BEMO-COFRA

Brazil-Europe Monitoring and Control Framework

(Project No. 288133)

D3.3 - Traffic Modelling for Industrial Applications

Published by the BEMO-COFRA Consortium

Dissemination Level: Restricted

Project co-funded by the European Commission within the 7th Framework Programme
and
Conselho Nacional de Desenvolvimento Científico e Tecnológico
Objective ICT-2011-EU-Brazil

Document control page

Document file: D3.3 - Traffic Modelling for Industrial Applications
Document version: 0.1
Document owner: Djamel Sadok (UFPE)

Work package: WP3 - Large-scale distributed system architecture
Task: T3.3 – Traffic Models for the Floor
Deliverable type: P

Document status: ▫ approved by the document owner for internal review

x approved for submission to the EC

Document history:

Version	Author(s)	Date	Summary of Changes made
0.1	Vinícius Fraga (UFPE) Diogo Falcão (UFPE)	20/08/13	First version
0.2	Djamel Sadok (UFPE)	28/08/13	Refined document structure and added more details
0.3	Djamel Sadok (UFPE)	29/08/13	Final version submitted to the European Commission

Internal review history:

Reviewed by	Date	Summary of comments
Judith Kelner (UFPE)	28/08/13	Accepted

Index:

1. Executive summary

This is the document for the prototype deliverable 3.3: Traffic Models for the Floor. This work is part of WP3: Large-scale distributed system architecture.

This document describes the methods and models considered to obtain the final traffic model for an industrial environment, through the analysis of traces of captured traffic data, resulting in a stochastic model. This model should assist the project developers in next tasks, providing an accurate guidance to achieve the main objectives of BEMO-COFRA, considering its previously described requirements.

2. Introduction

The BEMO-COFRA framework seeks achieving an effective monitoring and control methods of large-scale complex systems at manufacturing plants that by nature adopt a large number of heterogeneous smart devices networked. Thus, scalability and performance issues are both of utmost importance to the project's success. These parameters can be measured and generated based on a network traffic model, which is the purpose of this document to detail.

In order to provide a reliable network traffic model to support all future BEMO-COFRA development work, a packet capture process was performed in the the industrial environment available to our industrial partner - Comau. The statistical analysis of such data is the tool naturally employed to obtain a desired stochastic model; there are also some well known methods to achieve such a result, that will be briefly described in the this document.

2.1. Purpose, context and scope of this deliverable

Task 3.3 focuses on describing the network behavior in an industrial environment, in order to establish a reliable reference capable of assisting in future development phases of the project.

The concerned environment is a car manufacturing line, composed by hundreds of welding robots, assembly robots, painting robots and so on. There is also constant movement of metal parts everywhere.

The scope of the task includes the analysis of a previously captured trace, provided by our industrial partners. The metrics chosen provide visible results and correlations in terms of traffic priority, latency, packet loss and more, in order to, combined with spectrum measures, expose a reliable model to support the design of the wireless communication in terms of protocols and algorithms, considering scalability and performance.

2.2. Background

The results provided by D3.1:Robotics and sensor integration, and D3.2:Initial Architectural Design Specification, grant useful information about the specific characteristics of the robots which compose the industrial scenario and its communication limitations. These include the network architecture's basic structure allowing a more accurate metric selection and objective evaluation, resulting in a refined model in these terms. Therefore, the quality of the provided trace (when speaking of duration and details richness; this will be discussed in section 4) becomes the bottleneck of the model's accuracy and reliability.

2.3. Deliverable structure

This document is organized as follows: section 3 addresses the principal statistical methods used to model the network traffic data, highlighting the ones chosen for the task to which this document is related; section 4 describes the provided trace, used as source for the analysis, in terms of presented information and duration; section 5 presents the results of applying the methods mentioned in section 3; section 6 unfolds the noticed correlations between the metrics and proposes a stochastic traffic model; section 7 contains the external references.

3. Methodology

This section brings the description of the methods and techniques used to obtain the results presented in the current document, and the reasoning behind using each one of these as well. All the employed methods take as input a trace of captured traffic at the floor level of a car manufacturing line, saved as a "pcapn" binary file. The file is then analyzed with the WireShark tool version 1.8.2, in addition to other tools, such as the statistical tool R version 3.0.1 and other common spreadsheet manipulators.

3.1. Histogram

A well-known statistical technique, the histogram presents the frequency of occurrence in time of a determined parameter. It is useful to have a clear view of the predominance of length of the packets and presence of a specific protocol. In these cases, histograms can show the viewer conclusions regarding bandwidth requirements and exclusive control links, for example. On the other hand, a high contrast between quantities can make it hard to visualize a determined subset of data, i.e., a major subset with values in a higher scale.

3.2. Box Plot

The box plot is a graphical representation of data that shows a data set's lowest value, highest value, median value, and the size of the first and third quartile (25th and 75th percentiles). The box plot is useful for analyzing data sets that do not lend themselves easily to histograms. Because of the small size of a boxplot, it is easy to display and compare several box plots in a small space. A box plot is a good alternative or complements a histogram and is usually better for showing several simultaneous distributions, hence being very useful for comparisons.

3.3. Per Protocol Percentage of Packets

Considering that the purpose of a traffic model is to serve as a reference to the definition and evaluation of metrics and objectives to be targeted during the development of the BEMO-COFRA framework, it is important to know the occupation and relevance of each protocol used in an industrial communication system. As a results it would be possible to determine traffic priorities, for example, according to the protocols' role and volume of data related to it.

As a result, there a number of graphs relating to the quantity of packets and their volumes in bytes will be presented for each one of the protocols identified in the trace analysis.

3.4. Protocol Hierarchy Statistics

During the analysis of the trace, it was noticed that a number of different protocols different protocols were used at the same protocol stack layer; therefore, it is interesting to know which protocol is making use of what other protocol services and in which proportion this happens, for each specific protocol at a specific layer. This work leverages on the use of the WireShark tool which provides a "protocol hierarchy" window containing exactly this kind of information.

3.5. Flow Analysis of the Link Layer Conversations

Although most of the communication of a computer network occurs using upper layers protocols, there are important phases of neighborhood discovery which use only link layer functionalities, i.e., the information is passed in frames that are immediately processed when some specific link layer headers are identified. Actually, such events happen because of the need to obtain link and network addresses to establish the routing tables and then start communication using protocols of the upper layers. Thus, this document brings

results based on the analysis at the frames conversation level, gathering this useful knowledge and representing it as flows analysis graphs (where each conversation is seen as a flow).

4. Network Traffic Data

This section describes the characteristics of the analyzed trace. Since all the results originated from the analysis are obtained considering some determined parameters, it is of utmost importance to have the adequate fields available in the trace, therefore, the next subsections discuss these characteristics.

4.1. Trace Time Aspects

According to the trace timestamps, the WireShark tool indicates that the whole captured traffic corresponds to thirty three minutes of communication, between 3:40 A.M. and 4:13 A.M.

4.2. Trace Fields

The following fields have been captured in the trace: [Packet Number][Time of Arrival][Source Address] [Destination Address][Protocol][Length][Info]

- Packet Number refers to the order of capturing;
- Time of Arrival is the trans-curred time from the beginning of the capture to the moment the current packet was captured;
- Source and destination addresses and Protocol are self-explainable;
- Length is the packet length, in bytes;
- Info represents the "message" contained in the packet's payload.

4.3. Perceived Protocols and their Purposes

Several protocols were detected during the analysis; each one has its own functionalities and peculiarities, thus, it is relevant to have some knowledge about the role of each protocol. To achieve this, a brief description of all the detected protocols follows:

- TCP
  - Transmission Control Protocol. Provides reliable, ordered and error-checked delivery.
- SMB
  - Provides shared access to files, printers, serial ports and miscellaneous communications.
- NBNS
  - NetBios Name Service - similar to DNS but more limited, with names stored in a flat space and without support to IPv6.
- LLDP
  - Discovery purposes, over Ethernet frames. Vendor neutral.
- ARP
  - Telecommunications protocol used to resolve network layer addresses into link layer addresses.
- LLMNR
  - Provides name resolution for IPv4 and IPv6 hosts on the link, Based on the DNS packet format.
- DHCP and DHCPv6
  - Automatic IP hosts configuration (IPv4 and IPv6, respectively).
- PN-DCP
  - Configuration of IP addresses and station names, on subnets. It is normally used without the presence of a DHCP server.
- BROWSER

  - Typically, uses SMB as its transport protocol.
- NBSS
  - NetBios Session Service, designed for connection-oriented communications, lets two hosts establish a connection for a conversation, allows larger messages to be handled, and provides error detection and recovery.
- ICMP
  - Internet Control Message Protocol, commonly used to control purposes generated in response to errors.
- LANMAN
  - Provides hash functionalities to store users passwords.
- SSDP
  - Simple Service Discovery Protocol, provides advertising and discovery network services without server-based mechanisms or static configuration of a host. *Intended for use in small environments.*
- IGMPv3
  - Internet Group Management Protocol, used on IP networks to establish multicast group memberships.
- UDP
  - User Datagram Protocol, a well known transport layer protocol which aims for simplicity. It is not reliable, and thus does not provide error checking.
- LLC
  - Logical Link Control, the upper sublayer of data link layer. Provides multiplexing mechanisms to allow several network layer protocols to coexist.
- MDNS
  - Similar to DNS; receives a name request and multicasts the response over UDP to every node in its subnetwork.

5. Partial Results

This section presents the results for each aforementioned method in section 3.

5.1. Box plot Results

There are two box plot analysis made: one encompassing all the protocols in one set of packets (Figure 1), and a per protocol version (Figure 2).

In the first one, it's visible that the lower quartile and the median are coincident; 25% of the analyzed packets are under the lower quartile, in which the length threshold is around 62 bytes. Also, 50% are between 62 and 143 bytes (interval between the lower quartile and the median). Until the third quartile, the length of the packets starts to get bigger and dispersed. Finally, in the upper quartile, there are the other 25%, varying from 143 to 255 bytes. Many outliers were identified, indicating some possible conclusions:

- The employed measurement method is incorrect (this hypothesis can be discarded, since this method is largely used).
- Another possible reason for the outliers frequent occurrence is when multiple probabilistic distributions overlap each other, like in a mixture model.
- What we can surely conclude is that the packet length is directly related to the protocol; the prevailing protocol in this analysis is TCP, which packets are usually 60 bytes long in this trace. Nevertheless, some other protocols have significant packets amount, as SMB, which packets are 400 bytes long, explaining the presence of outliers.

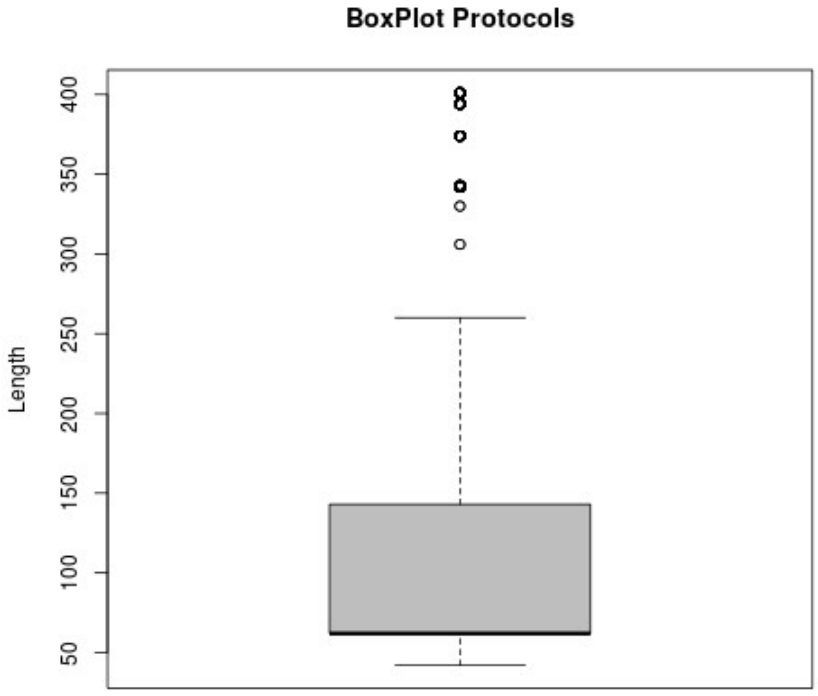

Figure 1: Box plot of captured packets, without protocol differentiation.

The second graph presents a per protocol box plot analysis; it's now possible to observe the protocols diversity and assume some parameters variation, led by the length differences between packets of a same protocol.

In spite of the noticed differences and variations, many of the observed protocols also present similarities in terms of data dispersion, such as ARP, DHCP, DHCPv6, ICMP, IGMPV3, LLC, MDNS, NBNS, PN-DCP and SSDP.

On the other hand, there are protocols with significant deviation, as SMB, with packet length variating from 93 bytes to 401 bytes, and also a 200 bytes distance between the lower and the upper quartile.

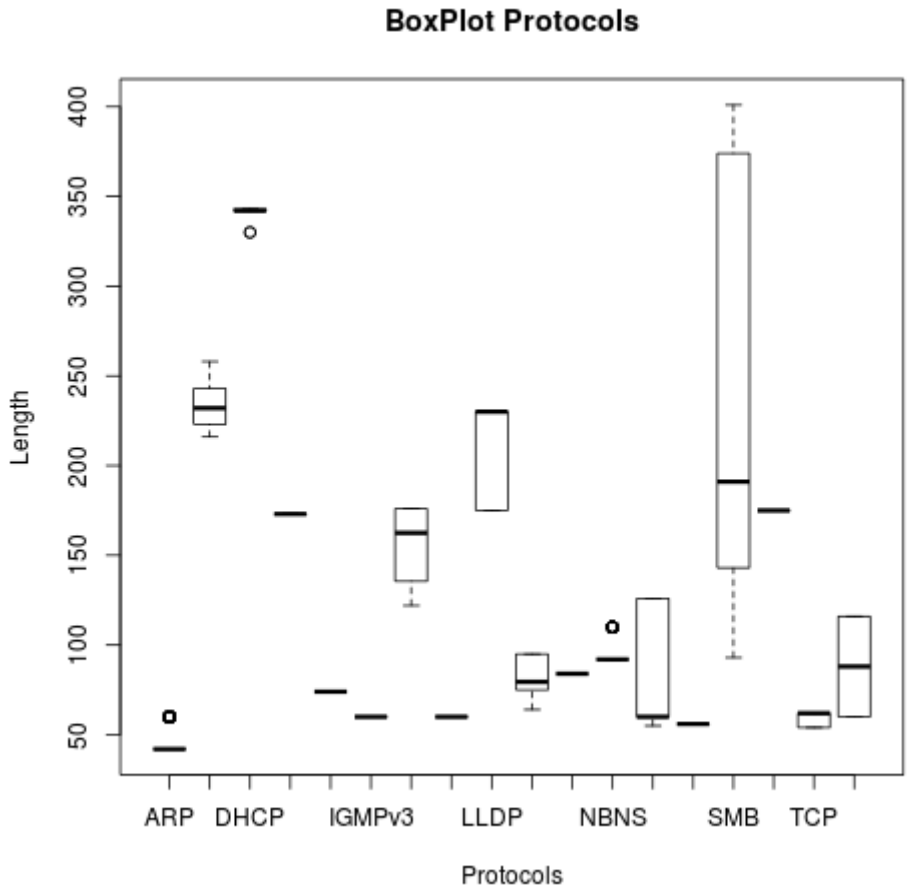

Figure 2: A box plot of packets per protocol.

5.2. Packets per Protocol

To base our conclusions about the outliers noticed in the previous box plot graphs, we present now a pie chart with the per protocol percentage of packets, revealing, as aforementioned, the TCP's predominance. Also, it's clear that SMB has a representative amount of packets, what enhances the cause theory of the outliers perceived in the first box plot.

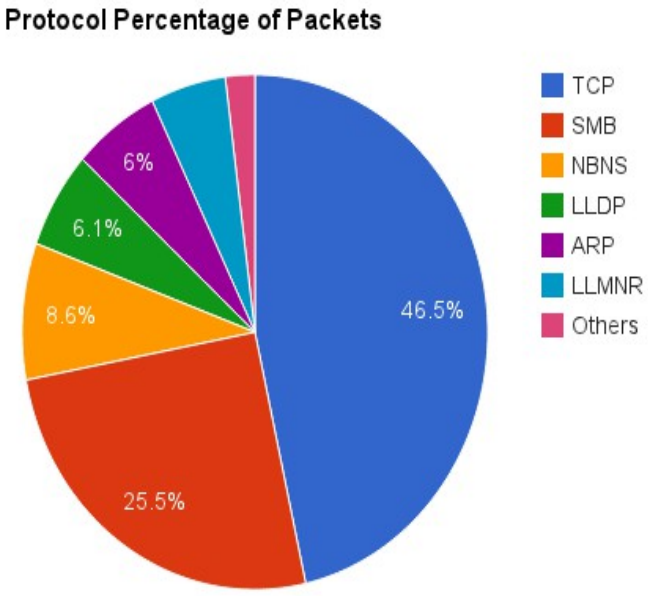

Figure 3: Per protocol percentage of packets.

5.3. Protocol Hierarchy

As the focus of BEMO-COFRA is, when speaking of link layer, to work with wireless connections, it's relevant to point that all the traffic captured is composed by Ethernet frames. No wireless traffic was analyzed, as it actually does not exist yet in the manufacturing line (our partners had some previous experience with wireless technology though, shortly abandoned due to several connectivity problems). Figure 4 is a set of results from the WireShark tool, presenting the protocol stack hierarchy of the trace.

Protocol	% Packets	Packets	% Bytes	Bytes	Mbit/s	End Packets	End Bytes	End Mbit/s
▾ Frame	100,00 %	12015	100,00 %	1458770	0,006	0	0	0,000
▾ Ethernet	100,00 %	12015	100,00 %	1458770	0,006	0	0	0,000
Address Resolution Protocol	6,01 %	722	2,27 %	33042	0,000	722	33042	0,000
Link Layer Discovery Protocol	6,13 %	737	10,34 %	150865	0,001	737	150865	0,001
▾ PROFINET Real-Time Protocol	0,56 %	67	0,26 %	3752	0,000	0	0	0,000
PROFINET DCP	0,56 %	67	0,26 %	3752	0,000	67	3752	0,000
▾ Internet Protocol Version 6	2,90 %	349	2,38 %	34663	0,000	0	0	0,000
▾ User Datagram Protocol	2,90 %	349	2,38 %	34663	0,000	0	0	0,000
Domain Name Service	2,61 %	314	1,96 %	28608	0,000	314	28608	0,000
DHCPv6	0,29 %	35	0,42 %	6055	0,000	35	6055	0,000
▾ Internet Protocol Version 4	84,39 %	10139	84,76 %	1236388	0,005	0	0	0,000
▾ User Datagram Protocol	12,08 %	1452	10,25 %	149533	0,001	0	0	0,000
Domain Name Service	2,62 %	315	1,54 %	22412	0,000	315	22412	0,000
NetBIOS Name Service	8,57 %	1030	6,53 %	95318	0,000	1030	95318	0,000
Bootstrap Protocol	0,56 %	67	1,57 %	22920	0,000	67	22920	0,000
▾ NetBIOS Datagram Service	0,29 %	35	0,56 %	8182	0,000	0	0	0,000
▾ SMB (Server Message Block Protocol)	0,29 %	35	0,56 %	8182	0,000	0	0	0,000
▾ SMB MailSlot Protocol	0,29 %	35	0,56 %	8182	0,000	0	0	0,000
Microsoft Windows Browser Protocol	0,29 %	35	0,56 %	8182	0,000	35	8182	0,000
Data	0,02 %	2	0,01 %	176	0,000	2	176	0,000
Hypertext Transfer Protocol	0,02 %	3	0,04 %	525	0,000	3	525	0,000
▾ Transmission Control Protocol	72,18 %	8673	74,44 %	1085847	0,004	5573	332340	0,001
▾ NetBIOS Session Service	25,68 %	3086	51,60 %	752737	0,003	16	1415	0,000
▾ SMB (Server Message Block Protocol)	25,55 %	3070	51,50 %	751322	0,003	3066	750699	0,003
▾ SMB Pipe Protocol	0,03 %	4	0,04 %	623	0,000	0	0	0,000
Microsoft Windows Lanman Remote API Protocol	0,03 %	4	0,04 %	623	0,000	4	623	0,000
Data	0,12 %	14	0,05 %	770	0,000	14	770	0,000
Internet Group Management Protocol	0,02 %	2	0,01 %	120	0,000	2	120	0,000
Internet Control Message Protocol	0,10 %	12	0,06 %	888	0,000	12	888	0,000
▾ Logical-Link Control	0,01 %	1	0,00 %	60	0,000	0	0	0,000
Data	0,01 %	1	0,00 %	60	0,000	1	60	0,000

Figure 4: WireShark's protocol hierarchy statistics.

Another relevant aspect of the network showed in these results is that 84,39% of the packets are using IPv4 protocol; we also noticed that some nodes work as gateways between subnetworks using different network layer protocols, indicating possible scalability problems that might be resolved by upgrading the whole network to IPv6, which currently represents a small part of it.

5.4. Packet Length

In this subsection follows a histogram of packets size in bytes; can be used to determine requirements of bandwidth.

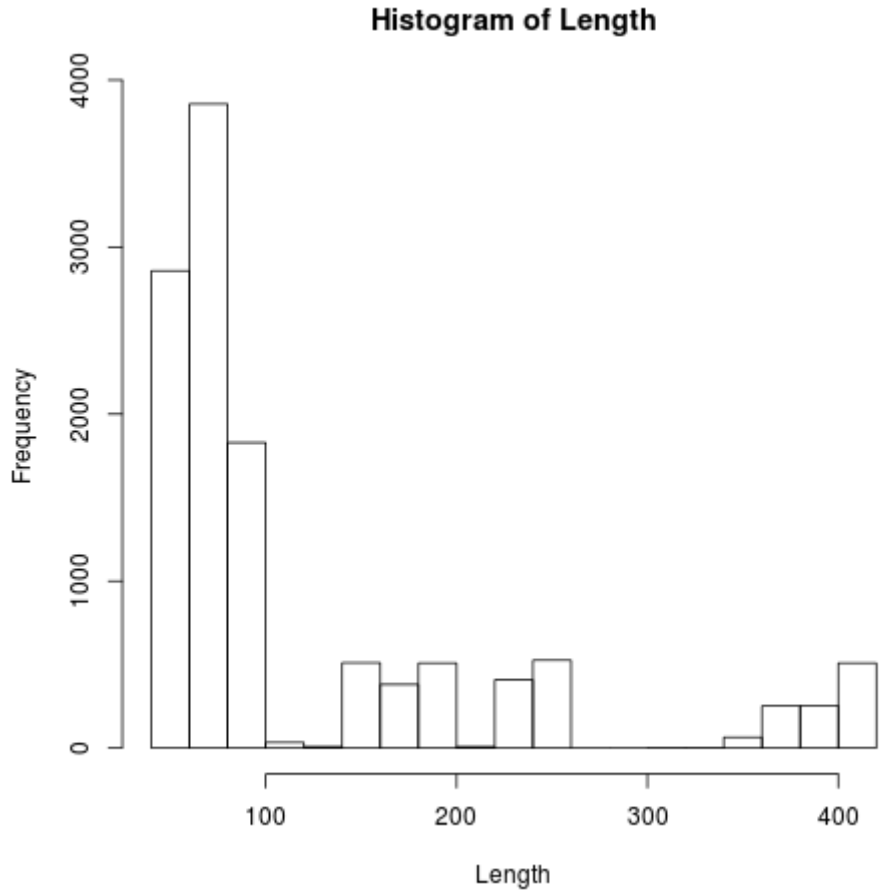

Figure 5: Histogram of packets length in bytes.

Complementary to BoxPlot, we have the histogram of the size of the packets evaluated. In this histogram one may observe that most of the packets have a size around 100 bytes, with small volumes between 140 and 250 bytes and elsewhere between 350 and 450 bytes approximately. As stated previously, the histogram was strongly influenced by some protocols, including TCP and ARP that use packets with an average length between 60 bytes and 40 bytes respectively. In the second group we have the protocols DHCPv6 , LLDP, Browser, LANMAN, whereas in the third group we find that the primary responsibility for determining packet size comes from SMB.

5.5. Link Layer Conversations

Last but not least, a number of link conversations were identified. At this point of the analysis, it is possible to have a view of the trace as a set of flows (in this case, each conversation between two addresses represents a flow). Until this section, all the graphs have showed us information at the packet level, denoting characteristics of data volume for both absolute and per protocol approaches. From now on, the flows' study can lead us to understand the nature of the network events perceived in the trace, e.g., it is

possible to infer metrics from noticed discrepancies between the flows, considering protocols and its already known respective volumes of data.

Ethernet Conversations: 21

Address A	Address B	Packets ▲	Bytes	Packets A→B	Bytes A→B	Packets A
SiemensA_ea:e4:f2	IntelCor_92:b1:90	8 659	1 085 574	4 586	603 721	
IntelCor_92:b1:90	Broadcast	937	96 215	937	96 215	
SiemensA_ea:e4:f2	Broadcast	577	24 658	577	24 658	
Siemens_00:bd:2f	LLDP_Multicast	398	91 540	398	91 540	
SiemensA_ea:e4:f2	LLDP_Multicast	339	59 325	339	59 325	
IPv6mcast_00:01:00:03	IntelCor_92:b1:90	314	28 608	0	0	
IPv4mcast_00:00:fc	IntelCor_92:b1:90	314	22 328	0	0	
Intel_61:d9:3d	Broadcast	190	29 267	190	29 267	
SiemensA_ea:e4:f2	Siemens_27:84:67	44	2 563	16	883	
IPv6mcast_00:01:00:02	IntelCor_92:b1:90	35	6 055	0	0	
Siemens_00:bd:a5	Broadcast	34	2 040	34	2 040	
Siemens_00:bd:2e	PN-MC_00:00:00	34	1 904	34	1 904	
Siemens_00:bd:a5	PN-MC_00:00:00	33	1 848	33	1 848	
Siemens_00:bd:2e	Broadcast	33	1 980	33	1 980	
Siemens_2f:50:6e	Broadcast	32	1 920	32	1 920	
SiemensA_a3:8d:e1	Broadcast	31	1 860	31	1 860	
Siemens_27:84:67	Broadcast	4	240	4	240	
Intel_61:d9:3d	IPv4mcast_7f:ff:fa	3	525	3	525	
Intel_61:d9:3d	IPv4mcast_00:00:16	2	120	2	120	
Siemens_51:18:ea	Broadcast	1	116	1	116	
IPv4mcast_00:00:fb	IntelCor_92:b1:90	1	84	0	0	

Figure 6: Ethernet conversations. Most of it have discovery purposes.

The flows as sorted by their quantity of packets. At first sight, it is possible to perceive a great difference of packets and, consequently, transmitted bytes, between the flow number 1 and all the remaining flows. This indicates that the link between these two nodes is a critical one, and the nodes might be running some important functionality to the network. Thus, this link needs special attention, leading us to the idea of a configurable and dynamic topology based on policies for example.

Looking at the rest of the flows, it is noticeable that most of these are broadcast conversations. This foments the aforementioned hypothesis that the traffic was captured during a discovery phase, an event in which this kind of communication mode among the nodes is very common to take place.

The figures 6, 7, 8 and 9 present this information. Particularly in figure 9, the aforementioned discrepancy becomes quite visible.

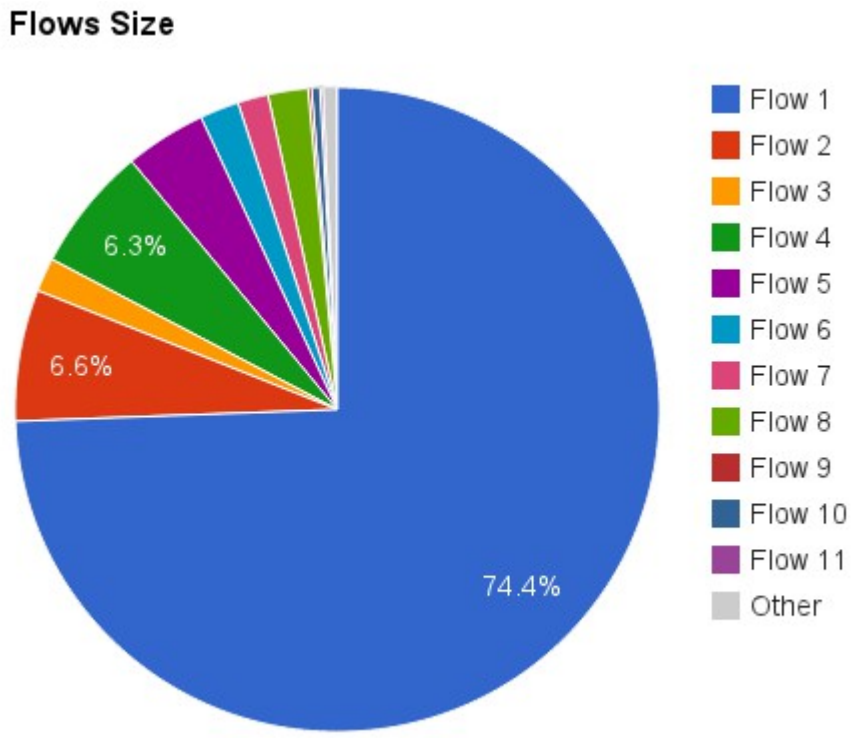

Figure 7: Size of the flows, in bytes.

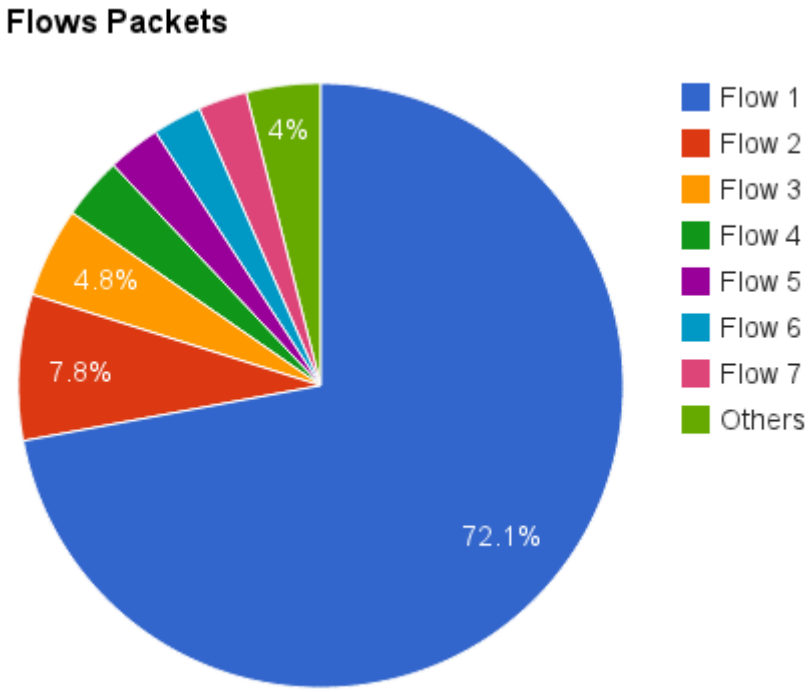

Figure 8: Quantity of packets per flow.

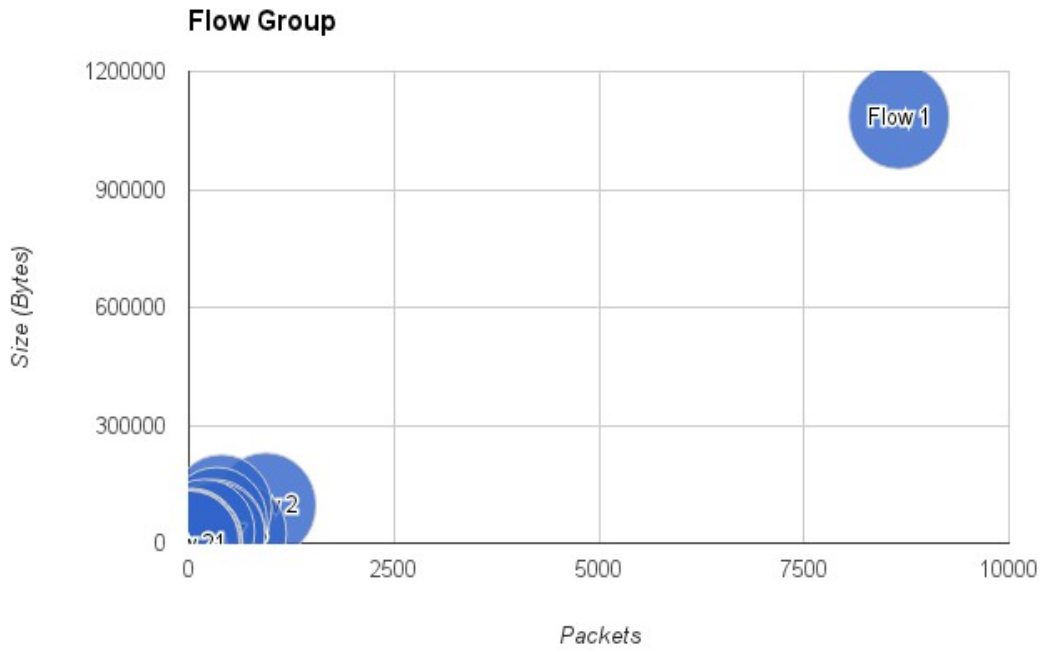

Figure 9: The vertical axis represents the transmitted bytes, while the horizontal axis represents the quantity of transmitted packets.

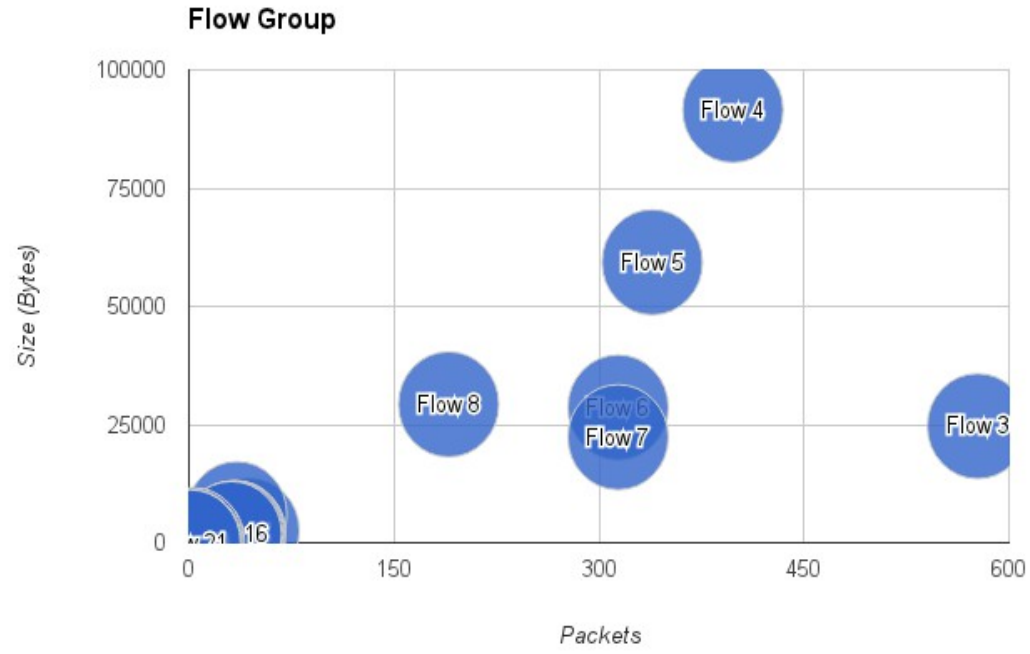

Figure 10: More detailed flow group, without the flow 1.

5.6. Perceived Network Events

At this point, a set of network events should be enumerated and described, in order to provide enough information for any conclusions to be assumed. On the other hand, as aforementioned in section 2, the limitation of the present analysis comes from the small and limited trace of captured traffic we gained access to. It is of utmost relevance to consider that the taken conclusions are strictly related to this trace. It is hoped that longer traces are obtained in the future and more general conclusions may be made.

According to the network behavior that could be observed, the studied trace is in its majority of a network discovery phase: it mainly contains broadcast flows, and many discovery protocols, such as LLDP, PN-DCP, LLMNR, NBNS, ARP and so on. As the trace time length is about thirty three minutes, this discovery phase is the only clearly and dominating perceived event occurring during the interval.

6. Conclusion

After the accounting and gathering of all the partial results, what we actually have in hands to construct a traffic model for the manufacturing floor can be organized as follows:

- In spite of having a vast set of protocols being used, the traffic is composed in its majority by TCP segments over IPv4 packets;
- There is no wireless communication due to connectivity problems, caused by external interference.
- In spite of most of the traffic being of small packets, there are some large packets to consider which are important when considering bandwidth.
- There is a heavy communication between a small set of nodes, which might overload some given links.
- Most of the traffic has control purposes, creating stringent requirements in terms of delay and packet loss.
- So far the only event clearly noticed is a discovery phase of the network, which probably does not reliably represents the network behavior during most of the time.

7. References

(BEMO-COFRA Consortium, 2012) D3.1: Robotics and Sensor Integration

(BEMO-COFRA Consortium, 2012) D3.2: Initial Architectural Design Specification