OP2: A richer network abstraction?

Idea

To provide Internet applications with the *choose path* feature I propose the introduction of a new **routing layer**. This layer between the network and the transport layer takes higher-level routing decisions which the network layer can't provide.

The routing layer is implemented by border routers of autonomous systems (AS), where each AS forms one routing unit identified by its AS address (ASA). In contrast to AS numbers, ASAs are hierarchically structured and - like IPs - encode regionality/locality, allowing ASA route aggregation.

In that architecture, an end-system interface is uniquely identified by the tuple of `ASA:IP`. Therefore, DNS requests for domains resolve to `ASA:IP` addresses and packets are also addressed like that.

Functioning

Applications have the existing interfaces for sending packets with two additional, optional arguments: *routing policy tag* and *ASA exclusion prefix*. As proposed by [1, p. 135], the routing policy tag can specify a performance metric, like latency or throughput, that the chosen path should optimise. Using hierarchical prefix aggregation, the exclusion prefix can define a set of ASes the packet should not enter.

Upon sending, the routing layer adds a *routing header* to the transport layer packet and sends it to the IP address of the local AS's border router. The routing header consists of source and destination `ASA:IP` address, routing policy tag and ASA exclusion prefix. In general, packets get from one border router to the next by being sent to the next router's IP address, similarly to TRIAD's relay agents [2, p. 2].

The border routers act as routing layer switches, taking coarse-grained routing decisions. Packets destined by their IP address to the border router are passed up. The router reads the *routing header* and chooses the next-hop AS. Among multiple options, the router selects the path that optimises the routing policy metric while avoiding next-hop ASes in the exclusion prefix. If all paths are excluded, the router returns the packet to the last router, adding its own ASA to the exclusion prefix. Subsequent packets of that communication copy the updated prefix.

To perform well-informed decisions, border routers participate in a more verbose version of BGP where they keep more than one path per destination in their forwarding table. Border routers also evaluate the paths' performances by monitoring the links with probe messages.

Discussion

Because an IP router only decides about the next IP-hop, I argue that a layer above the IP layer is better suited to provide the desired features. In IP, it's impossible to direct a packet along a path for more than one hop. Contrastingly, border routers know the local topology and their neighbours. Therefore, they are best positioned to provide those features while keeping the routing logic and performance monitoring overhead restricted to AS borders.

Implementing the features above the transport layer at end-hosts would reduce overhead even more but constitutes a layering violation. Indeed, the routing layer would need to assume a lower-level network structure which the transport layer is unaware of.

References

[1] D. Andersen, H. Balakrishnan, F. Kaashoek, and R. Morris, "Resilient overlay networks," in *Proceedings of the eighteenth acm symposium on operating systems principles*, 2001, pp. 131–145.

[2] D. R. Cheriton and M. Gritter, "TRIAD: A scalable deployable nat-based internet architecture," Computer Science Department Stanford University, 2000.