

Analysis of Proxy Server Log Files for User Prioritization Based on Data Usage to Control Internet Access within SEUSL

R.P.L.Y. Kumari, and P. Pirapuraj

Abstract Efficient management of shared internet resources remains a critical challenge in higher education institutions, particularly under limited bandwidth environments. This paper presents a data-usage-based user prioritization framework for adaptive internet access control at the Southeastern University of Sri Lanka (SEUSL). The proposed approach leverages proxy server log files generated by CC-Proxy to identify, classify, rank users according to real-time data consumption patterns and operational requirements, and control their access. A Python-based automation algorithm performs continuous log analysis and dynamically enforces prioritization policies to enable real-time monitoring and adaptive bandwidth allocation. The system is deployed within the Faculty of Technology's network infrastructure and evaluated using network utilization, congestion levels, and user experience metrics. Experimental results demonstrate improved bandwidth utilization efficiency, reduced network congestion, and enhanced service availability for high-priority users. The study further addresses practical constraints in academic network management, including bandwidth limitations, data security, and regulatory compliance, and discusses implementation strategies to balance control and usability. The findings indicate that proxy log-driven user prioritization provides an effective and scalable solution for intelligent internet resource management in university networks.

Index Terms— Proxy log analysis, User prioritization, CC-Proxy, Python automation, Real-time bandwidth management

I. INTRODUCTION

IN the rapidly evolving landscape of digital infrastructure management, higher education institutions face continuous challenges in allocating limited network resources while maintaining equitable and efficient access for a diverse user base. Campus networks must simultaneously support academic, administrative, and research activities, often under bandwidth and regulatory constraints. This pressure is exacerbated by growing demands for real-time services, multimedia content, and cloud-based applications, which place unprecedented strain on institutional networks. Effective resource and bandwidth management strategies are therefore critical for maintaining service quality and minimizing congestion in such environments.

Recent research has highlighted the critical importance of data-driven approaches to network resource management in educational environments. Web and proxy log analysis has been widely adopted to understand user behavior patterns and optimize bandwidth allocation policies [1], with studies demonstrating that adaptive bandwidth management systems can dynamically adjust network resources based on users' access to educational content [2][3]. These analytical

approaches have proven essential for reducing network congestion, improving response times, and enhancing overall service quality in campus networks [4]. Furthermore, implementing smart bandwidth allocation strategies ensures seamless online experiences and enhances overall productivity in educational institutions [5], while empirical studies have shown measurable improvements in network performance metrics when proxy servers are configured with optimized caching, filtering, and bandwidth allocation settings [4].

Resource management at network edges has also been recognized as a key component of scalable distributed systems [6]. Bandwidth management frameworks employing proxy servers and traffic monitoring have shown notable improvements in network throughput and efficiency [7]. Conceptual frameworks for intelligent campus resource allocation highlight the growing role of data analytics and adaptive decision-making in educational infrastructure [8]. Proxy server logs, when properly analyzed, provide rich behavioral insights that can facilitate more effective resource allocation beyond simple traffic accounting, including user behavior and usage trends derived from log metadata [9].

Despite advances in bandwidth allocation and traffic classification [10], there remains a substantial gap in research specifically addressing user-centric prioritization in educational networks based on real-time log analysis. Many existing approaches either rely on static policies or focus on network layer metrics without considering the diverse roles and requirements of users in an academic background. This gap underscores the need

R.P.L.Y. Kumari is a graduate from the Department of ICT, South Eastern University of Sri Lanka, Sri Lanka. (Email: kshani97yasanthika@gmail.com)

P. Pirapuraj is a Senior Lecturer at the Department of ICT, South Eastern University of Sri Lanka, Sri Lanka. (Email: pirapuraj@seu.ac.lk)

for mechanisms that dynamically rank and allocate resources based on context-aware usage metrics derived from proxy logs.

In this study, we propose a novel framework that integrates proxy server log file with real-time data analysis and a Python-based user prioritization algorithm to manage internet usage within the Southeastern University of Sri Lanka (SEUSL). The methodology leverages fine-grained proxy log analysis to identify usage patterns, enables adaptive resource allocation, and supports dynamic enforcement of data usage policies tailored to user priorities. By aligning resource allocation with actual user needs, this framework aims to maximize network utility, reduce congestion, and enhance the overall digital experience for students, faculty, and administrators. The key contributions of this work are:

- A real-time, user-centric prioritization framework that addresses the gap in context-aware resource allocation for educational networks
- A Python-based adaptive algorithm for dynamic bandwidth management based on fine-grained proxy log analysis and user role classification
- A practical implementation at SEUSL demonstrating improved network utility through data-driven resource allocation

II. RELATED WORK

Proxy server log analysis has been extensively employed to characterize user behavior and classify network traffic based on attributes such as IP address, bandwidth consumption, session duration, and file type [11]. Studies in educational environments have shown that proxy logs can identify user sessions and accessed websites, revealing predominant engagement with search engines, social media, news portals, and information resources [12]. These methodologies establish the foundation for data-driven network management by transforming raw log data into actionable intelligence for bandwidth optimization and policy enforcement [13].

Bandwidth allocation constitutes a primary challenge for university IT departments, directly impacting educational quality [14]. Universities employ diverse approaches from rate-limiting to monitoring anomalous consumption patterns. Adaptive systems dynamically adjust allocations based on educational content accessibility, while policy-based approaches prioritize academic applications during peak hours with flexible off-peak restrictions [15]. These strategies reflect the evolution from reactive to proactive bandwidth management in academic settings.

Role-based access control has been widely adopted in educational systems to improve security through user classification and role authorization [1]. Institutions require differentiated network provisioning for faculty, staff, students, and visitors based on roles and access times [16]. Network Access Control solutions enable granular policy enforcement aligned with individual responsibilities, defining hierarchical structures with differentiated permissions for academic records and enrollment data [17]. RBAC simplifies security management and enables regulatory compliance with FERPA and HIPAA [18], providing frameworks for user prioritization extending to dynamic resource allocation.

In addition to proxy log analysis, broader traffic classification techniques have been developed using deep packet inspection and machine learning to optimize quality of service (QoS) and traffic prioritization [10]. Recent work in software-defined networking (SDN) has demonstrated how centralized control and machine learning-based traffic classification can further optimize network performance in dynamic environments [19]. These advancements reflect a shift from reactive to predictive network management, essential for supporting cloud services, multimedia delivery, and real-time collaboration in modern educational environments. Table 1 summarizes and compares the related works with the proposed approach.

III. DESIGN AND METHODOLOGY

This section presents a comprehensive methodology for the design, implementation, and evaluation of a data-driven user prioritization framework for intelligent internet access control in an academic network environment. The proposed system integrates real-time proxy log analysis, adaptive prioritization algorithms, and dynamic bandwidth allocation mechanisms to optimize network resource utilization while ensuring equitable access across heterogeneous user groups.

A. System Architecture and Overview

The proposed framework follows a closed-loop architecture consisting of four core components: (i) a proxy server infrastructure for traffic interception and log generation, (ii) a real-time log collection and preprocessing module, (iii) a user identification and prioritization engine, and (iv) a dynamic resource allocation and policy enforcement module.

Figure 1 illustrates the overall system architecture and operational workflow, showing the flow of proxy traffic through CC-Proxy, continuous log extraction, feature processing, priority computation, and real-time enforcement of bandwidth and quota policies. This architecture enables continuous monitoring and adaptive control, allowing the system to respond dynamically to changing network conditions and evolving user behavior.

B. Proxy Server Configuration and Log Collection

The experimental deployment utilizes CC-Proxy as the primary proxy server within the Faculty of Technology network at the Southeastern University of Sri Lanka (SEUSL). CC-Proxy was selected due to its robust logging support, fine-grained access control, and seamless integration with existing campus infrastructure. The proxy server is configured to intercept all HTTP and HTTPS traffic and generate detailed log records for every transaction.

Each proxy log entry contains structured metadata including timestamps, source IP and MAC addresses, destination URLs, request and response sizes, protocol types, connection durations, and HTTP response codes. These attributes provide a comprehensive representation of user behavior and traffic characteristics. Log collection is performed continuously using a Python-based collection module that retrieves newly generated log entries at five-minute intervals. This near real-time acquisition minimizes processing latency while ensuring stability under high traffic volumes.

TABLE I
COMPARISON OF RELATED WORK AND THE PROPOSED APPROACH

Work	Data Source	Real-Time Analysis	User-Centric Prioritization	Dynamic Quota Enforcement	Deployment Complexity	Academic Network Focus
Proxy log analysis [11], [12]	Proxy logs	X	X	X	Low	Yes
Bandwidth control policies [14], [15]	Traffic statistics	Partial	X		Medium	Yes
RBAC-based access control [1], [16], [17]	User roles	X		X	Low	Yes
ML-based [10]	Packet	X	X	X	High	No
SDN-based network control [19]	SDN controller	✓		X	Very High	No
Proposed approach	Proxy logs	✓	✓	✓	Low	Yes



Fig. 1: System overview of proposed framework

C. Data Preprocessing and User Identification

Raw proxy logs undergo systematic preprocessing prior to analysis. Python scripts employing the Pandas library parse log entries, normalize timestamps, and extract relevant features. Invalid, incomplete, or corrupted records are filtered to maintain data integrity. User identification is performed primarily using validated source IP addresses through the IP address library. Active user sessions are maintained in a dynamic registry that tracks historical usage, session state, and current priority assignments. Session boundaries are defined using a 30-minute inactivity timeout, after which a session is terminated and a new session is initiated upon renewed activity.

Traffic metrics are aggregated over sliding time windows to derive session-level features, including cumulative data consumption, request frequency, average and peak bandwidth usage, total connection duration, number of concurrent connections, and diversity of accessed resources.

D. User Prioritization Algorithm

User prioritization is implemented using a multi-criteria scoring model that incorporates institutional roles, real-time usage behavior, and temporal context. Role-based prioritization assigns baseline scores according to predefined institutional categories such as faculty members, administrative staff, postgraduate students, undergraduate students, and visitors. These scores reflect operational importance and access requirements defined by university policy.

Dynamic adjustments are applied based on observed usage patterns. Moderate and consistent usage results in favorable priority adjustments, while excessive or bursty consumption triggers priority reductions to promote fairness. Temporal

weighting further adjusts priorities to favor academic hours and research-intensive periods.

The composite priority score for user u at time t is computed as:

$$P(u, t) = \alpha \cdot R(u) + \beta \cdot B(u, t) + \gamma \cdot F(u, t) + \delta \cdot T(t) \quad (1)$$

where: $R(u)$ is the role-based priority score, $B(u, t)$ is the bandwidth usage normalization factor, $F(u, t)$ is the session frequency factor, and $T(t)$ is the time-based weighting factor reflecting temporal importance. The coefficients $\alpha, \beta, \gamma, \delta$ are tunable parameters satisfying:

$$\alpha + \beta + \gamma + \delta = 1 \quad (2)$$

To penalize excessive bandwidth consumption during congestion periods, the bandwidth usage factor is defined as an inverse function of cumulative data usage:

$$B(u, t) = 1 + \frac{1}{1 + D(u, t)} \quad (3)$$

where $D(u, t)$ denotes the cumulative data consumption (in MB) of user u within the current observation window. This formulation ensures that users with higher bandwidth consumption receive lower priority scores. The session frequency factor $F(u, t)$ rewards distributed access patterns over bursty usage, while the temporal factor $T(t)$ assigns higher weights during academic and research-intensive periods.

Active user sessions are maintained in a priority queue implemented using Python's `heapq` library. Priority scores are continuously recomputed as new proxy log entries are processed, enabling real-time reordering of sessions.

Bandwidth allocation is performed dynamically using a weighted fair-share mechanism based on computed priority scores. The bandwidth allocated to user u at time t is given by:

$$BW(u, t) = BW_{\text{total}} \frac{P(u, t)}{\sum_{i=1}^N P(i, t)} \quad (4)$$

Where, BW_{total} is the available network bandwidth and N is the number of active users. High-priority users are guaranteed minimum bandwidth levels, while low-priority users may experience throttling during periods of congestion. In parallel, the system enforces role-based data usage quotas defined on daily, weekly, or monthly intervals. Automated responses are triggered at predefined thresholds, including warning notifications at 80% usage, bandwidth throttling at 100%, and temporary access suspension beyond 120%.

Policy enforcement is achieved through automated configuration updates deployed to CC-Proxy at five-minute intervals using Python scripts, ensuring low-latency and stable application of prioritization and quota policies.

E. Experimental Setup and Evaluation

The system is deployed within the Faculty of Technology network at SEUSL, serving approximately 1,200 users over a six-month period (January–June 2024). The dataset comprises approximately 15 million proxy log entries, representing more than 50,000 unique user sessions and approximately 2.5 TB of cumulative traffic.

Performance is evaluated using metrics including bandwidth utilization efficiency, Jain's fairness index, response time, session completion rate, quota violation frequency, processing latency, and computational overhead.

F. Ethical Considerations

User privacy is preserved through irreversible anonymization of IP addresses using SHA-256 hashing. The system complies with institutional data protection policies, limits data retention to 180 days, and provides transparency through user-accessible dashboards displaying individual usage statistics and quota status.

IV. RESULTS AND EVALUATION

This section presents the experimental results and performance evaluation of the proposed real-time user prioritization and internet usage control framework. The evaluation focuses on usage pattern identification, dynamic resource allocation effectiveness, congestion mitigation, user satisfaction, and overall network performance improvement.

A. Data Collection and Preparation

Prior to analysis, the collected proxy log data underwent systematic cleaning and preprocessing to ensure data quality and reliability. This process included the removal of duplicate entries, irrelevant records, and statistical outliers that could distort usage measurements. Missing values and incomplete session records were excluded from further analysis. The resulting dataset provided a clean and consistent basis for real-time monitoring, prioritization, and performance evaluation.

B. Real-Time Analysis and User Prioritization Results

The Python-based real-time analysis and user prioritization system produced several significant outcomes.

1) Identification of Usage Patterns

Real-time proxy log analysis revealed distinct and role-dependent usage patterns among different user groups. Academic staff and researchers predominantly accessed

research databases, academic repositories, and online collaboration platforms, exhibiting steady and moderate bandwidth consumption.

In contrast, students primarily engaged in course-related activities such as learning management systems, online assessments, and resource downloads, with usage intensity varying across academic periods. These findings validate the necessity of role-aware and behavior-driven prioritization in educational networks.

2) Dynamic Resource Allocation

Dynamic bandwidth allocation based on computed user priority scores was effective in practice. Users assigned higher priority levels consistently received a larger share of available network resources during congestion periods. This prioritization mechanism resulted in more equitable and efficient resource utilization by preventing excessive bandwidth consumption by non-critical activities while preserving access for essential academic tasks.

3) Congestion Mitigation and Latency Reduction

The adaptive prioritization and resource allocation mechanisms significantly reduced network congestion during peak usage hours. Real-time adjustments to bandwidth allocation ensured that critical academic services experienced lower latency and faster response times. As a result, essential services such as research portals and learning platforms maintained stable performance even under high load conditions.

4) Enhanced User Satisfaction

User feedback collected through periodic surveys indicated a noticeable improvement in overall user experience. Faculty members reported enhanced research productivity due to reliable access to online resources, while students experienced smoother and more consistent access to academic materials. Reduced service disruptions and improved responsiveness contributed to higher satisfaction levels across all user categories.

C. Data Usage Limitation Management

The implementation of data usage limitation mechanisms yielded positive outcomes in terms of policy enforcement and resource conservation. Users generally complied with assigned data quotas, and instances of quota violations triggered automated actions such as bandwidth throttling or temporary access suspension. These mechanisms ensured adherence to institutional policies and regulatory requirements. Additionally, the introduction of data limits encouraged responsible usage behavior, leading to reduced excessive consumption and fairer distribution of network resources.

D. Performance Metrics and Quantitative Evaluation

To quantitatively assess system performance, several key metrics were analyzed. Network efficiency was measured as the percentage of time the network operated within predefined congestion thresholds. Following system deployment, network efficiency consistently exceeded 90%, representing a substantial improvement in reliability and performance compared to the pre-deployment state. User satisfaction surveys further confirmed these improvements, with respondents reporting fewer disruptions, faster access times, and enhanced usability of digital services.

E. Multi-Tier User Prioritization Framework

To optimize resource allocation, a multi-tier prioritization model was implemented within the Python-based control logic. Users were classified into four priority levels based on historical usage patterns and application requirements, as summarized in Table I.

For example, users engaged in critical academic activities were assigned to the High 1 category, while users accessing non-essential multimedia services were assigned lower priority to preserve bandwidth availability.

F. Management of Heavy Data Users

To further control network congestion, dynamic reprioritization was applied to users exhibiting excessive data consumption. Table II illustrates how user priorities were adjusted based on observed behavior.

TABLE II
USER PRIORITIZATION LEVELS

Level	Description	Users
High 1	Users requiring the highest network resource priority.	Academic Requirement
High 2	User with high data need with low Bandwidth	Academic Requirement but not critical
Medium	User moderate data consumption	General browsing and email.
Low	User require HD and high data rate	Video, streaming,

TABLE III
MANAGEMENT OF HEAVY DATA USERS

Metric	Before Implementation	After Implementation
Network Congestion	High	Low
User Satisfaction	Mixed	Positive
Peak hour Disruptions	Frequent	Rare

TABLE IV
USER BLOCKING PATTERN

User	Original Priority	New Priority
User A	High 2	High
User B	High 2	Medium
User C	Medium	Low

For instance, User A initially categorized as High 2 was promoted due to reduced bandwidth demand relative to other active users, whereas User B and User C experienced priority reduction due to increased data consumption.

G. Exceedance Blocking Behavior

The system also enforced daily data usage thresholds to prevent excessive consumption. Table III presents sample exceedance blocking outcomes.

TABLE V
NETWORK PERFORMANCE COMPARISON

User	Daily data usage	Blocked?
User D	215	Yes
User E	180	No
User F	202	Yes

In this example, users exceeding the 200 MB daily threshold were temporarily blocked, while compliant users retained uninterrupted access.

H. Impact on Overall Network Performance

The combined effect of prioritization, dynamic allocation, and quota enforcement resulted in significant improvements in network performance, as summarized in Table IV. The results demonstrate that the proposed framework effectively improves network stability, fairness, and user experience in a real-world academic environment.

V. DISCUSSION

The experimental results demonstrate that real-time proxy log-driven user prioritization can significantly improve network performance and fairness in an academic environment. The observed reduction in congestion during peak hours and the improvement in user satisfaction highlight the effectiveness of combining role-based prioritization with adaptive bandwidth allocation. Notably, the results indicate that lightweight, proxy-based solutions can achieve performance gains comparable to more complex network architectures, such as software-defined networking, without requiring major infrastructure changes. However, the prioritization outcomes are influenced by predefined role weights and usage thresholds, suggesting that careful parameter tuning is essential to balance fairness and performance. These findings emphasize the practical viability of log-based prioritization while also revealing opportunities for further refinement through more adaptive and predictive mechanisms.

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

This study addresses key challenges in academic network management, including user prioritization, data usage control, and congestion mitigation, within the Faculty of Technology at

the Southeastern University of Sri Lanka (SEUSL). By integrating CC-Proxy for detailed traffic logging with Python-based real-time analysis, prioritization, and quota enforcement, the proposed framework demonstrates significant improvements in bandwidth utilization efficiency, network stability, and user experience.

Real-time user prioritization based on institutional roles and usage behavior enabled more effective resource allocation, reducing congestion and latency during peak periods. The implementation of data usage limitations promoted responsible consumption and ensured compliance with institutional policies and regulatory requirements. User feedback further confirmed enhanced satisfaction, with faculty members reporting improved research productivity and students experiencing more reliable access to academic resources. Overall, the sustained network efficiency exceeding 90% within congestion thresholds highlights the robustness and practicality of the proposed solution.

B. Future Work

While the proposed framework has proven effective, several directions remain open for future research. Fine-grained prioritization strategies can be explored by incorporating additional contextual factors such as specific research projects, departmental priorities, or application-level requirements. The integration of machine learning techniques for predictive traffic analysis and anomaly detection could further enhance adaptability and responsiveness. Extended user behavior analysis may support more personalized and proactive bandwidth management policies. Additionally, future work should address advanced cybersecurity and data privacy considerations, as well as scalability challenges, to ensure the framework remains effective as institutional networks continue to expand.

REFERENCES

- [1] Y. Lan, J. Qu, and J. Chen, "Research on Internet Surfing Behavior of College Students Based on Big Data," vol. 4, no. 6, pp. 21–34, 2021, doi: 10.25236/AJCIS.2021.040605.
- [2] R. K. Paredes and A. A. Hernandez, "Designing an Adaptive Bandwidth Management for Higher Education Institutions," vol. 2, no. 1, pp. 17–35, 2018, doi: 10.25147/ijcsr.2017.001.1.22.
- [3] M. M. H. Ibrahim, M. H. Omar, A. M. M. Habbal, and K. M. Zaini, "ANALYSIS OF INTERNET TRAFFIC IN EDUCATIONAL NETWORK BASED ON USERS' PREFERENCES," vol. 10, no. 1, pp. 99–105, 2014, doi: 10.3844/jcsp.2014.99.105.
- [4] J. L. B. Wenceslao and R. B. Wenceslao, "Network Performance of Proxy-Enabled Server Using Three Configurations," pp. 850–856, 2022.
- [5] G. Tilaye and L. A. Gojeh, "Use of Access Control List Application for Bandwidth Management Among Selected Public Higher Education Institutions in Ethiopia," vol. 8, no. 1, pp. 24–35, 2020, doi: 10.13189/csit.2020.080103.
- [6] S. Trindade, L. F. Bittencourt, and N. L. S. Fonseca, "Resource management at the network edge for federated learning," *Digit. Commun. Networks*, vol. 10, no. 3, pp. 765–782, 2024, doi: 10.1016/j.dcan.2022.10.015.
- [7] A. Gafur, "Bandwidth Management Analysis Using Proxy Server-Based Method Virtual Machine," no. Icaisd 2023, pp. 118–125, 2024, doi: 10.5220/0012444700003848.
- [8] C. Koukaras, E. Hatzikranielis, M. Mitsiaki, P. Koukaras, C. Tjortjijis, and S. G. Stavrinides, "Revolutionising Educational Management with AI and Wireless Networks: A Framework for Smart Resource Allocation and," pp. 1–30, 2025.
- [9] R. R. Abdalla and A. K. Jumaa, "Log File Analysis Based on Machine Learning: A Survey," vol. 6, no. 2, pp. 7–10, 2022, doi: 10.21928/uhdjst.v6n2y2022.77-84.
- [10] A. Azab, M. Khasawneh, S. Alrabaa, and K. R. Choo, "Network traffic classification: Techniques, datasets, and challenges," vol. 10, no. June 2022, pp. 676–692, 2024, doi: 10.1016/j.dcan.2022.09.009.
- [11] B. Williamson, "The hidden architecture of higher education: building a big data infrastructure for the 'smarter university,'" pp. 1–26, 2018, doi: 10.1186/s41239-018-0094-1.
- [12] W. Journal, C. Wang, and D. Wang, "Managing the integration of teaching resources for college physical education using intelligent edge - cloud computing," *J. Cloud Comput.*, vol. 1, pp. 1–14, 2023, doi: 10.1186/s13677-023-00455-1.
- [13] B. Alshemaimri, A. Badshah, A. Daud, A. Bukhari, R. Alsini, and O. Alghushairy, "Regional computing approach for educational big data," pp. 1–15, 2025.
- [14] B. Fei, J. Eloff, M. Olivier, and H. Venter, "Chapter 20 ANALYSIS OF WEB PROXY LOGS," pp. 247–258, 2006.
- [15] K. Bommepally, T. K. Glisa, J. J. Prakash, S. R. Singh, and H. A. Murthy, "Internet Activity Analysis Through Proxy Log".
- [16] X. Liu, "Heliyon The educational resource management based on image data visualization and deep learning," *Heliyon*, vol. 10, no. 13, p. e32972, 2024, doi: 10.1016/j.heliyon.2024.e32972.
- [17] "What is Network Behavior Analysis (NBA) - zenarmor.com." <https://www.zenarmor.com/docs/network-security-tutorials/what-is-nba> (accessed Dec. 20, 2025).
- [18] Z. Wang, C. Zhang, Y. Ding, H., "Applied Mathematics and Nonlinear Sciences," *Appl. Math. Nonlinear Sci.*, vol. 8, no. 2, pp. 3383–3392, 2023.
- [19] R. H. Serag, M. S. Abdalzaher, H. Abd, E. Atty, and M. Sobh, "Software Defined Network Traffic Classification for QoS Optimization Using Machine Learning," *J. Netw. Syst. Manag.*, vol. 33, no. 2, pp. 1–29, 2025, doi: 10.1007/s10922-025-09911-6.