Fantastic Traffic Models and Where to Find Them – A Literature Review

Manuel Kieweg
Computer Science Department
University of Applied Sciences
Darmstadt, Germany
manuel.kieweg@h-da.de

Tim Bender
Computer Science Department
University of Applied Sciences
Darmstadt, Germany
tim.bender@h-da.de

Martin Stiemerling
Computer Science Department
University of Applied Sciences
Darmstadt, Germany
martin.stiemerling@h-da.de

***Abstract*—Computer networks, especially the Internet, have grown to be an important aspect of the global economy and society. With their importance, their complexity also grew. In order to keep up with the rising demand for bandwidth and connectivity, network infrastructure is subject to continuous change. Outages and disruptions can easily cause financial or other damage. Network simulation is a viable tool to prevent disruptions caused by resource exhaustion and ill planned changes of infrastructure. This requires validated traffic models. This paper examines publications on network traffic modelling of the last 10 years. We identify proven methodologies as well as challenges that were encountered while creating these traffic models. We evaluate which traffic types and network types were focus of recent research and how the results do hold up with the rapid technological change over the past decade. Based on our findings we provide recommendations how to tackle the challenges we identified.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

The Internet has developed into a backbone of economy and society over the last decades. First, it has emerged from research institutions to the public domain providing access to the World Wide Web. With the advent of mobile computing and cloud services, it developed into an ubiquitous ether providing services and information anytime anywhere and enabling instant communication all across the world. This constant change creates challenges for network operators and service providers. In order to properly assess how current infrastructure will hold up to projected changes in network traffic, concise models about the traffic are required to simulate expected events. Since the operational scope of different network operators or service providers is not homogenous, scoped traffic models are needed, as there is no one-size fits all. This creates a huge challenge: The evolution of network traffic happens (a) at a fast pace and (b) is unevenly distributed. Network applications, such as video streaming but not limited to, significantly changed over the last two decades, whereas the underlying network and transport protocols barely did, e.g., TCP is still predominant though QUIC is on the rise [1]. The fact that HTTP established itself as the transport protocol for web applications and can be assumed as widely being encrypted changes the way "web traffic" needs to be modelled. How and if this change affects lower layer protocols needs to be investigated. This example illustrates the importance of state-of-the-art traffic models and a well-known toolkit to derive and validate these models. This paper analyses publications on the field of traffic modelling across the last 10 years and evaluates and categorises the presented models.

Contribution: This work aims to address the issue of a missing systematisation of knowledge for network modelling by analysing the methodologies and use-cases of recent traffic models, and identifying the main properties in terms of traffic types and network types reflected by these models. We provide an overview of the identified methodologies to derive certain traffic models as well as the limitations of given popular approaches.

A. Related work

To show that literature reviews are common in the areas of traffic modelling, network traffic, and traffic types, we list some publications that have conducted literature reviews in these areas. It is noteworthy that the works found always covered a very specific domain of research. As an example [2] shows the findings about research at Google Scholar about Network Traffic Classification by using Neural Networks. In the publication, they divide the literature into two different categories. A selection of the publications describes the classification of neural networks. The other selection describes algorithms for improvement. We will employ a similiar approach with this work. We divide the literature into the categories presented in Section IV. The publication [3] presents models for video traffic. They summarise and categorise these models based on their use. Publication [4] systematises the results found, presents definitions for them, and explains the development of programming paradigms. Great emphasis is placed on the presenting of knowledge. This paper employs a similar approach.

B. Structure of this paper

The structure of this paper is outlined here. The following section shows the background of mathematical formalisms, neural networks and the scopes of network traffic modelling. Then the research questions are presented in Section III. In the methodology section, we present the literature search for the scientific publications that we assembled for this work.

Furthermore, the criteria for answering the research questions are listed and explained. Section V lays out the results of this work. Section VI contains the discussion of our findings and in Section VII we draw conclusions based on our review.

II. Background

This section describes formalisms and concepts relevant to network traffic modelling. The section is subdivided into (i) background for mathematical formalisms, (ii) neural networks and (iii) the scopes of network traffic modelling.

A. Mathematical Formalisms

The creation of network traffic models adheres to specific mathematical formalisms. These formalisms vary depending on the scope of the model and the use-case of the simulations. Both will be discussed in more depth in Section II-B. The defining properties for network traffic modelling are statistic distributions as well as fractal properties. Poisson processes and other statistic distributions can be used to model parameters like inter-arrival times and the distribution of packet sizes. These parameters are not applicable to *network traffic* as a whole. They need to be precisely tuned to model a very distinct property of traffic. Identifying these parameters and corresponding use-cases is one of the major goals of our work. Fractal properties like scaling and self-similarity are a general property of network traffic and can be used to model a baseline behaviour to analyse traffic anomalies.

Statistical Distributions: The statistical properties of network traffic have been subjects of research for more than three decades. Several statistical processes and distributions have been found suitable to describe facets of network traffic behaviour. Articles published in the early and mid 90s disagreed on the applicability of Poisson modelling to network traffic. [5], [6] This disagreement stems from the fact that Poisson processes are insufficient to describe generic patterns of network traffic but are suitable to model traffic that is caused by user-interaction like SSH or HTTP web traffic. [6], [7] Other common distributions are the Pareto distribution, the Normal-distribution and the Lognormal-distribution. [8], [9] In order to define a traffic baseline for certain applications, entropy has proven a useful tool [10].

Markov Processes: Markov models have been used to characterise and simulate network traffic. [11]–[13] They have been used to model mobile data emitted from vehicles [14].

Self-Similarity: It has been shown that network traffic does exhibit self-similar behaviour. An analysis of the MAWI Repository [15] did show that random traffic samples show self-similarity [7]. Similar observations have been made by [16], [17].

1) Neural Networks: When it comes to modelling and simulation in recent years, neural networks were a tool introduced in multiple new disciplines. A substantial amount of effort to analyse network traffic using neural networks to derive models describing network traffic properties using Probabilistic Neural Networks [2], [18] and Parallel Neural Network Classifier Architectures (PNNCAs) [19].

Traffic Types	Web Traffic	Website traffic HTTP based APIs
	Real-Time Traffic	Audio streaming Video streaming Voice over IP Instant Messaging
Network Types	Wireless Networks	WiFi networks Cellular networks (3G, 4G, 5G)
	Optical Networks	Networks using optical fibres to transmit data
	Wide Area Networks	
	Local Area Networks	

TABLE I
Traffic Types and Network Types

B. Scopes of Network Traffic Modelling

One of the largest factors that have to be considered for traffic modelling is the scope of the model itself. Which statistical properties are exposed by networking traffic heavily depends on the type of traffic that is planned to be modelled and also on the network type the model is designed for. A traffic model focusing on HTTP traffic will contain a different set of metrics and parameters than a model for video streaming.

III. Research Question

The primary purpose of this literature review is to identify approaches to network modelling in the past 10 years and identify their different facets. In order to achieve this, we assess how the existing methodologies in traffic modelling and the resulting traffic models compare to our following research questions:

RQ1: Which traffic types and which network types are reflected in the traffic model? Network traffic can be highly diverse. The exact structure of a model highly depends on the traffic type and the network type the model is created for. Therefore it is highly important to take the traffic's specifics into account. This can be achieved by addressing properties of the underlying network type and attributes of the traffic type in question.

RQ2: To what extent do the resulting models demonstrate their suitability outside of the paper's Gedankenexperiment? Our goal is to understand existing methodologies and their fit for purpose. To draw valid conclusions from a theoretical model it is necessary to derive the assumptions for the model from appropriate data. This introduces the challenge to precisely extract metrics that model the data and the need to cross-validate the postulated model with real-life data.

Traffic Type:	A publication that presents at least one but possibly different types of traffic. For example video traffic or web traffic.
Traffic Model:	Traffic models are presented in the scientific publication.
Network Type:	This category includes network types like UMTS or wired networks.

TABLE II
TAGS USED TO CATEGORISE THE REVIEWED LITERATURE

RQ3: How do the assumptions hold up to technological development? Evolution in the field of Internet technologies is rapid. This poses a potential threat of obsolescence to all theoretical assumptions based on fixed technology stacks. We want to examine which models reflect state of the art internet technologies like anything-over-HTTP and app ecosystems brought by the ubiquity of mobile devices. We also want to determine if the networking stacks the models are based on are still a good fit for valid assumptions.

IV. METHODOLOGY

This section describes how we put together the literature for our review. We also introduce the criteria to answer the research questions here.

A. Literature Research

We conducted structured literature research to generate an overview of existing and researched traffic types, traffic models and network types. For this research, we designed a search string based on keywords from the scientific publications we have taken for background of our work. We searched through the digital libraries ACM and IEEE with the following filter:

```
years (2010−2020) and publication types (
    research−article)

(traffic AND (model OR simulation)) AND (
    network OR "computer network" OR
    internet OR "packet based networks")
    AND (traffic AND (type OR prediction
    OR "generating system" OR engineering
    OR distribution)) AND ((network OR www
    ) AND traffic) AND ("fractal modelling
    " OR "self similarity" OR "long range
    dependence") AND ("inter arrival time"
     OR "packet size") AND (simulation OR
    "synthetic workload" OR "queueing
    system")
```

Listing 1. Search string for collecting relevant publications

This search yielded 43 publications. The resulting publications were sighted by two people. During the sighting, the publications were tagged for the following tags that we chose. The tags are defined as follows:

The tags are related to the evaluation criteria which are presented in the following subsection to answer the research questions. We discuss our tagging after reading the publications to define which publications are relevant for reaching the goal of our work. We gathered a base of 19 publications that were used for a snowball-search and which are used as our dataset for this work.

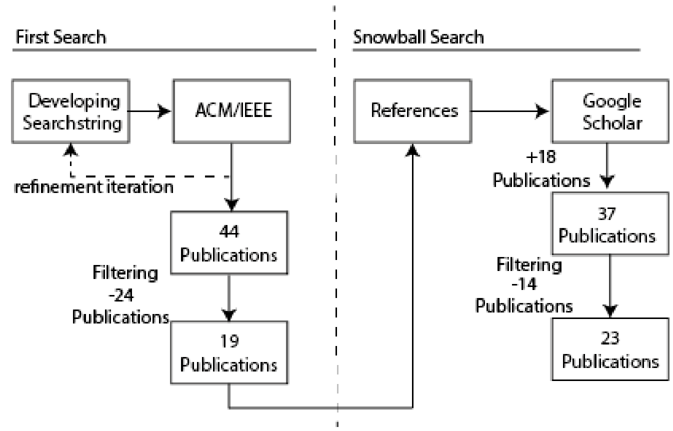

Fig. 1. Process of the Literature Search.

The snowball-search happened for one round as a reverse search. In summary, this means we have 33 publications after one round of reverse search from which information is drawn. Out of this we select four publications after discussing these. This leads us to a final dataset of 23 publications for this work.

The following subsection explains the criteria which we use to answer our research questions.

B. Evaluation Criteria

To assess the traffic models found during the data collection we established a catalogue of 6 criteria. Using these criteria we were able to answer the research questions [20] in Section III. We analysed the focus of a given publication in terms of the modelled network- and traffic types as well as the applicability when it comes to the transfer out of the academic field and technological recency.

To assess which traffic types and network types are reflected (*RQ1*), we created the following two criteria:

- **CR1: Traffic type.** This criterion is used to examine which publications model a specific kind of network traffic (table II-B). The kind of traffic and the goal of the traffic analysis heavily influence the parameters of the traffic model.
- **CR2: Network type.** With this criterion we assess if the characteristics of a certain network type are reflected in the model (table II-B). This is useful to determine if a given model is applicable.

To examine how traffic models hold up outside of their controlled setting (*RQ2*), we postulated the following three criteria:

- **CR3: Use of real world data sets.** This criterion allows us to assess the quality of the traffic model outside of its research constraints. The use of synthesized data sets is common practice to design a traffic model. However, if the datasets a traffic model is based on are fully synthesised the resulting model may not be valid outside of its experimental parameters.

- **CR4: Domain specificity.** Network traffic is highly diverse. A model claiming to be universal for all network traffic might only be valid for the dataset it was derived from, but does not hold up to other data sets from a similar domain. On the other hand, an overly narrow model might be very useful to model one specific aspect of traffic. For "*the IAT of TCP packets with a set RST flag on Monday afternoons seen at a specific RAN*" might be accurate to model but won't have much significance. Finding a modelling domain allows researchers to derive sound models that are applicable outside of the original experiment's scope.
- **CR5: Validation.** This criterion aims to assess the level of maturity of a model by verifying whether it has been tested and validated against real examples of the traffic it is supposed to simulate.

The evolution of technology is very fast-paced. To see if assumptions made in the examined publications still hold up to today's technology (*RQ3*) we defined the following criterion:

- **CR6: Technological recency.** Using this criterion we can assess if a traffic model reflects the current state of the art technologies. To account for the different categories of network traffic we implemented two sub-criteria:
  - **CR6.1: Web technologies.** To examine how a given web traffic model holds up, we review if the advent of modern web technologies and architectures has been considered. The key identifiers we use are the consideration of **HTTP based APIs**, **Content Delivery Networks (CDNs)** and **HTTPS** in the model. These verification parameters allow us to determine if *CR6* is fulfilled.
  - **CR6.2: RAN Technology.** This criterion allows us to evaluate how a model that is designed to model mobile networks reflects changes in the technological reality. Our key identifier here is the acknowledgement of different RAN generations (i.e. 3G, 4G, 5G) if a publication uses specific RAN properties in its model.

V. Results

After identifying the publications as described earlier, we extracted a set of traffic type definitions and existing traffic models. Whereas a number of the traffic types and models are universally applicable, some of them are only relevant for certain types of networks.

In this section, we will establish different methodologies of network traffic modelling and evaluate the publications we analysed in our literature review.

1) Overview: A variety of approaches have been suggested to model network traffic, each of them focusing on different aspects. We identified 3 main methodologies during our literature study: *self-similarity and Markovian processes*, *statistical distributions* and *Machine Learning techniques*. The histogram in figure 2 shows the distribution of the different methodologies over the year of their publication. An overview of the different methodologies is presented in section II-A. Next, we will discuss how we matched the reviewed publications and the methodologies.

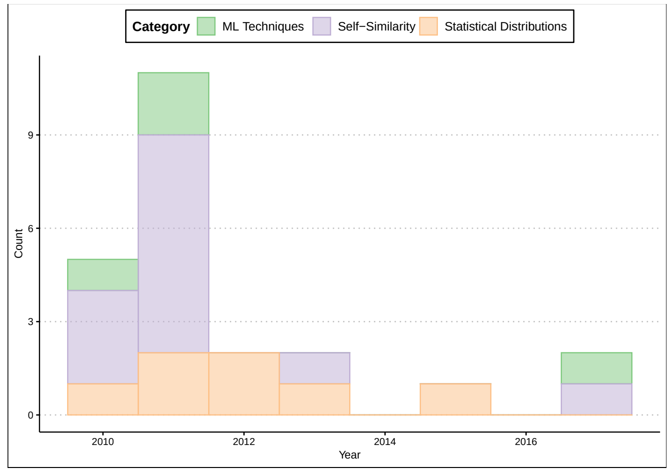

Fig. 2. Methods used for traffic modelling per year.

Self-Similarity and Markovian Processes: This category contains publications that primarily used the properties of self-similarity or entropy for their traffic models [7], [10], [11], [16], [17], [21]–[27]. These models usually address a very high-level abstraction of traffic. They are commonly used to create application-agnostic models of networking traffic.

Statistical Distributions: Publications in this category rely on statistical distributions to derive their models [8], [9], [28]–[32]. This modelling approach is usually very specific to a certain type of traffic. These models can be used to model a distinct application.

Machine Learning Techniques: This category encompasses works that use machine-learning techniques to model network traffic [14], [18], [19], [33]. They are very application-specific and used for both traffic simulation and security purposes.

A. Evaluation

Our evaluation is based on the criteria that were defined in Section IV-B to answer the research questions asked in section III. An overview of these results can be found in Table III, where we also visualise the fulfilment of a given criterion. We defined several levels of coverage to assess the fulfilment of criteria by a given publication. *Full coverage* (●) indicates that the authors conducted their experiments with this criterion in mind, *Partial coverage* ($\star$) that some aspects of a criterion are addressed, and *No coverage* (o) that the criterion is not reflected in the paper at all. Hereafter, we use *coverage* for both full and partial coverage unless explicitly stated if necessary.

CR1: Traffic type. We evaluated to which extent a publication addressed a certain traffic type. *Full coverage* was used if the authors established their model to simulate network traffic of a specific type. We used *partial coverage* if a publication mentioned one or multiple specific traffic types, but did not base the model on their specific properties, otherwise, the paper was marked *not* covering the criterion. We found 11 studies to fully incorporate this criterion, 7 studies to partially address it and 5 studies to not address it at all. Of the 18 publications covering this criterion 5 directed their work to web traffic and 8 focused on real-time media transmissions (A/V streaming and VoIP/Messaging).

CR2: Network type. We analysed the network types mentioned in each study. Again we considered *full coverage* if a model was designed to address specific properties of a given network type, *partial coverage* if one or several network types were acknowledged, but their specifics did not influence the design of the model and *no coverage* if neither was the case. The majority of traffic models (11) did not consider any attributes of the underlying network type. Of the 13 models that fulfilled this criterion 7 focused on wireless (WiFi and cellular) networks. 2 models addressed specifics of optical networks.

CR3: Use of real-world datasets. To assess if the traffic model is practically applicable we analysed the origin of the research datasets. If the model was based on a large corpus of traffic specific to the paper's domain *full coverage* was assumed. If the underlying data was captured by the authors for their specific experiment the criterion is *partially* fulfilled. Publications that based their models on fully synthesised traffic did *not* cover this criterion. It turned out that the overwhelming majority of publications did use either purely synthetic datasets or traces created by the authors. Only 4 publications did fully meet this criterion [7], [21], [22], [29].

CR4: Domain specificity. This criterion was found to be fulfilled by almost all publications. 13 were fully domain-specific, 6 partially. The area that received the most attention were mobile networks investigated from different angles. Especially streaming and loss prediction, but also user mobility, were prominent areas of interest. Given the rapid evolution of mobile networks and the large set of variables - especially compared to fixed networks - this is not surprising.

CR5: Validation. We found that most publications did not properly validate their data. Only 9 out of 23 papers presented sound data validation. A solid validation of data was assumed if the models were validated against real-world web traffic. If the traffic model was validated against synthetic or self-captured traffic we considered it to partially fulfil this criterion, which 8 publications did. Note that no claims on validation completeness can be inferred from the validation type. Validations might have intricate links to a particular context (or a range of contexts) regardless of their validation type. We want to point out that this criterion does not assess the validity of the individual contribution but its applicability outside of the experiment's scope. Due to redacted parameters in the work of Lu et al. [21], could not be assessed for this criterion.

CR6: Technological recency. Our analysis shows a very low coverage of this criterion, indicating a lack of research efforts in the last years. This is backed by the histogram in 2, which shows a lack of substantial research in the field of traffic modelling after 2011. Only 3 publications have taken either modern web technologies (*CR6.1*) or state-of-the-art RAN technologies (*CR6.2*) into account. We would like to point out, that some papers were not assessed for this criterion, since their models were based on physical principles that are independent of technological evolution.

VI. Discussion

Our work shows that the examined research landscape is divided into mainly two areas of interest, modelling of application-aware traffic and modelling of application-agnostic traffic. The third field of research is traffic models generation using Machine Learning techniques. In this section, we present our findings and answers to the research questions asked in Section III. We were able to identify multiple *research clusters (RQ1)* and we will give an overview of the *current state of network traffic modelling* (*RQ2* and *RQ3*).

A. Research Clusters in Traffic Modelling

We have identified several salient research topics. To answer *RQ1* we looked at the publications that fulfilled *CR1* and *CR2*. As we have shown in Section V-A the literature we reviewed heavily focused on the categories shown in Table II-B. It turned out that these categories often are mutually exclusive. This is not surprising since the traffic properties of these categories prohibit a unified modelling approach. This is backed by Figure 3 It clearly shows that traffic models focussing on the traffic type (*CR1*) are derived using a different methodology than models that are directing their work to network types (*CR2*).

Another emerging area of interest is the field of Machine-Learning (ML) based traffic models. Even though the number of ML-related papers in our literature review was rather small, the results look promising for very domain-specific traffic models.

B. Current State of Network Traffic Modelling

One key contribution of our work is the evaluation of the goodness-of-fit of network traffic models using the research questions specified in Section III. To be useful, traffic models need to be suitable to simulate network traffic that is valid outside of the initial experiments' scope. Only 7 publications used datasets containing real traffic from the authors' research domain. All validated models were only validated on the given input data, but none of them were validated against unseen real-world web traffic. Even though not all publications in our study allowed for real-world data validation due to their scope or other certain characteristics, this nevertheless shows the need for robust traffic models that are validated against arbitrary traffic.

	Statistical Properties							MLT				Self Similarity											
Criteria	[9]	[8]	[28]	[29]	[30]	[31]	[32]	[18]	[33]	[14]	[19]	[21]	[16]	[17]	[25]	[26]	[24]	[7]	[11]	[27]	[22]	[10]	[23]
Traffic Type	●	●	⋆	⋆	●	⋆	●	⋆	o	⋆	●	●	●	●	o	o	⋆	o	●	●	⋆	●	o
Network Type	●	o	●	o	o	●	o	o	o	●	o	●	⋆	●	●	●	o	o	●	⋆	o	o	●
Real Word Data Sets	⋆	⋆	⋆	●	⋆	⋆	⋆	⋆	⋆	o	⋆	●	⋆	⋆	o	o	o	●	o	o	●	⋆	o
Domain Specificity	●	⋆	●	o	⋆	⋆	●	●	o	●	⋆	●	●	●	o	o	●	⋆	●	●	●	⋆	●
Validation	⋆	o	⋆	●	o	o	o	⋆	o	⋆	●	NA	●	o	●	o	o	●	●	⋆	●	●	●
Technological Recency	⋆	⋆	●	o	⋆	⋆	o	⋆	o	⋆	⋆	⋆	o	o	NA	NA	⋆	●	⋆	o	●	o	NA

TABLE III

EVALUATION OF METHODOLOGIES FOR THE GENERATION OF NETWORK TRAFFIC MODELS; *Full coverage* (●), *Partial coverage* (⋆), *No coverage* (o)

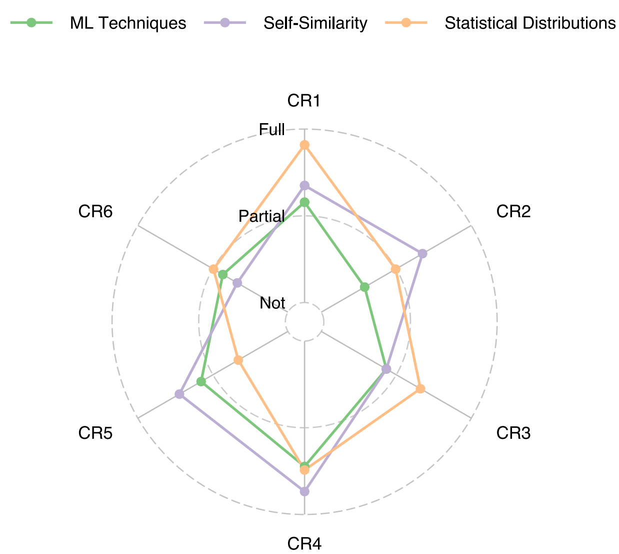

Fig. 3. Average coverage of criteria by method.

While this can partially be explained by the difficulty to obtain representative real-world traffic traces, it still taints the expressiveness of these traffic models. To solve this issue, we need to rethink the creation of traffic traces. Given today's volume of internet traffic, capturing traces and storing them for later analysis is not feasible. Large traffic datasets like CAIDA and MAWI [34], [35] usually retain the data of one hour for a certain weekday, which is very sparse data from a temporal point of view, covering only 0.006% of a week's time.

Another shortcoming we identified during our research is the staggering lack of technological recency when it comes to traffic models. Only 3 publications fully met the criterion of technological recency. Especially in the field of cellular networks, there were no publications mentioning 4G or 5G networks. While this is understandable for 5G, the first deployments happened in 2018, the complete lack of 4G traffic modelling is worrying given the start of 4G deployment in 2009. But the realms of web traffic modelling also showed a tremendous lack of technological recency. All publications that create web traffic models focus on plain HTTP/1.x web traffic only. No model did take HTTP/2 into account. This does not reflect the nature of modern web applications. Another important aspect of HTTP traffic are HTTP based APIs. RESTful web services are the predominant technology for mobile apps as well as many desktop applications. Their impact on HTTP related web traffic however is not reflected in any of the publications in our literature study. Similar issues arise in the field of real-time traffic. While Horváth et al. [9] models video streaming traffic via HTTP, no other traffic model in this field reflects technological realities.

Recommendation for research: This issue can be mitigated by developing methods that leverage state of the art data analysis tools to derive models directly from live traffic flows. We would recommended to agree on set of analysis tools within the community. At best, a well-defined methodology on how to collect, pre-process, store, and document live traffic captures and their models would be established. This would enable researchers to create and validate models in fast iterations with representative data without the need to store capture and store large amounts of data. This also addresses the lack of technological recency most papers did show. Having a validated and holistic tool-chain to derive traffic models on the fly potentially leads to an increase in traffic model generation. It also reduces the time and effort required to derive sound traffic models. This can be leveraged to overcome this deficiency.

VII. CONCLUSION

We conducted a comprehensive literature review based on a complex string-based search of the ACM Digital Library as well as the IEEE library and a following snow-ball search using Google Scholar, as described in Section IV, and evaluated the publications we found based on our research questions (Section III). We found that the current landscape of network traffic modelling shows two problems. Most traffic models lack applicability for real-life traffic simulations. They show a lack of validation and fail to keep up with technological change. We also discovered that the reason for this might be the logistical problems to capture and store a representative amount of real-life traffic. In our discussion, we proposed further research angles to address these shortcomings.

A. Future Work

We identified four areas of potential future research.

As discussed in Section VI past web traffic models strongly focused on plain HTTP/1 and HTTP/1.1 traffic. Since the technology stack of the World Wide Web has evolved an investigation of the properties of HTTP/2 and HTTP/3 is a very pressing research topic. Another HTTP related field of research is the modelling of traffic caused by HTTP-based APIs. It is widely accepted that the Poisson property that HTTP traffic exposes is caused by human interaction(cf. Section II-A). Given the prevalence of HTTP-based APIs a model of machine-generated traffic is needed.

When it comes to mobile networks we identified lack of modern RAN technologies in the publications we did review. Since 4G and especially 5G promise an evolution in bandwidth and latency their properties need to be taken into account for reliable traffic simulation.

In order to allow researchers to create proven traffic models a framework of best practices and sound methodologies is needed. Such a framework can give researchers guidance on how to collect data and validate their traffic models in a way that makes them reproducible and comparable. In addition a standardised open source toolchain could be part of said framework thus mitigating the issue of home brew solutions that make the reproduction of results harder or even impossible.

The third research opportunity we discovered is the generation of traffic models from live network traffic. This is very promising to solve the issue of models being based on sparse or skewed datasets, and is likely to improve the overall quantity and quality of traffic models.

REFERENCES

[1] M. Trevisan, D. Giordano, I. Drago, M. M. Munafò, and M. Mellia, "Five years at the edge: Watching internet from the isp network," *IEEE/ACM Transactions on Networking*, vol. 28, no. 2, pp. 561–574, 2020, publisher: IEEE.

[2] P. Bin and L. Ru, "Literature Review of Network Traffic Classification Using Neural Networks," in *Recent Progress in Data Engineering and Internet Technology*, F. L. Gaol, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 223–231.

[3] S. Tanwir and H. Perros, "A Survey of VBR Video Traffic Models," *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 1778–1802, 2013.

[4] R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," in *2010 IEEE Symposium on Security and Privacy*. Oakland, CA, USA: IEEE, 2010, pp. 305–316. [Online]. Available: http://ieeexplore.ieee.org/document/5504793/

[5] V. Paxson, "Empirically derived analytic models of wide-area TCP connections," *IEEE/ACM Transactions on Networking*, vol. 2, no. 4, pp. 316–336, Aug. 1994.

[6] V. Paxson and S. Floyd, "Wide-area traffic: the failure of Poisson modeling," *ACM SIGCOMM Computer Communication Review*, vol. 24, no. 4, pp. 257–268, Oct. 1994. [Online]. Available: https://doi.org/10.1145/190809.190338

[7] R. Fontugne, P. Abry, K. Fukuda, D. Veitch, K. Cho, P. Borgnat, and H. Wendt, "Scaling in Internet Traffic: A 14 Year and 3 Day Longitudinal Study, With Multiscale Analyses and Random Projections," *IEEE/ACM Transactions on Networking*, vol. 25, no. 4, pp. 2152–2165, Aug. 2017. [Online]. Available: https://doi.org/10.1109/TNET.2017.2675450

[8] R. Pries, Z. Magyari, and P. Tran-Gia, "An HTTP web traffic model based on the top one million visited web pages," in *Proceedings of the 8th Euro-NF Conference on Next Generation Internet NGI 2012*. Karlskrona, Sweden: IEEE, Jun. 2012, pp. 133–139. [Online]. Available: http://ieeexplore.ieee.org/document/6252145/

[9] G. Horvath and P. Fazekas, "Modelling of YouTube Traffic in High Speed Mobile Networks," in *Proceedings of European Wireless 2015; 21th European Wireless Conference*, May 2015, pp. 1–6.

[10] V. Petkov, R. Rajagopal, and K. Obraczka, "Characterizing per-application network traffic using entropy," *ACM Transactions on Modeling and Computer Simulation*, vol. 23, no. 2, pp. 14:1–14:25, May 2013. [Online]. Available: https://doi.org/10.1145/2457459.2457463

[11] P. Wang and I. F. Akyildiz, "Spatial correlation and mobility-aware traffic modeling for wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 19, no. 6, pp. 1860–1873, Dec. 2011. [Online]. Available: https://doi.org/10.1109/TNET.2011.2162340

[12] Shun-Zheng Yu and H. Kobayashi, "An efficient forward-backward algorithm for an explicit-duration hidden Markov model," *IEEE Signal Processing Letters*, vol. 10, no. 1, pp. 11–14, Jan. 2003.

[13] L. Muscariello, M. Meillia, M. Meo, M. A. Marsan, and R. L. Cigno, "An MMPP-based hierarchical model of Internet traffic," in *2004 IEEE International Conference on Communications (IEEE Cat. No.04CH37577)*, vol. 4, Jun. 2004, pp. 2143–2147 Vol.4.

[14] Y. Li, X. Hao, H. Zheng, X. Su, J. Riekki, C. Sun, H. Wei, H. Wang, and L. Han, "A two-level hidden Markov model for characterizing data traffic from vehicles," in *Proceedings of the Seventh International Conference on the Internet of Things*, ser. IoT '17. New York, NY, USA: Association for Computing Machinery, Oct. 2017, pp. 1–8. [Online]. Available: https://doi.org/10.1145/3131542.3131556

[15] C. S. L. Sony and K. Cho, "Traffic data repository at the WIDE project," in *Proceedings of USENIX 2000 Annual Technical Conference: FREENIX Track*, 2000, pp. 263–270.

[16] N. M. Markovich, "Modeling of dependence in a peer-to-peer video application," in *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference*, ser. IWCMC '10. New York, NY, USA: Association for Computing Machinery, Jun. 2010, pp. 316–320. [Online]. Available: https://doi.org/10.1145/1815396.1815470

[17] N. M. Markovich and U. R. Krieger, "Analyzing measurements from data with underlying dependences and heavy-tailed distributions," in *Proceedings of the 2nd ACM/SPEC International Conference on Performance engineering*, ser. ICPE '11. New York, NY, USA: Association for Computing Machinery, Mar. 2011, pp. 425–436. [Online]. Available: https://doi.org/10.1145/1958746.1958811

[18] R. Sun, Bo Yang, L. Peng, Z. Chen, L. Zhang, and S. Jing, "Traffic classification using probabilistic neural networks," in *2010 Sixth International Conference on Natural Computation*, vol. 4, Aug. 2010, pp. 1914–1919.

[19] B. Mathewos, M. Carvalho, and F. Ham, "Network traffic classification using a parallel neural network classifier architecture," in *Proceedings of the Seventh Annual Workshop on Cyber Security and Information Intelligence Research*, ser. CSIIRW '11. New York, NY, USA: Association for Computing Machinery, Oct. 2011, p. 1. [Online]. Available: https://doi.org/10.1145/2179298.2179334

[20] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and Software Technology*, vol. 64, pp. 1–18, 2015.

[21] H. Lu, C. Vu, and X. Gou, "Analysis of traffic model and self-similarity for QQ in 3G mobile networks," *IET Conference Proceedings*, pp. 131–135, Jan. 2011. [Online]. Available: https://digital-library.theiet.org/content/conferences/10.1049/cp.2011.1442

[22] T. Ma, Y. El-khatib, M. Mackay, and C. Edwards, "Characterising a grid site's traffic," in *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing*, ser. HPDC '10. New York, NY, USA: Association for Computing Machinery, Jun. 2010, pp. 707–716. [Online]. Available: https://doi.org/10.1145/1851476.1851581

[23] S. Askar, G. Zervas, D. K. Hunter, and D. Simeonidou, "Evaluation of Classified Cloning Scheme with self-similar traffic," in *2011 3rd Computer Science and Electronic Engineering Conference (CEEC)*, Jul. 2011, pp. 23–28.

[24] J. Lee, H. Hwang, K. Park, C.-G. Lee, and S. Lim, "Delay-bandwidth normalized service sharing with service rate guarantees," in *Proceedings of the 2011 ACM Symposium on Research in Applied Computation*, ser. RACS '11. New York, NY, USA: Association for Computing Machinery, Nov. 2011, pp. 118–123. [Online]. Available: https://doi.org/10.1145/2103380.2103406

[25] S. Ghani, "The impact of self similar traffic on wireless LAN," in *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference*, ser. IWCMC '10. New York, NY, USA: Association for Computing Machinery, Jun. 2010, pp. 52–56. [Online]. Available: https://doi.org/10.1145/1815396.1815409

[26] S. Ghani and F. Iradat, "Loss Probability in Networks with Pareto Distributed Traffic," in *Modelling and Simulation 2011 Second International Conference on Intelligent Systems*, Jan. 2011, pp. 355–360.

[27] K. Jain, A. Roy-Chowdhury, K. K. Somasundaram, B. Wang, and J. S. Baras, "Studying real-time traffic in multi-hop networks using the EMANE emulator: capabilities and limitations," in *Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques*, ser. SIMUTools '11. Brussels, BEL: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Mar. 2011, pp. 93–95.

[28] A. Vishwanath, V. Sivaraman, and G. N. Rouskas, "Anomalous loss performance for mixed real-time and TCP traffic in routers with very small buffers," *IEEE/ACM Transactions on Networking*, vol. 19, no. 4, pp. 933–946, Aug. 2011. [Online]. Available: https://doi.org/10.1109/TNET.2010.2091721

[29] F. Geyer, S. Schneele, and G. Carle, "RENETO, a realistic network traffic generator for OMNeT++/INET," in *Proceedings of the 6th International ICST Conference on Simulation Tools and Techniques*, ser. SimuTools '13. Brussels, BEL: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Mar. 2013, pp. 73–81.

[30] J. V. P. Gomes, P. R. M. Inácio, B. Lakic, M. M. Freire, H. J. A. Da Silva, and P. P. Monteiro, "Source traffic analysis," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 6, no. 3, pp. 21:1–21:23, Aug. 2010. [Online]. Available: https://doi.org/10.1145/1823746.1823755

[31] R. Goleva, D. Atamian, S. Mirtchev, D. Dimitrova, and L. Grigorova, "Traffic sources measurement and analysis in UMTS," in *Proceedings of the 1st ACM workshop on High performance mobile opportunistic systems*, ser. HP-MOSys '12. New York, NY, USA: Association for Computing Machinery, Oct. 2012, pp. 29–32. [Online]. Available: https://doi.org/10.1145/2386980.2386987

[32] J. Idziorek, M. Tannian, and D. Jacobson, "Modeling web usage profiles of cloud services for utility cost analysis," in *Proceedings of the Winter Simulation Conference*, ser. WSC '11. Phoenix, Arizona: Winter Simulation Conference, Dec. 2011, pp. 3323–3334.

[33] J. Domanska, A. Domanski, and T. Czachorski, "Internet Traffic Source Based on Hidden Markov Model," in *Smart Spaces and Next Generation Wired/Wireless Networking*, S. Balandin, Y. Koucheryavy, and H. Hu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, vol. 6869, pp. 395–404. [Online]. Available: http://link.springer.com/10.1007/978-3-642-22875-936

[34] The CAIDA UCSC anonymized internet traces dataset (april 2008 - january 2019). [Online]. Available: https://www.caida.org/catalog/datasets/

[35] MAWI working group traffic archive. [Online]. Available: https://mawi.wide.ad.jp/mawi/