least one of the h links on the route[3], *i.e.*, $P(B) = \left[1-(1-p)^h\right]^{TC}$.

In order to devise numerical results, Ofek *et al.* assume the TF unavailability probability density function to be a truncated Gaussian distribution and calculate the blocking probability for several standard deviation and average values of the TF load, *i.e.*, mean of the Gaussian distribution.

The work in [4] also devises the blocking probability for non-immediate forwarding in case (*i*) one TF and (*ii*) up to *TC* TFs of extra forwarding delay are introduced at each hop. The calculation is impractical in the former case (*i*) above when the number of TFs per time cycle *TC* is not very small; consequently, [4] provides some results for small *TC* values, *e.g.*, *TC*=5. Not being able to use the devised analytical equations, numerical results with larger numbers of TFs per time cycle, *e.g.*, *TC*=48, are obtained in [4] through simulation.

One major shortcoming of [4] when compared to this work is relying on the assumption that the distribution of the probability of unavailable TFs is Gaussian, which is not substantiated in any way. Also, the probability p, and consequently the blocking probability $P(B)$, are not related to usable system parameters, such as the traffic matrix describing the distribution of reservation requests through the network, the reservation request generation process by end systems at the network ingress, the holding time of the reservations, and routing on the network. As a result the devised blocking model cannot be easily deployed for network dimensioning purposes.

Further work on blocking analysis that does not rely on a specific reservation probability density function has more recently been presented in [8] for a single-hop scenario first and then extended to multi-hop scenarios. A combinatorial enumeration approach is adopted to derive the blocking probability of reservation requests through a single switch as $P(B)=C_{block}/C_{total}$, where C_{total} is the total number of possible schedules through a switch and C_{block} is the total number of schedules through a switch that would lead to blocked reservation requests. In this analysis the link reservation level is modelled in terms of number of unavailable TFs per time cycle, rather than the mean of the distribution of the probability p for a TF to be unavailable, as in [4]. However, when extended across several hops, as only sketched in [8], the combinatorial analysis of the blocking probability becomes impractical. In fact, it requires the computation of a set of conditional probabilities that grows with the number of TFs in the time cycle, which makes the solution not applicable to realistic time cycle dimensions and no numerical results are provided in [8].

To summarize, the analysis presented here follows a different approach for the blocking probability derivation compared to [4] and [8] and leads to a formulation that is actually usable. The blocking probability experienced by reservation requests is derived as function of the steady state distribution of the number of active reservations on each link it traverses calculated taking into account network routing, traffic matrix, and reservation request arrival process and duration, i.e., network and traffic quantities commonly taken into account when engineering a network and its traffic. This work also offers an improved multi-hop blocking analysis because it (*i*) does not assume the same mean link utilization on different links, (*ii*) provides computational feasible blocking models also for large numbers of TFs per time cycle, *e.g.*, 1000 TFs, and (*iii*) addresses the cases of both PF-unaware and PF-aware end-systems.

[3] The formulation ignores the propagation delay, which does not affect the result.

VI. Conclusions

This work presents in details the blocking problem stemming from the *pipeline forwarding* operating principles and improves upon previous blocking analysis for the case of *immediate forwarding* operation. The upper-bound blocking probability experienced by reservation requests issued by pipeline forwarding-aware and pipeline forwarding-unaware sources has been derived on general topology networks as function of the average utilization of the network links given the model of the reservation request arrival process at the network ingress, the model of reservation holding process, the traffic matrix describing the mean load of reservation requests from every ingress to every egress node and the network routing, that are in most practical cases known parameters useful for network dimensioning purpose.

Moreover the validity of the blocking models is verified by simulations in different network and traffic scenarios. Finally, the different analytical approaches of related works on the blocking problem are discussed showing the improvements offered by this analysis.

References

[1] J. C. R. Bennett, K. Benson, A. Charny, W. F. Courtney, J-Y Le Boudec, “Delay Jitter Bounds and Packet Scale Rate Guarantee for Expedited Forwarding,” *IEEE/ACM Transactions on Networking*, vol.10, no.4, pp. 529-540, August 2002.

[2] Cisco Systems, Inc., “The Zettabyte Era – Trends and Analysis,” May 29, 2013, [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni /VNI_Hyperconnectivity_WP.html

[3] M. Yuksel, K. K. Ramakrishnan, S. Kalyanaraman, J. D. Houle, R. Sadhvani, “Value of Supporting Class-of-Service in IP Backbones,” *Fifteenth IEEE International Workshop on Quality of Service*, Evanston, IL, pp. 109-112, June 2007.

[4] C-S. Li, Y. Ofek, A. Segall, K. Sohraby, “Pseudo-Isochronous Cell Switching in ATM Networks,” *Computer Networks and ISDN systems*, vol. 30, no. 24, pp. 2359-2372, December 1998.

[5] M. Baldi, G. Marchetto, Y. Ofek, "A Scalable Solution for Engineering Streaming Traffic in the Future