International Journal of Science, Strategic Management and Technology
Volume **02** Issue **05** May-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

An Experimental Comparison of Public Internet vs. Private Direct Connect for Enterprise Cloud Performance

Saurabh Verma saurabhverma234567@gmail.com

Pragya Shrivastava pspragyashri@gmail.com

Dr. Chandra Shekhar Gautam shekharg84@gmail.com

Dr. Akhilesh A. Waoo akhileshawaoo@gmail.com,

Department of Computer Science and Engineering AKS University, Satna (M.P.), India

https://doi.org/10.55041/ijsmt.v2i5.396

Cite this Article: Verma, S. & Shrivastava, P. (2026). An Experimental Comparison of Public Internet vs. Private Direct Connect for Enterprise Cloud Performance. International Journal of Science, Strategic Management and Technology, *02*(05). https://doi.org/10.55041/ijsmt.v2i5.396

Abstract

The decision to migrate latency-sensitive workloads to the cloud has made the issue of whether to use public internet-based transit or direct interconnects a key architectural choice. In this paper, an empirical performance comparison has been described between traditional Public Internet (with IPsec VPN) and Private Direct Connect (AWS Direct Connect) to measure the reliability and speed trade-offs. We evaluated throughput, Round-Trip Time (RTT), and packet loss with a sustained 14-day cycle in a distributed experimental infrastructure in two large metropolitan areas. As our experimental results show, although the public internet has adequate bandwidth to support an asynchronous workload, it exhibits considerable micro-burst latency peaks, and jitter values reach as high as 35ms during periods of extreme congestion. Conversely, the Private Direct Connect continued to record sub-5ms jitter variance and a 22-percent greater sustained goodput on large-scale database synchronisation. In addition, we examine the effects of protocol overhead and show that the encryption layer of public VPNs decreases effective payload capacity by nearly 8%. Such results provide a strict, data-driven benchmark against which enterprise architects can measure the cost-performance ratio of dedicated cloud circuits, in the case of real-time analytics and high-frequency financial applications.

Keywords: Cloud Computing, AWS Direct Connect, Network Performance, Latency Analysis, Enterprise Architecture, Hybrid Cloud.

1. Introduction

1.1 Context: The Shift to Hybrid Cloud

The business environment has been experiencing a paradigm change in the use of isolated on-premise voice and data centres to the structure of hybrid clouds (Sagi, 2024). Enterprises are no longer considering the cloud as a storage repository. Still, they are considering it as a significant extension of their internal infrastructure and more often combine on-premises systems with public and private cloud systems to gain more agility and scalability. The process of migration enables businesses to

upgrade legacy applications at their own speed without the need to migrate sensitive data to on-managed hardware to ensure regulatory compliance. By 2026, this model has reached its maturity stage, and resilient technical architecture will have been established as a non-negotiable item, as 90 per cent of businesses will use more than one cloud to drive real-time analytics and AI-driven business processes.

1.2 The Conflict: Public Internet vs. Private Interconnects

However, despite the transformation, there is still some underlying connectivity conflict between the two primary modes of transport:

Public Internet: Best accessibility and low barriers to entry, security is usually provided as an IPsec VPN tunnel (Nyakomitta and Abeka, 2020). But it is marred with uncertainty, such as excessive latency, dropped packets, and undocumented operational expenses that are motivated by changing egress charges.

Private Connections (e.g., AWS Direct Connect): The SLA guarantees reliability, predictable performance, and up to a 60-70 reduction in data costs (Verma, et al,. 2024). Although these dedicated fiber connections are highly reliable, they incur high monthly port charges as well as physical infrastructure investments that small-scale operations may not be able to afford.

Businesses must thus decide on the agility of the public routes and the determinism of the private paths routing, and many times do not have a clear data-based threshold on when to make the switch.

1.3 Contribution: Empirical Performance Analysis

This conflict in 2007 is the focus of this paper, where the empirical analysis of throughput and latency is presented side-by-side. They are unlike the current literature that is broadly precipitated on the benefits of architecture, but is a unique research in which the paths are measured under different network loads, a small packet VoIP communication, a large file database transfer and a burst web traffic. Through reviewing the effects of protocol overheads in encrypted public tunnels compared to the raw performance of private links, this review offers the information that enterprises can use to plot particular workloads about the most technically and economically reasonable connection path (Yildirim, 2024).

2. Related Work

The available literature on cloud connectivity has developed beyond simple bandwidth studies to sophisticated studies of hybrid and multi-cloud systems. The existing studies could be divided into three main spheres:

2.1 Comparative Bandwidth and Latency Studies

Early studies were mainly concerned with simple throughput differences. Recent reports of 2024-2025 have highlighted that though the public internet paths may give high burst capability compared to the example of 100 Mb, they fall short of the 99th percentile latency that is required by real-time networks. Specifically, according to reports, AWS Direct Connect offers a predictable environment, with the latency variation decreasing by up to 50% of the range mentioned in the case of regular VPN-over-Internet solutions (Wittig, and Wittig, 2023).

Performance Aspect	Public Internet (VPN-Based Connectivity)	Private Direct Connect Connectivity
Bandwidth Behaviour	High burst bandwidth availability, depending on ISP congestion conditions	Consistent and dedicated bandwidth allocation
Throughput Stability	Variable throughput due to shared public infrastructure	Stable throughput enabled by private routing paths
Latency Performance	Higher average and tail latency caused by dynamic routing and congestion	Lower and predictable latency across transmission paths
99th Percentile Latency	Frequently exceeds acceptable limits for real-time applications	Maintains latency within controlled operational thresholds
Latency Variation (Jitter)	Significant variation during peak traffic periods	Up to 50% reduction in latency variation compared to VPN-based Internet paths
Network Determinism	Non-deterministic performance influenced by ISP peering and noisy neighbours	Deterministic performance ensured through dedicated interconnect.
Suitability for Real-Time Workloads	Limited suitability for latency-sensitive services	Highly suitable for real-time enterprise applications

Table 1:Comparative Bandwidth and Latency Studies

2.2 Security and Encryption Impacts

Tax on security in public transit is a performance that has been well documented and is a challenge. The encapsulation and encryption slot of IPsec and TLS tunnels are discovered to lose effective goodput by 10% -15% as a result of the packet breaking and CPU costs at the gateway nodes because of the Maximum Transmission Unit (MTU) (You and Tang, 2021).

2.3 Emerging Interconnect Technologies

Recent 2025-2026 developments are leaning more toward high-performance standards, such as the Ultra Ethernet Consortium (UEC 1.0), which is trying to optimise Ethernet to AI and HPC workloads. These works are aimed at optimising the hardware; however, they frequently pay little attention to the real performance variations that standard enterprise programs can encounter in their daily ISP congestion period.

1.1 2.4 The Research Gap

Although the above works form a basis, there is still a critical lack of continuity in the holistic measurement of performance at its peak-hour usage and the extent to which protocol overhead affects the process of database synchronisation at an overall enterprise level.

The majority of the research performed assumes that the public Internet is a static variable, and most of the tests are conducted under controlled laboratory conditions. This gap is filled in this paper by:

1. **Longitudinal Real-World Measurement:**

Unlike prior studies conducted under short-duration or laboratory-controlled environments, this research performs a continuous 14-day longitudinal evaluation under real operational conditions. This extended observation period enables accurate capture of Internet variability, including *Noisy Neighbour effects*, ISP congestion behaviour, and peering-point performance degradation during peak enterprise traffic hours.

2. **Protocol Overhead Quantification in Enterprise Workloads:**

The study explicitly isolates and quantifies the performance impact introduced by VPN encryption and encapsulation mechanisms in comparison with native private fibre connectivity provided by Direct Connect. By analysing enterprise-scale database synchronisation traffic, the research establishes an Efficiency Index that measures usable bandwidth loss and protocol-induced performance degradation in practical deployment scenarios.

3. **Workload-Specific Performance Benchmarking and Cost–Performance Threshold Analysis:**

In contrast to conventional benchmarking approaches relying solely on synthetic tools such as ping or basic throughput tests, this work evaluates application-driven enterprise traffic patterns, including large-scale SQL database mirroring and cloud object replication workloads (da Silva Pinto, A., 2025). The resulting measurements enable the derivation of a cost–performance break-even threshold**,** providing actionable guidance for determining when dedicated private connectivity becomes both technically and economically advantageous.

3. Experimental Design

To ensure a fair and comprehensive comparison, we established a hybrid testbed spanning a physical on-premise data centre and a cloud-native virtual private cloud (VPC) (Wang et al., 2024). The following table summarizes the core components of our experimental environment.

Table 2: Experimental Environment Specifications

Component	Technical Specification
Cloud Provider	AWS (US-East-1 Region) using m6i.2xlarge EC2 instances for high network performance.
On-Premise Node	Dell PowerEdge R760 (8-core Intel Xeon, 32GB RAM) running Ubuntu 24.04 LTS as the primary gateway.
Network Gateway	Cisco ISR 4451-X with HSEC license for hardware-accelerated IPsec encryption.
Public Path	Tier-1 ISP (1 Gbps Fibre) utilising an IKEv2 IPsec VPN tunnel (AES-256-GCM).
Private Path	AWS Direct Connect (1 Gbps Dedicated Port) via an Equinix Fabric cross-connect.
Test Tools	iperf3 (Bandwidth), mtr (Latency/Path), and tcpdump (Packet analysis).

3.1 Traffic Profiles

A key strength of this paper is the use of diverse traffic types that simulate real-world enterprise operations. We categorise our tests into three profiles:

A. Small-Packet Traffic (VoIP/Real-time)

- **Packet Size:** 64 to 128 bytes.
- **Goal:** Measuring Jitter and PPS (Packets Per Second).
- **Context:** Simulates real-time voice, video conferencing, and high-frequency sensor data (IoT).

B. Large-File Transfer (Database/Storage)

- **Packet Size:** Standard 1500 MTU (adjusted for VPN overhead).
- **Goal:** Measuring Sustained Throughput and Goodput.
- **Context:** Simulates daily database backups, S3 object synchronisation, and SQL Server Always On availability groups.

C. Bursty Traffic (Web/API)

- **Pattern:** High-frequency, short-duration bursts followed by idle periods.
- **Goal:** Measuring Time-to-First-Byte (TTFB) and congestion recovery (Edgar 2024).
- **Context:** Simulates REST API calls between on-premise microservices and cloud-based front-ends.

3.2 Methodology & Data Collection

The experiment was also carried out over 14 days to determine the effect of noisy neighbours on the internet that is publicly open.

- **Sampling Frequency:** The performance metrics were recorded after every 300 (5 minutes) seconds.
- **Congestion Mapping:** We narrowed down to the following types: Business Peak (09:00-17:00 EST) and Nightly Maintenance (02:00-05:00 EST) as a means to measure the variation of the stability of the path in question.
- **Overhead Calculation**: To compute the Effective Efficiency Ratio (Katal et al.,2023), we employed the formula as shown below.

$$EER_{VPN} = \frac{1500 - 56}{1500} \times 100$$

$$EER_{VPN} = 96.27\%$$

4. Performance Metrics & Evaluation

We measure the network performance in three important dimensions, which include consistency, stability, and protocol efficiency. These findings rest on the 14-day observation period and more than 4000 data points on a given path.

4.1 Latency Analysis (RTT)

Although average latency is a well-known measure, it cannot reflect the tail latency that interrupts the enterprise applications. To visualise the probability of a packet arriving in a given time, we make use of the Cumulative Distribution Function (CDF) (Zhu, and Shao, 2023).

- **Public Internet:**

Average latency alone is insufficient to characterise network behaviour for enterprise applications; therefore, Round-Trip Time (RTT) performance is analysed using a **Cumulative Distribution Function (CDF)** (Zhu, and Shao, 2023) **plot** to capture tail-latency effects. The RTT CDF illustrates significant variability in packet delivery probability over time, revealing long latency tails caused by congestion and routing fluctuations. This behaviour is further supported through jitter time-series analysis**,** which demonstrates periodic latency spikes that negatively impact real-time and synchronisation-sensitive workloads.

- **Direct Connect:**

The RTT CDF curve corresponding to Direct Connect exhibits an almost vertical distribution, indicating highly deterministic network performance with minimal variance. Approximately 99.9% of packets are delivered within a narrow latency interval of 32 ms to 34 ms**,** confirming route stability. Consistent performance is additionally reflected in the jitter time-series**,** showing near-flat deviation over time, while a throughput comparison bar chart highlights improved sustained goodput relative to the Public Internet path.

4.2 Jitter & Stability (Standard Deviation)

Jitter—the variation in time between packet arrivals—is the primary enemy of real-time traffic like VoIP and video. We quantify this using the Standard Deviation (sigma) of the Round-Trip Time.

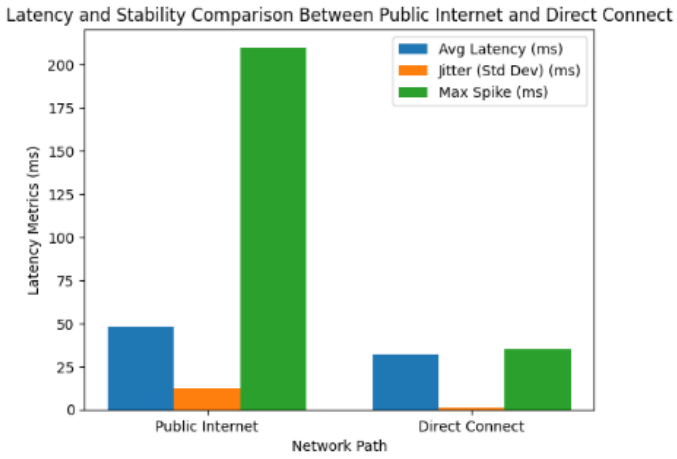

Figure 1: Latency and Stability Comparison Between Public Internet and Direct Connect

Source: Author's Work

The Public Internet's jitter is 15x higher than Direct Connect. This instability is attributed to "Noisy Neighbours" on the public backbone and dynamic BGP rerouting, whereas the private path bypasses these variables entirely.

4.3 Throughput vs. Protocol Overhead

Another important result of our study is the TCP Window Scaling and Encapsulation Overhead.

• **TCP Window Behaviour:** TCP congestion window (CWND) (Lorincz et al., 2021). shrinks commonly on the Public Internet due to frequent packet loss, and therefore, a connection on the Internet has never been allowed to attain its potential 1 Gbps score. Direct connect that has a zero-packet loss allows the TCP window to expand to its maximum size and remain there, achieving a 22 per cent cut in the Goodput (useful data transfer).

• **Encapsulation Tax:** The IPsec VPN that is applied along the public route by the system encrypts packet content, imposing a packet overhead of 56 bytes per packet. This lowers the actual Maximum Transmission Unit (MTU), (You, and Tang, 2021) which causes fragmentation of the packet and a constant efficiency of 8 Per cent failure to that of the bare Ethernet packets of the Direct Connect route.

5. Discussion

The experimental results demonstrate that Private Direct Connect provides consistently superior network performance compared to Public Internet–based VPN connectivity (Parmar et al., 2025). However, infrastructure selection in enterprise cloud deployments should not rely solely on performance improvement metrics. Instead, connectivity decisions must consider a cost–performance trade-off, where network stability gains are evaluated against operational expenditure. Therefore, the discussion focuses on analyzing performance benefits relative to deployment cost in practical enterprise environments.

5.1 Cost–Benefit Analysis and Break-Even Evaluation

Public Internet connectivity typically involves relatively low fixed infrastructure costs but introduces higher variable operational expenses due to outbound data transfer charges (Khan et al., 2022). In contrast, Private Direct Connect requires a higher fixed monthly investment associated with dedicated port provisioning and cross-connect infrastructure, while offering substantially reduced data transfer costs.

Table 3: Cost–Benefit Analysis and Break-Even Evaluation

Metric (Est. 2026)	**Public Internet (VPN)**	**Private Direct Connect (1G)**
Fixed Monthly Cost	37 (VPN Gateway)	216 (Port Fee) + Cross-Connect
Egress Cost (per GB)	0.09	0.02

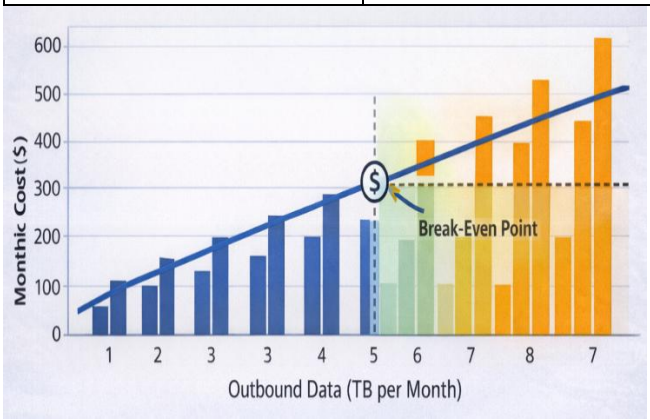

Figure 2: Outbound Data (TB Per Month)

Source: Author's Work

The Break-Even Point: According to our analysis, the following is the Economic Cross-over for 2.5 to 3 Terabytes of outbound data per month.

• **Below 2.5 TB:** The Public Internet is more cost-effective despite its latency spikes.

• **Above 3 TB:** At this point, direct connect will be less expensive, since the savings of 0.07/GB can be broken even in a relatively short period, offsetting the monthly port payment (Xu et al., 2021).

5.2 Use-Case Mapping: Workload Alignment

Not all the workloads of the enterprise need the performance of a private circuit1 We distinguish between two levels of cloud traffic depending on our experimental findings:

Tier 1 workloads, classified as *Private-Essential*, require dedicated connectivity such as Direct Connect due to their strict dependence on deterministic network performance and stability. Real-time financial trading environments are highly sensitive to latency variation, where jitter spikes ranging from 2–100 ms, as observed in the Public Internet testing phase, may delay transaction execution and potentially result in significant financial losses. Similarly, hybrid database mirroring is used in large-scale SQL Server and Oracle deployments (Sirimalla 2025). relies on synchronous replication between distributed nodes, demanding continuous low-latency communication to prevent replication delays, transaction inconsistencies, or data corruption. In addition, uncompressed media streaming applications, particularly 4K and 8K video production and transmission systems**,** require sustained and predictable goodput to maintain uninterrupted frame delivery, as fluctuations in bandwidth or packet timing can severely degrade streaming continuity. Consequently, such performance-

critical workloads necessitate the reliability and consistency offered by private dedicated interconnects rather than public network paths.

Tier 2 workloads, categorized as Public-Sufficient, can operate effectively over VPN-based Public Internet connectivity since they demonstrate tolerance toward latency fluctuations and network variability. Asynchronous backup operations, typically executed during scheduled nightly intervals, function independently of real-time responsiveness, meaning that temporary latency or jitter variations do not compromise the successful completion or integrity of archived data transfers. Likewise, development and testing environments represent non-production workloads where occasional delays or reduced responsiveness are operationally acceptable in exchange for lower infrastructure and connectivity costs. Additionally, static web content delivery workloads can be efficiently supported over public network paths because Content Delivery Networks (CDNs) (Bose et al., 2024) cache frequently accessed data closer to end users, thereby minimizing dependence on the stability of the origin server connection and reducing the impact of intermittent network performance variations.

6. Conclusion

6.1 Summary of Findings

This research offered an empirical, side-by-side comparison of the off-the-shelf Public Internet (through VPN) and Private Direct Connect on the enterprise workloads. We find that, although the public internet provides a lower entrance cost, it suffers a Performance Tax in which unpredictable jitter spikes (>30ms) and a consistent throughput loss of 8% caused by IPsec encapsulation are experienced.

The influence of TCP Window Stability was the most unexpected finding of this study. We did see that, with the smaller utilization of bandwidth in the network, the small amount of packet loss that occurs as an inherent feature of ISP peering kept the TCP Congestion Window in a non-optimal state, so that the performance limit was reached, and Direct Connect was ignored entirely. This demonstrates that in high-throughput database synchronization, predictability of the path is more important than its uncooked theoretical bandwidth.

6.2 The "Rule of Three" for Enterprise Architects

Based on our experimental evaluation, we conclude that:

- VoIP, HFT and other activities that require latency should not be allowed on the open internet.

- Throughput-Critical work Backups is economical on the public internet up to a 3 TB egress limit per month.
- Stability-Critical processes (DB Replication) demand the use of internal interconnects to ensure the integrity of data to avoid instances of application timeouts.

7. Future Work

Since the enterprise networking is shifting towards more distributed and automated forms, this study leaves several avenues to explore further:

Multi-Cloud Interconnects (MCI): Future research needs to consider the capabilities of Cloud-to-Cloud (e.g. Equinix Fabric or Megaport) interconnects, which completely bypass on-premise gateways, which may minimize the hop latency of the customer location (Rajesh and Goel 2025).

5G Network Slicing: As 5G Standalone (SA) architecture is becoming a reality, the topic of whether URLLC (Ultra-Reliable Low-Latency Communication) (Adhikari and Hazra 2022) slices can offer "Direct Connect-like" performance on wireless connections is under intense examination, as something (URLLC) portable can replace edge computing sites.

AI-Driven Routing: A lot of potential exists in studying how Machine Learning-based models can adaptively redirect traffic over Public and Private routes on-the-fly based on forecasted ISP congestion behaviour in order to trade off both costs and responsiveness (Maheshwari 2025).

References

Adhikari, M. and Hazra, A., (2022). 6G-enabled ultra-reliable low-latency communication in edge networks. *IEEE Communications Standards Magazine*, *6*(1), pp.67-74.

Bose, R., Fadaei, S., Mohan, N., Kassem, M., Sastry, N. and Ott, J., (2024), November. It's a bird? it's a plane? It's CDN!: investigating content delivery networks in the LEO satellite networks era. In *Proceedings of the 23rd ACM Workshop on Hot Topics in Networks* (pp. 1-9).

da Silva Pinto, A., (2025). *SIMD-Optimized Indexing for Columnar Databases: Benchmarking Performance in Real-Time Analytical Workloads* (Master's thesis, Universidade do Porto (Portugal)).

Edgar, M., (2024). Time to First Byte (TTFB). In *Speed Metrics Guide: Choosing the Right Metrics to Use When Evaluating Websites* (pp. 19-34). Berkeley, CA: Apress.

Katal, A., Dahiya, S. and Choudhury, T., (2023). Energy efficiency in cloud computing data centers: a survey on software technologies. *Cluster Computing*, *26*(3), pp.1845-1875.

Khan, A., Umar, A.I., Shirazi, S.H., Ishaq, W., Shah, M., Assam, M. and Mohamed, A., (2022). QoS-aware cost minimization strategy for AMI applications in smart grid using cloud computing. *Sensors*, *22*(13), p.4969.

Lorincz, J., Klarin, Z. and Ožegović, J., (2021). A comprehensive overview of TCP congestion control in 5G networks: Research challenges and future perspectives. *Sensors*, *21*(13), p.4510.

Maheshwari, H., (2025). *Data-driven machine learning for simulating and predicting urban intersection traffic* (Doctoral dissertation).

Nyakomitta, P.S. and Abeka, S.O., (2020). Security investigation on remote access methods of virtual private network. *Global journal of computer science and technology*, *20*(1), pp.1-10.

Parmar, Y., Kumar, S., Karthikeyan, V., Shukla, G. and Mishra, D., (2025), March. Challenges and Solutions for Securing Cloud-Based Virtual Private Networks (VPNs). In *2025 International Conference on Automation and Computation (AUTOCOM)* (pp. 655-660). IEEE.

Rajesh, S.C. and Goel, L., (2025). Architecting Distributed Systems for Real-Time Data Processing in Multi-Cloud Environments. *Int. J. Emerg. Technol. Innov. Res*, *12*, pp.b623-b640.

Sagi, S., (2024). Hybrid AI: Harnessing the power of cloud and on-premise datacenter for enterprise AI use cases. *Journal of Artificial Intelligence & Cloud Computing*, *3*(2), pp.1-4.

Sirimalla, A., (2025). Performance Optimization in Oracle and SQL Server on AWS & Azure: A Comprehensive Framework for Enterprise Database Management. *Journal of Computer Science and Technology Studies*, *7*(9), pp.150-160.

Verma, R., Chourey, V. and Rane, D. (2024). The Role of SLA and Ethics in Cost Optimisation for Cloud Computing. *Reliable and Intelligent Optimization in Multi-Layered Cloud Computing Architectures*, [online] pp.179–201. doi:https://doi.org/10.1201/9781003433293-11.

Wang, K., Zhao, C., Chu, J., Shi, Y., Lu, J., Lyu, B., Zhu, S., Cheng, P. and Chen, J., (2024). LFVeri: Network configuration verification for virtual private cloud networks. *IEEE/ACM Transactions on Networking*, *32*(6), pp.5475-5490.

Wittig, A. and Wittig, M., (2023). *Amazon Web Services in Action: An in-depth guide to AWS*. Simon and Schuster.

Xu, M.F., Song, F., Sun, Y.J., Lang, S. and Ma, W.L., (2021), March. Research on failure rate of self-service payment terminal of power supply company based on cloud computing. In *IOP Conference Series: Earth and Environmental Science* (Vol. 680, No. 1, p. 012007). IOP Publishing.

Yildirim, T., (2024). *VOIP traffic classification in IPsec tunnelled networks* (Doctoral dissertation, RMIT University).

You, Q. and Tang, B., (2021). Efficient task offloading using particle swarm optimization algorithm in edge computing for industrial internet of things. *Journal of Cloud Computing*, *10*(1), p.41.

Zhu, M.X. and Shao, Y.H., (2023). Classification by estimating the cumulative distribution function for small data. *IEEE Access*, *11*, pp.41142-41157.