FED4FIRE
Federation for FIRE Plus

Internet on Fire (IoF)

UNIVERSITY OF TRENTO - Italy

Ca' Foscari University of Venice

Leonardo Maccari,
Renato Lo Cigno, Michele Segata,
Mattia Milani, Lorenzo Ghiro,
Marco Nesler, Luca Baldesi

leonardo.maccari@unive.it

Table of Contents

BGP: Border Routing Protocol

- BGP is the de-facto standard routing protocol used between Autonomous Systems in the Internet:
  - business efficiency first, technical efficiency then
  - affected by known overhead and convergence speed problems
- Link/policy changes trigger storms of updates
- RFC introduces Minimum Route Advertisement Interval (MRAI)
  - minimum time between consecutive UPDATEs for the same destination to the same neighbor

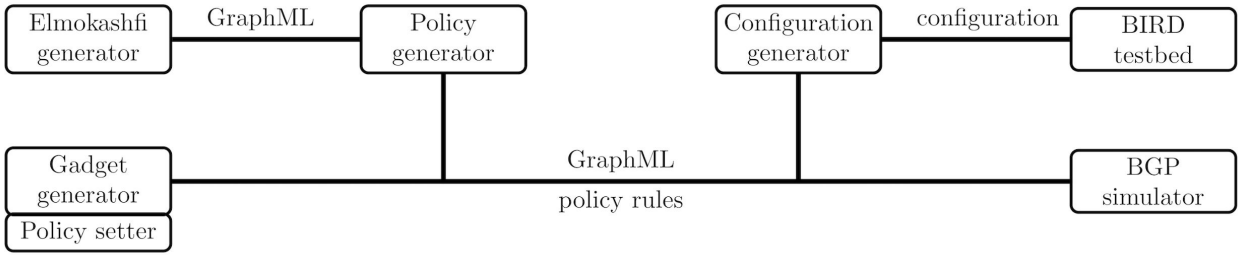
Issues with MRAI

- By default set to 30 seconds (as of RFC)
  - highly impacts convergence speed: may take minutes to propagate updates and reach routing table convergence
- There were proposals to reduce/eliminate MRAI
  - Fabrikant and Rexford (Infocom ‘11) showed that pathological cases may arise with non coordinated changes in MRAI
  - we cannot expect a coordinated/synchronized change...
  - ...research stopped but operators are moving independently, which exposes to future risks.
- **Goals of IoF:**
  - develop a BGP experimentation framework on Fed4FIRE+ testbeds
  - verify Fabrikant results
  - develop an incrementally deployable MRAI configuration mechanism

Project set-up

- We used and modified the BIRD open source BGP routing daemon
  - BIRD is one of the most used software for stub ASes
  - we modified it to introduce MRAI
  - we set-up virtual networks of BIRD instances with linux network namespaces
- We used both Virtual Walls testbeds to emulate large-scale networks running BGP
- We were able to emulate networks made of **up to 20.000 nodes.**
- Consider that CAIDA lists about 67.000 active AS in the Internet, so we are close to Internet scale
- To the best of our knowledge this is the largest documented BGP emulation with repeatable configurations and production software.

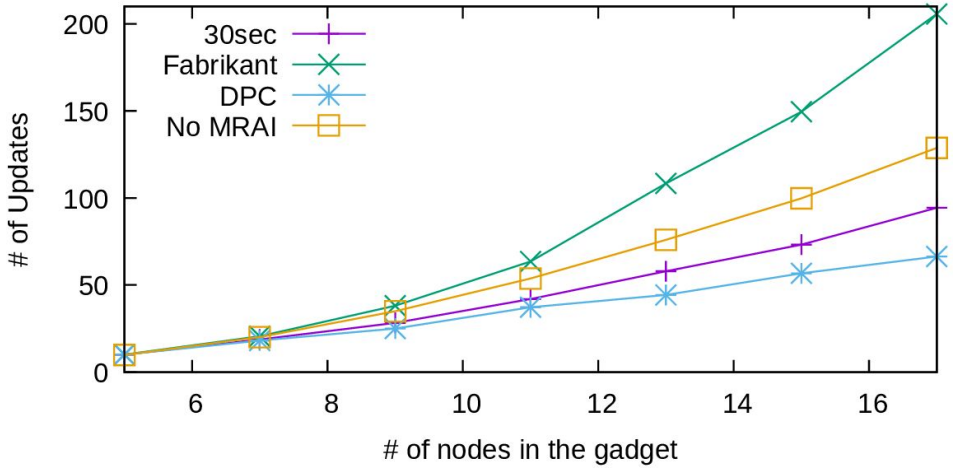
Results: Repeatable BGP Emulation set-up



- Realistic emulations require:
  - A topology generator with AS relationships (we implemented a generator by Elmokashfi et al.)
  - A policy generator (prefer-customer)
  - Computing power and throughput (2.300 cores for 20.000 nodes)
  - A huge machinery to configure nodes, run BIRD, monitor emulations, collect data and review them.
- Every detail is documented in the project website, to be easily reproduced: **https://iof.disi.unitn.it/software.html**

Results: Confirmed Fabrikant Hypothesis

- Fabrikant produced gadget topologies that should theoretically show an abnormal growth in the number of updates required for convergence
- We reproduced gadget topologies and confirmed that this effect exists



We can not simply remove/decrease MRAI in a non-coordinated way or we may create pathological situations.

But how to coordinate 67.000 Internet ISPs?

Results: Centrality-based MRAI

- Proposal: compute the centrality of each AS, using a graph centrality metric we designed to be computable in a fully distributed way: DPC (Destination Partial Centrality).
- Tune MRAI based on DPC.
- Result summary:
  - We were able to implement DPC, DPC is incrementally deployable
  - DPC converges! We tested networks with 1000 nodes and increasing penetration ratio.
  - We tried to tune MRAI based on DPC (no space for details, but essentially, increase MRAI on the path to the center of the network, decrease after information reached the center)

Results: DPC Vs Standard, 4000 nodes

	Strategy	mean	
updates	Fabrikant		
	No MRAI		+16%
	30 s	24806	
	DPC	29044	
conv. (s)	Fabrikant		
	No MRAI		-62%
	30 s	184.6	
	DPC	70.86	

Business Impact

- We are academics, so our impact is on research. We want to stress some real and potential impact aspects of our research:
  - BGP convergence is historically slow and the number of UPDATE messages/s is progressively increasing: there is a commercial need to be efficient.
  - BGP is used also out of the “Internet”, i.e. in data centers where convergence speed is essential.
- There is a lack of real-world results on BGP, aside observing BGP in the wild
- Accepted work at CNERT Infocom workshop (Reviews: 5-4-5), two more publication in preparation, one Ph.D thesis involved, one Bachelor thesis:

“The toolkit and experimental methodologies described appear likely to be useful for future research by the authors and others.”

Feedback to Fed4Fire+: Value Perceived

- With the availability of the testbed federation we were pushed to find a new line of research consisting in:
  - implement the features on real code, **improving an existing open source BGP implementation**
  - tested it in large networks
  - in the process, we developed a network generator that could produce graphs made of tens of thousands of nodes (now merged into networkx python library)
- We are now in the conditions of achieving results on BGP that were never achieved before, at this scale.
- The whole process is repeatable and documented.

Feedback to Fed4Fire+: Funding

- The value of the access to Fed4Fire+ infrastructure is way beyond the funding received.
- Even ignoring the resources needed to set-up a testbed management software like the one offered by Fed4Fire+, the budget we received for IoF **would have been totally insufficient for replicating similar experiments outside Fed4Fire+**
- The **scalability of the Fed4Fire+ federation was essential** to support research we carried on in IoF:
  - step 1: start with small scale experiments on one testbed
  - sted 2: replicate on thousands of cores in more than one testbed

Feedback to Fed4Fire+: What can be improved

- We could not use Grid 5000 testbed because it did not use the same reservation method at the time.
- JFed is handy for set-up of small-scale experiments, but we developed some custom command-line tools to create .rspec files for large simulations
- Visibility of running experiments and experiments results on Fed4Fire+ project website can be improved
- The information about the testbeds is scattered around several websites with different formats. Description of testbed details can be made more homogenous (useful during proposal preparation)

Conclusions

Conclusions

- We were able to reproduce Fabrikant Results and show that there can be a real practical issue with MRAI
- We set-up large scale emulation campaigns on the two testbeds (VW1 and VW2)
- We begun the research on a distributed MRAI applicable to BGP and to the Internet
- We opened research directions for our group and possibly for our research area
- We are still using the testbed after the end of the project to produce more publications.

Internet on Fire
-
Questions?

WWW.FED4FIRE.EU


Co-funded by the European Union


Co-funded by the Swiss Confederation

This project has received funding from the European Union's Horizon 2020 research and innovation programme, which is co-funded by the European Commission and the Swiss State Secretariat for Education, Research and Innovation, under grant agreement No 732638.