Enabling Traffic Management without DPI

"DPI Is Dead, Long Live Traffic Management"

by Mirja Kühlewind, Dirk Kutscher, Brian Trammell

Deep packet inspection (DPI) is a widely applied technology for network measurement, management, and security purposes. In mobile networks, it is often applied to traffic classification and the implementation of differential treatment based on these classes — in the absence of trustworthy (or deployed) explicit signaling about the type of traffic.

However, DPI has several disadvantages regarding reliability, performance, scalability (state maintenance), operational inefficiency -- and applicability in the presence of encryption.

More seriously, the use of DPI is threatened by the increased deployment of encryption in the Internet, itself driven by increased privacy awareness following the Snowden revelations. Future deployments of opportunistic encryption protocols will further increase the proportion of traffic that is encrypted.

The fundamental question here is: is DPI really needed for traffic management in mobile networks? Our position is "no". Traffic management is usually realized through relatively simple mechanisms like rate shaping, prioritization, and dropping packets. Compared to these mechanisms, the semantics of applications that can be exposed through DPI are much richer; traffic classification anyway maps these semantics down to a simple set of categories.

The question then arises whether operators are really helped by brittle, insecure and expensive mechanisms for gaining higher-fidelity information for the coarse traffic information for traffic management, or whether simple signaling would suffice for traffic classification for mobile network management purposes.

Obviously, when relying on endpoints to signal information about the underlying application which may be used to change the network's treatment of that application's traffic, questions of trust arise: how can the network be sure the endpoints are being honest, and prevent endpoints from gaming the system to their advantage (and the disadvantage of others); can these signaling approaches be used as an attack vector. Here the approach is to define the vocabulary of the signaling protocol to properly incentivize honest cooperation, while allowing the network to verify this cooperation [4].

Is this paper we discuss two application-independent approaches for traffic management that are based on network-compatible metrics: ConEx Policing and low latency support with SPUD.

Congestion Policing in Mobile networks using ConEx signaling

Congestion Exposure (ConEx) [2] is a mechanism that enables senders to inform the network about previously encountered congestion in flows -- thus enabling senders and network infrastructure to respond to congestion based on operator policies. This information are provided in the IP header and can still be accessed even if the payload is encrypted. ConEx information is auditable by comparing the congestion level at network egress to the ConEx signal which incentivizes the sender to state its congestion contribution correctly.

ConEx can be applied in the 3GPP Evolved Packet System (EPS), as described in more detail in [3], for traffic management (reducing the need for complex DPI) and other functions.

Using ConEx would allow for a bulk packet traffic management system that does not have to consider application classes. Instead, with ConEx accurate downstream path information on incipient congestion are visible to ingress network operators. This information can be used to base traffic management on the actual current cost (which is the contribution to congestion of each flow) and enable operators to apply congestion-based policing/accounting depending on their preference and independent of application characteristics. Such traffic management would be simpler, more robust (no real-time flow application type identification required, no static configuration of application classes) and provide better performance as decisions can be taken based on the real actual cost contribution at each point in time.

Note that traffic management is only necessary during periods of congestion — indeed, if there is enough capacity to satisfy all senders, there is no need to classify and manage traffic at all. As ConEx is only applied when congestion occurs, it avoids classification and management overhead inherent in “always-on” management approaches.

By enforcing overall limits for a user’s congestion contribution or by accounting/charging based on the congestion contribution, ConEx enables applications to decide on their response to congestion notification while incentivizing them to react (in general) appropriately. For example, adaptive video streaming could adjust its sending rate early to give preference to short-lived interactive communication (which would not need to respond). In addition a service provider can use the same mechanisms to offer different levels of QoS, e.g. premium service with larger congestion allowance.

ConEx enables a flexible and effective traffic management that does not require operators to apply DPI (and often outdated application flow policies) and/or to give up end-to-end-security. Instead, with ConEx, operators can incentivize users (and their senders) to implement capacity sharing in a cooperative way.

Low Latency Support using simple SPUD signaling

Real-time media traffic has hard requirements on the maximum end-to-end latency (and jitter). To provide appropriate quality of service for this kind of traffic today often DPI-based detection mechanisms are applied. However, due to the large and growing variety of interactive services, this is often not possible. In the absence of any other information, traffic is usually treated by the network as being more sensitive to loss than latency, whereas the provision of an additional more latency-sensitive service would be more required by such traffic.

The Substrate Protocol for User Datagrams (SPUD) is a new approach to selective information exposure designed to support transport evolution as described in [1]. SPUD is realized as a shim between UDP and an (encrypted) transport protocol. The basic SPUD protocol provides minimal sub-transport functionality by grouping of packets together into tubes and signaling of the start and end of a tube. This will assist middleboxes in state setup and teardown along the path. Further, SPUD provides an extensible signaling mechanism based on a type-value encoding for associating properties with individual packets or all packets in a tube.

The SPUD protocol can be used to signal low latency requirements from an endpoint to the network, or expose the existence of support for such services from the network to the endpoint. Therefore we propose to provide four SPUD signals: a latency sensitivity flag, a signal to yield to another tube, an application preference for a maximum single queue delay, and a facility to discover the maximum possible single queue length along the path.

Based on the latency-sensitivity flag a network operator can implement an additional service (as compared to today's best effort service) that uses smaller queues and/or different AQM parameters without changing the service that is provided today. Signaling of lower queue priority or maximum single hop delay can further be used to preferentially drop packets of the same sender or within one flow. Information about expected queuing delays on the path can be used for buffer configuration at the endpoints. These mechanisms provide an incentive to expose this information without introducing a benefit for lying.

Conclusion

Ubiquitous usage of DPI for traffic classification is problematic for several reasons. Most importantly, it is incompatible with a secure Internet design and will break in the presence of increasingly deployed end-to-end encryption.

Instead of compromising on security or starting an arms race between security & privacy enablers on one side and traffic analytics and classification schemes on the other side, we argue that it is possible to implement a cooperative approach that can provide both, security & privacy and hooks for operators to provide better traffic management. The two examples ConEx and SPUD-enabled traffic management illustrate that exposing application-independent management information can enable sufficient traffic management -- without requiring DPI.

References

[1] B. Trammell and J. Hildebrand. Evolving transport in the Internet. *Internet Computing, IEEE*, 18(5):60–64, September 2014.
[2] M.Mathis, B. Briscoe; Congestion Exposure (ConEx) Concepts, Abstract Mechanism and Requirements; Internet Draft draft-ietf-conex-abstract-mech-13, Work in Progress; April 2015.
[3] D. Kutscher et. al; Mobile Communication Congestion Exposure Scenario; Internet Draft draft-ietf-conex-mobile-04, Work in Progress; September 2014.
[4] B. Trammell; Thoughts a New Transport Encapsulation Architecture; Internet Draft draft-trammell-stackevo-newtea-01, Work in Progress; May 2015.