


# Towards the Generation of Virtualized Network Traffic According to Modern Data Centers

**Daniel Spiekermann**

(University of Applied Science and Arts, Dortmund, Germany)

 <https://orcid.org/0000-0003-4762-6062>, [daniel.spiekermann@fh-dortmund.de](mailto:daniel.spiekermann@fh-dortmund.de)

**Abstract:** The evolution of modern data centers from traditional hardware-based infrastructures to highly virtualized environments has introduced new complexities in network traffic analysis. Virtual networks, characterized by dynamic changes in both overlay and underlay architectures, necessitate sophisticated methods for accurate anomaly detection and network analysis. This paper investigates the real-world behaviour of network traffic within virtualized environments, identifying the key factors that impact packet dynamics, including VM operations, multi-tenancy, user customization, and hardware adjustments. By defining the frequency and nature of these events, this research provides further details for the creation of accurate packet generation tools. These tools must simulate the dynamic characteristics of virtual networks, enabling reliable and realistic synthetic data generation. The research explains the need for enhanced packet generation methodologies to provide valid training and testing data for digital investigations, anomaly detection, and network simulations. As a result, this paper emphasizes the importance of developing accurate synthetic network traffic that mirrors real-world conditions and provides a valid basement for traffic analysis in virtual networks.

**Keywords:** virtual networks, anomaly detection, data syntheses, packet generation

**Categories:** H.3.4, H.4.3, I.6.3

**DOI:** 10.3897/jucs.140463

## 1 Introduction

The analysis of packets in modern networks aims to identify attacks, anomalies and unwanted behaviour inside the network, help to troubleshoot the environment and are a valuable resource in digital investigations of suspicious events in the environment. A common way to perform such analysis is a process of capturing, recording and analysing the network packets. Based on this information, anomalies can be detected in the network traffic and these possible attacks can be averted beforehand [Bhuyan et al., 2013].

The evolution of nowadays networks to highly virtualized environments with virtual machines (VM) and virtual networks has not changed the need for this analysis, but heavily impact the well-known techniques. The virtual network provides a resilient basement for an isolation of the different VMs of the customers as well as interconnecting the systems belonging to the same customer. In addition to this, the increasing performance of modern networks hampers the mostly manual analysis, so modern implementations use machine learning algorithms to filter or classify network packets to detect anomalies or outliers. The training of these algorithms is based on the captured network data of the productive network. Although such trainings and detections are well researched in traditional networks, these attempts fail in highly dynamic environments [Spiekermann and Keller, 2020]. Hence, successful provision of usable algorithms and models for anomaly detection in virtual networks depends on applicable and valid network data, which are complex to collect. Because of this, data generation techniques are used to create accurate packet capture files by synthesizing the network data. This paper discusses

the creation of virtualized network traffic based on real-world behaviour of modern networks and addresses the gap by proposing a requirements-driven model for synthetic packet generation that explicitly incorporates the stochastic and dynamic characteristics of virtualized environments. By capturing frequent and unpredictable changes like VM migrations, QoS adaptations, and overlay-underlay interactions, the proposed framework enables the generation of traffic traces that better reflect real-world conditions. This is essential not only for anomaly detection and digital forensics but might also serve as a basement for evaluating the behaviour of novel resource allocation mechanisms under realistic traffic loads.

The main research contribution is the definition of requirements for network generators, in order to create accurate and valid packet captures. This encompasses

- the analysis of incidences in virtual networks, that effect the transferred network packets,
- the definition of frequencies for these events,
- a derivation of necessary requirements for traffic generation.

The remainder is structured as follows. Section 2 lists relevant research in the fields of virtual networks, anomaly detection, and traffic generation. Section 3 provides information on virtual networks and artificial generation of network traffic. In Section 4, the critical changes in virtual environments are discussed and classified. Section 5 discusses the results of the characterization process related to the need for proper network generation workflows. 6 concludes the paper and gives an outlook for future research.

## 2 Related Work

Virtual environments like SDN, Docker [Hausenblas, 2018] or Kubernetes [Botez et al., 2021, Spiekermann and Keller, 2021] are heavily based on encapsulating protocols [Spiekermann et al., 2017]. These protocols are able to tunnel given protocol information to transfer these information without any further interaction from a given point in a network to its intended destination. This is done for security reasons like IPsec and Encapsulating Security Payload (ESP) [Freed and Borenstein, 1996], virtual private networks [Worster et al., 2005] or virtual environments [Chowdhury and Boutaba, 2010].

The most relevant analysis technique is the use of network packet captures freezing the transferred network traffic. The capture process in virtual environments is complex and faced with different challenges [Spiekermann and Eggendorfer, 2016, Fernando, 2021]. For a real world application of network traffic, partly captured network traffic related to single protocols or applications is not useful, so datasets with a full packet capture are *CICDDoS2019* [Sharafaldin et al., 2019], *CSE-CIC-IDS 2018* [Sharafaldin et al., 2018] and *INSecS-DCSI* [Rajasinghe et al., 2018] more applicable, but to the author's knowledge there is no publicly available data-set with different virtual protocols. [Mohajer et al., 2024] identifies traffic fluctuations and resource heterogeneity as key challenges in modern networks, emphasizing that static and inflexible configurations often fail to address these dynamic conditions effectively. The framework proposed in [Zhou and Mohajer, 2024] addresses dynamic, traffic-aware resource allocation in UAV-assisted mobile edge networks, which aligns with the goal of this paper of modelling realistic traffic patterns under adaptive and virtualized network conditions.

Besides capturing live packet traces or using captured network data, network packet traces can be generated synthetically. The generation of network packets is defined as the result of time-stamped series of packets arriving and departing from particular network interfaces with realistic values [Vishwanath and Vahdat, 2006]. This process is performed in various disciplines like IT-security [Ghazanfar et al., 2020], network troubleshooting [Li et al., 2019], education and training [Padman and Memon, 2002, Son et al., 2012], testing [Voyiatzis et al., 2015] and network forensic investigation [Corey et al., 2002].

The process of generating data is not limited to network packets, [Göbel et al., 2022] describes *ForTrace* as a framework to create realistic forensic training data and [Voigt et al., 2024] present *Re-imagén* to create scenario-based forensic datasets. The generation of own packet captures is done for various reasons like testing of new network protocols, fuzzing, IoT design [Al-Hadhrami and Hussain, 2020] or security implementations [Belenko et al., 2018].

The synthetic crafting of network packets is connected to efficient and optimized strategies for packet analysis, transformation or generation. Recent research has made significant progress in dynamic resource allocation for next-generation communication systems, particularly in satellite-assisted and mobile edge networks [Yang and Mohajer, 2024]. This paper, even focusing on a different network model, details the need to manage dynamic, real-world network behavior—including fluctuating traffic loads and resource constraints—which underscores the relevance of generating realistic synthetic traffic for evaluating adaptive communication frameworks.

### 3 Background

#### 3.1 Virtual networks

The evolution of data centers is identified by a significant shift from traditional hardware-based infrastructures to software-defined, highly virtualized environments. To achieve this, modern networks are separated in an overlay- and an underlay network as shown in Figure 1. The underlay network is the hardware-based part of the infrastructure, whereas the overlay networks present by virtual networks. These virtual networks now play a central role in the architecture of modern data centers, enabling flexible, scalable, and efficient management of network resources.

To achieve the flexibility required in virtualized environments, several network virtualization protocols have been developed. These protocols provide the foundation for creating overlay networks that abstract the underlying physical infrastructure, enabling isolated and secure communication between virtual resources.

- VXLAN: XLAN (Virtual Extensible LAN): VXLAN is a widely adopted network virtualization protocol that extends Layer 2 networks over Layer 3 using UDP encapsulation. By incorporating a 24-bit VXLAN Network Identifier (VNI), VXLAN enables the creation of up to 16 million isolated Layer 2 segments, overcoming the limitations of traditional VLANs. The protocol format of VXLAN is shown in Figure 2.
- GENEVE (Generic Network Virtualization Encapsulation): is a protocol designed to unify the strengths of existing encapsulation technologies like VXLAN and NVGRE. It provides a flexible framework that supports customizable metadata, making it suitable for diverse use cases in virtualized environments. The protocol format of GENEVE is shown in Figure 3.

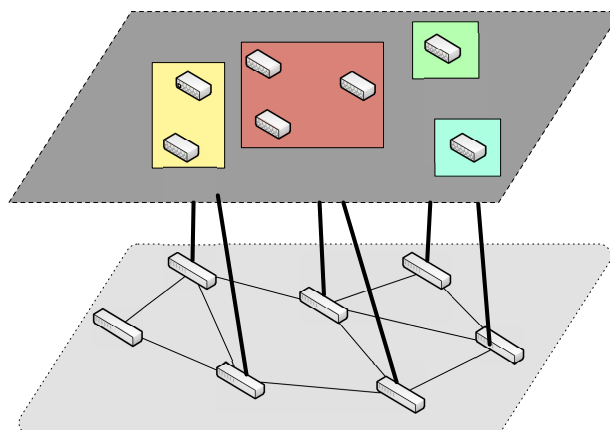


Figure 1: Overlay- and underlay network

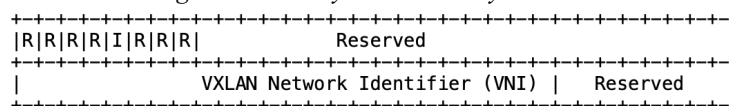


Figure 2: VXLAN protocol format

- NVGRE (Network Virtualization using Generic Routing Encapsulation): leverages the Generic Routing Encapsulation (GRE) protocol to encapsulate Layer 2 traffic over an IP network, using a 24-bit Tenant Network Identifier (TNI) for isolation. Although NVGRE has been less widely adopted than VXLAN, it remains an option for certain data centers that require specific GRE-based encapsulation capabilities. The protocol format of NVGRE is shown in Figure 4.

Tunnel endpoints (TE) like VXLAN Tunnel Endpoint (VTEPs) or NVGRE Endpoint are critical components in virtual environments that facilitate the communication between virtual and physical networks. They serve as the entry and exit points for the encapsulated traffic. Typically, they encapsulate Ethernet frames into VXLAN or NVGRE packets for transmission across an IP underlay network and decapsulating them upon arrival. Each tunnel endpoint is responsible for managing specific tunnels, mapping traditional MAC addresses to IP addresses using VNIs or TNIs. Figure 5 shows the encapsulation of a network packet.

In modern data centers, the leaf-spine architecture as shown in Fig. 6 has become the preferred standard due to its advantages over the traditional three-tier architecture. Unlike the hierarchical three-tier architecture, leaf-spine provides a flat and scalable network structure that minimizes latency and ensures consistent bandwidth across all connections. This architecture supports the high east-west traffic patterns typical in virtualized environments and cloud-based applications, making it ideal for handling the demands of modern data centers. Alongside with leaf-spine architecture, implementations like EVPN (Ethernet Virtual Private Network) are getting relevant [Singh et al., 2017]. EVPN is a modern control plane technology used in conjunction with protocols like VXLAN to enhance virtual network environments. By providing a scalable and robust

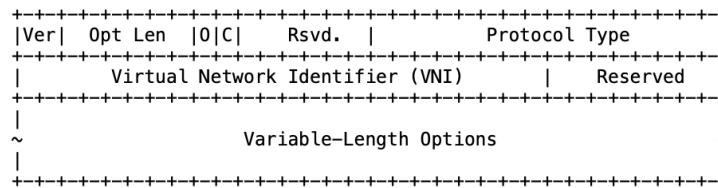


Figure 3: GENEVE protocol format

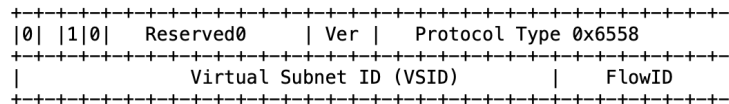


Figure 4: NVGRE protocol format

Layer 2 and Layer 3 VPN solution, EVPN enables the distribution of MAC and IP address information across the network using BGP (Border Gateway Protocol). This allows efficient multi-tenancy, optimized traffic engineering, and simplified operations in complex data center networks, which is crucial for modern network infrastructures.

### 3.2 Data syntese

Different use cases like machine learning, digital forensics, education and training, tool testing, performance measurement or troubleshooting demand valid packet captures in different formats and protocols. But the capturing of this information is a complex task, especially when performed in virtual networks [Spiekermann and Keller, 2021]. Because of this using existing datasets containing possible relevant data is a common methodology. There are different packet captures as well as datasets publicly released, but most of these files are limited in scope. Either they contain only a small number of packets, e. g. for a special protocol analysis or with removed meta information.

This is not a problem only in network forensics, but in nearly every branch of digital investigation [Garfinkel, 2007]. Hence, a common technique to eradicate this problem is the use of synthetical creation of data [Göbel et al., 2023]. This automated generation of network packets involves creating time-stamped sequences of packets that mimic real network traffic, allowing researchers and security practitioners to simulate various network conditions without the need for a live environment. This approach is particularly valuable when testing network security tools, such as intrusion detection systems (IDS)

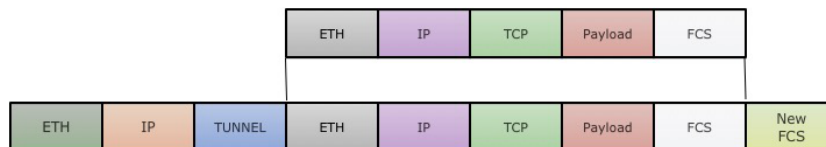


Figure 5: Encapsulation of network packets

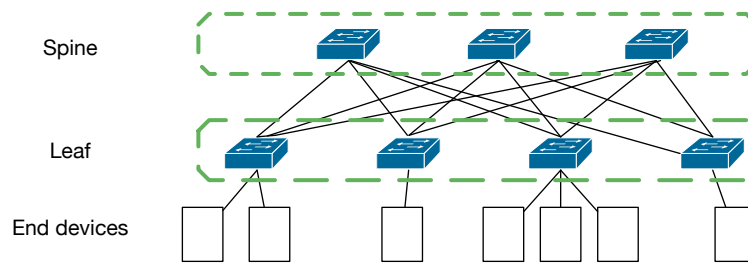


Figure 6: Leaf-Spine Architecture

or web application firewalls (WAFs), which rely on realistic traffic patterns to accurately identify and respond to threats.

Synthetic network data offers additional advantages. It allows for complete control over the environment, ensuring that specific scenarios—such as network attacks, protocol changes, or virtual machine migrations—can be reliably recreated. This reproducibility is essential for consistent testing of security systems, especially in modern virtualized environments where dynamic changes such as VM migrations or on-demand scaling introduce unpredictable factors.

However, the synthetic nature of this data poses risks. If the generated data fails to capture the intricacies of real-life network behaviour, such as the exact timing of packets, natural traffic patterns, or protocol anomalies, the testing process may produce inaccurate results. Tools trained on synthetic data that lack realism may underperform in real-world scenarios, leading to missed detections in network intrusion systems or other security tools.

Different studies have explored various aspects of synthetic traffic generation. [Yin et al., 2022] uses Generative Adversarial Networks, [Kholgh and Kostakos, 2023] uses Chat-GPT3, [Emmerich et al., 2015] presents *Moongen* as a solution to create network traffic. These existing approaches often focus on static network architectures and lack support for the dynamic behaviour inherent in virtualized data center environments. The majority of traffic generation tools are either designed for traditional networks or are protocol-agnostic, failing to capture the encapsulation overhead, tenant-specific routing changes, and bursty traffic patterns triggered by virtual machine (VM) operations.

#### 4 Dynamics of Virtual Network Traffic

Virtual environments as described in Section 3 are highly dynamic. Changes in the overlay as well as in the underlay network might have an impact of the transferred network packets. In this section, critical aspects of virtual networks, that occur in real-world environments, are discussed. To describe real world network traffic related to a virtual network, the following definition is used:

*A virtual network is a software-defined network that emulates the functionality of a traditional physical network, allowing devices, servers, and applications to communicate over a shared physical infrastructure. It uses virtualization technologies to create isolated network segments, manage traffic, and connect resources through overlay protocols, enabling scalable, flexible, and programmable connectivity independent of the hardware-based underlay network.*

Based on this definition, the following scenarios are derived and classified as overlay- or underlay-dependent:

#### 4.1 Overlay-dependent

The overlay-dependent scenarios appear in the virtual environment, either in a network of a specific customer or as an event related to an aspect inside the virtual network.

– VM-based scenarios

VMs are a basic component in virtualized environments, offering isolated compute resources that can be rapidly deployed, scaled, and migrated. Connected to the network by virtual networks interface cards, changes related to a VM can directly impact network performance and management. For instance, the frequent creation, deletion, and migration of VMs can lead to shifting IP addresses, altered traffic patterns, and changes in routing paths within the overlay network. The following events describe possible changes related to VMs, that might impact the network traffic inside the environment.

- VM-Creation

When a VM on a compute node (CN) is deployed, the overlay network must allocate necessary resources, including IP and MAC addresses, setting up the appropriate network, and updating routing tables and TE. Additionally, the overlay must enforce security policies and apply relevant firewall rules, security groups and Quality of Service (QoS) settings. The presence of a new VM can also affect load balancing configurations, potentially leading to the redistribution of traffic flows to accommodate the added compute power. Monitoring systems need to integrate the new VM into visibility tools to ensure consistent tracking.

- VM-Deletion

Similar to the creation of a VM, the process of deleting a VM affects the virtual network. When a VM is deleted, the overlay network must promptly update routing tables and remove any MAC address and IP mappings, reconfigure internal forwarding states to prevent routing inconsistencies and update TEs. The release of network resources, including allocated IP addresses and VNIs/TNIs can lead to changes in network traffic.

- VM-Migration

The migration of a VM is a process of moving a VM from one CN to another. During a migration, the network must maintain uninterrupted connectivity to the migrating VM, ensuring that ongoing connections are not disrupted. This requires real-time updates to the overlay network, including the adjustment of routing tables and recalculation of traffic paths to reflect the new location. Migration can introduce temporary network latency or jitter as traffic routes are adapted, especially if the migration occurs over a long distance or between data centers. Additionally, changes in network topology, such as the reallocation of IP addresses and the adjustment of MAC address entries in the virtual switches, must be handled correctly to prevent packet loss or duplicate traffic. In addition to this, new TEs are created, and the old TEs are terminated and removed from the network. Beyond the creation and removal of TEs, VM migrations inherently alter the network load by shifting traffic flows across different paths, potentially concentrating demand on specific links and introducing dynamic load imbalances that must be considered in realistic traffic modeling.

- Multi-tenancy
  - Multi-tenancy is a core feature of both cloud environments and corporate networks, enabling multiple users to share the same infrastructure while maintaining isolation. The need to provide secure and isolated virtual environments for different tenants can significantly impact the overlay network. It requires the creation of separate virtual segments, each with customized routing, security policies and TEs.
  - Adding new customers
    - When a new customer is added, the system typically provisions a dedicated virtual network segment with specific settings, including custom security policies, IP address allocations, and VNIs or TNIs mapped to new TEs. This process can introduce additional traffic routes and modify existing network paths, which may impact latency and resource utilization as the network adapts to accommodate the new user. Due to the dynamic nature of the environment, new customers can sign up for the service at any time. The automation inside the virtual infrastructure allows for seamless provisioning of resources without the need for manual intervention by an administrator or operator.
  - Deleting customers
    - On the other side, when a customer is deleted, the corresponding virtual resources are decommissioned, and associated network segments are removed. This can lead to a reduction in traffic load but may also trigger updates in routing tables and necessitate adjustments in the overlay structure. Some of the freed resources like VNIs/TNIs, ip addresses or hostnames might be used by new customers.
- Customization
  - User customization can lead to a variety of scenarios that impact the virtual environment. For instance, a customer may create or modify virtual network segments to accommodate new applications or adjust network policies to optimize performance for specific workloads. These changes can involve deploying additional Virtual Network Functions (VNFs), like virtual firewalls or load balancers, or modifying the QoS parameters to prioritize critical traffic. Adjustments at this level often result in shifts of traffic flows, routing paths, or bandwidth allocation, leading to altered network behaviour. Furthermore, user-driven modifications, such as expanding or merging network segments, may necessitate updates in tunnelling and encapsulation settings, directly impacting how traffic is managed and monitored. Additionally the overlay network has to adapt to this changes dynamically without compromising performance or security.
- Virtual protocol change
  - While the current work abstracts from the specifics of overlay encapsulation protocols, it is important to note that technologies such as VXLAN, GENEVE, and NVGRE introduce protocol-dependent encapsulation overheads. These additional header data encompass approximately 50-60 bytes when using VXLAN, resulting from 8 byte VXLAN-header, 8 bytes UDP header, and additional 20 bytes IPv4 and 14 bytes Ethernet header for VTEP communication can lead to MTU (Maximum Transmission Unit) fragmentation if not properly accounted for, thereby increasing latency and reducing transmission efficiency. Furthermore, encapsulation protocols differ in their flexibility and metadata handling: for example, GENEVE supports variable-length options, which may dynamically influence packet sizes and processing costs at tunnel endpoints. Such variations can significantly impact synthetic

traffic patterns, especially in scenarios involving high packet rates or diverse traffic mixes. In practice, variations in VNI or TNI space can affect lookup overheads, routing table sizes, and even impact flow distribution if certain ranges are densely populated.

## 4.2 Underlay-dependent

The events related to the underlay network encompass all possible events, that are typically initiated by the operator of the network. These changes are not executed automatically, but have an impact on the resulting network traffic inside the network.

- Protocol change  
Protocol changes in the underlay network refer to modifications in the network's communication protocols, such as switching or routing protocols, updating protocol versions, or altering protocol parameters (e. g., timers or priority settings). These changes are typically initiated by the network operator to optimize performance, enhance security, or accommodate evolving network requirements. Although not automatically triggered, protocol changes can significantly impact network traffic behaviour by modifying routing paths, influencing latency, or causing temporary disruptions as devices converge on the new configuration. Adjustments to protocols can lead to variations in packet handling, changes in traffic flows, and even the reallocation of bandwidth, all of which affect the overall network dynamics and visibility of traffic within the virtual environment.
- Topology change  
Modifying the underlay topology is uncommon in modern networks, as stability and scalability are prioritized. The prevalent architecture in modern data centers is the leaf-spine design, depicted in Figure 6, which offers superior scalability compared to the traditional three-tier (3Tier) architecture [Liu et al., 2021]. The shift from a 3Tier to a leaf-spine architecture is largely driven by the internal virtualization capabilities that modern data centers offer. Depending on the data center's configuration, various migration techniques can be employed [Ramakrishnan et al., 2007]. Such migrations are especially relevant when moving a data center to the cloud, as seen in scenarios like disaster recovery operations [Kokkinos et al., 2016].
- Hardware optimization  
Hardware optimization in the underlay network involves upgrading or adjusting physical network components to improve performance, reliability, and scalability. This can include deploying faster switches, high-performance routers, and advanced Network Interface Cards (NICs) that support hardware acceleration techniques or offloading specific tasks such as encryption, routing, and packet filtering. Enhanced hardware capabilities can reduce latency, increase throughput, and minimize packet loss, directly impacting the efficiency of the overlay network. In modern data centers, specialized hardware, such as ASICs (Application-Specific Integrated Circuits) and FPGAs (Field-Programmable Gate Arrays), are increasingly used to optimize specific network functions by providing significant improvements in processing speed. Adding or replacing links between the CNs or the switches is a simple form, which is commonly used in modern networks [Poutievski et al., 2022].
- Adjustments  
Underlay networks frequently employ QoS mechanisms to manage bandwidth and

prioritize critical traffic. Modifying QoS policies, such as adjusting traffic shaping, rate limiting, or prioritization rules, can impact the performance of different services within the underlay. These changes may result in shifts in latency, jitter, or throughput for certain applications, directly affecting the end-user experience and overall traffic patterns.

– Network dynamics

However, in addition to these virtualization-specific changes, traditional network dynamics—such as packet loss, bursty traffic, retransmissions, and fluctuating latencies—must also be considered. These classical networking phenomena occur independently of virtualization and are influenced by factors such as congestion, link failures, and transport-layer mechanisms like TCP retransmissions. Especially packet loss might result in noticeable effects, minor packet loss is common and often mitigated by transport-layer mechanisms.

### 4.3 Summary

The discussed effects are possible results of the dynamic and flexible behaviour of a virtual environment. Each of these situations occur randomized, or mostly nondeterministic. Some of the results might happen more frequent than others. Table 1 summarizes the scenarios and gives value for the possible frequency in virtual networks.

Area	Change	Frequency
Overlay	VM Migration	high
	VM Creation	high
	VM-Deletion	high
	Adding customers	high
	Deleting customers	high
	Customization	high
	Virtual protocol change	low
Underlay	Protocol change	low
	Topology change	low
	Optimization	medium
	Adjustments	high
	Network dynamics	high

Table 1: Frequency of changes

As a result, these dynamic and flexibility might hamper a valid packet capture. Such a capture process needs to be reconfigured after crucial effects occur, which might result in a packet loss during the reconfiguration of the capture process [Spiekermann and Eggendorfer, 2016].

Hence, a valid packet generation process might provide a better, i. e. entire packet list of encapsulated network traffic.

## 5 Results

The findings from Section 4 have outlined a set of key dimensions that must be captured to generate synthetic traffic that closely reflects the complexity of real-world virtualized networks. These include both virtualization-specific dynamics—such as encapsulation overhead, tenant migration, and overlay-underlay interactions—and classical network phenomena like packet loss, retransmissions, and bursty traffic behaviour.

Figure 7 summarizes the relevant aspects identified, structured into logical domains that reflect their origin and operational impact.

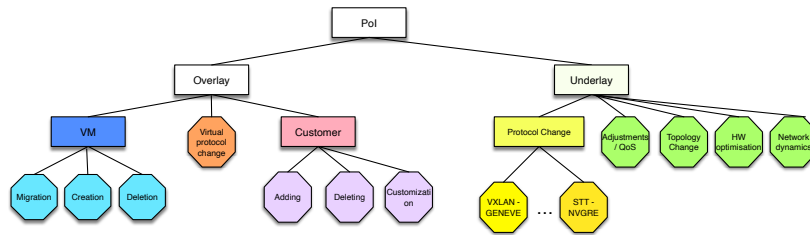


Figure 7: Points of Relevance

In the overlay network, frequent events like VM creation, deletion, and migration lead to shifting IP addresses, fluctuating traffic patterns, and the need for continuous updates to routing and forwarding tables. The high frequency of customer-related actions, such as adding or removing tenants, further intensifies this variability.

The underlay network, although less dynamic, still presents significant challenges. Changes in routing protocols, topology adjustments, hardware optimizations, and QoS or monitoring configurations can introduce temporary disruptions, alter packet flows, and affect network latency and jitter. These changes, while less frequent, can have profound effects on network stability, particularly during planned maintenance or upgrades.

The frequency and non-deterministic nature of these events, as summarized in Table 1, indicate that the generation of valid and representative packet data related to a virtual environment is a challenging task. Although it is possible to recreate network traffic patterns synthetically, elements such as traffic bursts, the behaviour of users, and network congestion can be difficult to simulate accurately. Real-world networks often provide unpredictable traffic transmissions, which must be managed by a packet generation process. In the field of anomaly detection, these special events have an impact on the detection rate of the algorithms [Spiekermann and Keller, 2020]. Therefore, possible changes in the infrastructure demand for a new training of the model, based on new data.

A packet generation tool needs to be aware of these effects and needs randomized as well as predefined parameters to create accurate network traffic according to virtual networks. In addition to [Spiekermann and Keller, 2022] the following requirements are relevant to be implemented by a packet generation tool:

- Creating packet bursts (e. g. by adding a high number of randomized or predefined packets)

- Creating packet loss (e. g. by removing packets during the generation process)<sup>1</sup>
- Delaying or speeding up packets
- Adding and deleting virtual systems (e. g. ip and mac-addresses, hostnames)
- Adding and deleting subnets and TEs (e. g. network addresses)
- Reusing of existing values after deletion

Each requirement must appear according to the frequency as shown in Table 1.

An additional aspect to consider is the preliminary quantitative characterization of the target environment, which inherently depends on the underlying infrastructure. In virtualized settings with moderate tenant activity and dynamic VM operations, realistic traffic modelling typically requires the collection of several hundred thousand to several million packets to comprehensively capture encapsulation effects, burst behaviour, and overlay–underlay interactions. In large-scale data center environments, where east–west traffic dominates over north–south communication [Ben Yoo, 2022], infrastructures are designed to sustain aggregated throughput rates of up to 13.1 Pb/s [Vahdat, Amin, 2024]. This results in billions of concurrent flows, exhibiting diverse lifetimes and heterogeneous transport-layer characteristics.

As a result, the generated data can be used to implement improved training data for digital investigation, anomaly detection or network simulation. Current tools often struggle to adapt to the frequent changes in traffic patterns and topology inherent to virtual networks, leading to inappropriate datasets. The limitation of live packet captures in virtual environments is eradicated due to the use of such a packet generation tool, additional more sophisticated and realistic synthetic datasets are possible.

## 6 Conclusion

This paper describes real-world scenarios, which occur in modern data centers, and furthermore provides an overview about the frequency of these changes. The defined classification is crucial for nowadays packet generators, used for the generation or transformation of network traffic in virtual networks. If implemented correctly, a synthesis of network packets through tools like *Encapcap* offers multiple advantages:

- **Accuracy:** By maintaining the integrity of the original network traffic while adding virtual network characteristics, the synthetic packets retain realism, which is essential for effective analysis.
- **Cost-Efficiency:** Generating synthetic data reduces the need for complex and costly real-world setups, allowing organizations to create scenarios that would be impractical to replicate in live environments.
- **Customization:** Synthetic packet generation allows the tailoring of packet characteristics to specific testing needs, such as simulating particular attack vectors, anomaly types, or virtual network configurations.

---

<sup>1</sup> Due to the inherent retransmission of network packets in various network protocols like TCP, the process of deleting packets might create further side-effects or unwanted anomalies in the network traffic.

- **Scalability:** Automatically generating large data sets of network traffic enables researchers to scale their testing and training environments without manual packet capture processes.

The synthesis of network packets is not merely a technical task; it is a theoretical necessity in the age of virtualized networks. Without accurate and scalable methods to generate realistic packet captures, the reliability of network security solutions is compromised, and the ability to conduct thorough digital investigations is hindered. The future research focuses on the interplay between overlay and underlay networks in modern data centers. Understanding how changes in one layer affect the other could lead to more accurate traffic generators. While this approach aims to offer a lightweight and flexible method for synthetic traffic generation, future work could extend this foundation by integrating stochastic traffic models and protocol-aware behaviour (e. g., retransmissions). Additionally, analysing overlay-underlay interactions and evaluating scalability in large-scale scenarios or the resource usage in resource-constrained environments would provide further insights into the robustness and applicability of the method in dynamic environments.

## References

- [Al-Hadhrami and Hussain, 2020] Al-Hadhrami, Y. and Hussain, F. K. (2020). Real time dataset generation framework for intrusion detection systems in iot. *Future Generation Computer Systems*, 108:414–423.
- [Belenko et al., 2018] Belenko, V., Krundyshev, V., and Kalinin, M. (2018). Synthetic datasets generation for intrusion detection in vanet. In *Proceedings of the 11th international conference on security of information and networks*, pages 1–6.
- [Ben Yoo, 2022] Ben Yoo, S. J. (2022). Prospects and challenges of photonic switching in data centers and computing systems. *Journal of Lightwave Technology*, 40(8):2214–2243.
- [Bhuyan et al., 2013] Bhuyan, M. H., Bhattacharyya, D. K., and Kalita, J. K. (2013). Network anomaly detection: methods, systems and tools. *IEEE Communications Surveys & Tutorials*, 16(1):303–336.
- [Botez et al., 2021] Botez, R., Costa-Requena, J., Ivanciu, I.-A., Strautiu, V., and Dobrota, V. (2021). SDN-based network slicing mechanism for a scalable 4G/5G core network: A Kubernetes approach. *Sensors*, 21(11):3773.
- [Chowdhury and Boutaba, 2010] Chowdhury, N. M. K. and Boutaba, R. (2010). A survey of network virtualization. *Computer Networks*, 54(5):862–876.
- [Corey et al., 2002] Corey, V., Peterman, C., Shearin, S., Greenberg, M. S., and Van Bokkelen, J. (2002). Network forensics analysis. *IEEE Internet Computing*, 6(6):60–66.
- [Emmerich et al., 2015] Emmerich, P., Gallenmüller, S., Raumer, D., Wohlfart, F., and Carle, G. (2015). Moongen: A scriptable high-speed packet generator. In *Proceedings of the 2015 Internet Measurement Conference*, pages 275–287.
- [Fernando, 2021] Fernando, V. (2021). Cyber forensics tools: A review on mechanism and emerging challenges. In *2021 11th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–7. IEEE.
- [Freed and Borenstein, 1996] Freed, N. and Borenstein, N. S. (1996). Multipurpose Internet Mail Extensions (MIME) part two: Media types. RFC 2046, RFC Editor.
- [Garfinkel, 2007] Garfinkel, S. (2007). Forensic corpora: a challenge for forensic research. *Electronic Evidence Inf. Cent. 1e10*.

- [Ghazanfar et al., 2020] Ghazanfar, S., Hussain, F., Rehman, A. U., Fayyaz, U. U., Shahzad, F., and Shah, G. A. (2020). IoT-flock: An open-source framework for IoT traffic generation. In *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*, pages 1–6. IEEE.
- [Göbel et al., 2023] Göbel, T., Baier, H., and Breitingner, F. (2023). Data for digital forensics: Why a discussion on “how realistic is synthetic data” is dispensable. *Digital Threats: Research and Practice*, 4(3):1–18.
- [Göbel et al., 2022] Göbel, T., Maltan, S., Türr, J., Baier, H., and Mann, F. (2022). Fortrace-a holistic forensic data set synthesis framework. *Forensic Science International: Digital Investigation*, 40:301344.
- [Hausenblas, 2018] Hausenblas, M. (2018). *Container Networking*. O’Reilly Media, Incorporated.
- [Kholgh and Kostakos, 2023] Kholgh, D. K. and Kostakos, P. (2023). Pac-gpt: A novel approach to generating synthetic network traffic with gpt-3. *IEEE Access*, 11:114936–114951.
- [Kokkinos et al., 2016] Kokkinos, P., Kalogeras, D., Levin, A., and Varvarigos, E. (2016). Survey: Live migration and disaster recovery over long-distance networks. *ACM Computing Surveys (CSUR)*, 49(2):1–36.
- [Li et al., 2019] Li, Y., Miao, R., Alizadeh, M., and Yu, M. (2019). DETER: Deterministic TCP replay for performance diagnosis. In *16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19)*, pages 437–452.
- [Liu et al., 2021] Liu, K., Zhao, A., and Pan, J. (2021). Data center architecture, operation, and optimization. *Future Networks, Services and Management: Underlay and Overlay, Edge, Applications, Slicing, Cloud, Space, AI/ML, and Quantum Computing*, pages 185–212.
- [Mohajer et al., 2024] Mohajer, A., Hajipour, J., and Leung, V. C. (2024). Dynamic offloading in mobile edge computing with traffic-aware network slicing and adaptive td3 strategy. *IEEE Communications Letters*.
- [Padman and Memon, 2002] Padman, V. and Memon, N. (2002). Design of a virtual laboratory for information assurance education and research. In *Workshop on Information Assurance and Security*, volume 1, page 1555.
- [Poutievski et al., 2022] Poutievski, L., Mashayekhi, O., Ong, J., Singh, A., Tariq, M., Wang, R., Zhang, J., Beauregard, V., Conner, P., Gribble, S., et al. (2022). Jupiter evolving: transforming google’s datacenter network via optical circuit switches and software-defined networking. In *Proceedings of the ACM SIGCOMM 2022 Conference*, pages 66–85.
- [Rajasinghe et al., 2018] Rajasinghe, N., Samarabandu, J., and Wang, X. (2018). Insecs-dcs: a highly customizable network intrusion dataset creation framework. In *2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE)*, pages 1–4. IEEE.
- [Ramakrishnan et al., 2007] Ramakrishnan, K., Shenoy, P., and Van der Merwe, J. (2007). Live data center migration across wans: a robust cooperative context aware approach. In *Proceedings of the 2007 SIGCOMM workshop on Internet network management*, pages 262–267.
- [Sharafaldin et al., 2018] Sharafaldin, I., Lashkari, A. H., and Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In *ICISSp*, pages 108–116.
- [Sharafaldin et al., 2019] Sharafaldin, I., Lashkari, A. H., Hakak, S., and Ghorbani, A. A. (2019). Developing realistic distributed denial of service (ddos) attack dataset and taxonomy. In *International Carnahan Conference on Security Technology (ICCST)*, pages 1–8. IEEE.
- [Singh et al., 2017] Singh, T., Jain, V., and Babu, G. S. (2017). Vxlan and evpn for data center network transformation. In *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–6. IEEE.
- [Son et al., 2012] Son, J., Irrechukwu, C., and Fitzgibbons, P. (2012). Virtual lab for online cyber security education. *Communications of the IIMA*, 12(4):5.

- [Spiekermann and Eggendorfer, 2016] Spiekermann, D. and Eggendorfer, T. (2016). Challenges of network forensic investigation in virtual networks. *Journal of Cyber Security and Mobility*, pages 15–46.
- [Spiekermann and Keller, 2020] Spiekermann, D. and Keller, J. (2020). Impact of virtual networks on anomaly detection with machine learning. In *6th IEEE Conference on Network Softwarization (NetSoft)*, pages 430–436.
- [Spiekermann and Keller, 2021] Spiekermann, D. and Keller, J. (2021). Wiretapping pods and nodes—lawful interception in Kubernetes. *Electronic Communications of the EASST*, 80.
- [Spiekermann and Keller, 2022] Spiekermann, D. and Keller, J. (2022). Requirements for crafting virtual network packet captures. *Journal of Cybersecurity and Privacy*, 2(3):516–526.
- [Spiekermann et al., 2017] Spiekermann, D., Keller, J., and Eggendorfer, T. (2017). Network forensic investigation in openflow networks with forcon. *Digital Investigation*, 20:S66–S74.
- [Vahdat, Amin, 2024] Vahdat, Amin (2024). Speed, scale, and reliability: 25 years of data center networking at google. Accessed: 2025-03-27.
- [Vishwanath and Vahdat, 2006] Vishwanath, K. V. and Vahdat, A. (2006). Realistic and responsive network traffic generation. In *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, SIGCOMM '06*, page 111–122, New York, NY, USA. Association for Computing Machinery.
- [Voigt et al., 2024] Voigt, L. L., Freiling, F., and Hargreaves, C. J. (2024). Re-imagien: Generating coherent background activity in synthetic scenario-based forensic datasets using large language models. *Forensic Science International: Digital Investigation*, 50:301805.
- [Voyiatzis et al., 2015] Voyiatzis, A. G., Katsigiannis, K., and Koubias, S. (2015). A Modbus/TCP fuzzer for testing internetworked industrial systems. In *20th Conference on Emerging Technologies & Factory Automation (ETFA)*, pages 1–6. IEEE.
- [Worster et al., 2005] Worster, T., Rekhter, Y., and Rosen, E. (2005). Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE). RFC 4023, RFC Editor.
- [Yang and Mohajer, 2024] Yang, J. and Mohajer, A. (2024). Multi objective constellation optimization and dynamic link utilization for sustainable information delivery using pd-noma deep reinforcement learning. *Wireless Networks*, pages 1–21.
- [Yin et al., 2022] Yin, Y., Lin, Z., Jin, M., Fanti, G., and Sekar, V. (2022). Practical gan-based synthetic ip header trace generation using netshare. In *Proceedings of the ACM SIGCOMM 2022 Conference*, pages 458–472.
- [Zhou and Mohajer, 2024] Zhou, G. and Mohajer, A. (2024). Blind reconfigurable intelligent surfaces for dynamic offloading in fixed-noma mobile edge networks. *International Journal of Sensor Networks*, 46(3):142–160.