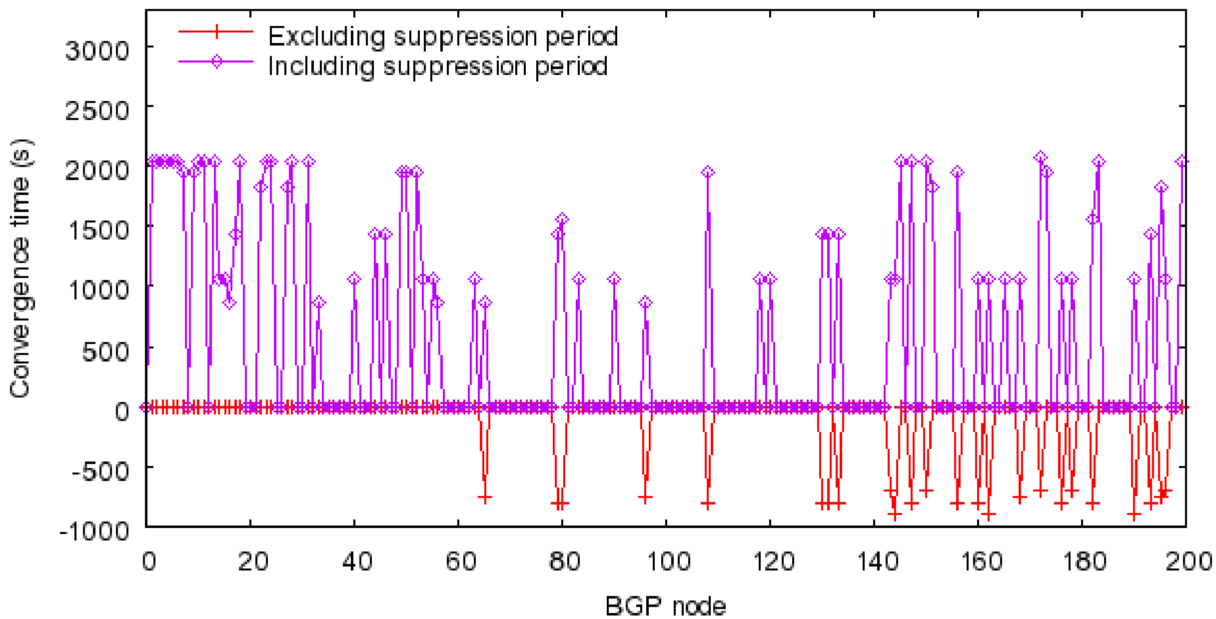
- Original RFD: withdrawal phase





Occasional flaps

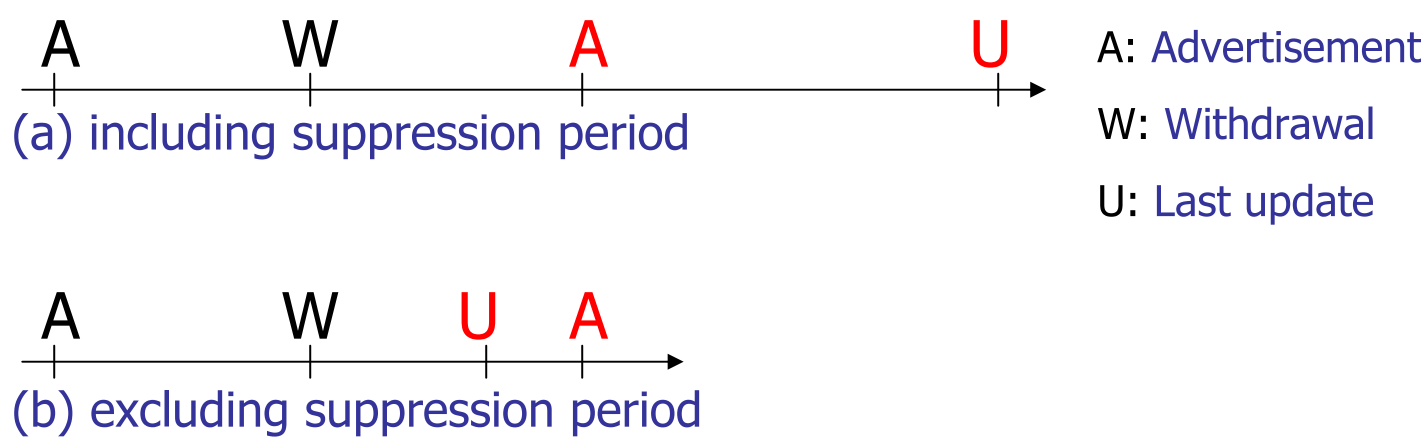
- Original RFD: one flap may cause many network nodes to suffer from a significant delay in convergence



Negative values imply that the nodes do not receive the route re-advertisement after withdrawal and will wait until other nodes become reused and start to advertise

Negative convergence time?

- Convergence time: time difference between the re-advertisement (second A) and the last update (U)
  - U occurs after A when including suppression period (a)
  - U occurs before A when excluding suppression period (b)



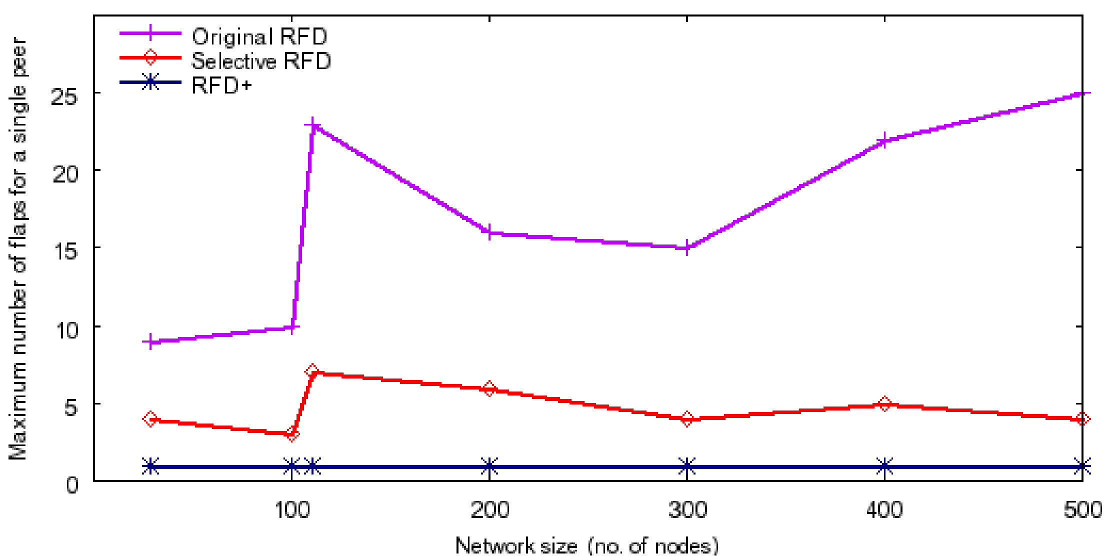
Occasional flaps

- Selective RFD performs better than original RFD in terms of the number of flaps and suppressions
- RFD+ has the best behavior because it does not mistake path explorations for route flaps
- Maximum number of flaps associated with a single peer:

RFD algorithm	BRITE (no. of nodes)					Routing tables (no. of nodes)	
	100	200	300	400	500	29	110
Original RFD	10	16	15	22	25	9	23
Selective RFD	3	6	4	5	4	4	7
RFD+	1	1	1	1	1	1	1

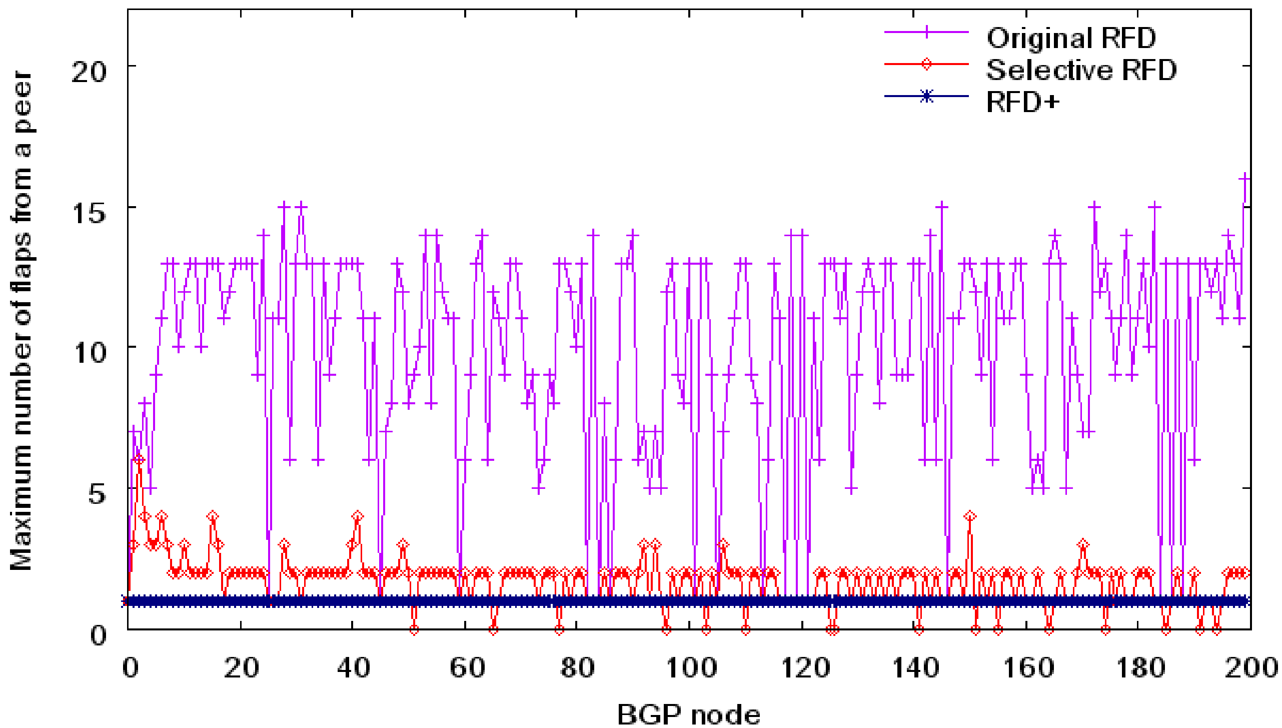
Occasional flaps

- Selective RFD performs better than original RFD in terms of the number of flaps and suppressions
- RFD+ has the best behavior because it does not misinterpret path explorations as route flaps



Occasional flaps

Original RFD: 16
Selective RFD: 6
RFD+: 1



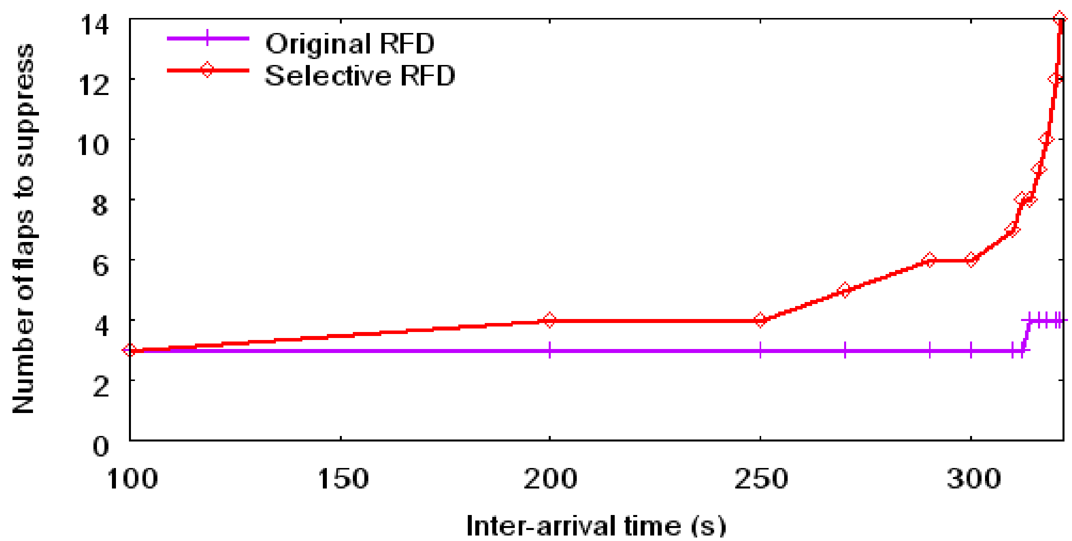
Persistent flaps

- Original RFD prevents the spread of routing oscillations as early as possible
- Selective RFD may require additional flaps in order to suppress a flapping route
  - the number depends on inter-arrival time
- Number of flaps required to suppress a route:

RFD algorithm	Inter-arrival time (s)												
	100	200	250	270	290	300	310	312	314	316	318	320	321
Original RFD	3	3	3	3	3	3	3	3	4	4	4	4	4
Selective RFD	3	4	4	5	6	6	7	8	8	9	10	12	14

Persistent flaps

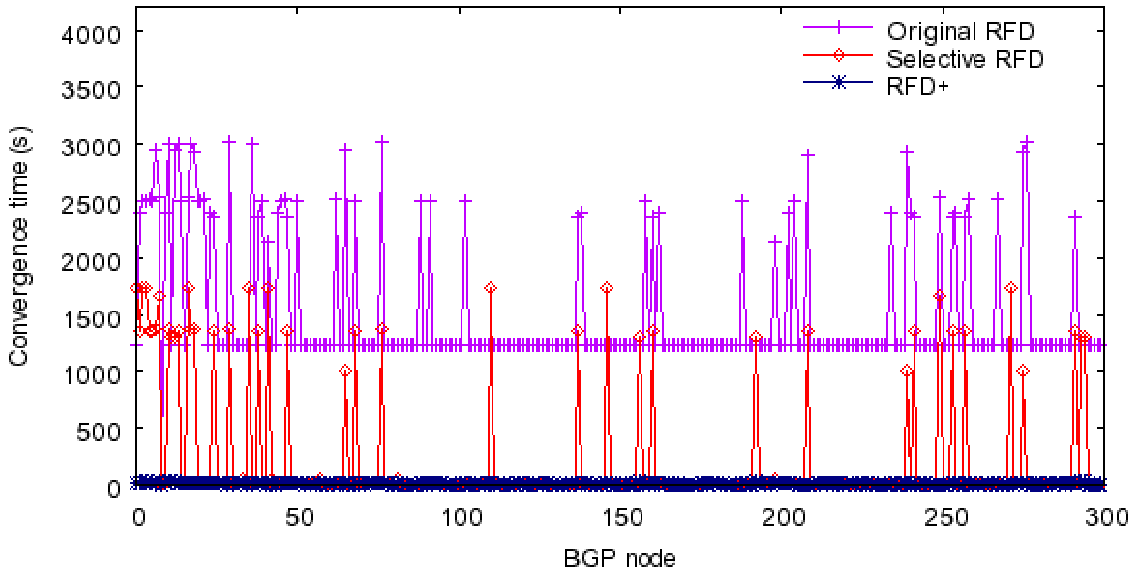
- Original RFD prevents the spread of routing oscillations as early as possible
- Selective RFD may require additional flaps in order to suppress a flapping route
  - the number depends on inter-arrival time



Based on default Cisco settings

Persistent flaps

- RFD+ underestimates the number of genuine flaps, causing a delay in route suppression



- Selective RFD and RFD+ are less aggressive than the original RFD in suppressing persistently flapping routes (not desirable)

Persistent flaps

- Selective RFD and RFD+ are less aggressive than the original RFD in suppressing persistently flapping routes (not desirable)
- Advertisement of a route after reuse may cause new route suppressions in the network, causing a cascading effect

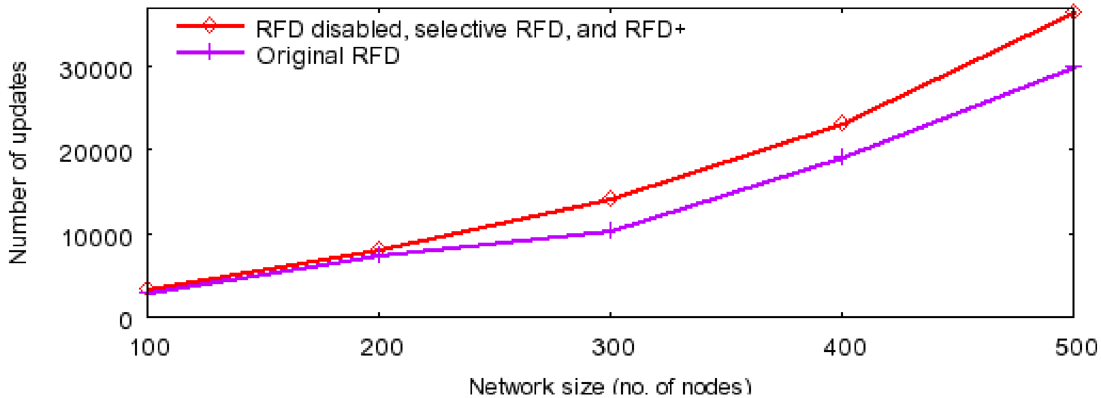
Network stability

- Original RFD: optimal in achieving network stability because it always leads to the least number of updates

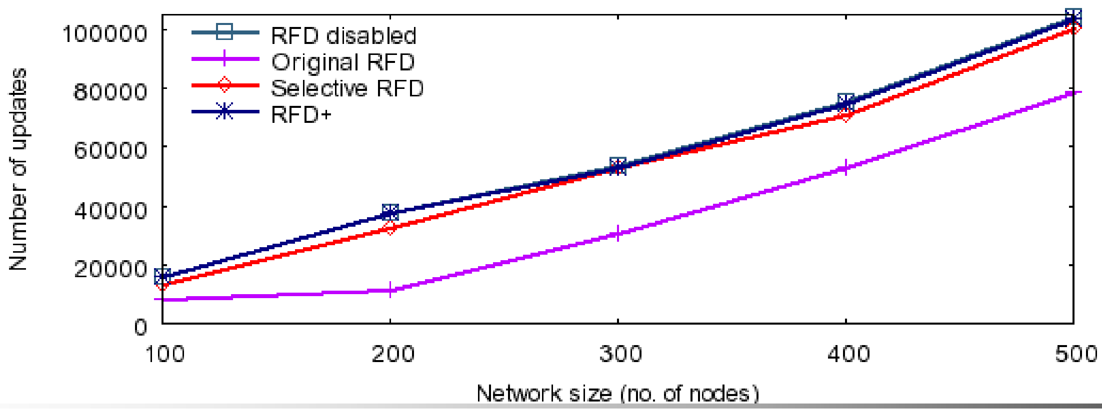
Nature of flaps	RFD algorithm	Network size (no. of nodes)					Routing tables (no. of nodes)	
		100	200	300	400	500	29	110
Occasional	No RFD	3407	8056	14126	23344	36642	1209	17621
	Original RFD	2984	7380	10464	19179	29924	1149	11514
	Selective RFD	3407	8056	14126	23344	36642	1211	17616
	RFD+	3407	8056	14126	23344	36642	1209	17621
persistent	No RFD	15935	38052	53789	75674	104236	5109	45607
	Original RFD	8016	11784	30822	53271	78519	1675	18308
	Selective RFD	13251	32535	52852	70924	100726	3291	44568
	RFD+	15850	37569	53452	75205	103892	5106	45694

Network stability

- Original RFD: optimal in achieving network stability because it always leads to the least number of updates



Occasional flaps



Persistent flaps

Roadmap

- Introduction to Route Flap Damping (RFD)
- ns-2 implementations of RFD
- Simulation scenarios
- Performance analysis
- Improvements to RFD algorithms
- Conclusions and references

Weaknesses: selective RFD and RFD+

- Selective RFD:
  - does not always correctly identify flaps
  - may require additional flaps to suppress a flapping route
    - situation worsens when inter-arrival time increases
- RFD+:
  - underestimates the number of genuine flaps:
    - reporting floor((N+1)/2) flaps if the origin router experiences failure and then recovery for N times
  - potentially large memory consumption

Improving RFD+

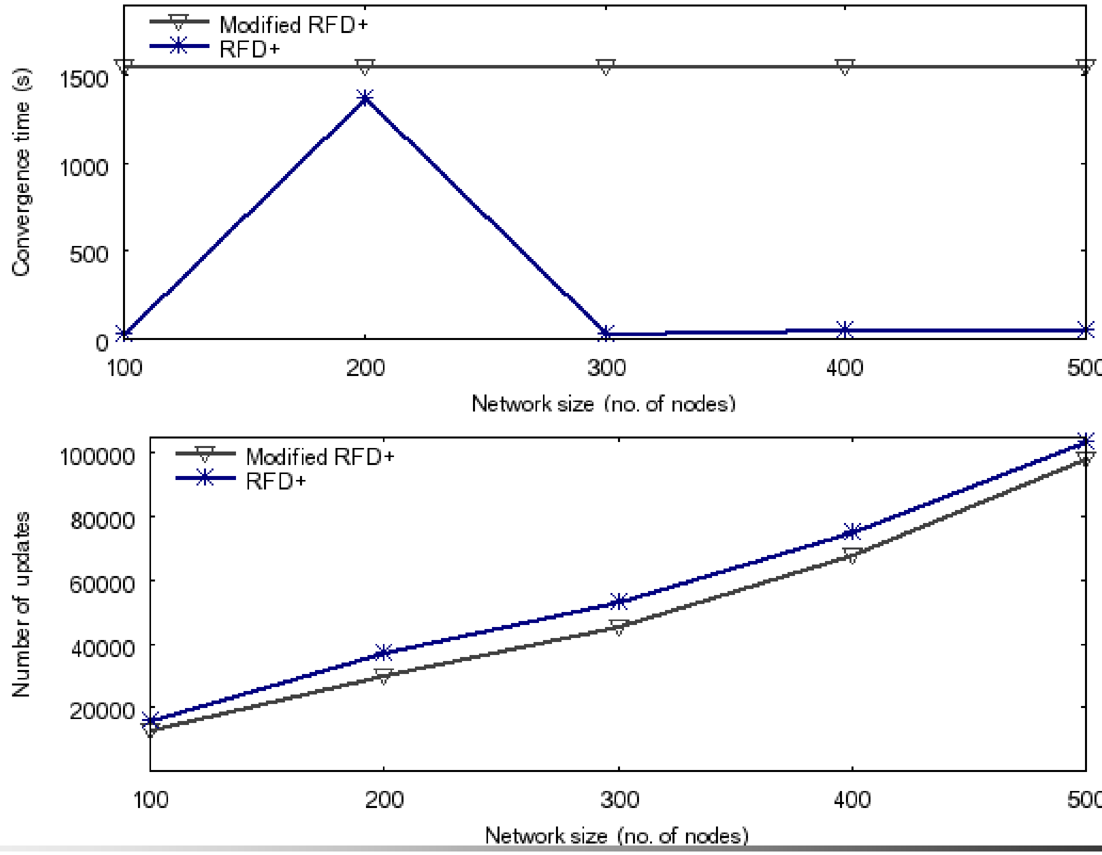
- Simple remedy (modified RFD+):
  - keep track of the "up-down-up" state of a route: advertise, withdraw, and re-advertise
  - report a flap either when identified by RFD+ or when a route is advertised, withdrawn, and advertised again:
    - can identify all N flaps if origin router fails and then recovers for N consecutive times
    - in rare cases, may report additional flaps
  - hash interim routes into a simpler data type:
    - reduce memory consumption and processing time
- Modified RFD+: identifies genuine flaps better than other RFD algorithms in both occasional and persistent flaps

Modified RFD+ vs. RFD+: persistent flaps

Algorithm	Evaluation parameters	Network size (no. of nodes)				
		100	200	300	400	500
Modified RFD+	Convergence time	1555.71	1555.71	1555.71	1555.71	1555.71
	No. of updates	12805	30541	45859	68038	98194
	No. of flaps	692	1335	2016	2684	3417
	No. of suppressions	5	22	31	22	23
RFD+	Convergence time	27.02	1369.21	27.02	51.02	51.01
	No. of updates	15850	37569	53452	75205	103892
	No. of flaps	856	1642	2557	3391	4417
	No. of suppressions	44	91	100	121	201

Modified RFD+ vs. RFD+: persistent flaps

- Modified RFD+ suppresses persistent flaps and leads to fewer updates



Combined RFD: adaptive approach

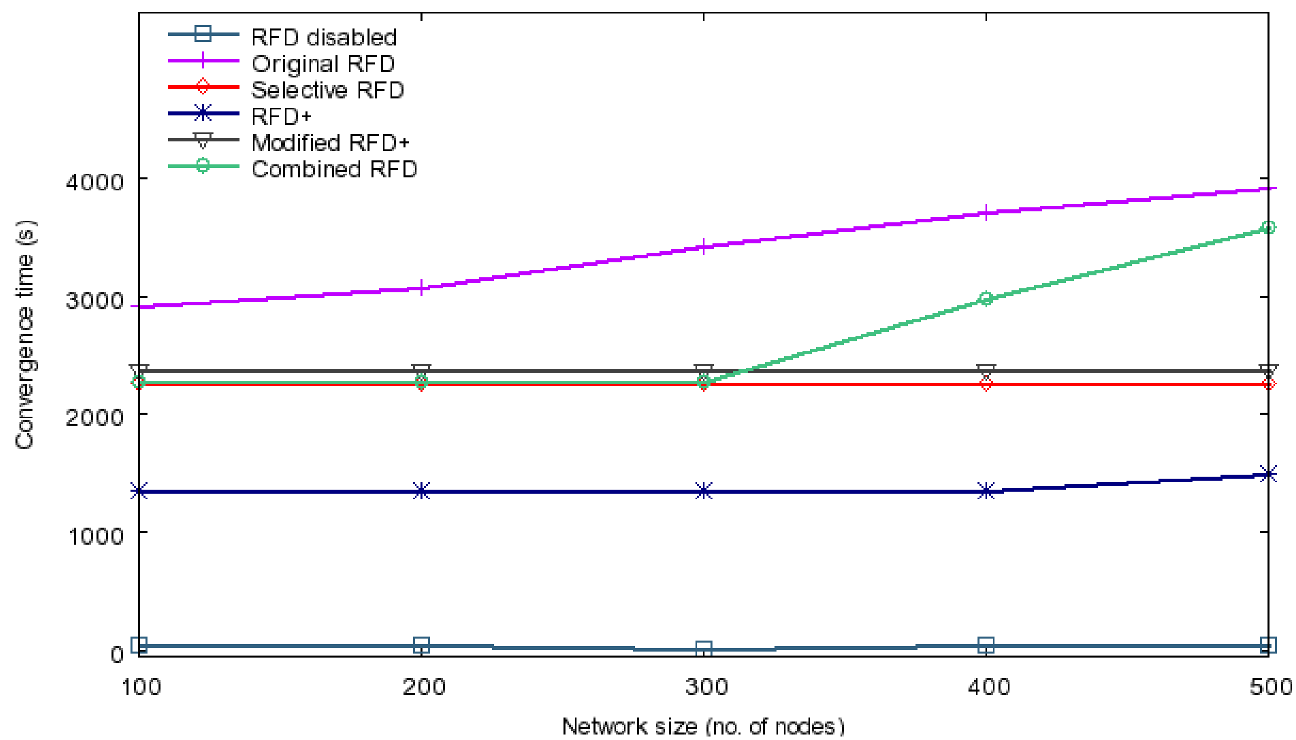
- Trade-off between network stability and availability of routes
- Combined RFD approach:
  - integrates original RFD and modified RFD+
  - uses modified RFD+ for first two flaps within a certain period of time, and switches to original RFD starting from the third flap
- Combined RFD:
  - does not suppress a relatively well-behaved route
  - suppresses persistently flapping routes efficiently
  - tends to generate fewer updates than selective RFD, RFD+, and modified RFD+

Persistent flaps: comparison

Algorithm	Evaluation parameters	Network size (no. of nodes)				
		100	200	300	400	500
No RFD	Convergence time	42.02	42.02	15.21	42.1	42.02
	No. of updates	15675	30904	45823	59441	78906
	No. of flaps	/	/	/	/	/
Original RFD	Convergence time	2908.06	3067.35	3426.2	3705.34	3910.4
	No. of updates	7519	16062	28054	37699	53436
	No. of flaps	4553	10235	19103	25563	36896
Selective RFD	Convergence time	2254.92	2254.92	2254.92	2254.92	2254.92
	No. of updates	8468	16207	33859	40549	60068
	No. of flaps	1874	3803	5952	7610	10479
RFD+	Convergence time	1349.38	1349.38	1349.38	1349.38	1502.64
	No. of updates	14344	28673	50167	66713	87916
	No. of flaps	1100	2154	3329	4326	5729
Modified RFD+	Convergence time	2374.92	2374.92	2374.92	2374.92	2374.92
	No. of updates	9190	19347	30958	45119	65630
	No. of flaps	676	1270	1975	2552	3300
Combined RFD	Convergence time	2271.41	2271.41	2271.41	2971.71	3583.96
	No. of updates	8202	17204	28608	37716	55477
	No. of flaps	2006	4622	9012	12137	19476

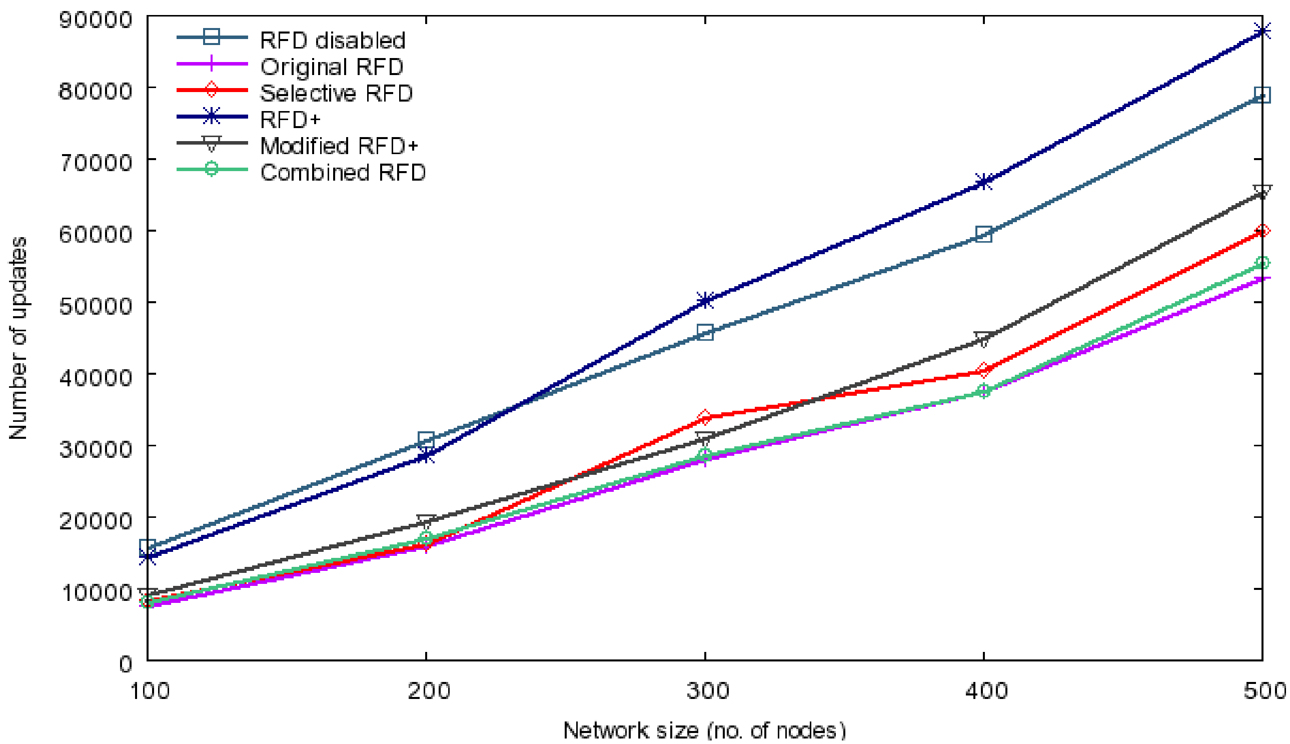
Persistent flaps: convergence time

- Convergence time of Combined RFD is between selective RFD and original RFD and depends on network topology

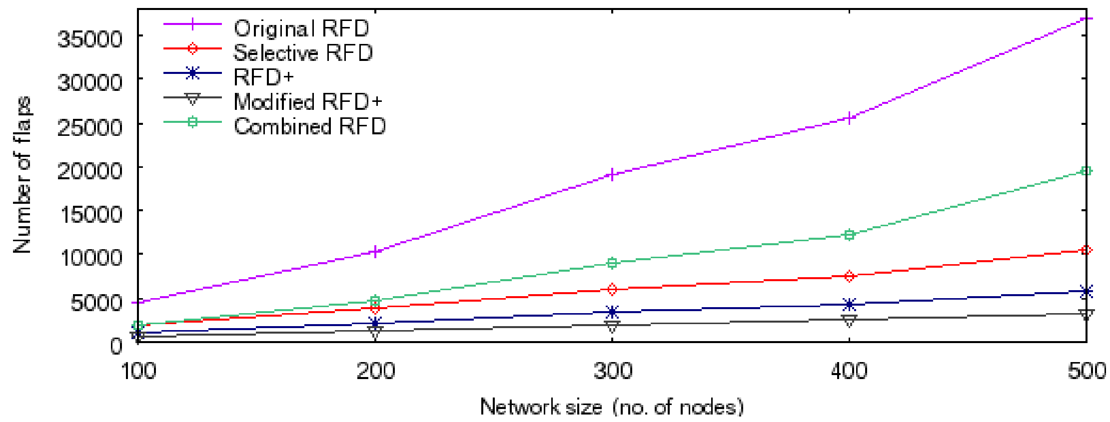


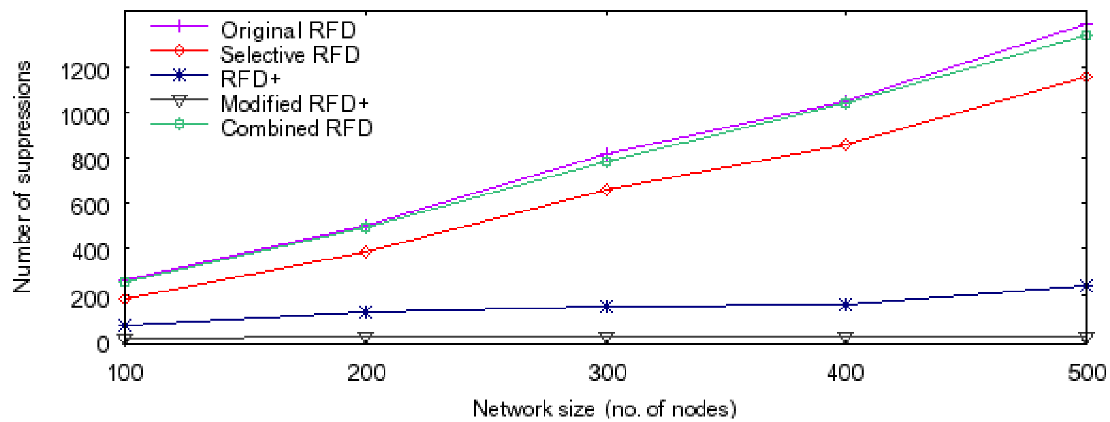
Persistent flaps: number of updates

- Combined RFD is the second best in reducing update messages (close to original RFD)



Persistent flaps: number of flaps/suppressions





Conclusions

- Compared RFD algorithms using realistic network topologies:
  - no algorithm performs optimally in all circumstances
  - Original RFD:
    - efficient in suppressing persistently flapping routes and achieving network stability
    - may cause significant convergence delay in the case of occasional flaps
  - Selective RFD:
    - identifies route flaps better than the original RFD
    - does not always correctly identify genuine flaps
    - may require additional flaps to suppress a flapping route, causing a delay in route suppression

Conclusions

- RFD+:
  - reports no additional flaps
  - may underestimate the number of genuine flaps, causing a delay in route suppression
- Proposed improvements:
  - modified RFD+ (modification to RFD+):
    - identifies genuine flaps better than other RFD algorithms in cases of occasional and persistent flaps
  - combined RFD (adaptive approach):
    - efficiently suppresses routes that flap persistently
    - does not suppress a route that flaps only once or twice
    - tends to generate fewer updates than selective RFD, RFD+, and modified RFD+

References

- C. Villamizar, R. Chandra, and R. Govindan, “BGP route flap damping,” *IETF RFC 2439*, Nov. 1998.
- Z. Mao, R. Govindan, G. Varghese, and R. Katz, “Route flap damping exacerbates Internet routing convergence,” in *Proc. SIGCOMM 2002*, Pittsburgh, PA, Aug. 2002, pp. 221–233.
- Z. Duan, J. Chandrashekar, J. Krasky, K. Xu, and Z. Zhang, “Damping BGP route flaps,” in *Proc. IPCCC 2004*, Phoenix, AZ, Apr. 2004, pp. 131–138.
- ns-2 [Online]. Available: http://www.isi.edu/nsnam/ns.
- SSFNet [Online]. Available: http://www.ssfnet.org/.
- BRITE [Online]. Available: http://www.cs.bu.edu/brite.
- Multi-AS topologies from routing tables [Online]. Available: http://www.ssfnet.org/Exchange/gallery/asgraph.
- T. Bu and D. Towsley, “On distinguishing between Internet power law topology generators,” in *Proc. INFOCOM 2002*, New York, NY, June 2002, pp. 638–647.
- T. G. Griffin and B. J. Premore, “An experimental analysis of BGP convergence time,” in *Proc. ICNP 2001*, Riverside, CA, Nov. 2001, pp. 53–61.
- T. D. Feng, R. Ballantyne, and Lj. Trajkovic, “Implementation of BGP in a network simulator,” in *Proc. ATS’04*, Arlington, VA, Apr. 2004, pp. 149–154.