VIPNANO: Monitoring of **Vi**rtual **P**rivate Cloud **N**etworks for Automated **Ano**maly Detection

Marleen Sichermann*, Katharina Dietz*, Jochen Kögel**, Sebastian Meier**,
Stefan Geißler*, Tobias Hoßfeld*
*Chair of Communication Networks, University of Würzburg, Würzburg, Germany,
**Isarnet Software Solutions GmbH, Munich, Germany,
*{marleen.sichermann, katharina.dietz, stefan.geissler, tobias.hossfeld}@uni-wuerzburg.de,
**{jochen.koegel, sebastian.meier}@isarnet.de

***Abstract*—Anomaly detection in enterprise networks is crucial for cybersecurity, system monitoring, and identifying outages. Despite extensive academic research, practical deployment of proposed mechanisms remains rare. The VIPNANO project investigates key shortcomings in academic approaches, focusing on two major obstacles: (1) reliance on unrealistic datasets that fail to reflect real-world complexity, and (2) overly complex machine learning models with impractical computational overhead. Additionally, we highlight a critical gap – the lack of rigorous real-world validation. Through systematic analysis, we emphasize the need to prioritize realistic data, scalability, and verifiable solutions to bridge the gap between theory and deployment.**

***Index Terms*—Anomaly Detection, Practical Validation, Real-World Systems.**

I. Introduction

The shift to cloud-based, API-driven, and software-defined networking is fundamentally changing how networks are built and managed. The drivers of softwarization are SDN, NFV, and cloud-native architectures. In recent years, the rapid evolution of these concepts has made their management increasingly complex, especially as systems scale in size and functionality. As enterprises start to transparently use both on-premise and cloud infrastructure in heterogeneous virtual private cloud (VPC) deployments, both operational and connectivity aspects need to be taken into account. Especially, connecting infrastructure segments across cloud provider boundaries and integrating VPC resources with legacy on-premise services is a demanding challenge. To this end, researchers and practitioners are constantly working to develop mechanisms that not only meet current demands but also anticipate future developments in dynamic, real-world environments. The same applies to monitoring in such heterogeneous deployments, especially for detecting anomalies, outages, or malicious attacks, where established approaches often fall short in applicability, scalability, or adaptability [1].

The key challenge is effectively monitoring and representing the state of a large-scale deployment, especially its service interconnections, for fast, accurate issue detection. Naturally, anomaly detection, intrusion detection, and network monitoring are well-researched fields. However, we argue that many currently proposed mechanisms are ill-suited for real-world large-scale systems due to their requirements, scalability, or adaptability. Despite significant advances, many solutions rely on unrealistic input data (e.g., full packet traces, labeled data), lack scalability (e.g., per packet analysis), or struggle to adapt to changing systems (e.g., pre-trained ML models) [1].

To this end, in the project *Monitoring of **Vi**rtual **P**rivate Cloud **N**etworks for Automated **Ano**maly Detection of Enterprise Applications in heterogeneous Networks* (VIPNANO), we work towards novel mechanisms for the detection of network anomalies that can explicitly be applied to large scale, real-world enterprise networks. This means working with limited or highly aggregated data, often without labeled data. In this extended abstract, we highlight the problem and the gap in current literature regarding practical mechanisms for real systems. Following this, Section 2 describes the scenario investigated in the project, while Sections 3 and 4 provide an overview of existing research and highlight gaps in literature.

II. Scenario

A. Heterogeneous Cloud Scenarios

In heterogeneous cloud environments — characterized by a mix of on-premise resources and virtual private clouds across public providers (Fig. 1) — monitoring and anomaly detection present significant challenges. The network infrastructure connecting services, users, and data spans different architectures, complicating the creation of a unified monitoring framework. Variations in data formats, logging standards, and asynchronous data collection across these systems hinder timely anomaly detection and response. Moreover, the

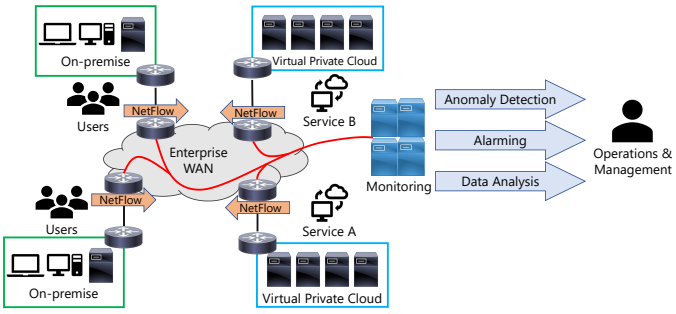

Fig. 1. Multicloud monitoring scenario: unidirectional aggregate NetFlow logs captured at gateway routers between network subnets.

dynamic nature of these environments, with frequent configuration changes and fluctuating workloads, further complicates the detection of subtle performance degradations and security breaches [2]. Addressing these challenges requires flexible monitoring solutions that integrate diverse data sources and offer cohesive insights into service and network infrastructure.

A key challenge in detecting anomalies from outages, misconfigurations, or attacks is obtaining data that enables actionable insights. In large-scale deployments, monitoring data is often available only at specific points. To minimize storage and processing, traffic captures are often aggregated over time. In extremely large scale deployments, not even flow level logs may be available for longer timeframes or when looking at historical data. Instead, traffic streams may be aggregated even further, for example into pairs of communicating subnets.

When it comes to distributed monitoring, to avoid duplicate records, traffic is usually only captured either in ingress or egress direction, necessitating merging and post-processing different data sources (e.g., for the identification of bidirectional flows). This aggregation of different data sources poses a further challenge for the detection of anomalous traffic, as an anomaly might only be visible in some parts of the network and hence only reflect in ingress or egress direction. Finally, distributed measurements come with the inherent challenge of synchronizing timestamps across different capture points, as variations in clock skew and latency can lead to misaligned records that obscure the true sequence of events [3]. Note, however, that due to the traffic aggregation and coarse timescales, synchronization errors are less impactful when it comes to aggregate traces than for packet level traces.

Lastly, the heterogeneity of monitoring tools and hardware across network segments can cause inconsistencies in data granularity and format, complicating event correlation. The dynamic nature of network configurations further complicates the establishment of consistent baselines for normal activity, potentially masking or distorting the signature of an anomaly [2]. These factors, when combined with the inherent lack of fine-granular data, create a complex landscape in which subtle and localized anomalies may easily be overlooked.

B. Exemplary Multicloud Application

We consider an exemplary application running in a multicloud environment at a large German transportation company. In the VIPNANO project, we have access to flow data and are in contact with both the network operation and the application team. This gives us valuable insights into the application setting as well as traffic behavior and anomalies.

The application provides virtual desktops via Citrix to thousands of users, connected either via the enterprise WAN or externally via VPN (red lines Fig. 3). The underlying service runs in a VPC at cloud provider 1 on several dynamically spawned VMs. When users work on their virtual desktops, they access services such as enterprise-specific web applications, file shares, Internet browsing, etc. from other VPCs of the same cloud provider, VPCs of cloud provider 2 or on premise (dashed purple lines). Furthermore, application servers in

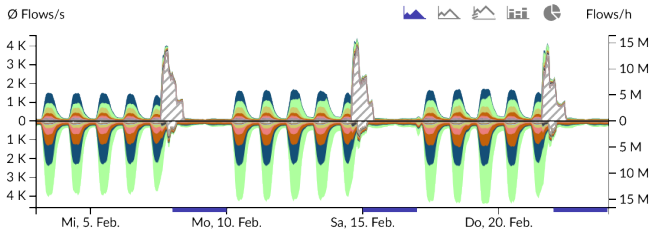

Fig. 2. Exemplary three weeks.

VPCs may access cloud resources (e.g., object storage) of the cloud provider directly. The virtual desktops must offer low response times, as users expect instant feedback for interactions (mouse, keyboard). Client-to-desktop interactions create an asymmetric traffic pattern, where both directions are delay-sensitive, and downstream video requires high bandwidth. Combined with traffic heterogeneity, these factors emphasize the complexity and need for early anomaly detection.

To address the identified shortcomings in anomaly detection, we aim to leverage the dataset and insights from the virtual desktop application monitored at the enterprise WAN edge routers to explore this field under real-world constraints. These include, but are not limited to, the high aggregation of real-world data as a result of modern monitoring techniques, data protection laws, the absence of labeled data, and the need for reasonable computational effort. Therefore, we currently exclude supervised methods that require labeled data and overly complex models. To achieve this, we plan to assess existing methods for their applicability to our real-world dataset in an upcoming survey. We also aim to develop new mechanisms within these constraints. Currently, the focus is on baselining approaches due to the seasonal nature of the underlying data. This is further illustrated in Fig. 2, which shows three weeks of flow data for an application server subnet. In the stacked chart, each of the top five protocols is indicated by a different color. Every Friday a large hatched peak is visible, which is related to scan traffic originating from the security department across a wide port range. This highlights a major challenge in enterprise networks: traffic on weekdays is almost perfectly periodic, and deviations in the form of large peaks might seem to be anomalies at first. Therefore, advanced algorithms are required that can learn traffic patterns from such scheduled events. Regarding our monitoring infrastructure, we will investigate which features can be extracted from our dataset and explore ways to enhance the NetFlow monitoring and aggregation infrastructure to capture the most valuable information. Furthermore, we intend to implement a human-in-the-loop mechanism to address the challenge of missing labels.

III. Related Work

Several publications have criticized the limited real-world adoption of academic anomaly and network intrusion detection approaches. Maseer et al. [4] systematically review anomaly network intrusion detection systems, analyzing notable recent

work and revealing that none are designed for real-world deployment. Most rely on datasets from controlled environments and lack real-world performance testing. Expanding on this, the authors of [5] highlight that despite numerous machine learning-based intrusion detection methods outperforming traditional approaches, they remain rarely used due to incompatibility with existing infrastructures, high computational costs, and usability issues. Additionally, Arp et al. [6] systematically examine pitfalls in applying machine learning to security, emphasizing the widespread reliance on lab-generated data. They warn that this reliance may lead models to incorporate information unavailable in real-world scenarios.

The objective of VIPNANO is to highlight these shortcomings, illustrate them with concrete use cases, and advance research on anomaly detection under realistic constraints. This includes methods like VITALFlow [7], FACT [8], and other flow-based network anomaly detection approaches.

IV. Gap in Literature

To investigate the disparity between academia and practical environments, we conducted a comprehensive literature survey on intrusion and anomaly detection ourselves [1] (166 papers in total). Therein, we established 17 hypotheses – based on proto-typical monitoring tool users – why academic research might not be adopted in practice. These hypotheses concern a broad variety of topics, such as usability or practicability.

Figure 4 illustrates five of those hypotheses which are especially relevant in the context of VIPNANO. For example, the first hypothesis (H1) highlights that roughly two thirds of the evaluated papers do not depict a generic solution, e.g., were only applied in a single scenario (i.e. on only one dataset) or do not take into account network or traffic changes. The second (H2) and third (H3) hypotheses showcase that the data is infeasible or even impossible to obtain in a non-negligible portion of papers, e.g., due to legal requirements or hardware constraints. We also found that in the vast majority of papers the utilized models exhibit a high degree of complexity (H4), which may limit their trustworthiness, causing users to remain skeptical. Lastly, the incorporation of domain knowledge is often ignored (H5), e.g., additional internal or external con-

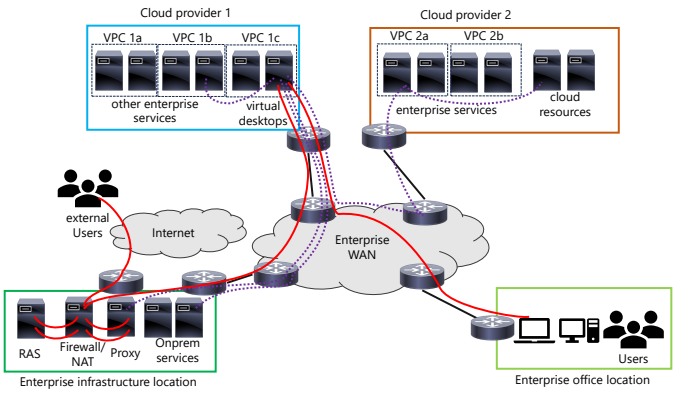

Fig. 3. Exemplary multicloud application.

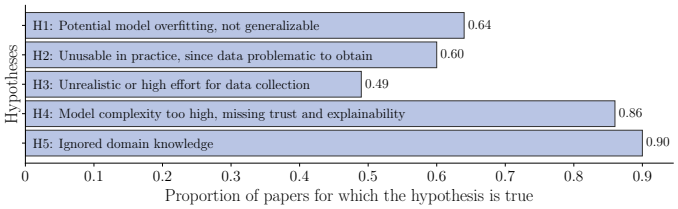

Fig. 4. Excerpt of analyzed hypotheses [1] (renumbered).

text information or even admin feedback via active learning approaches is overlooked. Therefore, in VIPNANO we aim to bridge the above gaps and focus further on the practicability aspect of our prior literature review, e.g., by conducting a more pragmatic survey similar to the previously mentioned related works [4], [6], [9], wherein we pay special attention to applicability in real-world scenarios.

Acknowledgment

This work is funded by the Bavarian Ministry of Economics, Regional Development and Energy (StMWI) within the project VIPNANO as part of the R&D program for information and communication technology under research grants DIK-2307-0005 and DIK-2307-0006. The authors alone are responsible for the content.

References

[1] K. Dietz, M. Mühlhauser, J. Kögel, S. Schwinger, M. Sichermann, M. Seufert, D. Herrmann, and T. Hoßfeld, "The missing link in network intrusion detection: Taking AI/ML research efforts to users," *IEEE Access*, 2024.

[2] C. Nwachukwu, K. Durodola-Tunde, and C. Akwiwu-Uzoma, "Ai-driven anomaly detection in cloud computing environments," *International Journal of Science and Research Archive*, 2024.

[3] J. Kögel, *One-way delay measurement based on flow data in large enterprise networks*. Inst. für Kommunikationsnetze und Rechnersysteme, 2013.

[4] Z. K. Maseer, Q. K. Kadhim, B. Al-Bander, R. Yusof, and A. Saif, "Meta-analysis and systematic review for anomaly network intrusion detection systems: Detection methods, dataset, validation methodology, and challenges," *IET Networks*, 2024.

[5] M. Husák, D. Manoj, and P. Kumar, "Machine learning in intrusion detection: An operational perspective," in *IEEE International Conference on Network and Service Management (CNSM)*. IEEE, 2024, pp. 1–7.

[6] D. Arp, E. Quiring, F. Pendlebury, A. Warnecke, F. Pierazzi, C. Wressnegger, L. Cavallaro, and K. Rieck, "Dos and don'ts of machine learning in computer security," in *USENIX Security Symposium*, 2022, pp. 3971–3988.

[7] T. Tremel, J. Kögel, F. Jauernig, S. Meier, D. Thom, F. Becker, C. Müller, and S. Koch, "Vitalflow: Visual interactive traffic analysis with netflow," in *IEEE/IFIP Network Operations and Management Symposium (NOMS)*. IEEE, 2022, pp. 1–6.

[8] D. Schatzmann, S. Leinen, J. Kögel, and W. Mühlbauer, "FACT: Flow-based approach for connectivity tracking," in *International Conference on Passive and Active Network Measurement (PAM)*. Springer, 2011, pp. 214–223.

[9] G. Apruzzese, P. Laskov, and J. Schneider, "SoK: Pragmatic assessment of machine learning for network intrusion detection," in *IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2023, pp. 592–614.