Modularized BGP for Decentralization in a Distributed Router

Introduction
Routing in the Internet is a growing challenge since Internet grows with more hosts, people demand QoS, and mobility. When Internet was designed and routing over Internet computers where big machines that were rarely moved due to power and cooling demands and their sheer weight. Nowadays people have PDAs and laptops with wireless communication. People want to stay online wherever ever they are.

Distributed router
Traditionally routers have been seen as monolithic units. The ForCES [2] project have outlined a structure that allows the router to be distributed. The distributed router consists of Forwarding Elements(fe), Control Elements(CE) and glue in a form of an internal network connecting the different parts. The forwarding elements are fast and cheap units that mainly do forwarding according to a forwarding-table. Control Elements do processing and distribute the forwarding table to all forwarding elements. Advantages with this design is cheaper parts, and easier adaptation to our changing routing world, with new services, IPv6 and such. Such a design can also solve different problems within an Autonomous System(AS).

BGP
BGP[1] is the de-facto standard routing protocol on the Internet. BGP was originally designed to run on a single router, here we have investigated how BGP can be divided in different parts to fit the distributed structure, and also which problems can be solved with the distributed BGP.

Figure 1: Network element, with control and forwarding elements connected by an internal network

Figure 2: Simplified overview of a router, BGP protocol, in- and out- policies and the three databases, RIB-In, RIB-out and Loc-RIB (routing table)

Figure 3: Forwarding problem, dashed lines are BGP sessions and solid lines are physical links. A’s shortest path out of the AS is through B and C, while B’s shortest is through A and D.

Figure 4: The number of prefixes (Y-axis) in each block (first 8 bit of IP-address) received during start-up from LTU main router

Figure 5: The number of bursts (Y-axis), and size of bursts (X-axis) received during a day from LTU main router

Figure 6: The number of bursts (Y-axis) and time between bursts (X-axis) received during a day from LTU main router

Figure 7: N Session managers connected to M decision processes

Figure 8: Multiple decision processes with suggested structure for managing the FIB; each decision process is responsible for a part of the normal FIB and the backup

Statistics
The CIDR-report [3] claims that there is currently around 160000 prefixes. We have defined a prefix block as all prefixes that start with the same 8 bits. In figure 4 are the number of prefixes for each block. As we can see they are not evenly spread over the prefix universe. Updates on these prefixes arrives at a rate of 1.4 updates every second. The burstiness and size of bursts can be seen in figure 5 and 6.

Ideas
We suggest placing a BGP-proxy in the forwarding element to allow several BGP processing units in CEs to work in parallel, and handle a part of the forwarding table each. This will increase stability in the AS and also protect against single points of failure.

Conclusions
Our suggested modularization of BGP for the distributed router solves some internal AS problems such as, I-BGP oscillation, I-BGP forwarding loops, internal starvation and related flaps.

References