applied sciences

MDPI

Article

From Brown-Field to Future Industrial Networks, a Case Study

Mehrzad Lavassani [1,2,*], Johan Åkerberg [1] and Mats Björkman [1]

[1] Division of Networked and Embedded Systems, Mälardalen University, 721 23 Västerås, Sweden; johan.akerberg@mdh.se (J.Å.); mats.bjorkman@mdh.se (M.B.)

[2] Division of Industrial Systems, RISE—Research Institutes of Sweden, 852 33 Sundsvall, Sweden

* Correspondence: mehrzad.lavassani@ri.se

Abstract: The network infrastructures in the future industrial networks need to accommodate, manage and guarantee performance to meet the converged Internet technology (IT) and operational technology (OT) traffics requirements. The pace of IT–OT networks development has been slow despite their considered benefits in optimizing the performance and enhancing information flows. The hindering factors vary from general challenges in performance management of the diverse traffic for green-field configuration to lack of outlines for evolving from brown-fields to the converged network. Focusing on the brown-field, this study provides additional insight into a brown-field characteristic to set a baseline that enables the subsequent step development towards the future's expected converged networks. The case study highlights differences between real-world network behavior and the common assumptions for analyzing the network traffic covered in the literature. Considering the unsatisfactory performance of the existing methods for characterization of brown-field traffic, a performance and dynamics mixture measurement is proposed. The proposed method takes both IT and OT traffic into consideration and reduces the complexity, and consequently improves the flexibility, of performance and configuration management of the brown-field.

Keywords: converged networks; brown-fields characteristics; network performance measurement

check for updates

Citation: Lavassani, M.; Åkerberg, J.; Björkman, M. From Brown-Field to Future Industrial Networks, a Case Study. *Appl. Sci.* **2021**, *11*, 3231. https://doi.org/10.3390/app11073231

Academic Editor: Fabio La Foresta

Received: 18 February 2021
Accepted: 1 April 2021
Published: 3 April 2021

1. Introduction

The future industrial networks are expected to accommodate both Internet technology (IT) and operational technology (OT) traffic in a unified infrastructure. The expectation begun with the vision of Industry 4.0 and efforts towards digitization in industries to increase productivity and to enable innovation. In order to reach these visions, it is needed to facilitate the availability of data from everywhere to anywhere, which requires effort to unlock the data in automation pyramid communication. Furthermore, IT and OT systems need to be integrated to have a bridge for utilizing the unlocked data. Many research works and under development technologies, such as Time Sensitive Networking, are trying to predict the requirements of converged IT–OT networks and provide a collection of mechanisms to address requirements of various traffic classes. Despite these efforts, there is still a lack of insight and results from the existing systems, brown-fields, that can be used for a smooth transition from today's system toward the digitalized industries and Industry 4.0 vision.

The importance of IT–OT convergence for IIoT systems has been discussed during recent years, and it is considered a mandatory element of future industries. Nonetheless, the progress in developing IT–OT networks has been difficult to advance because of several challenges that need to be overcome to reap the benefits of IT–OT integration [1,2].

From a practical perspective, these challenges can roughly be categorized into three groups. The first challenge set concerns with different characteristics of IT and OT traffic and their diverse, and sometimes contradicting, performance requirements. OT systems are the combination of communication and control components with a relatively long life-cycle, whereas the IT technological spectrum commonly has a shorter life-cycle and are expected to go through more changes in shorter periods of time [3]. From the performance perspective, OT networks focus on availability, reliability, real-time and determinism, whereas IT networks emphasize privacy, confidentiality, authentication and integrity, with reduced emphasis on real-time and determinism [2,4].

IT–OT converged networks are yet to become de facto infrastructure in the industrial environment and have a long path to reach technological maturity. The existing research and development efforts still lack the knowledge for configuration and performance management purposes. Hence, another challenge set hindering the development progress is concerned with uncertainties surrounding characteristics and performance requirements that IT–OT networks need to address. Among the more discussed challenges, complexity, flexibility and scalability of the converged networks are standing out. The increased number of smart devices and their demand for communication resources both in vertical and horizontal integration pose serious difficulties on configuration and performance management, and require more investigations for identifying what would be the vital parameters to consider.

Another challenge set, adjacent to the ones stated, is that the new network infrastructure cannot be realized in one revolutionary act. In practice, this approach faces a two-fold problem. Firstly, there are no preliminary information or data that can be used to build and to configure the system from scratch, known as green-field implementation, with parameters that can ensure essential performance criteria. Secondly, it is unexpected that a whole functioning system with ongoing processes is stopped and replaced by a new system with a high risk of downtime. There needs to be considerations for integration with legacy systems, the machinery and infrastructure that already exist, also known as brown-field [5].

The limited knowledge of IT–OT consolidated networks' performance and configuration requirements partially derives from a lack of understanding about the characteristics as well as the complexity of configuration, resources and performance managements of IT–OT converged system of systems. Many of the existing industrial networks are also handling IT traffic, generated from IT services such as system updates and anti-viruses, back-up servers and remote access servers. The two traffics are commonly handled by a division of time for accessing the shared link [6]. While the IT traffic in the brown-fields cannot be a just representative of the IT traffic in the future consolidated IT–OT networks, it still can provide partial knowledge about IT characteristics and be used as a basis for knowledge extraction about traffic behavior. One requirement for innovation is to revisit the existing concept and to gain additional insight to build the evolution to the next stage. Investigation of the brown-field traffic can provide a better understanding of the existing networks and facilitate the estimation of future requirements, performance criteria and guidelines for next step integration.

Industrial networks are usually designed and configured for a specific application domain with considerations on environment and service requirements. It is impractical to generalize the knowledge gained from one use-case functionality to all the others. However, studying brown-fields can help to establish a basis of network functional properties, and to distinguish the divergence from initial setups and common theoretical assumptions. In other words, studying a brown-field can help with the interpretation of the evolved network to the current stage, and therefore granting insight for partial estimation of the requirements in the next possible stages.

In recent years, industrial network traffic has been investigated for purposes such as intrusion and anomaly detection, network management and traffic monitoring. In [7], industrial IP network traffic is investigated for intrusion detection by modeling each pair of traffic flows based on profiled communication patterns. Traffic patterns of SCADA networks were studied in [8] and inter-arrival time and correlation models were proposed and defined as additional key parameters (KP) for traffic characterization and flow modeling. The same KPs showed promising results when applied on emulated network traffic for anomaly detection [9], but the performance was not as satisfactory on the real traffic of SCADA systems [10]. The reason for inaccurate results can be motivated by the complexity of brown-field communication patterns and virtualized architecture. To find the capabilities and limitations of brown-field networks for managing the demands of new use-cases in terms of inter-arrival time and packet size, a methodology is proposed in [11]. A matrix representation captures the state of the network with respect to the selected key performance indicator (KPI), and the effect of one data flow on the other flows is arbitrary ignored. The methodology is validated using emulated SCADA traffic where no logical grouping for flexible communication between physically distributed devices are applied.

Even with the literature's valuable works, there is still a gap for insight from the brown-field to identify the current state and further be applied to develop a methodology for the evolution of the brown-fields to the expected future industries. Recognizing the challenges mentioned above and existing related literature, this work presents the results of a case study on brown-field network traffic collected from the Iggesund paperboard production network. To the best of the authors' knowledge, this is the first presented study of shop-floor network data. The contributions of this paper then expand to the following:

- A comparative study highlights the disparities between the common assumptions of network traffic flow characteristics in the literature and what the collected brown-field traffic projects. Extending deficiency of the existing flow characterization and traffic modeling for network configuration and performance management, with respect to the scalability and flexibility requirements of the converged networks, are discussed.
- A new measurement method for characterization and analyzing the network traffic is proposed that can potentially reduce the complexity of resource allocation and management of the scaled-up converged networks, and enables flexible technological integration into brown-fields.
- Ensuing the findings and the proposed method, possible future research directions considering the existing challenges and open issues in evolution to the consolidated IT–OT networks are discussed.

In what follows, the basis of network traffic analysis and modeling are briefly reviewed in Section 2. The brown-field of this case study, measurement environment, collected data and network characteristics are introduced in Section 3. The findings from the brown-field study, comparative analysis with respect to related works, and the proposed performance and dynamic mixture measurement for brown-field characterization are detailed in Section 4. Finally, the gaps and potential future research direction based on the findings of this study are discussed in Section 5.

2. Network Analysis and Modeling

Design, management and configuration of network traffic have decades of support in theory and practice. Monitoring and characterization of network traffic is the essential step in various network management functions, namely configuration, performance, fault, accounting and security managements [12].

2.1. Objectives and Measurement

Network traffic monitoring and analysis are generally categorized into four types of problems depending on the context of the network management objectives [6]. The state of resources and the relationship between them is the main focus in configuration management. The process follows a detailed view analysis, where communication patterns are studied for each pair of communicating components, i.e., source and destination. The purpose is to estimate the traffic volume for resource provisioning and capacity planning. Traffic matrix measurement with the general approach of network tomography and direct management is the analysis approach for this objective. This approach, with a detailed view of the system, can even assist in understanding the impact of fault on network capacity.

A more abstract approach measures traffic volume to determine the total traffic of a network flow. This measurement is of use in network performance and security management to understand changes in traffic and identify abnormal flows. This type of measurement monitors traffic generated by any of the sources, using IP addresses and independent of possible destinations, and analysis packet counts and sizes for heavy usage.

Networks are dynamic systems where the temporal variation of traffic represents their various operational states. Traffic dynamic measurement aims to monitor the dynamics and extract knowledge that can be used in configuration management for link capacity estimation and allocation. This type of measurement is also used in performance measurement to test the stability of the network. The monitored parameters usually include packet delay, packet loss and available bandwidth [6].

A more recent approach in network monitoring and analysis aims to measure traffic mixture, i.e., aggregation of the detailed view of source and destination with the abstract view of traffic and temporal dynamics. The emphasis here is on the importance of extracted features in aggregated traffic data over time for performance and security managements. The extracted features with respect to performance are flow attributes such as delay, throughput or packet drop rate. In security management, frequent patterns in inter-arrival time and correlation attributes are extracted to model the flows and to detect anomalies [7,9,10,13]. A comprehensive survey of research works for measurement and characterization of traffic network based on the objectives of analysis is presented in [6].

2.2. *Characterization and Modeling*

Similar to the various management purposes that network traffic can be analyzed for, there are different scales and measures for setting the analysis framework. A network can be considered as a number of sub-systems grouped together to carry out a pre-defined process. The network traffic can be studied separately for each flow or as accumulated traffic of each sub-system. Equivalently, the network traffic can also be studied from the topological structure, either physical or logical, that is sub-systems and monitoring points or VLANs and virtual communication groups. The choices on the level of abstraction for the scope of the analysis can set the relevant methods to be applied in the process, while itself is derived from the objectives of the study.

2.2.1. Characterization with Matrix and Dynamics Mixture

The process follows a down-top approach, where communication patterns are studied with respect to network components, i.e., source and destination. The initial efforts are to either classify or model the characteristics of the traffic flows in order to create flow profiles. The dominant communication patterns of flows are modeled using various probabilistic and mathematical models, and the anomalous activities are identified where deviation happens from the tuned parameters of the initial learned model. To learn the normal communication in the network, a packet-level investigation is applied. All packets are considered, and values of various fields in the header and payload are extracted. These values are then used to profile individual communication flow based on either source or destination [7], or protocols [8–10,14]. The created profiles are used to learn and classify each flow and further model the network traffic communication pattern to find abnormal behaviors.

2.2.2. Modeling with Volume and Dynamics Mixture

Another approach that reduces the complexity of the flow-level characterization and modeling aims for a more abstract view of the network and targets different types of traffic, and their grouped flow characteristics. This type of approach profiles traffic either based on the type of services and traffic class or based on network virtualization traffic. The focus here changes from individual communication flow to general communication patterns from sub-systems in the network. The network traffic is usually classified into three categories: cyclic or time-triggered, stochastic and burst. The profiling of traffic classes considers the characteristics of each class, such as packet size and time intervals between transmissions. The cyclic traffics are small size packets with fixed generation time intervals, where stochastic traffics are event-triggered with small packet size and short duration and unknown transmission interval. The burst traffic flows are defined with large data transmissions for short intervals and self-similar transmission patterns in longer time intervals.

2.3. Methods and Tools

The methods applied for analyzing network traffic and developing monitoring systems are highly dependent on the scope of the study and available data. More traditional approaches analyze the data using statistical and mathematical methods. In [15], a survey of methods found in the literature for network traffic analysis is presented. Recent years also witnessed applying various machine learning methods on network data to learn the behavior of the system from the data. Machine learning methods are used for flow-based characterization, or modeling [6,16]. Classification methods for network flow analysis are surveyed in [17]. The objectives are commonly directed to anomaly and intrusion detection by flow characterization and traffic profiling. Supervised learning and classification methods have been more favorable since the results show higher accuracy in comparison to unsupervised learning. Nonetheless, due to the unavailability of data, the methods are usually carried out on simulated and emulated data.

3. Case Study: Iggesund Paperboard

The Iggesund paperboard is a typical process automation factory, and the automation network can be considered as an example of production networks in manufacturing. Reliability of the production network is ensured by hot-standby redundancy consisting of two networks: primary and secondary. The communication backbone for two paperboard machines comprises three marshaling rooms and a number of switches located across the factory. The communication between different systems is provided by the configuration of several virtual LANs (VLANs). This case study covers a part of the operational network consisting of 5 control systems, with 43 stations connected to the server network and 32 process controllers on various VLANs. Figure 1 illustrates the network topology at the paperboard machine where the data were collected. In this study, the Client–Server Network traffic with connectivity to the Internet is considered IT traffic, and the traffic from the control systems is the OT traffic.

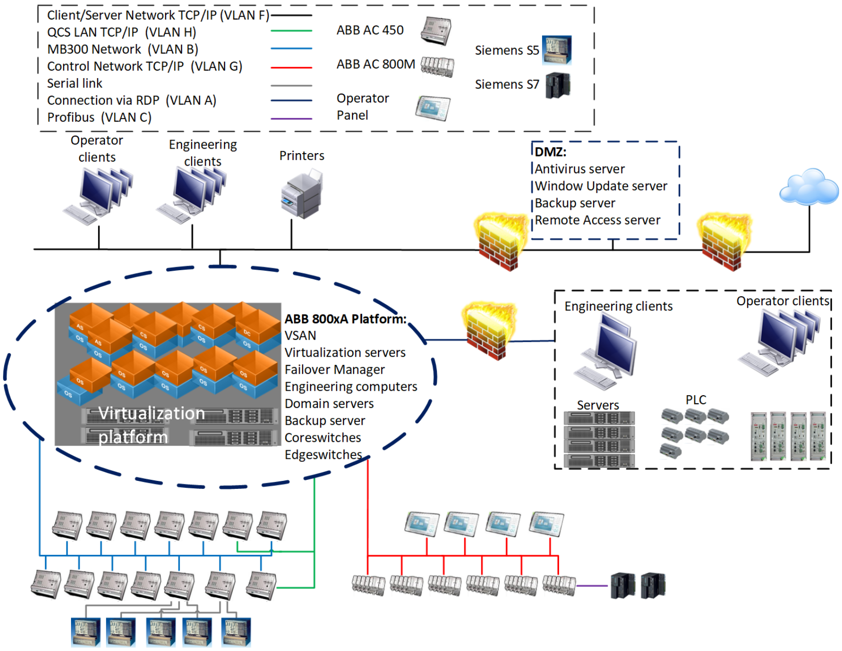


Figure 1. Network topology at one of the paperboard machines at Iggesund based on the documentation [18].

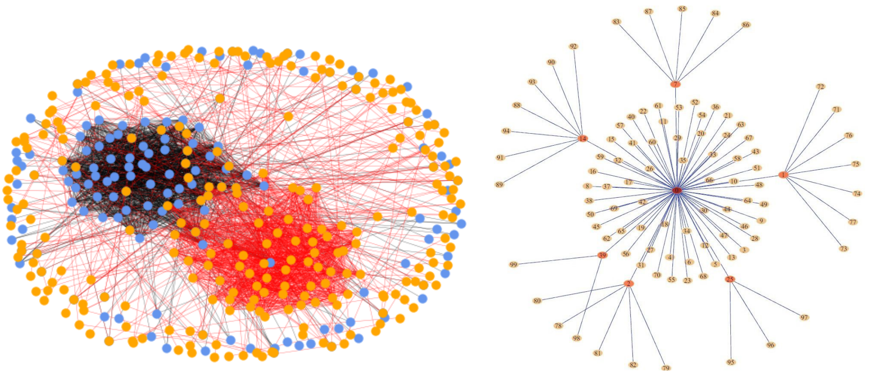
3.1. Data Collection and Data Set

The network traffic was captured by enabling mirroring of the traffic recorder port connected to one of the switches in the production network. The reason for limiting the data collection to one switch was due to the physical distance between control rooms. Consequently, the captured data does not represent all the traffic of the production network, but the communication of one of the machinery. The captured traffic contains both primary and secondary networks traffic, not simultaneously, for a duration of almost 12 h. The result of the data collection was 60 files, and each file contains more than 3 million records and is 0.99 GB in size. The initial captured traffic flows are huge files containing packet dumps from the network, in *.pcap* format. After a preliminary packet-level investigation, to create a suitable format for further exploratory traffic analysis and characterization in python [19], each file was converted to .csv format. Each of the files consists of header fields, and payload data of more than 3 million packet and each record was formatted as [$Timestamp_{\mu s}$, $VLANID$, IP_{source}, $IP_{destination}$, $Protocol$, $Length$, $Hardware_{source}$, $Hardwaredestination$, $Port_{source}$, $Port_{destination}$, $info$, $delta$].

Production networks work in cycles expanding over time. The network traffic dynamics can be captured more accurately when it is studied over several operational cycles. Hence, 10 consecutive files were selected for the analysis and experiments carried out in this study. The final data set consists of more than 30 million packet records over 90 min. Due to the resource-intensiveness of the processing the huge data set, and to enhance the clarity of illustrations in this study, a simplified data set with record tuples of [$Timestamp_{\mu s}$, $VLANID$, IP_{source}, $IP_{destination}$, $Protocol$, $Length$] are selected.

3.2. Traffic Structure and Communication Patterns

The network comprises 337 devices, and 6,554,498 recorded transmissions were studied for illustrating the results of this experiment. The collected data consist of both IT and OT traffic, with statistics summarized in Table 1. Figure 2a illustrates the connectivity structure of the brown-field that represents the communication between network components; IT and OT traffics are color-coded in blue and orange, respectively. A larger number of packets are transmitted between devices belonging to each of the traffics. Six communication clusters comprise the network, and devices communicate with 1 to 108 peers. That is, the sub-systems have a maximum of 108 communication destinations for both IT and OT. To see the dynamics of the traffic, a 25-min window is illustrated in Figure 3. As expected, OT traffic shows a less volatile data transmission pattern compared to IT traffic. The IT traffic comprises the majority of the transmitted packets as well as the larger bandwidth consumption. The ratio between the number of transmitted packets and the transmitted data indicates that, on average, larger packets belong to IT.



(**a**) (**b**)

Figure 2. Connectivity structure graph: (**a**) brown-field case study and (**b**) literature [7].

Table 1. Summary of the collected traffic.

	Devices#	VLAN#	Bandwidth%	Transmissions%	Protocols#
OT	124	7	0.35	0.42	15 (MMS > TCP > UDP)
IT	213	2	0.65	0.58	40 (TCP > UDP > SMB2)
Total	337	9	1	1	43

The previous knowledge about the network configuration and structure identifies 9 VLANs, amongst which two are dedicated to IT and seven to OT. The number of devices in each VLAN varies from 4 to 100. Studying the data set revealed 7205 packets with missing VLAN tags, and further analysis shows they belonged to OT traffic. Moreover, the traffic flows of two of the VLANs traffics flagged as OT show communication traces with IP addresses that were associated with the IT VLANs. In other words, the IT and OT traffics are not isolated in this network, and there are communication clusters with both IT and OT traffics.

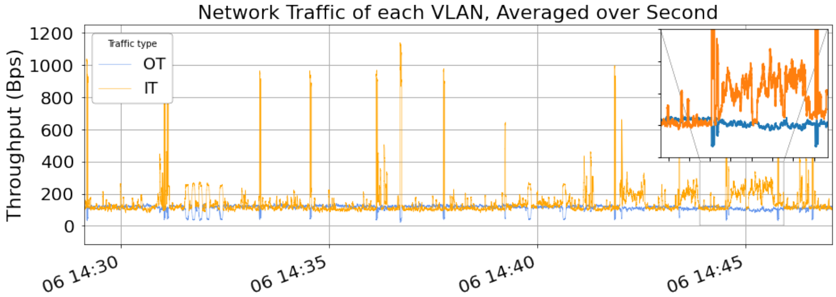


Figure 3. Data transmission dynamics in the network for Internet technology (IT) and operational technology (OT) traffics.

4. Brown-Field Traffic Analysis

In this section, we first present a comparative analysis between common assumptions in the literature and data from a brown-field to demonstrate the differing outcomes. We then suggest a characterization approach that can potentially reduce the complexity of the configuration and performance managements of converged IT–OT networks.

In analyzing the network traffic, we first set a framework for the characterization process. Network modeling and characterization play an important role in various network management functions, including configuration and performance management. It is a prerequisite to estimate traffic demands on the network capacity for bandwidth allocation and to identify potential congestion and delays caused by exceeding allocated bandwidth [6]. Accordingly, we focus on the transmitted data in the network and throughput. Network traffic dynamics is an important aspect to consider for configuration and performance evaluation. Therefore, we apply a mixture measurement and study the throughput dynamics. The scope of measurement for the comparative analysis covers connectivity structure and topological structure (VLANs). The reason for this choice is to find points of connections with the body of related works and common practice. The comparisons are motivated regarding the complexity and flexibility required for the realization of the converged IT–OT networks.

4.1. Flow Characterization

The matrix measurement is commonly used to characterize network flows for configuration, fault and security management purposes. The connectivity structure of the network provides the basis to identify all of the communication pairs for matrix measure and flow characterization. This meticulous approach provides a detailed view of all network components, but it scales poorly for larger size networks with more complex flow characteristics. An example of a commonly considered network connectivity structure is shown in Figure 2b. The brown-field network graph, Figure 2a, consists of 6 communication clusters with 2 main communication hubs. The degree of the graph connectivity of a single source varies from 1 to 108, and more than 70% of the sources communicate at least with 2 separate destinations. As it is evident from the figure, the connectivity patterns in brown-field are by far more complex than of those considered in the literature. Characterizing brown-field flows using matrix and dynamic mixture then becomes an even more tedious task, which cannot deliver the level of accuracy required for network configuration management due to the imposed complexity.

The matrix and dynamic measurement is also a method used for anomaly detection in the network. The dynamics of each pair of communicating components are modeled using various probabilistic and mathematical methods. The initial efforts are to model the dominant communication pattern and classify the future traffic to identify the anomalies or security bridges. Figure 4 illustrates the outcome of applying Discrete Time Markov Chain (DTMC) to model communication between a pair of components, associated with the connectivity structure in Figure 2a. The model is applied on 40,000 packet results in 37 states and 277 transitions [7].

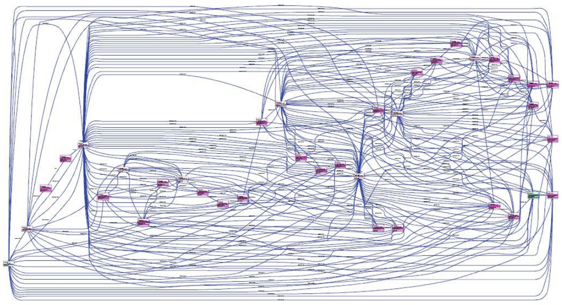
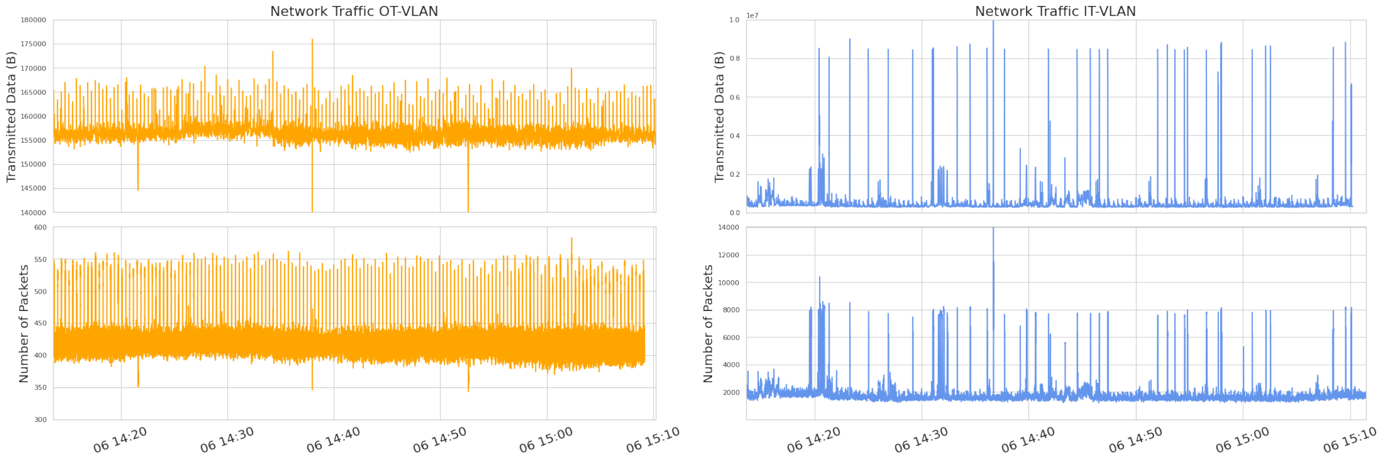


Figure 4. Complexity of modeling communication flow between two devices with Discrete Time Markov Chain (DTMC) and packet-level details [7].

It is evident that the complexity and computational costs associated with new rounds of characterization and re-learning the dynamic model after any communication or structural changes are not going to be negligible. It also affirms scalability and complexity problems of matrix and dynamic characterization for the brown-field, where communication patterns are more complex, and more changes are envisioned from the current state of industrial networks towards the IT–OT converged networks.

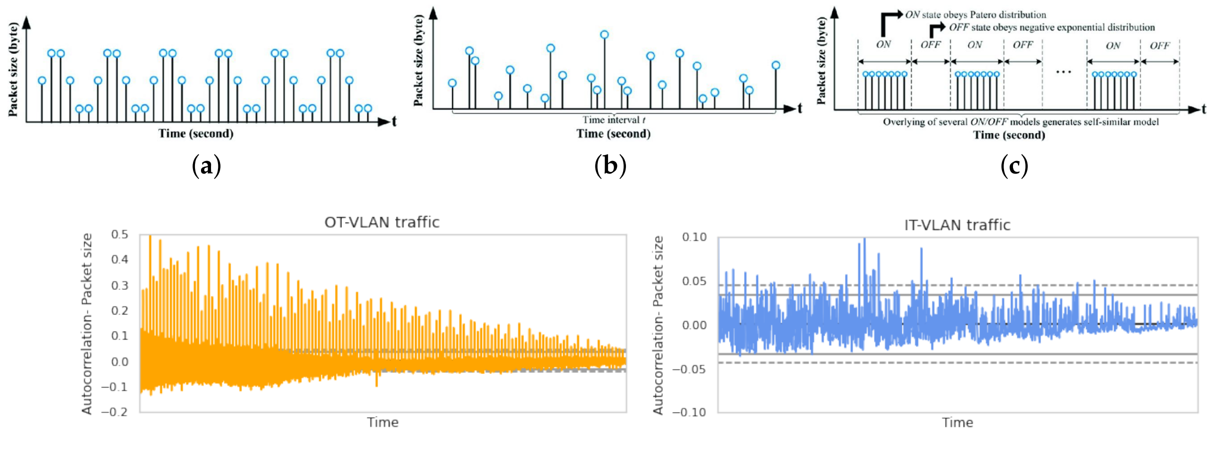
4.2. Traffic Pattern Modeling

The volume and dynamic mixture expands the scope of measurement from individual flows to groups of flows with a specific traffic class that is generated from a VLAN. The IT traffic is commonly assumed to have the burst traffic type, and OT traffic is generated more periodically. The sporadic traffic can be generated by both IT and OT from alarm and critical applications. The brown-field traffic was collected under normal operation with no emergency or alarm, so the data are not a proper representative for the sporadic class. The engineering documentation, along with exploratory analysis, were used to map VLANs in the network to the suggested data flow classes in the literature. Figure 5a,b show traffic dynamics from one of the OT-VLANs and one of the IT-VLANs, respectively. The OT-VLAN shows a smaller variance in accumulated data per second as well as the number of transmitted packets in unit time intervals when compared to the IT-VLAN. The smaller variance, however, does not reflect the expected cyclic pattern in the literature, Figure 6a,d. This difference is also true for the burst traffic of the IT-VLAN, Figure 6e, and the assumed pattern in literature, Figure 6c.



(**a**) (**b**)

Figure 5. Traffic dynamics in the network for transmitted data and the number of transmitted packets in OT-VLAN (**a**) and in IT-VLAN (**b**).



(**a**) (**b**) (**c**)

(**d**) (**e**)

Figure 6. Data packet generation models in literature: (**a**) cyclic, (**b**) stochastic and (**c**) burst [20] vs. brown-field: (**d**) OT-VLAN (cyclic) and (**e**) IT-VLAN (burst).

These dissimilarities can be reasoned as the result of a more complex operation of the brown-field, which differs from the theoretical models. It can be argued that flows can be decomposed into their components to achieve the same representation as in the theoretical view; the counter-argument points at a simultaneously increased level of complexity that this approach initially wants to avoid. The same reasoning is also standing correct for the burst traffic. Moreover, the analysis of the flows from the brown-field IT-VLAN shows traces of IP addresses associated with OT-VLANs. This insight contradicts the common assumption of IT and OT isolation in the networks and the suitability of a single model per VLAN.

While volume and dynamics mixture can reduce the complexity of flow characterization and flow-based profiling, by removing some levels of details and providing a more comprehensive level of abstraction for network modeling, it also falls short in terms of accuracy or completeness of input parameters when applied to the brown-field traffic. The sub-system view over the network flow, considering isolation due to virtualization, can simplify the modeling; nonetheless, it does not by any means simplify the classification of data flow within each of the VLANs.

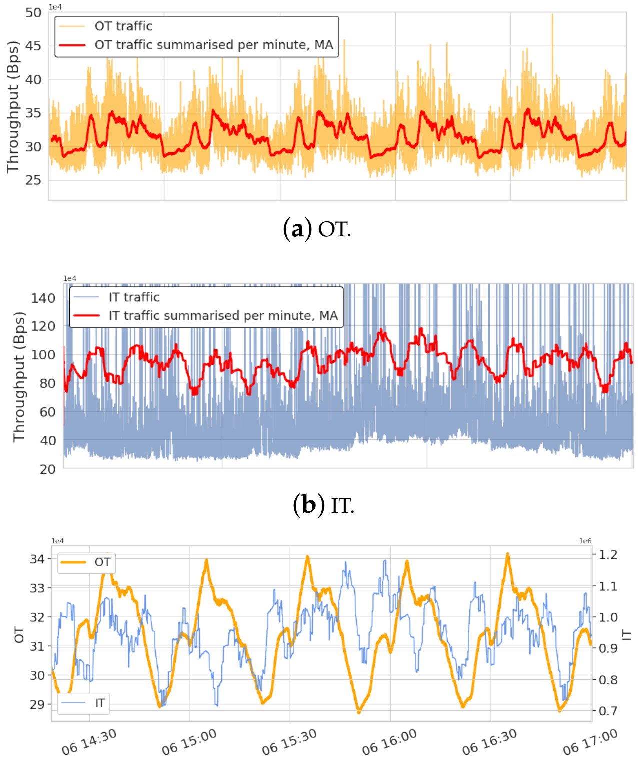
4.3. Performance Dynamics Characterization

It is evident that the existing characterization and profiling methods are not well suited for performance and configuration management of the IT–OT converged network for two main reasons. Firstly, the detailed tomography in flow monitoring and characterization do not scale well, which can result in the high complexity and lack of flexibility that is required for performance management of IT–OT converged networks. Secondly, the assumption of uniformity in flow patterns based on traffic classes does not apply when brown-field traffic is considered. Furthermore, the existing methods and approaches are either addressing IT or OT since the majority of the exiting networks manage IT and OT traffic by the division of time to isolate the two traffic types. In the converged network, both traffic types ought to be generated from devices in the network, which troubles the flow and traffic isolation. The unexplored performance criteria of IT–OT networks cannot be identified by studying the two traffic types separately, nor can it be derived by summing up the two sets of requirements. However, a reasonable starting point can be a traffic-type-agnostic characterization approach. In other words, it sets a broad monitoring perspective that unveils the overall network behavior and performance bottlenecks for performance and configuration management, i.e., resource allocation and provisioning. Hence, we propose a new measurement and characterization approach to provide additional insight into the network behavior with respect to bandwidth consumption and network throughput, and measure the dynamics of available resources, i.e., utilization and dynamic mixture.

As mentioned in Section 2, data are collected from a production network, and consist of both IT and OT traffics. The underlying traffic consists of both control network traffic for carrying out regular processes and server network traffic for communication with file servers or similar. Since the production networks follow regular workflow cycles to maintain the manufacturing process, it is credible to assume periodic patterns being generated by the sum of communication flows. This assumption is also partially supported in literature for IT and OT traffics, but separately. Self-similarity in Ethernet IT traffic and the Internet traffic was studied for many years [20–22]. Recently, similarity in inter-arrival time and the correlation between traffic flows in SCADA systems are also being studied and illustrated on simulated and emulated industrial data [8–10]. We expect that the network dynamic with respect to network utilization, i.e., accumulated throughput of all traffic, should also show a cyclic pattern, which in itself can show the performance level of operational states of the network with respect to the utilization parameter.

Dynamics of the network with respect to network utilization can be formulated as a time-series $X = x_1, x_2, x_3, ..., x_i$ where each x_i is the transmitted data of recorded packets at time i. The resolution of the collected data is in microseconds, and the number of transmitted packets and their length varies at each unit time interval. Accordingly, $X = x_i$ for $x_i = \sum_{j=1}^{n} l_j$, where l_j is the length of the $j - th$ packet, with resolution of x set to seconds, gives the dynamics of throughput for the network. The fluctuated pattern of throughput is the result of diversity observed in the network transmissions in terms of protocols, packet size and the number of transmissions. Downsampling the data and applying moving average filter with $window\_size$ = 60$\sim$1 min smooths the signal and mitigates the effect of random and short term fluctuations and anomalies.

Figure 7 illustrates the results of this process on the brown-field traffic. The smoothed OT throughput dynamics, Figure 7a shown in red, reveals a repeating cyclic pattern in each 20 min time window. The cyclic pattern in IT traffic is not as evident as in OT traffic, but similar patterns are observable in the smoothed pattern, Figure 7b. One reason for the higher level of fluctuations observed in IT traffic is the higher number of diverse impacting factors, such as the different number of protocols in specific time intervals.



(**a**) OT.

(**b**) IT.

(**c**) OT and IT.

Figure 7. Production cyclic work-flow projected from network traffic.

It is also notable that IT dynamics clearly show seasonality and trend, which, when removed, can provide more insight and discussion points to draw a more accurate conclusion. A closer look at the generated throughput dynamic patterns shows some level of similar dynamics in IT and OT behavior, Figure 7c. This similarity can be reasoned with the observed shared traffic between OT and IT in some of the VLANs. As presented in the comparative analysis of volume and dynamic mixture, the assumption of exclusive traffic class and flows patterns based on the pre-defined VLANs does not hold for brown-field traffic, i.e., IT and OT communications are not necessarily isolated in VLANs. The IT–OT flows can potentially show their effect on IT dynamics by similar patterns repeated in longer time intervals. Finding the correlated patterns between IT and OT bandwidth dynamics can further assist in characterizing the IT–OT flows and their requirements in terms of communication intervals and the amount of transmitted data. This insight subsequently can provide a baseline for integrating traffics from new technologies or replaced network components.

5. Discussion and Future Directions

IT–OT converged networks ought to accommodate traffics with different characteristics and requirements. There are ambiguities on the specific obligatory performance metrics, other than the general network performance criteria, i.e., delay throughput and packet drop rate, as well as lack of recommendations for configuration. It cannot be a fair assumption to classify IT traffic patterns in future networks as burst flows, that is considering the current trend in industry and academia for further digitalization of the factories and migration of some of the control systems functionalities to the fog or the cloud, namely Virtual Programmable Logic Controllers (VPLCs). Hence, some of the IT traffic might project expected periodic communication patterns of OT traffic. The gap in the literature is even more palpable for the transformation of brown-fields, where there is not enough insight from the existing networks, nor any recommended approach for moving forward. Lack of available data from brown-field also leaves the researchers with simulated and emulated data for traffic profiling and flow modeling, where the models do not perform well when applied to real traffic data. As it was shown, the assumption of singular traffic class communication pattern on the logical division of OT traffic does not hold for brown-field traffic.

Characterization of the network traffic is the first step for configuration and performance management of the network. The existing methods for characterization and profiling the traffic fall short of meeting the scalability requirements of IT–OT converged traffic. A sophisticated network management system requires an accurate and thorough analysis of the network components about the resources and criteria for performance management, among other important parameters. Affected by the previously discussed increased complexity, the accurate analysis of the network performance with flow-level measurement will potentially be even more difficult.

Performance dynamic mixture characterization can potentially reduce the complexity of network management for the integration of new traffic flows. The integration of new technologies and consequently adding traffic to the network needs to be carried out without, or with a minimum, interruption of the ongoing regular workflows. A sensible approach would be a configuration that minimizes the possibility of traffic congestion which has a direct negative effect on network performance in terms of delay and packet drop rate. Following the discussion in Figure 7, the peaks of the patterns happen when a larger amount of data is transmitting in the network, e.g., larger packets, a higher number of transmissions or a combination of both. Hence, integrating any new IT–OT traffic or flow will increase the risk of congestion, even if the flow configuration of the intended communication does not show any restrictions caused by simultaneous flow transmissions in source or destination. On the contrary, a flow can be configured and added to the network as long as throughput dynamics show the availability of resources. The same approach can be applied as a batch replacement, or integration, to ensure the network's uninterrupted operation by estimating the available resources in time interval where the anticipated resource is proportionally lower than the dynamic peaks.

Modeling of IT–OT networks for monitoring and anomaly detection purposes is one of the future challenges since the huge number of flows in converged network traffic will introduce the curse of dimension, and its high computational costs, to network monitoring and performance management. The proposed method set the monitoring scope on the performance of the network, with a level of abstraction that reveals regular workflows. The abstract performance dynamic formulation provides the opportunity to take the operational states of the network into account and reduce the number of parameters for network modeling proportional to the number of existing states. In other words, the regular workflow of the network can be divided into operational states, and the flows can be monitored based on the performance requirements of the associated state. This approach can reduce the computational complexity and improve the model accuracy by reducing the sample space and variance of the outcomes. However, further investigations are needed to identify and estimate the impacts of other important network parameters such as the number of transmissions, transmission intensity, acceptable delay and transmission intervals. There still is a need for formulating a KPI that can quantify the performance dynamics and the impact of the important parameters on the overall performance of the network.

One of the difficulties when analyzing traffics of the existing networks is that the brown-fields have evolved during the years by the integration of new technologies and adopting new business models. Additionally, brown-fields are narrowly designed for specific production use-cases and scenarios. The proposed approach was evaluated and tested on the data collected from one brown-field, and the aim is not to generalize the model and the method at this stage to all scenarios. Nonetheless, it is a plausible discussion that due to similar features in traffic patterns and similarities between traffics of different classes, this approach can have relevance when applied to similar scenarios. For instance, it is an acceptable assumption to observe cyclic patterns in production networks with regular workflows, human intervention or planned maintenance. The performance dynamics can investigate these similarities to exploit insight about communication patterns and the configuration requirements to meet a set of functionality guarantees, even if not with the most optimized performance and highly detailed application-dependent configuration. For all the aforementioned discussion, more investigation into brown-field networks and analysis of real-data for profiling, characterization and modeling is needed to procure a path to the future industrial networks.

6. Conclusions

In future industries, communication barriers between IT and OT systems are going to be removed. The network infrastructures need to accommodate, manage and guarantee the performance of the merged IT–OT traffic. The pace of developments so far has been disproportionate to all the presumed benefits of converged networks for enabling innovative applications in future industries. Lack of clarity over brown-field configurations and operations, performance criteria and requirements of the converged networks and contradicting requirements of IT and OT traffics are some of the causes behind the lagging advances.

The presented case study showed the unsatisfactory performance of the common assumptions and theoretical methods when applied to brown-field data. The comparison between the results of exploratory analysis of the brown-field and the theoretical approaches in the literature showed dissimilarities among the common assumptions and the real deployments. This additional insight is a step towards improving the accuracy of the brown-field characterization for future integration and performance management. The unsatisfactory performance of the commonly used methods for network configuration and performance management in terms of scalability and flexibility also questioned the suitability of common methods for brown-fields and further for converged networks.

The proposed network-level measurement and characterization demonstrates suitability for brown-field data analysis for performance and resource management purposes, specifically when the scalability and flexibility of the converged networks are considered. The network-level view enables performance and resource estimation with the minimum footprint on the ongoing operations. This method needs to be further investigated to identify the important characteristics of the networks traffic, and quantify the impacts of these parameters into measurable performance criteria, as it is set for our future work.

There are still many challenges in the path for the realization of converged networks in brown-fields, from identifying probable use-case scenarios where this realization provides the most benefits to finding the best technical and engineering approach for actualizing them. Nonetheless, the road map for moving toward the next evolutionary stage of industrial networks from brown-fields cannot be forged without capturing the essence of brown-field operational states.

Author Contributions: Conceptualization, J.Å., M.B. and M.L.; methodology, experiments and visualization, M.L.; writing—original draft preparation, review and editing, M.L.; review, supervision, J.Å.; project administration, funding acquisition, J.Å. and M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been partly financed by the Future Industrial Networks project, grant number 2018-02196, within the Strategic innovation program for process industrial IT and Automation, PiiA, a joint program by Vinnova, Formas and Energimyndigheten.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We would like to thank the reviewers for their constructive comments that helped to improve the context of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tian, S.; Hu, Y. The role of opc ua tsn in it and ot convergence. In Proceedings of the Chinese Automation Congress (CAC), Hangzhou, China, 22–24 November 2019; pp. 2272–2276.
2. Garimella, P.K. It-ot integration challenges in utilities. In Proceedings of the 3rd International Conference on Computing, Communication and Security (ICCCS), Kathmandu, Nepal, 25–27 October 2018; pp. 199–204.
3. Shi-Wan, L.; Bradford, M.; Jacques, D.; Graham, B.; Amine, C.; Robert, M.; Brett, M.; Mark, C. The Industrial Internet of Things Volume g1: Reference Architecture. 2019. Available online: https://www.iiconsortium.org/pdf/IIRA-v1.9.pdf (accessed on 18 February 2021)
4. Vitturi, S.; Zunino, C.; Sauter, T. Industrial communication systems and their future challenges: Next-generation ethernet, iiot, and 5 g. *Proc. IEEE* **2019**, *107*, 944–961. [CrossRef]
5. Al-Hawawreh, M.; den Hartog, F.; Sitnikova, E. Targeted ransomware: A new cyber threat to edge system of brownfield industrial Internet of Things. *IEEE Internet Things J.* **2019**, *6*, 7137–7151. [CrossRef]
6. Mahmood, A.N.; Leckie, C.; Hu, J.; Tari, Z.; Atiquzzaman, M. Network traffic analysis and scada security. In *Handbook of Information and Communication Security*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 383–405.
7. Faisal, M.A.; Cardenas, A.A.; Wool, A. Profiling communications in industrial ip networks: Model complexity and anomaly detection. In *Security and Privacy Trends in the Industrial Internet of Things*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 139–160.
8. Lin, C.-Y.; Nadjm-Tehrani, S. Understanding iec-60870-5-104 traffic patterns in scada networks. In Proceedings of the 4th ACM Workshop on Cyber-Physical System Security, Incheon, Korea, 4–8 June 2018; pp. 51–60.
9. Lin, C.-Y.; Nadjm-Tehrani, S. Timing patterns and correlations in spontaneous {SCADA} traffic for anomaly detection. In Proceedings of the 22nd International Symposium on Research in Attacks, Intrusions and Defenses, Beijing, China, 23–25 September 2019; pp. 73–88.
10. Lin, C.-Y.; Nadjm-Tehrani, S. A comparative analysis of emulated and real iec-104 spontaneous traffic in power system networks. In Proceedings of the 1st International Workshop on Cyber-Physical Security for Critical Infrastructures Protection, Guildford, UK, 18 September 2020.
11. Soós, G.; Ficzere, D.; Varga, P. Investigating the network traffic of Industry 4.0 applications–methodology and initial results. In Proceedings of the 16th International Conference on Network and Service Management (CNSM), Izmir, Turkey, 2–6 November 2020; pp. 1–6.
12. Sloman, M. *Network and Distributed Systems Management*; Addison-Wesley Longman Publishing Co., Inc.: Boston, MA, USA, 1994.
13. Weissenberg, M.; Głąbowski, M.; Hanczewski, S.; Stasiak, M.; Zwierzykowski, P.; Bai, V. Traffic modeling in industrial ethernet networks. *Int. J. Electron. Telecommun.* **2020**, *66*, 145–153.
14. Matoušek, P.; Ryxsxavỳ, O.; Grégr, M.; Havlena, V. Flow based monitoring of ics communication in the smart grid. *J. Inf. Secur. Appl.* **2020**, *54*, 102535. [CrossRef]
15. Mohammed, A.M.; Agamy, A.F. A survey on the common network traffic sources models. *Int. J. Comput. Netw.* **2011**, *3*, 103–115.
16. Soysal, M.; Schmidt, E.G. Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison. *Perform. Eval.* **2010**, *67*, 451–467. [CrossRef]
17. Pacheco, F.; Exposito, E.; Gineste, M.; Baudoin, C.; Aguilar, J. Towards the deployment of machine learning solutions in network traffic classification: A systematic survey. *IEEE Commun. Surv. Tutor.* **2018**, *21*, 1988–2014. [CrossRef]
18. Åkerberg, J.; FurunäsÅkesson, J.; Gade, J.; Vahabi, M.; Björkman, M.; Lavassani, M.; Gore, R.N.; Lindh, T.; Jiang, X. Future Industrial Networks in Process Automation: Goals, Challenges and Future Directions. *Appl. Sci.* **2021**, under review.
19. Rossum, G.V.; Drake, F.L., Jr. *Python Tutorial*; Centrum Voor Wiskunde en Informatica Amsterdam: Amsterdam, The Netherlands, 1995.
20. Cao, Y.; Li, Y.; Liu, X.; Rehtanz, C. Modeling and simulation of data flow for vlan-based substation communication system. In *Cyber-Physical Energy and Power Systems*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 75–101.
21. Willinger, W.; Taqqu, M.S.; Leland, W.E.; Wilson, D.V. Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements. *Stat. Sci.* **1995**, *10*, 67–85. [CrossRef]
22. Willinger, W.; Taqqu, M.S.; Sherman, R.; Wilson, D.V. Self-similarity through high-variability: Statistical analysis of ethernet lan traffic at the source level. *IEEE/ACM Trans. Netw.* **1997**, *5*, 71–86. [CrossRef]