SINERGI Vol. 26, No. 2, June 2022: 145-154
http://publikasi.mercubuana.ac.id/index.php/sinergi
http://doi.org/10.22441/sinergi.2022.2.003

The process mining method approach to analyze users' behavior of internet in the Local Area Network of Sriwijaya University

Muhammad Irfan Jambak[1], Amrifan S. Mohruni[2], Muhammad Ihsan Jambak[3*], Endy Suherman[4]
[1]Department of Electrical Engineering, Faculty of Engineering, Universitas Sriwijaya, Indonesia
[2]Department of Mechanical Engineering, Faculty of Engineering, Universitas Sriwijaya, Indonesia
[3]Department of Informatics Management, Faculty of Computer Science, Universitas Sriwijaya, Indonesia
[4]Information & Communication Technology Technical Implementation Unit, Universitas Sriwijaya, Indonesia

Abstract
The Sriwijaya University internet network management unit does not yet have a standard formulation for implementing Bandwidth Management & Bandwidth Allocation. To provide the best service, they apply the Best-Effort Service concept. As a result, it requires a relatively large network capacity and bandwidth provision so that it has an impact on costs. Therefore, it is necessary to know how users use internet bandwidth as the basic principle of Bandwidth Management & Bandwidth Allocation. This study has completed how to determine the behavior of internet users on the campus LAN as a reason to evaluate internet bandwidth usage. With the Process Mining method, process mapping has been carried out for all access to internet usage from all faculties. As a result, the factors that characterize and need to be considered in bandwidth management are obtained. In order of significance are Total Access Length, Average Variance, Number of User Case IDs, Number of Non-Academic Ports, Number of Academic Ports, Number of Access Frequency, Number of Events, and Number of Ports of Service.

Keywords:
Bandwidth Management;
Business Process;
Internet;
Process Mining;

Article History:
Received: June 26, 2021
Revised: October 22, 2021
Accepted: October 29, 2021
Published: June 2, 2022

Corresponding Author:
Muhammad Ihsan Jambak
Department of Informatics Management, Faculty of Computer Science, Universitas Sriwijaya
Email: jambak@unsri.ac.id

INTRODUCTION

Quality of Service (QoS) of the internet network is determined mainly by network management, defined as the various activities, methods, and knowledge required to manage computer networks. QoS parameters consist of Bandwidth, Throughput, Latency/Delay, Jitter, and Packet Loss [1, 2, 3, 4]. QoS is also a term used to define the ability of a network to provide different levels of service assurance. Or it can be said that QoS is a network mechanism that allows services to operate according to the needs expected by users. Thus, QoS can make bandwidth, latency, and jitter predictable and tailored to the needs of the network users. There are three levels of QoS that can be used:

a. Best-Effort service: makes all efforts to deliver a packet to a destination;
b. Integrated service: provides applications with a level of service assurance that parameters have been negotiated end-to-end;
c. Differentiated service: provides a set of classification tools and a queuing mechanism for protocols with specific priorities on different networks.

Bandwidth is the maximum amount of data transferred from one point to another within a certain period, so the unit is bits per second (bps). Bandwidth management on the network is essential because bandwidth on a computer network is valuable. Several bandwidth management methods include Traffic Shaping (Rate Limiting), Scheduling Algorithms, Congestion Avoidance, Bandwidth Reservation Protocols, and Traffic Classification.

Traffic Shaping or bandwidth limitation based on data flow or data stream by increasing or decreasing the priority of the flowing packets is the easiest and most frequently used way. However, the 3rd level of QoS is Differentiated Service, so management should be the most effective and efficient QoS level. However, this service requires accurate and comprehensive data and information about the needs and behavior of network users, especially on bandwidth parameters due to the application of Traffic Classification techniques. Therefore, Bandwidth Management becomes essential because it will impact computer network performance.

Sriwijaya University (Unsri) also has an internet network in the form of a Local Area Network (LAN) that serves all the needs of the academic community, which are students, lecturers, and employees, including the increasing demand for high-speed computer interconnection between university and laboratory systems. Overall, the total bandwidth capacity of Unsri for both the Bukit Besar and Indralaya campuses is 5.5Gbps. Therefore, the availability of internet access services at Unsri should be utilized well by the academic community.

The distribution of bandwidth in Unsri, which has been done so far, is by way of Best-Effort-Service, where users can access in the form of sending and receiving extensive data at any time without the need to ask permission from the network manager. However, this has the consequence that it requires a relatively large network capacity and bandwidth provision, thus impacting the cost. In this case, the Unsri campus internet network manager, the Information & Communication Technology Technical Implementation Unit (UPT TIK), does not have a standard formulation for implementing Bandwidth Management & Bandwidth Allocation.

It is necessary to know how the users utilize internet bandwidth usage from Unsri's campus LAN to fulfil principles of bandwidth management & bandwidth allocation. Moreover, this utilization provides an overview of how users use internet network facilities in the Unsri LAN. Thus, the formulation of the problem that is solved in this research is: How to know the behavior of internet users on the Unsri LAN as rational to evaluate that internet bandwidth usage is effective and efficient?

METHOD

Regardless of the type of business operation, business is dynamic and must continue to evolve to keep up with the changing needs of the business environment to remain relevant. One of the attributes that do a great business is efficient business processes. Business processes define the tasks that must be performed to achieve predetermined business goals and objectives, such as producing products and providing services to end-users. Therefore, streamlining business processes is essential to make them more adaptable to the changing needs of the business environment, gain a competitive advantage, achieve operational excellence and improve customer experience.

Business Process Management (BPM) is a discipline in operations management in which various methods are used to discover, model, analyze, measure, improve, optimize, and automate business processes [5, 6, 7, 8]. Any combination of methods used to manage a company's business processes is BPM [9]. Operations can be structured and iterative or unstructured and variable.

BPM looks at how the business processes in an organization can be assessed and improved, and it involves continuous evaluation of the process with actions being taken to optimize it. In BPM itself, two primary areas consist of Business Process Improvement (BPI) and Business Process Modeling (BP Modeling). Both of them are aim to create an organization to look again at how organizational elements (such as people, processes, and technology) are aligned; also optimize core business processes. By mapping and analyzing business processes, they can be redesigned, improved, and better managed.

As an approach, BPM sees processes as an essential asset of an organization that must be understood, managed, and developed to announce and deliver value-added products and services to clients or customers. This approach is very similar to other total quality management or continuous improvement process methodologies. In addition, the BPM approach can be supported or activated through technology [10, 11, 12]. Because of this, BPM was often discussed from one of two perspectives: people and technology.

Business process modeling allows an organization to confirm the current state of its processes and provide opportunities to improve them, then make it more effective and efficient. With every detail of the process mapped, showing end-to-end activities, it will be easier to align the process with the goals and values of the organization. In addition to process improvement, organizations model business processes, among others: complies with regulatory bodies; assists in employee orientation/training; and facilitates internal audit.

Many businesses have adopted modeling business processes for improvement. Business Process Modeling allows organizations to graphically document their business processes, including business activities, events, flow control, stakeholders, and relations [13][14]. While a business process, in general, is a combination of operational steps and management control that, together, produce a product or provide a service.

BP Modeling involves describing the current state of business processes ("as-is") as well as analyzing and improving them (process "as-is") to create more efficient business processes ("to-be"). In addition, diagrams are used for better visualization to facilitate easier understanding to capture current business processes.

Although many organizations have used diagrams such as flowcharts, data flow diagrams, etc., to describe their processes, it is not a model. A model is an abstraction that contains all the elements needed to clarify the intent of the process being modeled. In contrast, a diagram is a specific view of what we try to understand in a particular context [15].

Behaviors are actions that are always used or performed. The process is the action stages of activity; thus, modeling a process will reflect the behavior. For example, one way to find out the behavior of internet users on the Unsri LAN is by knowing the processes carried out by these internet users. Then these processes are modeled to provide a complete picture of how these activities take place.

Process Mining is a technical group that supports business process analysis based on event logs. Using specific algorithms applied to the event log data can identify trends, patterns, and details in the event logs recorded by the information system. Process Mining aims to increase process efficiency and understand processes [16, 17, 18]. The term Process Mining comes from the data mining field. The concept is "mining" data for understanding, answering questions, or solving problems. The search is usually specific to identified challenges or obstacles in data mining. Data mining has some similarities, particularly in analyzing big data for business decision support. Process Mining uses specific algorithms against event log data to identify trends, patterns, and details of how the whole process is going.

Process Mining is an analytical discipline for discovering, monitoring, and improving actual processes by extracting knowledge from event log data available in today's information systems. Process Mining offers an objective, fact-based understanding of the existing event logs, which helps audit, analyze and improve business processes by answering questions related to compliance and performance.

There are many ways to describe Process Mining: direct, comprehensive, visual, objective. The focus is usually on extracting data from system event logs, providing complete transparency about how the business operates. With the insights gained from Artificial Intelligence and Machine Learning, organizations can then identify opportunities for process optimization. Process Mining bridges the gap between traditional model-based process analysis and data-centric analysis techniques such as Machine Learning and Data Mining [17, 19, 20, 21].

Process Mining technology isn't just about seeing where things can be improved. Figure 1 explains that, which generates a process model from event log data with no additional, non-empirical input, there are two other types of Process Mining: Conformance and Enhancement [17]. In Conformance, an existing process model is compared with an event log of the same process to check for alignment. In other words, confirming whether the process carried out follows the process model. While Enhancement will improve or refine the model using data derived from the event log. Rather than disclosing or comparing process operations, improvement aims to modify or enhance existing models. We tend to think of models as ideal scenarios of how something functions. But in the context of Process Mining, the final model should be viewed less as a fixed state and more like a map.

Its purpose is to guide the user to a destination using the best route, knowing that things will change over time. So, Process Mining technology uses a model with empirical data derived from event logs in organizational systems [22]. The research unit used was the Unsri Local Area Network Management System in this study. The research stages are as shown in Figure 2 with six steps [23], namely:

a. Planning; at this stage, the object data events log is planned to be downloaded by determining the school year and semester.
b. Data Extraction; is the process of downloading data from the server and preprocessing.
c. Data Processing; is a data processing algorithm selected by the Process Mining application.
d. Mining & Analysis; is the stage to produce a research output in the form of a Process Model.
e. Evaluation; the stage of diagnosis and review of the Process Model generated in the previous step.
f. Process Improvement; the stage of compiling follow-up suggestions.

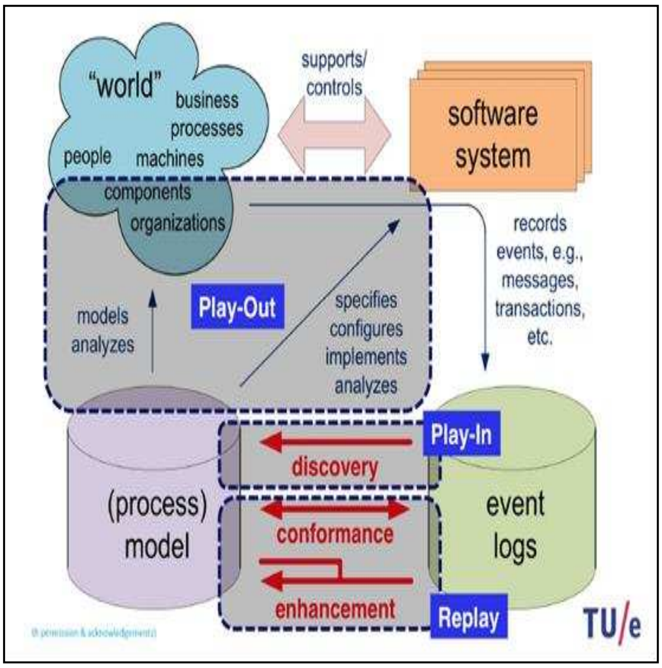


Figure 1. Process Mining: Discovery, Conformance, & Enhancement [17]

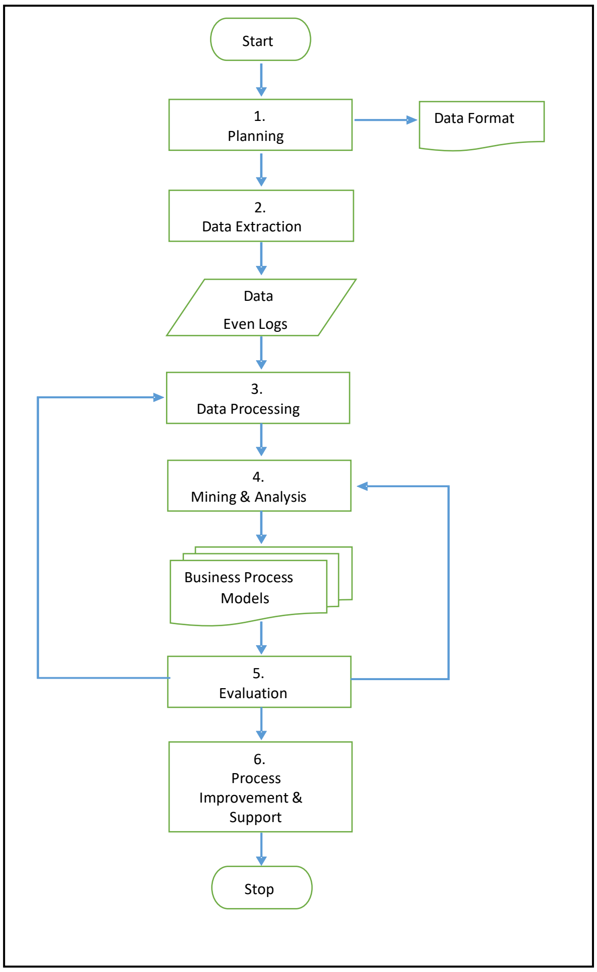


Figure 2. Stages of Research

In this study, the even log data were extracted from records on the Sangfor firewall server in the Network Operating Control room, recording from 1-7 October 2020, of all internet access users per faculty at the Unsri campus Bukit Besar, Palembang. Each user access even is logged and tracked with 12 attributes consisting of Date and Time; Services/Applications used; Protocol; Service Zone; Group; source IP; Source Port; Destination Zone; IP destination; Destination port; Policy Name; and Description. Thus, it can be known where and what the purpose of each access is; the obstacles faced are the large data size and the limited number of rows of data for recording downloads in the Microsoft Excel application. For this reason, each faculty is recorded separately.

RESULTS AND DISCUSSION

Following the process mining method using Fluxion Disco software, the event log data is then processed to produce a model that shows how the processes of using internet bandwidth all take place. However, because the data used is an event log that traces all activities carried out from the beginning of access to the end, the resulting model is a complex "spaghetti map" that is too complicated to interpret and use.

Fluxion Disco software miner is based on Christian's Fuzzy Miner, which is the first mining algorithm to introduce a "map metaphor," highlighting frequent activities and paths through color and thickness [24]. The complete picture of the process is shown, the same as when it happened. This is very important to understand the reference point of the process map because it shows a one-to-one match of data.

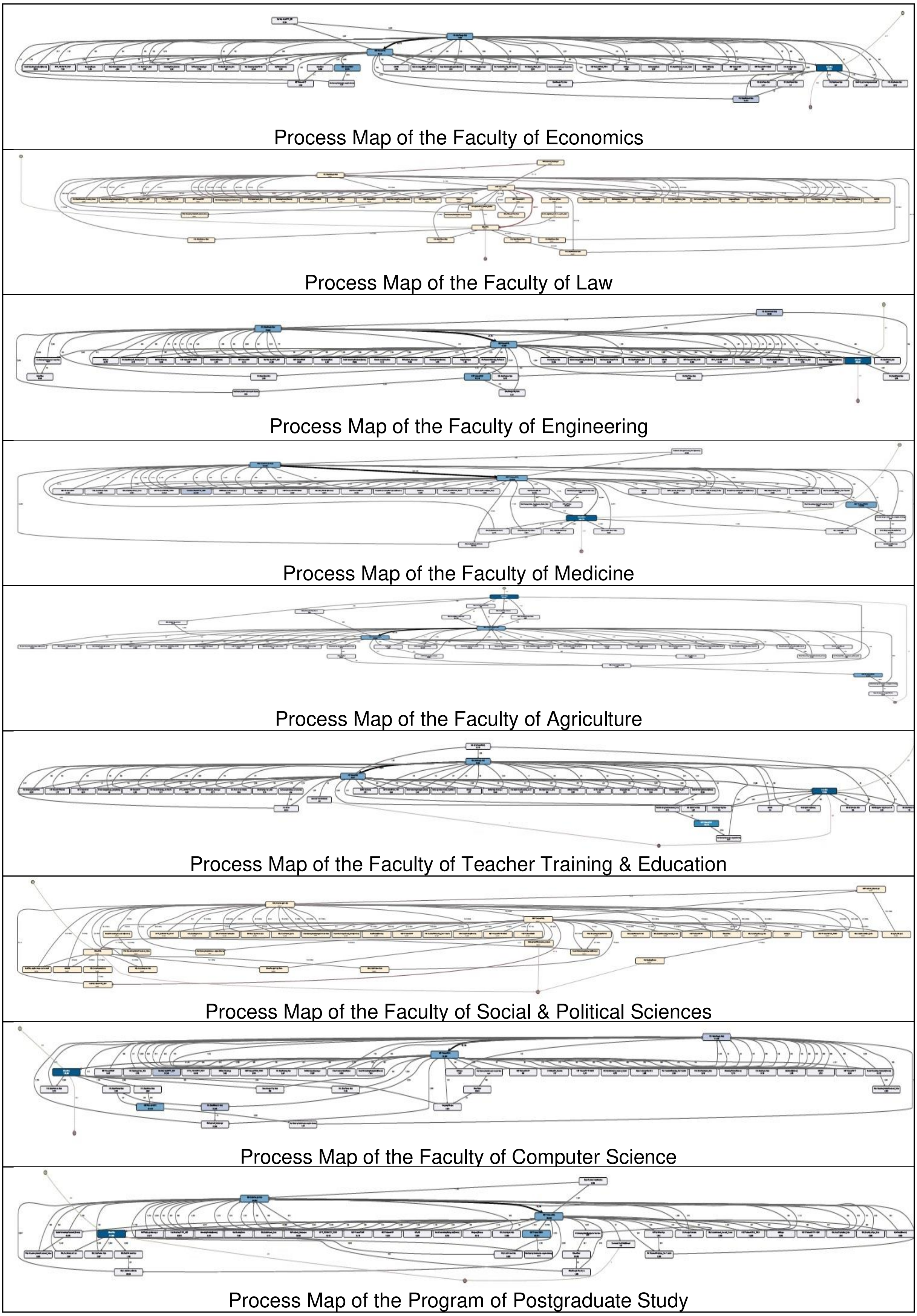
However, if no simplification strategy is applied, the entire process is usually too complicated to view in 100% detail. To get more manageable pieces, we need to break this down and simplify the process map. The resulting model as the output of data processing in the Fluxion Disco software for each faculty is shown in Figure 3. The numerical data derived from the process maps are shown in Table 1.

Each activity on the process map is accompanied by information about how often the process is repeated (frequency) as well as the length of access time. Thus, it can be clearly seen which processes or activities dominate bandwidth usage. The Faculty of Engineering has the highest number of users and access by defeating 22.6% of the total. Followed by the Faculty of Medicine 15.7%, the Postgraduate Program 14.3%, and the Faculty of Computers Science 10.7%. Pearson's statistical Test with a coefficient of 0.905 and a significant value of 0.001 shows that there is a positive correlation between the number of users and the number of accesses, except for an anomaly in the Faculty of Economics which has a large number of users. Still, the number of accesses is relatively low.

Meanwhile, the Faculty of Medicine became the internet access user with the longest usage time with a total time of 320,076,300 seconds or an average of 159,639.052 per user. The Faculty of Engineering follow them with a total time of 301,956,120 seconds (average 104,338.673) and the Faculty of Computers with a total time of 215,272,809 seconds (average 157,709.018). Therefore, frequency of use and length of access time will be the initial consideration factors in bandwidth allocation management. Comparison of the average length of access time provides a proportional comparison in analyzing the level of bandwidth utilization within Unsri campus.

The number of Event data is recorded based on the destination port accessed by each user. The number of events shows where and what site the user is going to. One event means one destination, which is processed by one-time internet usage access by the user. So, the total number of events is the number of destination ports accessed.

Fluxion Disco software also provides Variants, which shows us the extent to which we tend to underestimate the complexity of the process. However, we cannot see how individual cases move through the process or how many cases pass through this additional loop once, twice, or even more frequently. In order to grasp the typical process execution pattern from beginning to end, we need to look at the variants. Process variants are about changes in a process flow. Process variants are the only path from the beginning to the end of the process. In other words, a process variant is a specific sequence of activities, just like "tracking" from the beginning to the end of the process. Because variant is how many ways or paths are accessed to reach the destination port, a high or low number of variants will have an impact on internet bandwidth usage.



Process Map of the Faculty of Economics

Process Map of the Faculty of Law

Process Map of the Faculty of Engineering

Process Map of the Faculty of Medicine

Process Map of the Faculty of Agriculture

Process Map of the Faculty of Teacher Training & Education

Process Map of the Faculty of Social & Political Sciences

Process Map of the Faculty of Computer Science

Process Map of the Program of Postgraduate Study

Figure 3. Process Maps Model for Each Faculty

Table 1. Numerical Data Derivate from Process Maps

Faculty of	Number of User Case ID	Freq. of Access	Length Time of Access (second)		Number of Events		Number of Variants Per Case	
	Sum	Sum	Mean	Sum	Mean	Sum	Mean	Sum
Economics	1,532	616,532	102,129.832	156,462,902	402	616,532	666	10,208,980
Law	902	533,685	112,514.126	101,487,742	592	533,685	422	380,683
Engineering	2,894	1,612,540	104,338.673	301,956,120	557	1,612,540	1,224	3,541,290
Medical	2,005	1,284,058	159,639.052	320,076,300	640	1,284,058	942	1,888,856
Agriculture	524	359,063	116,584.405	61,090,228	685	359,063	255	133,397
Education	1,307	729,692	95,931.725	125,382,764	558	729,063	518	677,397
Social Political	442	186,326	86,034.980	38,027,461	422	186,326	203	89,709
Computer	1,365	1,272,670	157,709.018	215,272,809	932	1,27,2670	639	871,981
Postgraduate	1,838	1,291,689	94,941.197	174,502,880	703	1,291,689	848	1,558,791
Total	12,809	7,886,255		1,494,259,206				

Table 2. Purposes of Access

Faculty of	Service	Academic	Non-Academic
Economic	486,819 (78.96%)	26,736 (4.34%)	102,977 (16.70%)
Law	421,966 (79.07%)	19,902 (3.73%)	91,817 (17.20%)
Engineering	1,275,232 (79.08%)	68,345 (4.24%)	268,963 (16.68%
Medical	1,005,074 (78.27%)	46,619 (3.63%)	232,365 (18.10%)
Agriculture	286,615 (79.82%)	12,160 (3.39%)	60,288 (16.79%)
Education	584,244 (80.07%)	22,181 (3.04%)	123,267 (16.89%)
Social & Political	142,118 (76.27%)	7,102 (3.18%)	37,106 (19.91%)
Computer Science	1,058,319 (83.16%)	46,462 (3.65%)	167,889 (23.19%)
Postgra-duate	1,036,392 (80.24%)	44,299 (3.43%)	210,998 (16.34%)

The number and type of event data and variants vary greatly for each faculty, based on the destination port address. Therefore, we have classified it as Purposes of Access into the Service category (for the network machine destination port), Academic category, and Non-Academic category, shown in Table 2.

The data in Table 1 and Table 2 are used as data sources for clustering in order to know the behavioral characteristics of users of Unsri LAN internet bandwidth access. Clustering is done using the Centroid Linkage algorithm on the Hierarchical Cluster method and the k-Means algorithm on the Partition Cluster method. Ideally, the results show that on the Unsri campus, users are categorized into three groups. They are shown in Figure 4 and Figure 5.

The first group is users from the Faculty of Law, Faculty of Agriculture, Faculty of Education, and Faculty of Social Politics. The characteristics of this group are shown by all data attributes of low internet usage. Next is the second group which has a high level of internet usage on all attributes used, consisting of the Faculty of Medicine, Faculty of Computers, Postgraduate Programs, and Faculty of Engineering. While the Faculty of Economics, in particular, has its own characteristics, which are high in the number of users, low in time of access, low in events of access objectives, but has a very high variant of access.

In the clustering process, factors that significantly affect cluster formation were also tested. To get a statistical test has been carried out with the Analysis of Variance (ANOVA) test, the results of which are shown in Table 3. Of the eleven attributes used, there are eight attributes that have a significant influence, namely attributes with a significant value less than or equal to 0.05.

The eight have been sorted by the calculated F value. Therefore, the determination of bandwidth management will use these eight factors as the basis for setting it. In comparison, the remaining three attributes are considered not to affect bandwidth usage.

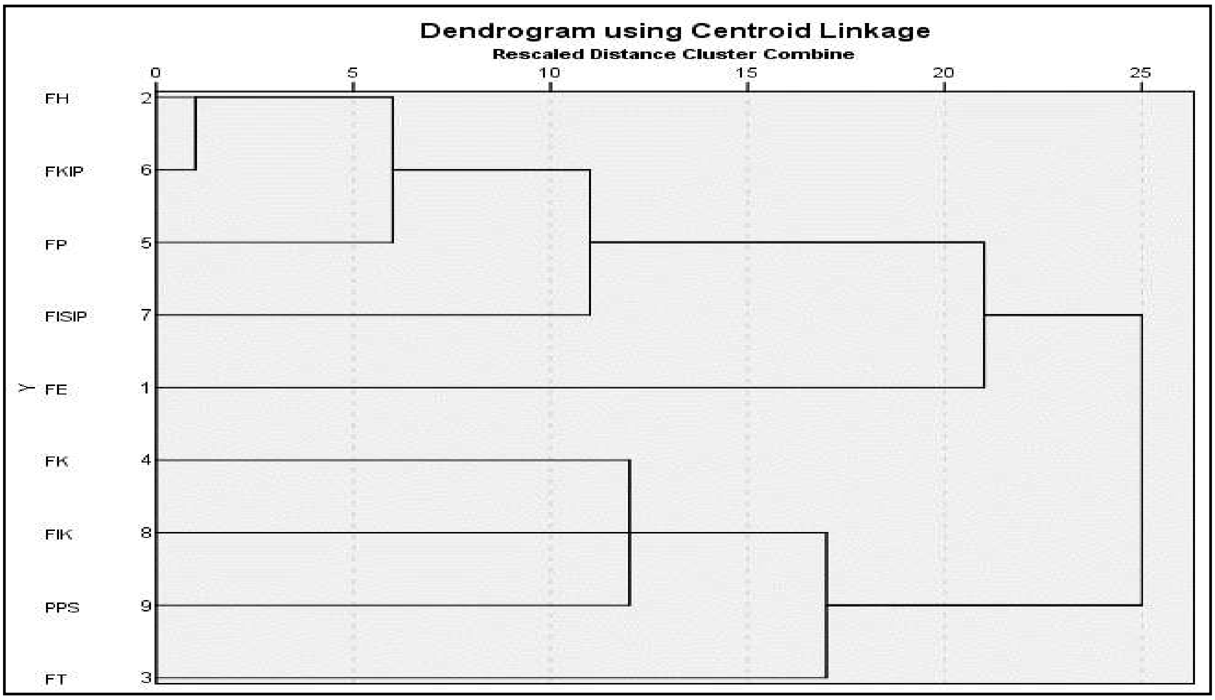


Figure 4. Clustering result using the Centroid Linkage algorithm

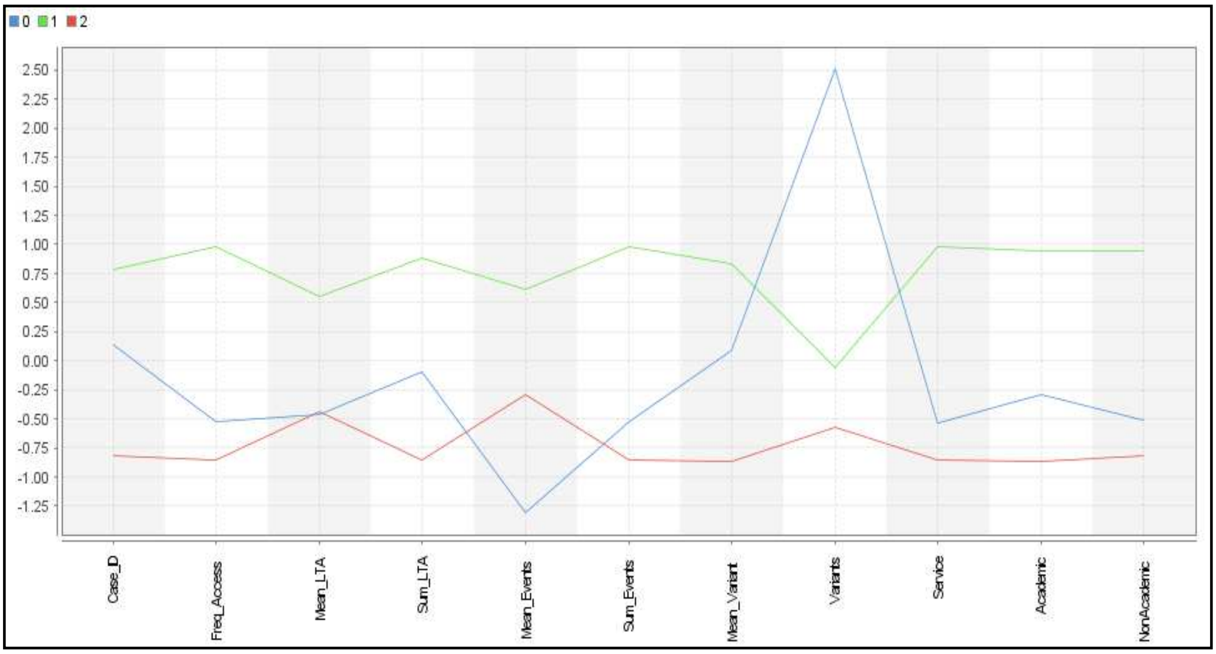


Figure 5. Clustering result using the k-Means algorithm

Table 3. Analysis of Variance (ANOVA) Test Result

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Length of Total Access	35777546340065100.000	2	1075693307791980.000	6	33.260	.001
Mean of Variant	377953.069	2	20281.903	6	18.635	.003
Sum of User Case ID	2029814.694	2	114310.694	6	17.757	.003
Number of Port of Non-Academic	21307696161.292	2	1522995330.569	6	13.991	.006
Number of Port of Academic	1202895009.611	2	128262260.278	6	9.378	.014
Sum of Freq. Access	748051538798.403	2	81971102557.236	6	9.126	.015
Number of Events	747987788696.167	2	82023139648.611	6	9.119	.015
Number of Port of Service	470583412513.611	2	57102347423.278	6	8.241	.019
Number of Variant Per Case	8797108108826.730	2	10825035982908.500	6	.813	.487
Mean of Length Time of Access	446584249.210	2	802246811.115	6	.557	.600
Mean of Number of Events	5995.486	2	31885.319	6	.188	.833

CONCLUSION

As the formulation of the problem solved in this research is: How to find out the behavior of internet users on LAN Unsri rationally to evaluate the effective and efficient use of internet bandwidth? This research has succeeded in modeling the behavior of internet users who use bandwidth from the Unsri LAN network by mapping the user access process from the beginning of access to the end of access, with the algorithm contained in the Mining Process Method. The behavior in question is reflected in the map of the processes carried out by users in accessing the internet.

From the Analysis of the process map obtained, the authors found that there are factors that characterize the behavior of internet users, which can be a standard formulation for the implementation of Bandwidth Management and Bandwidth Allocation on the Unsri campus LAN. The proposed formulation divides users into three categories: users with high bandwidth requirements, low bandwidth requirements, and special specifications. The factors that characterize and need to be considered in bandwidth management, in order of significance, are Length of Total Access, Mean of Variant, Sum of User Case ID, Number of Port of Non-Academic, Number of Port of Academic, Sum of Frequency of Access, Number of Events, and Number of Port of Service.

ACKNOWLEDGMENT

The research and publication of this article were funded by the DIPA of Public Service Agency of Universitas Sriwijaya 2020. SP DIPA-023.17.2.677515/2020, on March 16, 2020. In accordance with the Rector's Decree Number: 0685/UN9/SK.BUK.KP/2020, on July 15, 2020. We thank our colleagues from Malaysia Institute of Information Technology (MIIT), Universiti Kuala Lumpur, who provided insight and expertise that greatly assisted the research.

REFERENCES

[1] M. Aboubakar, M. Kellil and P. Roux, "A review of IoT network management: Current status and perspectives," *Journal of King Saud University - Computer and Information Sciences*, April 2021, doi: 10.1016/j.jksuci.2021.03.006

[2] J. Ding, *Advances in network management,* CRC Press (Taylor & Francis Group), London, UK, 2016, doi: 10.1201/9781420064551

[3] A. Clemm, *Network management fundamentals*, Indianapolis, Cisco Press, Indianapolis, USA, 2007.

[4] F. A. Bohani, S. R. Yahya and S. N. H. S. Abdullah, "Microgrid Communication and Security: State-Of-The-Art and Future Directions," *Journal of Integrated and Advanced Engineering (JIAE)*, vol. 1, no. 1, pp. 37-52, 2021, doi: 10.51662/jiae.v1i1.10

[5] J. Jeston and J. Nelis, *Business process management: Practical guidelines to successful implementations*, Butterworth-Heinemann, London: Routledge, UK, 2014.

[6] T. Panagacos, *The Ultimate Guide to Business Process Management: Everything you need to know and how to apply it to your organization*, CreateSpace Independent Publishing Platform, Carolina Selatan, USA, 2012.

[7] P. Renna, C. Izzo, and T. Romaniello, "The Business Process Management Systems to Improve the Performance of Universities: Integrated Performance Plan," in *Handbook of Research on Operational Quality Assurance in Higher Education for Life-Long*

Learning: IGI Global, pp. 56-80, 2020, doi:10.4018/978-1-7998-1238-8.ch003
[8] R. Muwardi, H. Gao, H. U. Ghifarsyam, M. Yunita, A. Arrizki, J. Andika, "Network Security Monitoring System Via Notification Alert," *Journal of Integrated and Advanced Engineering (JIAE)*, vol. 1, no. 2, pp. 113-122, 2021, doi: 10.51662/jiae.v1i2.22
[9] N. Palmer, *What is BPM*, in Bpm.com, 2014.
[10] F. Reher, "Adaptation of business process management to requirements of small and medium-sized enterprises in the context of strategic flexibility," in *International Conference on Business Process Management*, vol. 202, pp. 561-566, 2014, doi:10.1007/978-3-319-15895-2_48
[11] F. J. F. d. Silva, "Integração de Internet of Things num processo de negócio," *Master Thesis*, 2017, doi: http://hdl.handle.net/1822/54493
[12] D. Thiault, *Managing Performance Through Business Processes*, CreateSpace, Seattle, WA, 2012.
[13] B. Brandall, "Why You Should Bother With Business Process Modeling." [Online] Available: https://www.process.st/business-process-modeling/ [accessed December 18, 2020]
[14] Y. Nagm-Aldeen, M. A. Abdel-Fattah, and A. El-Khedr, "A literature review of business process modeling techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, no. 3, pp. 43-47, 2015.
[15] R. A. Maksimchuk and E. J. Naiburg, *UML for mere mortals*, Addison-Wesley Professional, Massachusetts, USA, 2005.
[16] M. De Leoni, W. M. van der Aalst, and M. Dees, "A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs," *Information Systems*, vol. 56, pp. 235-257, 2016, doi: 10.1016/j.is.2015.07.003
[17] W. M. van der Aalst, *Process Mining: Discovery, Conformance, and Enhancement of Business Processes*, Springer: Berlin, Germany Heidelberg, Germany, 2014.
[18] W. M. van der Aalst, "Intra-and inter-organizational process mining: Discovering processes within and between organizations," in *IFIP Working Conference on The Practice of Enterprise Modeling,* Springer, Norway, Canada, 2021, vol. 92, pp. 1-11, doi: 10.1007/978-3-642-24849-8_1
[19] A. A. Kolosova and I. A. Lomazova, "Detection of Anomalies in the Criminal Proceedings Based on the Analysis of Event Logs," in *International Conference on Analysis of Images, Social Networks, and Texts*, Springer, 2019, pp. 401-410, doi: 10.1007/978-3-030-37334-4_36
[20] H. Horita, H. Hirayama, T. Hayase, Y. Tahara and A. Ohsuga, "A Method for Goal Model Repair Based on Process Mining*," 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 2019, pp. 121-126, doi: 10.1109/SNPD.2019.8935771.
[21] S. H. Jayady and H. Antong, "Theme Identification using Machine Learning Techniques," *Journal of Integrated and Advanced Engineering (JIAE)*, vol. 1, no. 2, pp. 123-134, 2021, doi: 10.51662/jiae.v1i2.24
[22] NN, *Whitepaper: An Introduction to Process Mining Technology*, Celonis, [Online] Available:https://www.celonis.com/process-mining/process-mining-white-paper/ [accessed December 18, 2020].
[23] M. L. Van Eck, X. Lu, S. J. Leemans, and W. M. Van Der Aalst, "PM $$^ 2$$: a process mining project methodology," in *International Conference on Advanced Information Systems Engineering*, Springer, 2015, vol. 9097, pp. 297-313, doi: 10.1007/978-3-319-19069-3_19
[24] C. W. Günther, "Process mining in flexible environments," *Ph.D Thesis*, Technische Universiteit Eindhoven, 2019, doi: 10.6100/IR644335