Enabling a forwarding plane for future data-centric networks

Christian Esteve Rothenberg
School of Electrical and Computer Engineering
University of Campinas - São Paulo, Brazil
chesteve@dca.fee.unicamp.br

1. INTRODUCTION

A lot of attention is being paid to overcoming a number of present Internet limitations, mainly w.r.t. mobility, security, address space exhaustion, and routing table size growth. The continuous patching of the TCP/IP suite with ad-hoc protocol extensions and overlay solutions is regarded as a complex and costly solution in the long term. As a consequence, the so-called "clean-slate" approach to future Internet research has gained a lot of interest. One key question is to what extent a new networking paradigm is really necessary, e.g., as packet switching was to circuit switching in the 70s. The reasoning is based on the large scale use of the Internet for dissemination of named pieces of data [3]. A myriad of devices generate and request content, without caring about the actual data source as long as integrity and authenticity are assured [5]. Such an networking approach has been baptized as "data-oriented" [5], "content-centric" [3] or "information-oriented" networking [1], and aims at shaping the future Internet to interconnect information at large rather than evolving the host-centric architecture. This sets the motivation of my PhD work[1], the contributions of which are an outgrowth of the efforts and vision of the PSIRP project[2].

2. APPROACH

A shift in the orientation of network architecture design implies rethinking many fundamentals, for instance, defining a new identifier space for information objects of different granularities (documents, channels, packets), enabling more expressive communication patterns (e.g., pub/sub), efficient transmissions (e.g., multicast, caching, netw. coding) and increased resilience (e.g., security, data replication). Our contribution to an overarching solution starts at the lower layer by re-thinking the forwarding plane in a way that it could accommodate different information-centric control planes (topology mgm., pub/sub). High level design objectives of the data-centric forwarding plane include:

- **Generality** w.r.t the identifiers on which switching decisions are taken, the network indirection points and the specific control and management planes.
- **Simplicity** of the forwarding plane, "outsourcing" the intelligence and tussle resolution to centralized controllers in the spirit of OpenFlow [6] and 4D [7].
- **Efficiency** by reducing state (memory reqs.) and yielding line speed decisions at forwarding elements.
- **Adaptability** to the actual network usage, the scale-free characteristics of networks and the long-tail distribution of content demands.

At a lower level, envisioning information-oriented applications, we prioritized the following capabilities:
(1) Multicast: mode of communication as the standard primitive, with unicast being a special case.
(2) Beyond id/loc separation: implies a naming and addressing scheme that departs from the identification of nodes, separating routing from forwarding and having link identities in a pivotal role.
(3) Security: not as an afterthought but *ab initio* in the forwarding plane, in tune with the overarching goal of a DDoS resistant architecture.
(4) Caching: support functions to enable a (distributed) caching system that exploits the data oriented naming, e.g., by explicitly including caching way-points.
(5) Service-oriented policies: Assuming that middlebox-like services (e.g., WAN accel., DPI, load balancers, encoders) will be continuingly demanded, the forwarding plane should provide a means for policy-based traffic indirection to services boxes without hardwiring the service specifics in the forwarding operations.

3. CONTRIBUTIONS

Following the premises above, we have developed several enablers for the forwarding plane, including:
Bloom-filter-inspired port forwarding: The SP-Switch [1] leverages a packet classification technique to behave as an abstract switching element with one programmable Bloom filter per output (physical/ virtual links, internal processes). Due to its hashing-based nature, the switching decisions can be taken at line rate and accommodates (*generality*) various types of

[1]The concepts in this paper are an interpretation and responsibility of the author.
[2]http://www.psirp.org

packet identifier spaces (e.g., 256-bit content IDs, flat forwarding labels). Acting as a probabilistic hash table, it returns always the inserted output value (or forwarding ID mapping entry) and, additionally, in rare cases (false positive $\approx O(10^{-6})$) it incurs in extra (non-programmed) multicast-like operations. We believe that the small, multiplicative false positive rate and the data-oriented approach justify the trading of over-deliveries for state reduction and line speed operations (*efficiency*).

Forwarding on Bloomed link identities: We developed a forwarding fabric [4] based on the idea of placing a small in-packet Bloom filter (iBF) containing the links involved in the path(s) between communicating entities. For each point-to-point link, we assign per direction a Link ID (e.g. $\overrightarrow{AB}$, $\overleftarrow{AB}$) without requiring a common agreement between the nodes. Link IDs take the form of a single element Bloom filter of length m (256) and with k (5) bits set to 1 and form thereby a probabilistically unique link naming space ($m!/(m-k)! \approx 10^{12}$).

Source-routing: Assuming enough network topology information, upon a request for a routing identifier, a delivery tree can be constructed by inserting the required Link IDs between source(s) and sink(s) in the in-packet Bloom filter (iBF). On packet reception, each forwarding node checks its outgoing Link IDs against the iBF. On match, the packet is forwarded along that link. False positives will cause packets delivered over unrequested links.

Scalable multicast: The forwarding approach enables moving state (forwarding information) between packets (iBFs) and network nodes. For instance, with 256-bit iBFs, stateless multicast can be supported by including around 35 links, which is is enough to reach up to 20 users in a typical WAN [4]. If larger multicast groups are required then network state can be installed by defining virtual links spanning multiple hops or by adding entries in the SPSwitch. [8]

Secure, path-dependent forwarding identifiers: The source-routing-based approach makes forwarding independent from routing and basically hides every node location information. Without explicit authorization of the receiver and involvement of the topology system, packets are not forwarded in the network due to the absence of static host addressing mechanisms. Thus, *security* is provided by virtue of "encrypted" source routes (aka capabilities) compactly represented in fixed-size iBFs, which maintain the link identities undisclosed and are meaningful (routable) only for nodes *en-route*.

Extended in-packet Bloom filter capabilities: We have studied in-depth the design space of iBFs, including three novel extensions (1) to increase the practicality and performance of iBFs by exploiting the power of choices at hashing time to have d candidate iBFs, (2) to enable false-negative-free element deletions by encoding collision-free iBF regions, and (3) to provide a secure method for iBF constructions by coupling packet-specific information and a time-based hashing mechanism to the iBF set/check operations. As a general data structure, iBFs can be useful for networking designs that tolerate false positives and decide to move state to the packets themselves.

4. DATA CENTER NETWORKING

In parallel to the work towards a fresh Internet-scale architecture [8], we are exploring the data center (DC) networking environment as a breakthrough application of our information-oriented forwarding approach. With cloud computing transforming Internet service delivery, the optimization of the underpinning network architectures of (geo-)distributed data centers has gained increased interest. As DC communication infrastructures scale out adding more commodity components, they start to suffer from limitations comparable to the global Internet, including scalability (at L2), management complexity and the need for traffic engineering.

Although, we are early to contribute to DCN research agendas [2], we have some hints as to where the solutions may lie. One line of research is to leverage the iBF-based source-routing approach to include deletable Service IDs representing middlebox processes and enable thereby explicit (policy-based) routing via location-independent services in the data path. Another work in progress regards a direct network control approach for network management [7]. As a first application, we envision a topology system that resolves information (service/host connectivity) requests to multi-paths towards candidate servers in order to achieve oblivious routing in data centers. Validation work includes building a prototype based on OpenFlow [6] and NOX, which provide an appealing abstraction to develop novel networking applications in a realistic and high performance environment.

5. REFERENCES

[1] C. Esteve, F. Verdi, and M. Magalhaes. Towards a new generation of information-oriented internetworking architectures. In *ReArch' 08*.

[2] A. Greenberg and et al. The cost of a cloud: research problems in data center networks. *SIGCOMM CCR.*, 2009.

[3] V. Jacobson and et al. Content-centric networking: Whitepaper describing future assurable global networks. Response to DARPA RFI SN07-12, 2007.

[4] P. Jokela, A. Zahemszky, C. Esteve, S. Arianfar, and P. Nikander. LIPSIN: Line speed publish/subscribe inter-networking. In *SIGCOMM'09*.

[5] T. Koponen and et al. A data-oriented (and beyond) network architecture. In *SIGCOMM '07*, 2007.

[6] N. McKeown and et al. Openflow: enabling innovation in campus networks. *SIGCOMM CCR.*, 2008.

[7] H. Yan and et al. Tesseract: A 4d network control plane. In *NSDI'07*, 2007.

[8] A. Zahemszky, A. Csaszar, P. Nikander, and C. Esteve. Exploring the pubsub routing/forwarding space. In *ICC Workshop on the Network of the Future*, 2009.