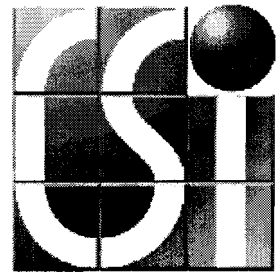
Reliability and availability of material handling systems

de Mey, D.W.

Award date:
2003

TU/e



Reliability and Availability

of

Material Handling Systems

Student	**: Ing. D.W. de Mey**
Student number	**: 460893**
University	**: Eindhoven University of Technology**
Department	**: Electrical Engineering**
Section	**: Information and Communication Systems (ICS)**
Research Chair	**: Computers, Networks & Design (CND)**
Support TU/e	**: Prof. Ir. F. van den Dool**
Company	**: CSi Industries B.V.**
Support CSi	**: H. Leusink**
Date	**: February 2003**

Summary

This paper describes a graduation assignment, which is done to end the study Electrical Engineering, specialization Information Technical Science at the Eindhoven University of Technology. This graduating assignment is done at CSi Industries B.V. at Raamsdonksveer, a supplier of Material Handling Systems. These systems take care of the internal logistics of a company.

Because the companies, which buy such a system, look more and more at the price of such a system and not at the performance of the system in itself, it becomes harder and harder to sell a system.
To make good systems you have to use good parts. Good parts are most of the time more expensive than parts of lesser quality. CSi's vision is; they want to sell good and reliable systems and so the systems are most of the time more expensive then the systems of the competitor.
To justify the higher price they want to find a way to calculate the reliability and availability of a system. This because if you can justify the higher price, for example by delivering a better performance of the system than the competitor's system has, you can explain this to the customer.

The customers of CSi also know that when a system fails and as a consequence there is no production, this costs the customer a lot of money. The customer then can decide by himself where to place his priorities; on purchasing a cheaper system or on purchasing a more expensive, but more reliable system.

So it is important to find out how the reliability and availability of systems can be investigated and calculated. Therefore this research and graduating assignment have been done on how reliability and availability of Material Handling Systems can be investigated.

Contents

Acknowledgements

I would like to thank everyone who contributed directly or indirectly to this project. However, some people deserve a special 'thank you'.

First of all I would like to thank my supervisor of the Eindhoven University of Technology, Prof. Ir. F. van den Dool. He supported this project from the beginning till the end from the Eindhoven University of Technology.
Furthermore I would like to thank dhr. H. Leusink of CSi Industries B.V. at Raamsdonksveer. He started this project and gave me the chance to do my graduating assignment at CSi.

A number of people helped me to do this project and gave me some practical tips and advice about how to do research to realize the goals of this project.
These people are Prof. Dr. Ir. A.C. Brombacher and Dr. Ir. J.L. Rouvroye of the Eindhoven University of Technology, Department of Technology Management.

Chapter 1

Introduction

This graduating assignment is done to investigate of Reliability and Availability of Material Handling Systems. In this chapter an introduction of CSi Industries B.V., the supplier of the systems, a short summary about the study Electrical Engineering and a description of the graduating assignment are given

1.1 Introduction of CSi Industries B.V.[1]

CSi serves as a worldwide partner in designing, producing, supplying and implementing integrated logistical systems for material handling and physical distribution. The market they serve consists of producers of fast moving consumer goods (FMCG) and their distributors.

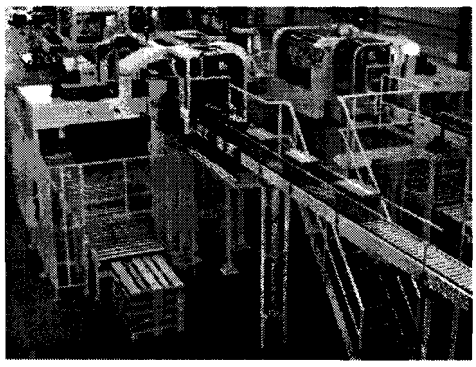
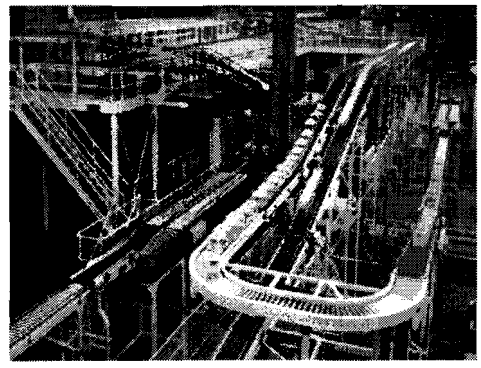
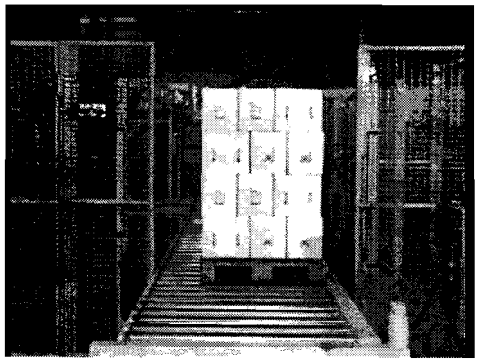
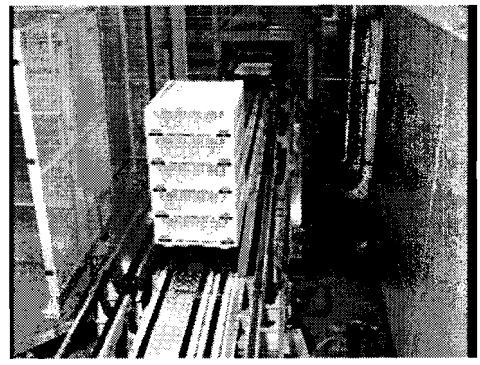
They support the customers in achieving their business objectives by designing, engineering and manufacturing sophisticated material handling systems with innovative and reliable production lines by implementing state-of-the art logistic and IT solutions. Reliability and availability are therefore a primary criterion in designing solutions.

The scope of supply comprises robotized secondary packaging, palletizing, conveying and automatic truck loading equipment including all required data control systems, translating Enterprise Resource Planning (ERP) data into physically operating packaging and internal transportation processes.

The correct handling of pallets, whether loaded or not-loaded, is an extremely important part of the logistical process. CSi pallet handling offers many kinds of solutions.
One example is that our conveyors can carry your loaded pallets automatically from your palletisers to the truck loading system. The control system can also closely monitor the pallets and ensure that the information is up-to-date and is available at any given time.

[1] For more information about CSi Industries B.V. see www.csiweb.nl

Some examples of the activities of CSi can be seen in the figures below:

1.2 Eindhoven University of Technology

The Eindhoven University of Technology provides engineering, postgraduate design, teacher-training programs, and post-academic courses. Education is based on the university's own research activities with a focus on design. Lecturers and students make use of modern information and communication means.

The Department of Electrical Engineering occupies with hardware and software. The interaction between hardware and software becomes more and more important, because more and more devices contain a microprocessor, which can be directed through software. Think of devices like the mobile phone, the television set etcetera. All these devices contain hardware and software components.

Educational and research activities within the Department of Electrical Engineering are carried out by four sections, each consisting of a number of research chairs.

The four sections with there research chairs are:

1. Electrical Power Engineering (EVT)
    - Electrical Power Systems (EPS)
    - Electro mechanics and Power Electronics (EPE)
2. Information and Communication Systems (ICS)
    - Computer, Networks and Design (CND)
    - Design Automation (ES)
    - Mixed-signal Microelectronics (MsM)
3. Measurement and Control Systems (MBS)
    - Control Systems (CS)
    - Signal Processing Systems (SPS)
4. Telecommunication and Electromagnetism
    - Electro-optical Communication (ECO)
    - Radio communication (ECR)
    - Opto-electronic Devices (OED)
    - Electromagnetics (EM)

The section, which supported this graduating assignment, is Information and Communication Systems, research chair Computer, Networks and Design. They focus on the overall design and implementation trajectory of analog and digital hardware and software of electronic systems.

1.3 Graduating Assignment

This graduating assignment consists of two aspects. The first aspect deals with the fact that it is harder nowadays to sell a system than it was in the past, because the negotiator between the client and the provider is nowadays a purchaser instead of a technician. Therefore, the price is the primary interest today.
The second aspect of this graduating assignment deals with the fact that devices are becoming more and more intelligent, enlarging the strain on the system and rendering some protocols unable to handle the data stream

CSi's vision is to sell reliable systems, that is: to produce a system with reliable parts. The price of these reliable parts is most of the time higher than less reliable parts, making the system more expensive.
CSi's competitors can sell their systems at a lower price, but these systems are more reliable. But this is hard to explain to potential customers

When the performance of a system could be proven, the client could be convinced to buy a more reliable system, diminishing the costs of system-failure. When the reliability and availability of the system are investigated, the higher price of the more reliable system could be justified.

An investigation of the reliability of a system is also important with the upcoming trend of 'variable payment'. Here, the costumer does not pay for the system but for each handled pallet. Therefore, it is important for the provider of such a system to have as little downtime as possible.

This project has been started to investigate different ways in which reliability and availability can be calculated. Stated shortly, this aspect consists of two parts:

1. To do research on reliability and availability
2. To implement reliability and availability in Material Handling Systems

When measures of reliability and availability are done in a system, changes will occur in the architecture of this system. These will be taken into account in this assignment.
To realize this part of the project, some changes will also occur in the mechanical part of the system, but they do not belong to this graduating assignment because this is out of the range of Electrical Engineering.

Because no real data was available yet, some errors were hypothesized during this graduating assignment to create a realistic situation. The next errors are hypothesized:

- Material Errors
- Operator Errors
- System Errors

Now, the starting situation has been created.

In the second part of this graduating assignment, the architecture of Material Handling systems and the different levels of communication will be investigated. Because the devices generate more and more information, the different communication protocols have to be evaluated on whether they can handle this increasing load on the network. Therefore the different communication protocols will be assessed on several important criteria, for example availability and response time.

Chapter 2

Availability and Reliability in general

Research has been done to find out whether previous research was available on reliability and availability, especially on Material Handling Systems.

2.1 General information about Availability and Reliability

Firstly, information was searched at CSi. After that, literature and norms or standards to express reliability and availability was searched for, in libraries as well as on the Internet. A summary of the results of this search can be seen below.

2.1.1 Information at CSi

A document was found at CSi about the availability of systems, based on the VDI 3581 norm of April 1983. In this document, the following terms are defined:

- Availability
- Capacity
- Working Time
- Breaks
- Operating Time
- Breakdown Time
- Available Time

For the exact definitions and the relation between the definitions, see *appendix A*.

In the case of a single machine, availability is easy to define and measure. If these machines form part of a system, it does not necessarily mean that the entire system will be unserviceable if one part of it breaks down.
It also specifies the effect of a breakdown requirement; stating that the availability is not time-related and that some time is needed to start the installation, the so-called running-in time.

The following formula for the availability of one part of a system is given:

$$\text{Availability} = \frac{\text{Operating Time} - \text{Breakdown Time}}{\text{Operating Time}}$$

The next step is the availability of a system, a serial as well as a parallel one.

For the availability of serial systems the availabilities of both parts can be multiplied. For example, when two parts have an availability of 0.98 (= 98%) then the availability of the system in series is 0.98 X 0.98 = 0.9604 (= 96.04 %).

The same goes for parallel systems, keeping in mind the distinction between parallel systems with undetermined paths or parallel systems with predetermined paths.
For the availability of a parallel system with undetermined paths, the formula is:
(1 - ((1 - 0.98) X (1 – 0.98))) = 0.9996 (= 99.96%).
The formula for the availability of a parallel system with predetermined paths is:
(1 - ((1 - 0.98) + (1 – 0.98))) = 0.96 (= 96.00%).

Now the availability of a system can be calculated.
However, the values found using this formula do not resemble the values found in practice, as will be explained later on in this chapter.

2.1.2 Literature about reliability and availability

While literature on reliability and availability in Material Handling Systems is scarce, literature on reliability and availability in general was searched.
The results of this search can be found in a paper that was done for the course 'Bibliotheek practicum (5J053)'.

Because from this paper it became clear that there is a lot of incongruence about the following terms, a book by M.J.P. van der Meulen proved very useful in getting clear definitions of them:

- Availability
- Operational Availability
- Reliability

These definitions can be found in *appendix B,* except the definition of VDI. The definitions that have been used in this assignment can be found later on in this chapter.

2.1.3 Norms and Standards on reliability and availability

Research was conducted to find out if there are some norms or standards for calculating reliability and availability of systems. Special attention was paid to information which had with a relation to Material Handling Systems, like the previous called VDI 3581 norm or other information which can be useful to calculate reliability and availability.

There are different authoritative sources that publicise norms. Some of these sources are:

- Nederlandse Normalisatie-Instituut (NNI)
- Insurance Services Office (ISO)
- Verein Deutscher Ingenieure (VDI)
- Deutsches Institut fur Normung (DIN)
- IEEE

At these sites was searched for documents related to reliability and availability. The searching terms were:

- Availability (= Beschikbaarheid (Ned.), Verfügbarkeit (Du))
- Reliability (= Betrouwbaarheid (Ned.), Zuverlässigkeit (Du))

Next, a short summary of these authoritative sources will be given and the results of the searching for documents can be found in Appendix C.

2.1.3.1 Nederlandse Normalisatie-Instituut (NNI)[2]

Normalisation is the process by which appointments between parties can be made, on national, European or global level, about the (technical) specifications of a product, service or company process.
These parties can be companies, government or consumer organisations. The document in which the appointments are documented, is called a norm.
The goal of normalisation is to try to make the processes more efficient or to try to increase the quality of the products.
Norms will be edited on national, European or global level. National normalisation institutes accompany the (inter)national normalisation process and record the appointments.
The 'Nederlands Normalisatie-instituut' at Delft is the national normalisation institute for the Netherlands. The 'Nederlands Normalisatie-instituut' represents the Netherlands on European level as a member of the Comité Européen de Normalisation (CEN) and on global level as a member of the International Organisation for Standardisation (ISO).

The Nederlands Elektrotechnisch Comité (NEC) is for the Netherlands responsible for normalisation on the area of electrical engineering, information technology and telecommunication. The NEC represents the Netherlands on European level as a member of Comité Européen de Normalisation

[2] www.nni.nl

Electrotechnique (CENELEC) and on global level as a member of International Electrotechnical Commission(IEC).
The execution of the NEC normalisation activities is delegated on the section 'Elektrotechniek, Telecommunicatie en Informatievoorziening (ETI)' of the 'Nederlands Normalisatie-instituut'.

2.1.3.2 Insurance Services Office (ISO)[3]

The International Organisation for Standardisation (ISO) is a world-wide federation of national standards bodies from some 140 countries, one from each country.
ISO is a non-governmental organisation established in 1947. The mission of ISO is to promote the development of standardisation and related activities in the world with a view to facilitating the international exchange of goods and services, and to developing co-operation in the spheres of intellectual, scientific, technological and economic activity.
ISO's work results in international agreements which are published as International Standards.

2.1.3.3 Verein Deutsches Ingenieure (VDI)[4]

The VDI is an organisation, independent from all economic and political concerns, which serves the common good, with around 126.000 engineers and natural scientists, and 12.000 craftsmen involved. The VDI, founded in 1856, is today the largest technical-scientific organisation in Europe. In Germany, it serves as a:

- Speaker for the engineers and craftsmen in the business-circles as well as in public.
- Leading institution for further education of and the exchange of knowledge between craftsmen and executives.
- Competent partner in the front lines of technological-political decisions as well as in all questions that an engineer can ask himself in his work-environment.

Some of the most important goals of the VDI are:

- The co-operation between al the intellectual forces in engineering.
- The furtherance of technological research and development.

[3] www.iso.ch
[4] www.vdi.de

- The education and in-service training of engineers and their furtherance in economics, politics and community.
- The development of acknowledged rules of engineering.

First was made sure that the version of VDI 3581 of April 1983 was up-to-date. A newer version of VDI 3581 was found, however, and this was published in March 2001. The difference with the older version was that *only* the availability of a system was mentioned and not, like in the previous version, availability *and* reliability. So this was a big difference and this could be important.

2.1.3.4 Deutsches Institut für Normung (DIN)[5]

DIN, the German Institute for Standardisation, is a registered association, founded in 1917. Its head office is in Berlin. Since 1975 it has been recognised by the German government as the national standards body and represents German interests at international and European level.
DIN offers a forum in which representatives from the manufacturing industries, consumer organisations, commerce, the trades, service industries, science, technical inspectorates, government, in short anyone with an interest in standardisation, may meet in order to discuss and define their specific standardisation requirements and to record the results as German Standards.

Mission statement of the DIN Group

1. DIN, the German Institute for Standardisation, is a technical and scientific association forming together with its subsidiary companies the DIN Group. The DIN Group is throughout a modern services organisation, dedicated to the creation of technical rules and the promulgation of their application.

2. Standardisation in Germany is organised by DIN. They have the largest share of the work in developing European and International Standards.

3. Their work is based on consensus and transparency, and is committed to the public benefit. It is their aim to serve industry and society by our efforts.

4. They determine the state of the art in technology, document the results, and make these available to the public world-wide.

5. Their competence derives from their close co-operation with experts from all sectors of trade and industry, from the scientific community, and from government.

[5] www.din.de

6. Their relations with all parties that have an interest in our work are constructive, durable, and maintained in a spirit of mutual trust.

7. Their work is characterised by a keen sense of responsibility, by quality awareness, market relevance, efficiency, and the employment of modern technology.

8. Their services and the results of their work enhance the competitive position of the companies that use them, increase job security, promote deregulation, and hence reduce government expenditure, all of which contribute to the benefit of standardisation for the economy as a whole.

9. The DIN Group represents the most comprehensive source of information on the to generate and distribute market-relevant, knowledge-intensive services and products that contribute to the financing of our standardisation activities.

10. All members of their staff are committed to these tasks and objectives and contribute with competence and high motivation to their achievement

2.1.3.5 IEEE[6]

The IEEE (Eye-triple-E) is a non-profit, technical professional association of more than 377,000 individual members in 150 countries. The full name is the Institute of Electrical and Electronics Engineers, Inc., although the organization is most popularly known and referred to by the letters I-E-E-E.

Through its members, the IEEE is a leading authority in technical areas ranging from computer engineering, biomedical technology and telecommunications, to electric power, aerospace and consumer electronics, among others.

Through its technical publishing, conferences and consensus-based standards activities, the IEEE

- produces 30 percent of the world's published literature in electrical engineering, computers and control technology,
- holds annually more than 300 major conferences and
- has more than 860 active standards with 700 under development.

[6] http://ieeexplore.ieee.org

2.1.4 Internet

Searching on the Internet for some publications about reliability and availability, the site of the section Quality of Products and Processes of the department of Technological Management was found, a department of Eindhoven University of Technology.
At this section they do some research projects, specifically research with respect to business processes and they concentrate on methods, tools and systems to analyse, design and control the business processes. The research of the section PPK will be carried out along two lines: a theoretical one and an empirical one. The theoretical research line focuses on the development of formal process models and the development of formal control tools with respect to Quality, Reliability, and Re-usability.
The empirical line focuses on validation and application of these formal process models and control systems in practice. Thus the two research lines are strongly interrelated; the theoretical line feeds empirical research with models and rules to predict and improve the performance of operational processes while the empirical line feeds the theoretical research with research questions from business practice, and performance data enabling the validation of formal models and systems.

2.1.4.1 Projects with respect to Reliability

Currently the possibility to predict the reliability of a product during its early design phases is very limited. In these phases only few (if at all any) tangible products exist. As a result the prediction of the expected reliability behaviour is very difficult. Tests in these phases are usually too time consuming. An alternative could be the usage of reliability models. However a lot of classical reliability models are based on the assumption of physical/chemical degradation of components. Usually this approach covers only a minor part of the real reliability problems. Purpose of this research is to define a method which enables the optimisation of the products in their early design stages for all reliability relevant factors by using new generation reliability models.

2.1.4.2 Conversation with Prof. Dr. Ir. Brombacher

Contact has been searched with Prof. Dr. Ir. Brombacher, the head of the section and a conversation followed. In that conversation he makes clear that the standards or norms found are not useful to calculate the reliability of Material Handling Systems.
He gave the tip to do it in another way, because his experience was that the outcomes of the standards are not correct, that there exists a gap between the theory and the practice, just like the experience of CSi was. His preference was to make a Markov Analysis and with this analysis calculate the Reliability.

Afterwards, the decision was made to try this new approach.
Another tip was to investigate the reliability of *a part of the system* first, because in respect to the reliability you want to know when a system goes down. For a complete system there are a lot of reasons why a system goes down. To keep these reasons clear it is better to first consider a part of the system.

2.1.5 Conclusion on research of documents

The choices that were made using the documents described above will now be enunciated.

In a lot of documents, the difference between reliability and availability was not very distinct. In the book by M.J.P. van der Meulen, several of the different definitions were mentioned, confirming this finding
For CSi, availability and reliability are two distinct terms. Therefore, a clear difference between the definitions is important in this project.

When comparing the version of VDI 3581 at CSi and a more recent version, the striking difference was that reliability was not mentioned anymore; the newer version only dealt with availability.

In the other documents that were found, approximately the same things about availability are mentioned as in the latest version of VDI 3581. Therefore, the VDI 3581 is taken as starting information, but a special note here is that CSi has the experience, using the older version of VDI 3581, that the calculated values with this norm were not the same as the values found in practice. So this gap between theory and practice was investigated and ways were searched to diminish this gap.

Little information was found on the Markov model, the tip from Prof. Brombacher, but the little that was found was used and evaluated.

2.2 Research

The experience was that the standards and norms gave not the expected results. The question is why there is a gap between the expected value and the calculated value. Therefore, the latest version of the VDI 3581 norm, which was published in March 2001, was examined.

2.2.1 Examination of VDI 3581

Inspecting the VDI 3581 it was interesting to see that the most recent version speaks only about Availability and **not** about Availability AND Reliability.

VDI 3581 specifies the availability of a component, or a part, of a system, as

> *the probability, which tries to express the time that the system is actually available within the total operation time.*

For the availability, in this graduating assignment, only the technical components are inspected, including the errors, because this particular problem is part of the research domain in electrical engineering.
In practice, the system has to be operating continuously, which means there has to be as little as possible downtime. This is necessary to be reassured of a good, practical measurement.

2.2.1.1 A single Transport conveyor

The availability of a single Transport conveyor will be the fraction between the downtime and the operation time. In formula:

$$\eta = \frac{T_e - T_a}{T_e} = 1 - \frac{T_a}{T_e} \qquad (2.1)$$

$$\eta = \frac{MTBF}{MTBF + MTTR} \qquad (2.2)$$

Formula 2.1 is very useful for the practical measurement of the availability of a single Transport conveyor. The meaning of the parameters is as follows:

T_e Operation Time:
the time when the system should be functioning. This is not dependent of the system having to do a job or not

T_a Downtime:
the time, within the operation time, in which the system is not functioning correctly.

Formula 2.2 is the mathematical version of formula 2.1, and this is useful for the theoretical measurement of the availability of a single Transport conveyor.

The parameters have the following meanings:

MTBF Mean Time Between Failure:
the expected average time between failure events, which causes the system to go down.

MTTR Mean Time To Repair:
the expected average time to restore a system after failure events, or, as a routine, for example scheduled maintenance action. This has to do with whether or not the system is down during restoration.

The general formulas of a single Transport conveyor are known and the next step is to use the formula and check if the result of the formula is equal to what is expected.

Example 1: An example of an availability calculation, using VDI on a single Transport conveyor.

This example will focus on one particular Transport conveyor, and it can not be used on multiple conveyors at the same time.

A system with an operation time T_e of 2 days (=2880 minutes) and the next two errors are given:

1. Photocell failure = 7 min.
2. Motor failure = 14 min

In the VDI norm the downtime will now be: 7 + 14 = 21 minutes, and so the availability will be:

$$\text{Availability} = \frac{2880 - 21}{2880} = 0.99271 \quad (= 99.271\%)$$

Taking a closer look at this, this is only correct when the failures do not occur at the same time. If more than one failure happens at one time, there have some overlap between the two failures. When taking the sum of all downtimes, the calculated down time is to long. In practice, there are two possible extremes; the *minimal-downtime-extreme*, which occurs when the two failures totally overlap. The other extreme occurs when the two failures do not overlap at all, the *maximal-downtime-extreme*.

For a single Transport conveyor the calculation is good, because it is here impossible that there are two or more failures at the same time. When a failure occur the Transport conveyor stops and gives an error signal. The next step is to do calculations for a system, serial and parallel.

2.2.1.2 A serial system

VDI deals with the calculation of the availability of a serial system by simply multiplying the separate availabilities of the Transport conveyors. In formula:

$$\eta_{ser} = \prod_{i=0}^{s} \eta_i \qquad \text{with s = Number Of Devices} \quad (2.3)$$

Availability η could be interpreted as a kind of probability, but when formula 2.3 is inspected the conclusion can be made that the availabilities are independent. This means that the failures of the Transport conveyor are independent. This is a assumption, which not always could be supported in practice. There are failures, which are dependant of other failures, but this is not always true.

But when this assumption is supported there still exists a gap, which could be simply be proven by an example. First the VDI calculations are done.

Example 2: An example of an availability calculation, using VDI on serial Transport conveyors.

The following data are given:

1. Photocell failure of Transport conveyor I:
   - Duration : 7;
   - operation time T_e of 2 days (=2880 minutes)
   - Availability = (2880 – 7) / 2880 = 99,757%
2. Motor Failure of Transport Conveyor II:
   - Duration : 14;
   - operation time T_e of 2 days (=2880 minutes)
   - Availability = (2880 – 14) / 2880 = 99,514%

VDI now calculates the availability of the serial system containing Transport conveyor I and II as follows:

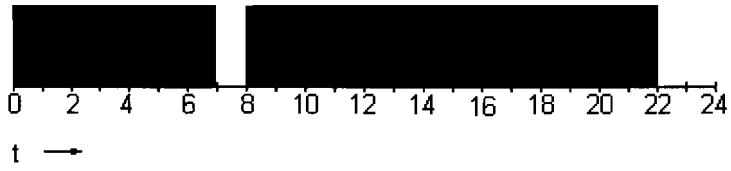
Availability System = 0,99757 x 0,99514 = 0,99272 (= 99,272%)

When the practical side of the problem is inspected, one found that the availability as calculated using VDI is not accurate.

Because this system contains more than one conveyor it is possible that both Transport conveyors have failures. These failures in a system could have overlap. It could easily be so that a failure in conveyor I and a failure in conveyor II both occur, but have no overlap-time at all.

When the errors have overlap this has to be calculated so this can be subtracted from the sum of all downtimes of the Transport conveyors. This will be clarified by an example.

Example 3a: An example of an availability calculation of serial Transport conveyors (without overlap).



A system with an operation time T_e of 2 days (=2880 minutes) and the next two errors:

1. Photocell failure of Transport conveyor I
   - start-time: t=0;
   - end-time: t=7;
   - duration-time: 7.
2. Motor Failure of Transport conveyor II
   - start-time: t=8;
   - end-time: t=22;
   - duration-time: 14.

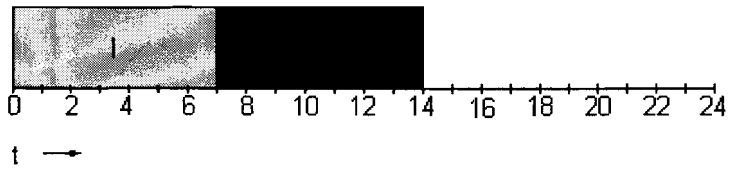
When these failures do not occur at the same time, which is the case here, the downtime will be 7+14=21 minutes.

So the availability is:

$$\text{Availability} = \frac{2880 - 21}{2880} = 0.99271 \quad (= 99.271\%)$$

Example 3b: An example of an availability calculation of serial Transport conveyors (with overlap).

When the two failures do have overlap-time, the procedure is as follows:



The system still has an operation time T_e of 2 days (=2880 minutes), but we now hypothesize the next two errors:

1. Photocell failure of Transport conveyor I
   - start-time: t=0;
   - end-time: t=7;
   - duration-time: 7.
2. Motor failure of Transport conveyor II
   - start-time: t=0;
   - end-time: t=14;
   - duration-time: 14.

The overlap-time is here from t=1 till t=7, so it occupies 7 minutes.

Now the availability is calculated as follows:

$$\text{Availability} = \frac{2880 - ((7+14) - 7)}{2880} = 0.9951388 \quad (= 99.514\%)$$

So when the failures mentioned above occur, the availability lies in the range of 99,271% and 99,514%. These values are the *minimal-downtime-extreme* (99,514%) and the *maximal-downtime-extreme* (99,271%).

VDI calculates 99,272% and is almost equal to the maximal-downtime-extreme and will not necessary be true. The maximal difference between the VDI-calculation (99,272%) and the minimal-downtime-extreme (99,514%) is than 0,242%. This value of 0,242% is the biggest possible difference between the value calculated in theory and the value found in practice. This difference is not acceptable for CSi, they want the calculation to resemble more the value found in practice. They thus want to reduce the difference between the calculated value and the practical value.

2.2.1.3 A parallel system

The same can be done for a parallel system. VDI deals with the calculation of the availability of a parallel system by taking 1 minus the multiplication of the unavailabilities of the two Transport conveyors. In formula:

$$\eta_{par} = 1 - \prod_{i=0}^{s} (1 - \eta_i) \qquad \text{with s = Number of devices} \quad (2.4)$$

The solution of VDI will be clarified by an example.

Example 4: An example of an availability calculation, using VDI on parallel Transport conveyors.

The following data are given:

1. Photocell failure of Transport conveyor I:
   - Duration : 7;
   - Operation time T_e of 2 days (=2880 minutes)
   - Availability = (2880 – 7) / 2880 = 99,757%
2. Motor Failure of Transport Conveyor II:
   - Duration : 14;
   - Operation time T_e of 2 days (=2880 minutes)
   - Availability = (2880 – 14) / 2880 = 99,514%

VDI now calculates the availability of the parallel system containing Transport conveyor I and II as follows:

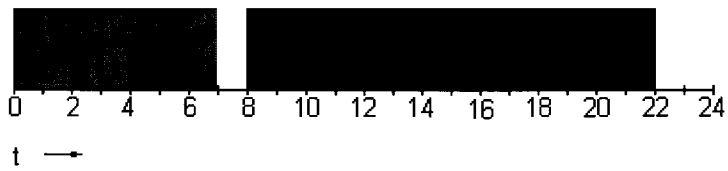
Availability System = 1-((1-0,99757) x(1- 0,99514)) = 0,9999 (= 99,99%)

Inspecting the practical side of the problem, the availability as calculated using VDI is not accurate. To make calculations more congruent with the practical findings, the overlap of the failures will be analyzed.

The calculation of the availability of a system with 2 parallel conveyors will be clarified in the following example.

Because this system contains more than one conveyor, there exist a possibility that there occur two failures at the same time. This does not mean that the failures in a system have to overlap. It could easily be so that a failure in conveyor I and a failure in conveyor II both occur, but have no overlap-time at all In case of a parallel system, the system is only down when all the Transport conveyors are down. When this is not true the system can do its function and will not be down. So it is important that the overlap-time when all Transport conveyors are down can be calculated. This will be clarified with an example.

Example 5a: An example of an availability calculation, using option two on parallel Transport conveyors, without overlap.



A system with an operation time T_e of 2 days (=2880 minutes) and the next two errors:

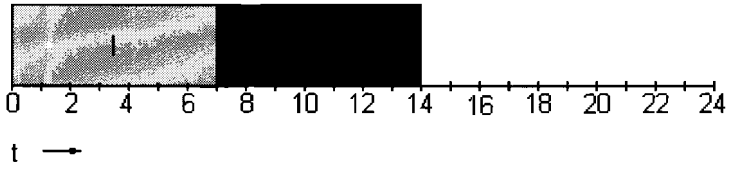
1. Photocell Failure of Transport conveyor I :
   - start-time: t=0;
   - end-time: t=7;
   - duration: 7;
2. Motor Failure of Transports conveyor II :
   - start-time: t=8;
   - end-time: t=22
   - duration: 14;

When these failures do not occur at the same time, which is the case here, there is no downtime because there is no overlap. The parallel system is never completely down and thus is the availability:

$$\text{Availability} = \frac{2880 - 0}{2880} = 1{,}0000 \quad (= 100{,}00\%)$$

Example 5b: **An example of an availability calculation, using option two on parallel Transport conveyors, with overlap.**

When the two failures do have overlap-time, the procedure is as follows:



The system still has an operation time T_e of 2 days (=2880 minutes), but the next two errors will be hypothesized:

1. Photocell failure of Transport conveyor I
   - start-time: t=0;
   - end-time: t=7;
   - duration-time: 7;
2. Motor failure of Transport conveyor II
   - start-time: t=0;
   - end-time: t=14;
   - duration-time: 14.

The overlap-time is here from t=0 till t=7, so it occupies 7 minutes.

Now the availability is calculated as follows:

$$\text{Availability} = \frac{2880 - 7}{2880} = 0{,}99757 \quad (= 99{,}757\%)$$

So when the failures mentioned above occur, the availability lies in the range of 99,757% and 100,00%. These values are the minimal-downtime-extreme (100,00%) and the maximal-downtime-extreme (99,757%).
Example 4 shows that VDI calculate 99,99% and is almost equal to the minimal-downtime-extreme which will not necessary be true. The maximal difference between the VDI-calculation (99,99%) and the maximal-downtime-extreme (99,757%) is 0,233% and this is not acceptable for CSi, who want this difference to be as small as possible.

2.2.2 Conclusion on research of VDI 3581

Looking at the examples, the gap between the expected value and the calculated value can still be reproduced. Even when it is supposed that the failures are independent. This has no influence on the result.
So one can conclude that the present version of VDI 3581 not suffices for what CSi expects, and thus is not acceptable for them. A possible cause for this gap could be that the Availability is strongly time dependant and this could not be said of probability theory. Therefore, another way has to be found to calculate the availability more accurately.
To do this more information about the errors than only the down time is needed. Especially the starting- and ending-time of an error is important, because then the overlap-time can be calculated. Using this, the availability calculation would become more accurate.

A note here is that when a single Transport conveyor is inspected, the calculations of VDI 3581 are fairly accurate, but when more than one conveyor is taken into account, a gap exists.

The conclusion that can be made from this paragraph is that the outcomes of the formulas of VDI 3581 are not accurate enough for CSi, especially when considering multiple Transport conveyors.

2.3 Definitions and interpretations

The project has started to analyze the Availability and Reliability. Much information is no clear about the difference between these terms, which is necessary for this project. Therefore a clear difference will be created before going on with the project.

2.3.1 Definitions

First, have to be clear what is expected if there is talking about the terms Availability and Reliability. The terms with a direct relation to the terms Availability and Reliability will also be mentioned.

2.3.1.1 Availability

In order to calculate availability it is necessary to consider each aspect in turn. This means that capacity must be ignored when considering availability, that is to say it is important to determine whether the system is capable of performing specified function, not whether it can achieve specified throughput. Likewise, it is assumed that all capacity data applies to systems where availability is 100%.

This may appear to be an over-simplification, but as a general rule capacity decreases considerably when the availability of a system falls. It is usually better to have simple, unambiguous rules and take their disadvantages into account rather than to have complex rules, which can lead to misunderstandings.

Availability can be defined in terms of a generic system.

> *The fraction of time that a (sub)system is actually operating (if it is an active (sub)system) or operable (if it is a standby (sub)system)*

First thing to notice in this definition is that availability is a fraction. This is the fraction between the time that the system is actually operating and the time that the system should be operating.
(To simplify calculations a relative availability of 1 is used instead of 100%.)

2.3.1.2 Operation Time

Operation time is the time that the system should be working. Operation time is divided into available time (the time that a system is actually operating) and downtime.

2.3.1.3 Down Time

Downtime covers all periods of interruption due to all *stoppages within operation time*, irrespective of whether they were caused by the material, by operating failures or by the supplier. It is difficult to make a distinction between the downtime caused by the operator, that caused by the material and that caused by the supplier.

The downtime is always attributable to a component in the system and comprises:

- waiting time for specialist or service personnel
- troubleshooting time
- waiting time for spares
- repair time
- time for testing and restarting

2.3.1.4 Reliability

Reliability can be defined in terms of a generic system.

> *The reliability of a system is the probability that the system will perform a specified function under specified operational and environmental condition at and throughout a specified time.*

The first thing to notice in this definition is that reliability is a probability, so we are dealing with the laws of random chance as they appear in nature. Indeed, occurrences of inopportune interruption in functionality in a system are random events, the expected frequencies of which we aim to reduce.
The next thing to notice is that the definition depends on a specified function, operating conditions, environment, and time. So before we can deal with reliability, the producer and the customer must reach formal agreement on what the product is to do, how the user is to use the product, the range of environments under which the product is expected to perform satisfactorily, and the instant or duration in time that the performance of the product is demanded.

2.3.1.5 Failure rate

Reliability for any system depends upon successful events or successful operations within a given time. So reliability may be regarded as the absence of failures.

The definition of the failure rate (λ):

> *The rate of occurrences of failures within a time of measuring.*

2.3.1.6 Repair rate

Reliability for any system depends upon how fast a component can be repaired. If this is soon than it has less influence on the reliability than when it takes a long time to repair the component.
So to get the best reliability it is useful to get a high as possible repair rate. That means that the repair time is as short as possible.

The definition of the repair rate (μ):

> *The rate of components which can be repaired within the time of measuring and depends on the mean duration of the repair of a component.*

Knowing the definitions of the terms Availability and Reliability, the difference between these terms has to be clear. This is very important so this will be clarified emphatically.
Availability, the time that the system was really operating or ready to operate, thus containing no errors within this time. Reliability, the fraction between number of times the system has done its function well and the total number of times that the system has to do its function, so this is functional and **not** time-related.

2.3.2 Theoretical availability analysis

In part 2.2 is proven that knowing the availability of the different conveyor each, the availability of the system (serial or parallel) can not be calculated accurately. Therefore, more information about the errors is needed to calculate the availability of the system more accurately.
One option to do this is to create a log-file in which more details about the errors are saved. If the starting- and the ending-time of an error are saved, a better analysis can be made, because there can be find out whether there is any overlap between the errors, which could have a big influence on the result of the availability-calculation for systems. The overlap-time is defined as the time that two or more errors occur at the same time.

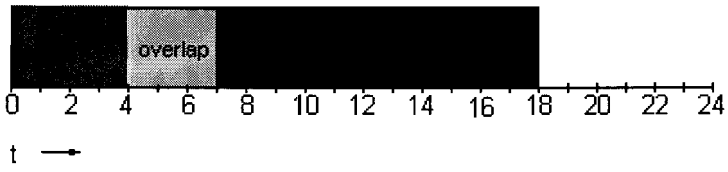
How the availability for systems could be calculated will now be explained, for a system, a serial as well as a parallel system.

2.3.2.1 A serial system of transport conveyors

Taking two or more Transport conveyors and place them serial, it is possible that an error occurs at position x and at the same time there occurs an error at position y. So now two or more errors occur at the same time in the system. In a serial system, it does not matter whether all the Transport conveyors in a system go down or just one Transport conveyor goes down, because as soon as one conveyor fails, the entire system goes down.
If the sum of the downtimes of each Transport conveyor in the system is taken the downtime of the complete system is overestimated. To calculate a more realistic downtime the overlap-time of these errors have to be known. This overlap-time can be subtracted from the sum of downtimes of the entire system and than the result is more realistic downtime-approximation.
To calculate the overlap-time between two errors, specific information about the errors is needed, especially the starting-time and ending-time.

Example 6: An example of an availability calculation of a serial Transport conveyors (with overlap).

When the two failures do have overlap-time, the procedure is as follows:



The system still has an operation time T_e of 2 days (=2880 minutes), but we now hypothesize the next two errors:

1. Photocell failure of Transport conveyor I
   - start-time: t=0;
   - end-time: t=7;
   - duration-time: 7;
2. Motor failure of Transport conveyor II
   - start-time: t=4;
   - end-time: t=18;
   - duration-time: 14;

The overlap-time is here from t=4 till t=7, so it occupies 3 minutes.

Now the availability is calculated as follows:

$$\text{Availability} = \frac{2880 - ((7+14)-3)}{2880} = 0.993750 \quad (=99.375\%)$$

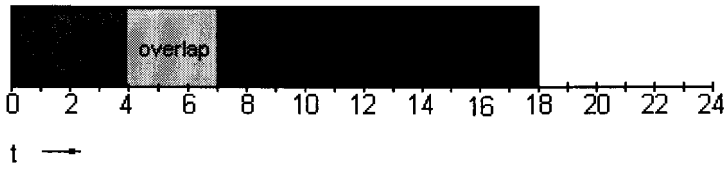
2.3.2.2 A parallel system of Transport Conveyors

Just like in the serial systems, it is possible that in a parallel system an error occurs at position x and at the same time there occurs an error at position y. So there will be a possibility that there exists an overlap between these errors. The difference with serial systems is that, in the case of parallel systems, when there occurs an error at one conveyor the other conveyor can still be operating. So the pallet could be routed about the other conveyors and take a redundant route. The system can be seen as a black box and it is not important **how** the pallet will be routed through the system, but **that** it will be routed through the system.

So only information about the situation in which all the all the possible routes are impossible is interesting and thus when all Transport conveyors are down at the same time. This is the overlap-time of the errors.
To calculate the real downtime the overlap-time is interesting and therefore the starting- and ending-time of the errors have to be known. With an example this will be clarified:

Example 7: An example of an availability calculation of a parallel Transport conveyors (with overlap)..

When the two failures do have overlap-time, the procedure is as follows:



The system still has an operation time T_e of 2 days (=2880 minutes) and the two next errors will be hypothesized:

1. Photocell failure of Transport conveyor I
   - start-time: t=0;
   - end-time: t=7;
   - duration-time: 7;
2. Motor failure of Transport conveyor II
   - start-time: t=4;
   - end-time: t=18;
   - duration-time: 14;

The overlap-time is here from t=4 till t=7, so it occupies 3 minutes.

Now the availability is calculated as follows:

$$\text{Availability} = \frac{2880 - 3}{2880} = 0{,}99896 \quad (= 99{,}896\%)$$

2.3.3 Theoretical reliability analysis

Earlier the definition for Reliability is given and a note here is, that the difference between Availability and Reliability was that the Availability is more time-related and that the Reliability is more count-related.

So for the Reliability it is desirable to know when the Transport conveyor has to do a specified function. In practice, it is not so easy to detect automatically when the Transport conveyor has to do its function. Therefore there is searched for an alternative method, with which the Reliability can be calculated with the logged errors.

In the next part a number of analysis-techniques, which have been found, are considered and one of these techniques in particular; the one, which is chosen to try to calculate the reliability of systems.

2.3.3.1 Available analysis techniques

Some of these techniques can be characterized as being of a more qualitative nature; others can be characterized as being of a quantitative nature.

2.3.3.1.1 Introduction

At this moment a number of reliability analysis techniques are used. All techniques have their pros and cons, use different assumptions, and are specialized on different aspects of system behavior. Another difference resides in the fact that completely different measurement scales are used, some merely giving the elements a ranking score, others making exact probability statements based on a careful calculation, with all sorts of gradations in between.

2.3.3.1.2 Available analysis techniques

A number of analysis techniques can be used, as shown in figure 2-1. The used analysis techniques can be grouped into quantitative techniques (techniques that can be used in quantifying effects of system behavior by calculating probabilities of system failure) and qualitative techniques.

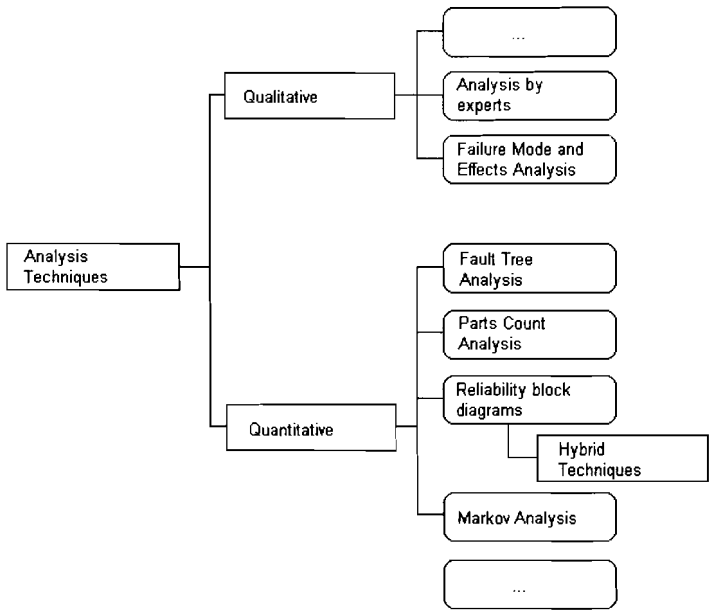


Fig 2-1 A number of available techniques

2.3.3.1.3 Comparison of the available reliability-analysis techniques

One way of looking at the analysis techniques is to interpret them as a kind of analysis process, in which input is converted into output by means of actions performed during the analysis process. These actions are performed according to a certain analysis methodology.
If this approach is used, the analysis techniques can be compared on necessary input information and the resulting analysis output.

A completely different approach for comparing the analysis techniques is by comparing aspects covered by each analysis technique.

The following tables came from the dissertation written through Dr. Ir. J.L. Rouvroye and they were used to make a choice between the different analysis techniques.

Analysis Technique / Information needed for analysis process			Qualitative		Quantitative				
			Experts analysis	FMEA	Parts count Analysis	Reliability block diagrams	Hybrid techniques	Fault tree analysis	Markov analysis
System failure to be analyzed						√	√	√	
Structure of system	Components		√	√	√				
	Functional	Partly	√			√	√	√	
		Complete	√						√
Failures modes components/ functions			√	√		√	√	√	√
Component Failure Data	Point Values			√	√	√	√	√	√
	Uncertain Data								
Testing procedures							√	√	√
Repair procedures							√	√	√

Table 2-1 : Comparison of information needed by analysis techniques

Analysis Technique / Aspects covered during analysis process	Qualitative		Quantitative				
	Experts analysis	FMEA	Parts count Analysis	Reliability block diagrams	Hybrid techniques	Fault tree analysis	Markov analysis
Effects of redundancy	✓			✓	✓	✓	✓
Common cause Failures	✓			✓	✓	✓	✓
Systematic failures	✓	✓		✓		✓	✓
Effects of diagnostics	✓	✓			✓	✓	✓
Effects of on-line test and repair					✓	✓	✓
Effects of off-line test and repair					✓	✓	✓
Time / sequence dependent aspects							✓
Effects of uncertain data							

Table 2-2 : Comparison aspects covered by different analysis techniques

With these tables a decision is made about which technique to use. CSi wanted to calculate the Reliability of a Transport conveyor during a period of time. Therefore, time was an important issue and you see in table 2-2 that this issue was only covered by a Markov analysis.

Prof. Brombacher also advised this technique, and so we independently came to the same conclusion.

2.3.3.1.4 Markov analysis

A Markov model describes a system using a set of mutually exclusive states and transitions between these states. The system can be in only one state at a time, and from time to time a transition is made from one state to another. One important property of Markov processes is that the transitions out of a certain state are independent of the way the system got into that state (no memory). This property requires to the application of constant failure rates.

For systems, examples of these states are:

- 'OK' (every component working perfectly),
- 'intermediate' or 'degraded' (one or more components have failed but the system is operational due to redundancy), or
- 'failed' (system is not operational).

Transitions between different states represent failure or repair of a component. For example; if there is a transition from an 'ok' state to a 'failure' state, the transition represents a failure. If the opposite occurs, from a 'failure' state to an 'ok' state, than the transition represents the repair of a component.
The Markov model is often presented by a number of circles, representing the states, and arcs between the circles, representing the transitions between the states. The arcs are often labeled by the failure and repair rates for the associated transition, although the mathematical meaning of an arc is a transition probability. The mathematical description of Markov models is a set of coupled differential equations for the probability for the system to be in each state. In a number of situations this set of equations can be solved analytically. If that is not possible, numerical approximation can be used.

Advantages of Markov analysis:

- Very detailed
- Complete system description in one model
- Can model different repair scenarios
- Can model sequence dependencies

Disadvantages of Markov analysis:

- Analysis is complex
- Models can become very large
- Generally, to describe a system, an entirely new model has to be created.

- Because constant failure rates are used in the model, factors like wear and fatigue of the system, which are not constant factors, are not taken into account, and they require another way of modeling.

With a small example the working of Markov model will be explained and later in this paper there will be made a Markov model of the Transport conveyor.

The easiest way to create a Markov model is to begin with the state in which the system is working perfectly (state 1, OK state). When a failure occurs (probability λ) the system will go into a failure state. The system stays in this state until the failure has been solved (probability μ), then it goes back to the Ok state. This is visualized in figure 2-2.

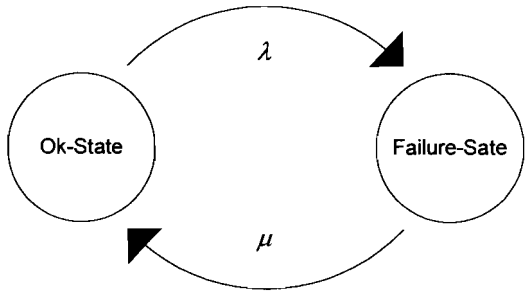


Figure 2-2 : Example of a Markov - model

So the information in general is interpreted and is it time to specify the just given general information to the specific area, namely Material Handling Systems.

Chapter 3

Material Handling Systems

Material Handling Systems are systems that take care of the internal logistic process within a company. This kind of system does not have a rigid configuration; it is company dependant. But to give a general impression we will give an example of what such a system might look like.

3.1 The physical architecture

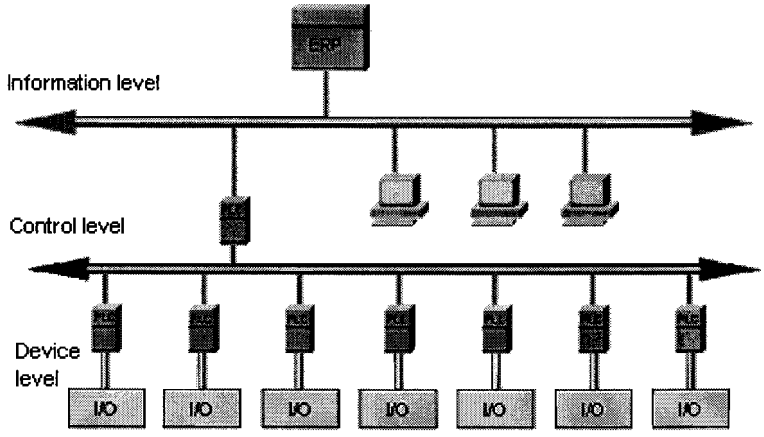


Figure 3-1 Architecture of Material Handling Systems

In chapter 1 in the paragraph on 'the graduating assignment' it was mentioned that this graduating assignment consists of two parts, namely a part on investigating and calculating the Availability and Reliability of the Material Handling System and a part on the communication of Material Handling systems.

Both goals will be accomplished, but in separate ways.

3.2 The Transport conveyor

Because the great complexity of the system not the whole system can be taken in consideration. Therefore a part of the system, the Transport conveyor, will be taken to investigate the Availability and Reliability.
More information about the Transport conveyor and the calculations can be found in Chapter 4.

3.3 Communication of Material Handling Systems

Analyzing the architecture of a Material Handling System three levels of communication become apparent. Because the architecture is variable it is a most logical to describe these levels of communication functionally. This will been done in chapter 5.
In this chapter several communication protocols will be mentioned and analyzed on several criteria.

Chapter 4

Transport Conveyor

CSi supplies intelligent material handling systems. The market in which they operate demands precision and conscientiousness. Therefore their product range consists of:

- Bin handling equipment
- Pack handling equipment
- Tote handling equipment
- Robots
- Palletisers
- Pallet handling equipment

For this graduating assignment, the pallet handling equipment has been studied. The correct handling of pallets, whether loaded or not loaded, is an important part of the logistical process.
An example is, that conveyors can carry loaded pallets automatically from the palletisers to the truck loading system. The control system can also closely monitor the pallets and ensure that there is up-to-date information at any given time.
Because there are so many factors that influence how the complete Pallet Handling System is functioning, the choice has been made to take a smaller part of the system and investigate the factors that influence this part. The part, which is chosen for this graduating assignment, is the transport conveyor, but before its factors can be described, an introduction to the transport conveyor will be given.

4.1 Description of the Transport conveyor

To transport a pallet, a transport conveyor will be used. There are two types of transport conveyors, namely the 'Roller Transport conveyor' and the 'Chain Transport conveyor'. A visualization of both types is given in figure 4-1.

The difference between these transport conveyors lies in the way of transporting a pallet. The 'Roller Conveyor' transports a pallet through rollers, which will be directed by the motor and a 'Chain Conveyor', transports a pallet through chains, which are continuously rotating and thus transporting the pallet.

From now on if the transport conveyor is mentioned, we are referring to the 'Roller conveyor '.

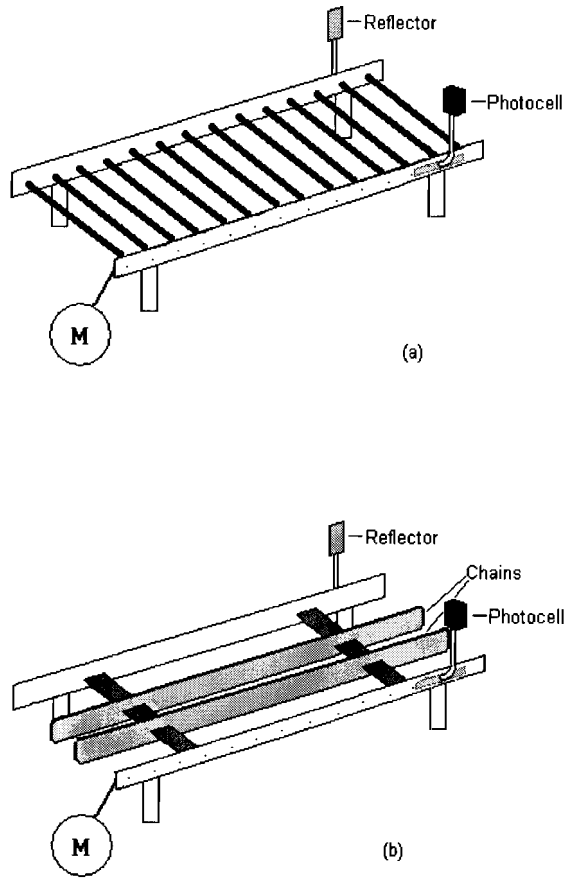


Figure 4-1 (a) 'Roller Conveyor' , (b) 'Chain Conveyor'

Taking a closer look, the transport conveyor consists of five major parts, namely:

- Frame
- Rollers
- Chains
- Motor
- Photocell(s)

The next step to be taken is to investigate the factors that influence the working of the transport conveyor. This means, that the transport conveyor can function when all parts are working correctly, and cannot do its function when one or more parts are not working.

4.2 Factors of influence

The factors that influence the functioning of the transport conveyor can be divided into three sections, namely:

- Maintenance
- System Factors
- Non-System Factors

This division is made, because all the factors have their influence on the functioning of the transport conveyor, but maintenance of the transport conveyor is preventive. This does not mean that there has to be a failure in the system, it is merely a check-up. The other factors are not preventive, but corrective. This means that the system has to be fixed first, otherwise the transport conveyor cannot function correctly.

These corrective factors can be divided into system and non-system factors. By non-system factors you have to think about things like the operator taking the wrong action or bad material on the transport conveyor. This division between system factors and non-system factors is important, because non-system factors, which are difficult to manage, are mostly not the responsibility of CSi.

4.2.1 Non System Factors

To get a correct view, there is tried to inspect all possible factors, which means that the non-system factors also will be inspected. The following division of these factors is made:

- Operators Action
- Temperature
- Relative Humidity
- Infrastructure

An operator can have a big influence on the functioning of the transport conveyor, because he has to react correctly when the system is going down. Dependent of his or her education and/or experience, he or she takes action and this is not necessarily the right one.

The temperature in the environment can also have influence. For example when the temperature is higher, the oil in the transport conveyor is thinner so the working of the oil is not as good as it could be. To guarantee the correct working of the oil, the temperature can be the best between a certain range, so the influence on the functioning of the system is minimal. This also goes for the relative humidity.

4.2.1.1 Operators Action

The action of the operator depends not only on his education or experience, but also on other factors. Now a list will be given of the different factors that have there influence on the action of the operator.

- Experience
- Education
- Mood
- Training
- Messages of the Transport Conveyor

On training and system message CSi has some influence, and when these are optimized, the probability that the operator takes a wrong action will be as minimal as possible.

4.2.1.2 Infrastructure

The infrastructure can be divided into two sections, namely:

- Bus
- Power of the system

The bus takes care of the communication between the transport conveyor and the PLC (Programmable Logic Controllers), but there are more conveyors communicating with one PLC. So this item can be better measured when more transport conveyors in a system are analyzed.
The customer has to take care of the power of the system and this is therefore the responsibility of the customer, not of CSi.

4.2.2 System Factors

Now the possibility remains that the transport conveyor goes down because of a component of the transport conveyor that is not working correctly; a system factor. These components can be divided into the following fractions.

- Electrical Components
- Mechanical Components
- Software

The priority of this graduating assignment lies at the electrical components of the transport conveyor, because this is the area of the Electrical Engineering study. But to get a complete view, the other areas will also be examined.

4.2.2.1 Mechanical Components

Starting with examining the mechanical components the following list is generated:

- Frame
- Rollers
- Axles
- Chains

There are different ways to influence these factors. The state of rollers, axles and chains for example can be influenced through the temperature, because when the temperature is not right the oil cannot do his work well and the strain on the system becomes higher. Thus, the probability that the transport conveyor will go down becomes higher. But not only the temperature has its influence, also the relative humidity and the qualities of the material have their influence. This can also be an extra strain on the system.

4.2.2.2 Software

The software will be written specifically for a customer, but the system will be tested before operating and in this time the most bugs will be fixed.

4.2.2.3 Electrical Components

Now the electrical components of the transport conveyor will be analyzed. The transport conveyor consists of the following electrical parts:

- Motor
- Photocell
- (Field distributor)

These components cause the transport conveyor not to be able to function correctly, but the field distributor is an optional component. Conveyors without a field distributor do exist and here the motor and photocells are directed with relais.

A complete diagram of all this can be seen in appendix D.
The next step is to find out the reason of these failures and how to detect them. In practice, this will and cannot all be done, because it is not worth the effort to do it, but it is always useful to examine possibilities and maybe it will be in the further more worthwhile.

4.3 Reasons of failure and detection

The factors with influence on the functioning of the transport conveyor are known. Now the question is, when these factors will occur and so have there influence on the functioning of the transport conveyor.
Research will be done on the electrical components, because this is in the range of electrical engineering and a check on what the reasons could be for the incorrect functioning, will be executed.

4.3.1 Motor Failure

The motor has to direct the rollers of the transport conveyor, so when this does not happen anymore, nothing will be transported. Therefore the motor is analyzed and the following reasons were found for the not correctly functioning of the motor:

- Motor gets no power
- Motor is broken
- Thermal overload of motor

Now the failures have to be detected, and especially the reasons why a failure occurs. With the present architecture this is not possible and the architecture of the system has to be changed to make this possible.
One of the ideas was, that two things about the Transport conveyor are necessary to be able to say which failure has occurred.
The needed information is:

- Detection by, for example, an encoder, if rollers are moving
- Current used by the Transport conveyor

With this information, the following differentiation can be made:

Failure	Moving of Rollers	Current
Motor gets no power	No	Low
Motor is broken	No	Normal
Thermal overload of motor	No	High

For example; when the rollers are not moving, and there is low current, the conclusion is that the motor gets no power.

So with some changes to the architecture a more detailed reason of the failure could be given. The same will be done for the photocell.

4.3.2 Photocell Failure

The following electrical component is the photocell, which has two functions. It has to send a light signal out (send function), which will be reflected through a reflector. Then the reflected light signal will be received through the photocell (receive function).
Because each function has its own failure reasons each function will be examined separately. First, the send function will be examined.

The photocell has to send a light signal to detect if there is something between the photocell and the reflector. This send function is examined, and the following failure-reasons are found:

- Photocell gets no power
- Pollution of the photocell
- Laserdiode failures

When there is nothing between the photocell and the reflector the sent light signal will be reflected. If the photocell receives the reflected signal it knows that there is nothing on the Transport conveyor. This is the receive function and has its own failure-reasons:

- Photocell gets no power
- Photocell out of position
- Photocell pollution
- Material reflection*
- Operator's hand in front of photocell*
- Reflector pollution
- Material on transport-conveyor*
- Receiver failures

Taking a closer look at these reasons it is clear that not all reasons (marked by a *) are situated on the Transport conveyor itself. For example, because of the material on the transport conveyor or because of the operator, failures can occur. So the photocell sends the wrong value, while the photocell itself is functioning correctly. It is difficult to differentiate whether the failure came from the photocell itself or it has another reason. Therefore we have to accept that this cannot be detected.

The other failures are inflicted by the photocell itself. The following failures are also hard to detect:

- Laserdiode failures
- Photocell out of position
- Receiver failures

For the other failures, there are there some solutions, but the architecture has to change before you can apply these solutions.
One of these solutions is to use photocells which can detect the pollution. They give a signal when the photocells are polluted. Using these photocells, a better differentiate of the failures can be made.
Some other changes are necessary to detect whether the photocell is getting power.

It is possible that the Transport conveyor contains another electrical part, namely the field distributor, which will be discussed below.

4.3.3 Field Distributor Failure

The field distributor functions as an interface between the Transport conveyor and a PC. The communication between the Field distributor and the PC happens via PROFIBUS and using this you can communicate with the Transport conveyor, for example to read out the status of a photocell. When it gives the signal that it is 'busy' the field distributor has to take care that this signal comes on the PROFIBUS so it can be read out.

During the research the following failures of the field distributor are found:

- Field distributor gets no power
- Field distributor broken
- No communication with the environment

Taking a closer look at the failures of the field distributor you can see that they are hard to detect and up to this point no solution has been found to fix this.

The electrical parts of the Transport conveyor are inspected now, because this is in the range of electrical engineering. To get a more accurate view this has to be done also for the mechanical parts, but this is another discipline and so has to be done in another study.
The next step will be to clarify the difference between Availability and Reliability, this is what the terms really mean and what it means when one is talking about these terms.

4.4 Markov-model of a Transport conveyor

The easiest way to create a Markov model is to begin with the state in which the system is working perfectly (state 1, OK state). After that, imagine that the system goes down. The reasons why the system can go down were mentioned previously, but the main reasons are:

- Maintenance
- Motor Failure
- Photocell Failure
- Field Distributor Failure
- Non System Failure

This all will lead to a new set of system states in which each failure creates a new state and so we now have six different states (the Ok state, and a new state for each of the above mentioned failures). New states could be created by a combination of these failures. If we consider this, than we get 33 (2^5 + 1) different states. Thinking on the combinations the expectation was and is that the chance that they occur is very, very small. Therefore, it is not useful to examine all these new states and that is why they are not considered in the Markov model.
The Markov model is used to calculate the reliability of a transport conveyor. So we can divide the system states into three groups.

- The first group of states, actually only holds one state; the state where the transport conveyor does not go down.
- The second group contains the states which result in the transport conveyor going down, but due to reasons having their grounds in normal functioning, like preventive maintenance, or to failing through non system influences. These states still belong to the reliability.
- The third group of states consists of the states which result in the system going down because there has to be done some corrective maintenance. One of the components fails and have to be repaired or replaced. These are so called corrective maintenance states and they do not belong to the reliability.

In all states except the OK state repair of failures can occur if the reason for the failure is known. This repair action leads to a transaction back to the earlier state, in this case that always is the ok-state.

The assumption is made that *each component will fail in only one failure mode. Once the component has failed, no other failures will occur in that component.* :

This leads to the following model:

1. **State 1 :**
   OK state. The system works correctly and does not go down.

   State 2:
   The system goes down because there has to be some preventive maintenance of the motor and the photocell of the transport-conveyor.

   State 3:
   The system goes down because there have to be some corrective maintenance of the motor of the transport-conveyor because of the following reason: 'No Input Voltage'.

   OR

   The system goes down because there has to be some corrective maintenance of the motor of the transport-conveyor because of the following reason: 'Thermal Overload '.

   OR

   The system goes down because there has to be some corrective maintenance of the motor of the transport-conveyor because of the following reason: 'Motor Broken'.

   State 4:
   The system goes down because there has to be some corrective maintenance of the photocell of the transport-conveyor because of the following reason: 'No Input Voltage'.

   OR

   The system goes down because there has to be some corrective maintenance of the photocell of the transport-conveyor because of the following reason: 'Laser diode failure'.

   OR

   The system goes down because there has to be some corrective maintenance of the photocell of the transport-conveyor because of the following reason: 'Receiver failure'.

   OR

The system goes down because there has to be some corrective maintenance of the photocell of the transport-conveyor because of the following reason: 'Pollution of photocell'.

OR

The system goes down because there has to be some corrective maintenance of the photocell of the transport-conveyor because of the following reason: 'Pollution of reflector'.

OR

The system goes down because there has to be some corrective maintenance of the photocell of the transport-conveyor because of the following reason: 'Photocell out of position'.

OR

The system goes down because there has to be some corrective maintenance of the photocell of the transport-conveyor because of the following reason: 'Hand in front of photocell'.

OR

The system goes down because there has to be some corrective maintenance of the photocell of the transport-conveyor because of the following reason: 'Material reflection'.

OR

The system goes down because there has to be some corrective maintenance of the photocell of the transport-conveyor because of the following reason: 'Material on transport conveyor'.

State 5:
The system goes down because there has to be some corrective maintenance of the temperature under which the transport-conveyor is functioning. So the temperature has to be corrected.

OR

The system goes down because there has to be some corrective maintenance of the humidity under which the transport-conveyor is functioning. So the humidity has to be corrected.

OR

The system goes down because there has to be some corrective maintenance of the power under which the transport-conveyor is functioning. So the power has to be corrected.

2. Preventive maintenance of the motor and the photocell of the transport-conveyor. The only possible state to go to now is back to the OK state (state 1).

3. Corrective maintenance of the motor of the transport-conveyor. The only possibility is to return to the OK state (state 1).

4. Corrective maintenance of the photocell of the transport-conveyor. The only possibility is to return to the OK state (state 1).

5. Corrective maintenance of non-system influences such as temperature, humidity and/or power. The only possible state to go to is back to the OK state (state 1).

A visualization of the Markov state model you can find in appendix D.

4.4.1 Set of differential equations of Markov model

The Markov model describes the Transport conveyor by using a set of mutually exclusive states and transitions between these states. With this model a set of differential equations can be generated, which can be used to calculate the Reliability.

First the equations are shown, which will be solved later on, to get a representation of the reliability of the transport conveyor.

Differential equations:

$$\frac{\partial P_1(t)}{\partial t} = -(\lambda_{1,2} + \lambda_{1,3} + \lambda_{1,4} + \lambda_{1,5} + \lambda_{1,6})P_1(t) + \mu_{2,1}P_2(t) + \mu_{3,1}P_3(t) + \mu_{4,1}P_4(t) + \mu_{5,1}P_5(t) + \mu_{6,1}P_6(t)$$

$$\frac{\partial P_2(t)}{\partial t} = -\mu_{2,1}P_2(t) + \lambda_{1,2}P_1(t)$$

$$\frac{\partial P_3(t)}{\partial t} = -\mu_{3,1}P_3(t) + \lambda_{1,3}P_1(t)$$

$$\frac{\partial P_4(t)}{\partial t} = -\mu_{4,1}P_4(t) + \lambda_{1,4}P_1(t)$$

$$\frac{\partial P_5(t)}{\partial t} = -\mu_{5,1}P_5(t) + \lambda_{1,5}P_1(t)$$

$$\frac{\partial P_6(t)}{\partial t} = -\mu_{6,1}P_6(t) + \lambda_{1,6}P_1(t)$$

4.4.2 Techniques for solving differential equations

There are different ways to solve these differential equations. The next options were considered:

- General appearance
- Laplace transforms
- Eigenvalues and eigenvectors
- Matrix multiplication

In appendix F the different ways are clarified. The first method that was tried to solve the differential equations was using Laplace transforming method, because in the study Electrical Engineering this method is a widely used method. But soon it became clear that this was not the way to solve the differential equations, because the formulas you got after the transformation to the S-domain were not easy to handle. This is especially so, when the model becomes quite large (for example, when it contains more than 10 states). This is a situation, which occurs reasonably often in practice

Therefore another method, the matrix multiplication method, was used to solve the equations.
Using this method a transaction matrix have to be created, a matrix with the probabilities from state x to state y ($\lambda_{x,y}$ = failure rate, $\mu_{y,x}$ = repair rate), and a current probability matrix, a matrix with the current probabilities of a specified state (P_x). General expressions for the two matrices will be given next:

$$\text{Transaction matrix} = \begin{pmatrix} \lambda_{1,1} & \lambda_{1,2} & \lambda_{1,3} & \lambda_{1,4} & \lambda_{1,5} & \lambda_{1,6} \\ \mu_{2,1} & \lambda_{2,2} & \lambda_{2,3} & \lambda_{2,4} & \lambda_{2,5} & \lambda_{2,6} \\ \mu_{3,1} & \mu_{3,2} & \lambda_{3,3} & \lambda_{3,4} & \lambda_{3,5} & \lambda_{3,6} \\ \mu_{4,1} & \mu_{4,2} & \mu_{4,3} & \lambda_{4,4} & \lambda_{4,5} & \lambda_{4,6} \\ \mu_{5,1} & \mu_{5,2} & \mu_{5,3} & \mu_{5,4} & \lambda_{5,5} & \lambda_{5,6} \\ \mu_{6,1} & \mu_{6,2} & \mu_{6,3} & \mu_{6,4} & \mu_{6,5} & \lambda_{6,6} \end{pmatrix}$$

$$\text{Current Probability Matrix} = \begin{pmatrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \end{pmatrix}$$

The ok-state is always the starting situation and has in this situation always probability 1 and all the other states get probability 0. Through matrix multiplication the probability that you are in a specified state will be calculated. So now the expressions for the reliability by multiplying the matrices are available.
This method was easier to handle in practice and it gave a result, so this option was used in this graduating assignment. Because a lot of calculations have to be done for calculating the Reliability with this matrix multiplication method, it is easier to handle when this is automated, using for example MATLAB and Visual Basic, as shown next.

4.4.3 Using the mathematical program MATLAB to solve differential equations

To fill the matrices the (following) parameters, failure rate and repair rate, have to be known. These have to be calculated from an error log-file.
With this information, a program in MATLAB is written to solve the equations automatically. MATLAB was chosen in this stage of the graduating assignment because MATLAB can handle matrices easily, making it easier to do the matrix-multiplications. Also, while doing the study Electrical Engineering at the Technical University of Eindhoven, there was some experience with this program, which made using this program an attractive option.
After doing the multiplications, the probability that one is in a defined state is known and this is the reliability of the component.

4.4.4 Using Visual Basic to solve differential equations

MATLAB is difficult to use in real industrial environments and the most commonly used programming language at CSi is Visual Basic (VB), so the decision was made to try to solve the equations using VB.

An attempt was made to do the same calculations as in the MATLAB program, but this time using Visual Basic. This was more difficult because MATLAB knows matrices and can calculate with them. Visual Basic does not know matrices, but using a multidimensional array solves this.

Visual Basic can also communicate better with other programs, so when the errors are logged in a Database (Microsoft Access), than Visual Basic can read them and calculate the failure rate and repair rate automatically.

4.5 Pilot project

After the theoretical examination, a practical part of this graduating assignment was begun. By using a pilot project, the theory, which was mentioned previously, will be tested.

First, a description of the pilot project is given. Furthermore there will be discussed what had to happen before the pilot project was working so we were able to get the right information out of it. Without this information it is not possible to calculate the availability or reliability at all.
If the necessary information is available, one can calculate the availability and reliability with the theory mentioned above. So than some results with practical parameters are available and can be compared to the expectations.

Finally, during this pilot project a new feature was created, which gave more information and creating the possibility to calculate more accurately.

4.5.1 Starting situation

In the starting situation, there is a row of three transport conveyors, one roller conveyor (TC 1) and two chain conveyors (TC 2 and 3), which transport the pallet continuously from left to right (CW, Clock Wise) and from right to left (CCW, Counter Clock Wise).

At each transport conveyor there are two photocells (Ph) for the detection of the position of a pallet and a motor (M), which directs the rollers.

The out-coming signals of the motor and photocells are routed to a BECKHOFF terminal, a PLC in which a program is located, that takes care of transporting the pallet continuously from left to right and back. An example of a BECKHOFF terminal is to see in fig 4-2.

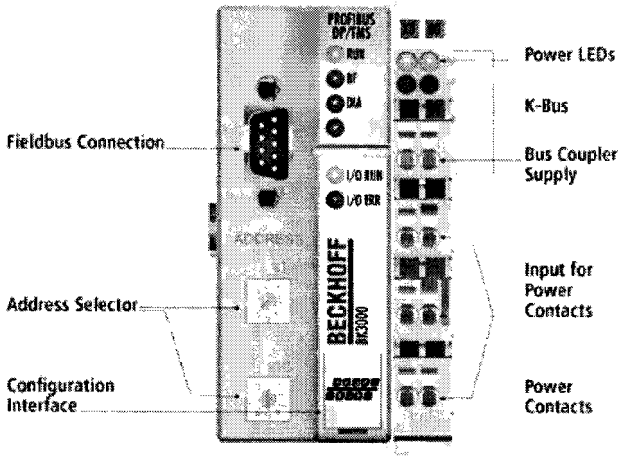


Fig. 4-2 Beckhoff Terminal

The next step was to create the communication between a pc and a BECKHOFF terminal. To take care of this I wrote a Visual Basic program which has the two following functions; 1) routing the BECKHOFF terminal and 2) receiving and logging the messages of the Transport conveyor generated through the BECKHOFF terminal. To clarify the situation a visualization of this is given in figure 4-3.

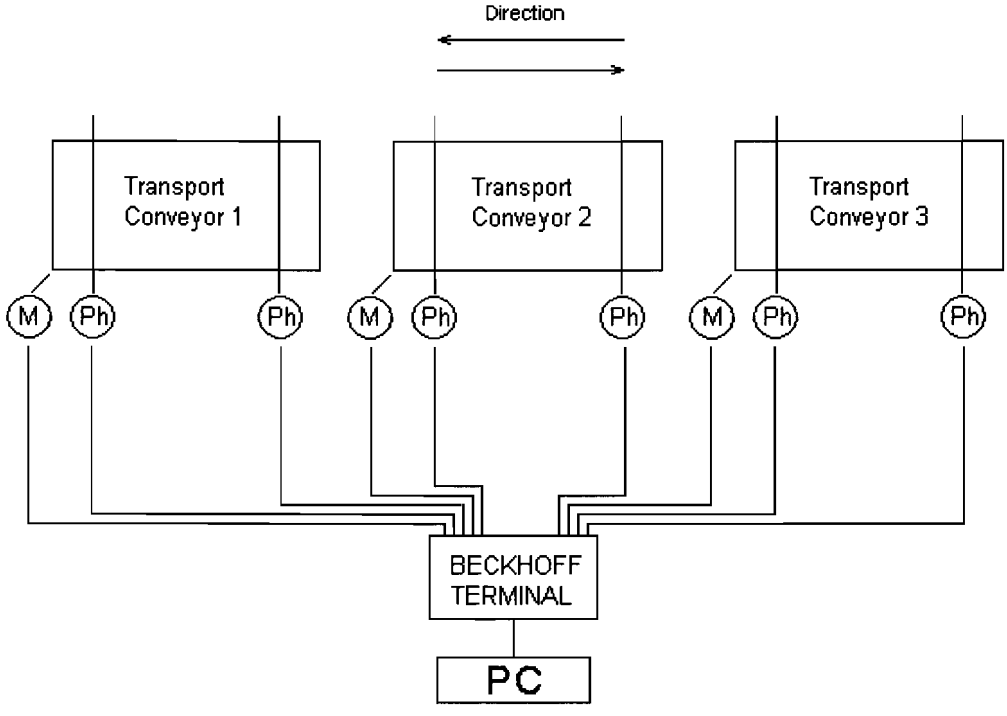


Figure 4-3

First the functions for routing the BECKHOFF terminal will be discussed and this will also be done for the receiving and logging of the messages of the Transport conveyors.

4.5.2 Routing the BECKHOFF terminal

The BECKHOFF terminal holds the program that directed each Transport conveyor. But the following functions have to be done outside the BECKHOFF terminal:

- Starting the system
- Stopping the system
- Resetting the errors of the system
- Giving an emergency stop
- Recovery of a Transport conveyor.
- Switching between Manual Mode/Automatic Mode

The first four functions are understandable, but the recovery function needs some explanation. The recovery function is needed to trace a pallet after an error message has occurred and the program does not know where the pallet is, because the pallet is not in front of a photocell.
Sometimes it is desirable to run the motors of the Transport Conveyors manually to route a pallet to a specific place. This is possible by switching from the Automatic Mode to the Manual Mode.

4.5.3 Receiving and logging the messages

This is an important part of the project. The BECKHOFF terminal notices that there is a change in the system and sends a signal with the number from the byte that has changed. By reading the byte and by comparing this with the last known value of that byte the status change is inspected.
Each bit has a unique meaning and could have two meanings; 1) that the status of the Transport conveyor has changed (for example the system has started or stopped) or 2) an error has generated.
I have created a function for the receiving and logging of the messages in the same program I wrote for the routing of the BECKHOFF terminal, so that all of this can be done using just one program.

To do this right all messages of the BECKHOFF have to be known and one category of messages next will be discussed; the error messages.
With the used architecture, it was not possible to detect all the errors, which are mentioned in theory above. So it can happen that an error message occurs, but the reason why this message occurs, is not clear. For example, the error message ' Photocell illegal occupied' occurs when the photocell is broken AND

when some operator holds his hand in front of the photocell. There is no way of knowing which option is the correct one.

Because the theoretical reasons why a Transport Conveyor does not function anymore are not exactly the same as the practical reasons, the **practical reasons** are given next.

4.5.3.1 Logging of the errors

The logging of all the errors will later be used for the calculations of the availability and reliability. Above there is noticed that practice and theory are not completely equal, because not all the error reasons can be detected. For a more detailed **registration of reasons** why the transport conveyor cannot function anymore, the architecture has to be changed, but with the existing architecture **at CSi,** this is not possible.

The following reasons can be detected with the present architecture, and with these errors will first be calculated:

- Emergency stop SCADA; The emergency stop in the Visual Basic application is activated.
- Emergency stop Field; The emergency stop in the field is activated
- Thermal overload of motor;
- Contactor CW (Clock Wise)
- Contactor CCW (Counter Clock Wise)
- Photocell illegally occupied; The photocell is occupied while it has to be free.
- Photocell occupied too soon
- Timeout in feed; It takes too long before the pallet is detected at the photocell.
- Timeout to end
- Data without pallet
- Pallet without data
- Photocell out feed too long occupied*
- Recovery CW
- Recovery CCW

The availability and reliability have to be calculated afterwards. This means that the system can operate during a period of time, and the calculations can be done after this period of time.
To make this possible, the errors will be logged in a Database, which I have designed, of Microsoft Access 2000 during the time of functioning. But not only the error events have to be logged, also other information like the start- and end-

* The time to transport the pallet is too long, so there is probably an error

time is important. To have a clear overview of all this, different tables are designed to log the errors.

4.5.3.2 Microsoft Access 2000 Database

I created four tables in the Database of Microsoft Access 2000, namely:

- Table with all possible errors (*tblErrors*)
- Table with all error events (*tblErrorEvents*)
- Table with starting- and ending-time of analyzing time (*tblProgramEvents*)
- Table with all transactions (*tblTransactionEvents*)

ErrorID	Description	MB	Active	ComponentID	Availability
1	POS.X,NOODSTOP SCADA	29	No	0	Yes
2	POS.X,NOODSTOP FIELD	29,1	No	0	Yes
3	POS.X,SPARE	29,2	No	0	Yes
4	POS.X,SPARE	29,3	No	0	No
5	POS.X,SPARE	29,4	No	0	No
6	POS.X,SPARE	29,5	No	0	No
7	POS.X,SPARE	29,6	No	0	No
8	POS.X,SPARE	29,7	No	0	No
9	POS.1,THERMAL OVERLOAD	30	No	1	Yes
10	POS.1,CONTACTOR CW	30,1	No	1	Yes
11	POS.1,CONTACTOR CCW	30,2	No	1	Yes

Table 4-1: tblErrors

In the *tblErrors* all the errors are mentioned with an unique ID-number, so it is always clear which error has occurred, and at which position.

The BECKHOFF terminal sends a Marker Byte (MB), when the status of the system changes. The present status can be read out of the MB. These MB's are also found in the table, so the connection between MB and Error is made.
The last status of the error is written in the field 'Active '. This is necessary because when the program that analyzes the system stops, and an error is active, you have to check if it is still active when the program will be started again. **If it is not active,** a new record in the database has to be made and is necessary for correctly analyzing.
The field 'Component ID' is, in this project, connected with the position. There are three Transport conveyors with three different positions. This is saved in 'Component ID'. When an error has influence on the whole system, like 'Noodstop Field', it gets 0 as general 'component ID' .
With the last field 'Availability' the user can choose whether an error will be taken in the calculation for availability and reliability or not. So the influence of this error on the availability and reliability can be analyzed.

Event ID	Transaction Nmb	Date Time	Arrived	Error ID	Error Description	Calculation
583	1	27-6-2002 14:31:25	Yes	9	POS.1,THERMAL OVERLOAD	Yes
584	0	27-6-2002 14:31:30	Yes	163	SCADA RESET BUTTON	Yes
585	1	27-6-2002 14:31:30	No	9	POS.1,THERMAL OVERLOAD	Yes
586	0	27-6-2002 14:31:31	No	163	SCADA RESET BUTTON	Yes
587	2	27-6-2002 14:32:15	Yes	9	POS.1,THERMAL OVERLOAD	Yes
588	0	27-6-2002 14:32:19	Yes	163	SCADA RESET BUTTON	Yes
589	2	27-6-2002 14:32:19	No	9	POS.1,THERMAL OVERLOAD	Yes
590	0	27-6-2002 14:32:20	No	163	SCADA RESET BUTTON	Yes

Table 4-2 tblErrorEvents

In the field 'Transaction Nmb' the transaction number is linked with the event. For reliability the number of fault transactions is important, so when more than one error appears in one transaction, it is still one transaction fault.
Whether an error appears or disappears is noticed in the field 'Arrived' and in the field 'calculation' the user can select whether it is desirable to take the error into the calculation for availability and reliability or not.

ProgramEventID	Description	StartDateTime	EndDateTime
112	Analyzing Time	27-6-2002 14:30:22	27-6-2002 14:35:00
113	Analyzing Time	28-6-2002 8:22:48	28-6-2002 8:22:54
114	Analyzing Time	28-6-2002 12:56:51	

Table 4-3 tblProgramEvents

This table is necessary because it is important to know when the system is analyzed and when not. The availability and reliability can only be calculated only within this time of analyzing. When there is no analyzing the errors will not be logged and nothing can be said about the Availability and Reliability.

During the pilot project, a new feature was designed. With this feature the Transport conveyor was able to send a signal when it has to perform a function, which has to be logged. This gives us some additional information, which will be used together with the Markov model to calculate the Reliability in yet another way.

Transaction EventID	PosNmb	Description	DateTime	Transaction Nmb
27	1	Pos.1,Transaction	27-6-2002 14:31:22	1
28	2	Pos.2,Transaction	27-6-2002 14:31:33	1
29	3	Pos.3,Transaction	27-6-2002 14:31:44	1
30	2	Pos.2,Transaction	27-6-2002 14:32:02	2
31	1	Pos.1,Transaction	27-6-2002 14:32:13	2
32	2	Pos.2,Transaction	27-6-2002 14:32:22	3
33	3	Pos.3,Transaction	27-6-2002 14:32:32	2

Table 4-4 tblTransactionEvents

In this table the number of transactions will be saved so that per position the number of transactions can be analyzed. This is important for expressing the reliability of the transport conveyor.

4.5.4 Calculations of Availability and Reliability.

Now the necessary information is available the real calculations can be done. First, the availability will be discussed and what the results were during the pilot project.

4.5.4.1 Parameters for Availability

For the Availability, the 'real' downtime and the analyzing time are the important parameters, as can be seen in Chapters 2 and 4. By logging the errors, the real downtime could be calculated more accurately and from the *tblProgramEvents* you can see the time of analyzing.
A Visual Basic program was written to calculate these parameters. To do so, first, the time of analyzing is taken from the *tblProgramEvents*.
From the *tblErrors*, the relevant errors of the right position will be filtered, which will be used to get the right error events, so that they can be analyzed for the downtime-parameter of the availability.

Now the downtime is known, the availability can be calculated using the formula:

$$\text{Availability} = \frac{\text{Analyzing Time - Downtime}}{\text{Analyzing Time}}$$

This is the way in which the Availability is calculated during this pilot project. The results of the calculations look very promising; they are more accurate than when the calculations are done using the VDI 3581 norm. This means that a step is made towards a better way of calculating.

The next step in the process towards calculating the Availability more accurately is to find a way to determine the difference between the different reasons because of which an error can happen, whether it was because of the material, the operator, or the system itself. But this is material to be covered in future research.

4.5.4.2 Parameters for the Reliability

There was already noticed that during this pilot project a new feature was designed, namely a signal from the Transport conveyor, which is send when it is doing its function. This was not possible before and the Markov model was introduced to calculate the Reliability in another way. Both ways, the functional counter by the signal and the Markov model, are implemented during this pilot project and later they will be evaluated to determine which is the best way.

4.5.4.2.1 Markov analysis

For the Markov analysis the following important parameters have to be known: the failure rate, how many times the error occurs and the repair rate and the mean duration of these errors.
Because there is already noticed that it is better not to have much different states in the Markov model, errors will be clustered, namely:

- Emergency stop Error
- Motor Error
- Photocell CW Error
- Photocell CCW Error
- Recovery Error

So the Markov model gets six states, namely one for the five sections each and one for the ok state. Because the assumption is made that no more than one error at the same time can occur, there become no other states in the model.

For each cluster the failure rate and the repair rate have to be calculated. To determine the value of the failure rate and repair rate the error log will be used. Therefore, I wrote a Visual Basic program to determine the values automatically.

When these parameters are known they can be used in the matrix multiplication method. So they are placed in the transaction matrix and with the current probability matrix the matrix multiplication method can be started. Now there is a representation for the Reliability.

There have been noticed that the test pilot is still functioning and that the discussed results are only about the part for so far and are not definite.

The results of the Markov model were not exactly as expected. At CSi, a very strict difference exists between reliability and availability, the first being count-related (from the x times, the system worked correctly x – y times) and the latter being time-related (from x minutes/hours, the system worked correctly x – y minutes/hours). The calculated values from the Markov model were more similar to the values of the availability than of the reliability as they were found in the pilot project. A possible explanation for this could be that the values, lambda and mu, are expressed in time-units. More research has to be done to find out the usefulness of Markov Models at CSi.

4.5.4.2.2 Transactions

During this project, I more and more came to realize that at CSi, I needed a count-related set-up to express the reliability. In cooperation with CSi, I tried to design a set-up that could meet these demands. I altered the Transport conveyor in such a way that it gives a signal when it is doing a function.
With this possibility, I designed a table with transactions, in which all the transactions will be logged with the time that the transaction happens. This gave us extra information and another probability to calculate the Reliability.
The Reliability can later be calculated by the fraction of transactions that go well during the time of analyzing. This is a more functional way to calculate the Reliability and this was also tried so both methods can be analyzed now more functional.

For this method the number of transactions has to be calculated. This can be done from the *tblTransactionEvents* and from the *tblErrorEvents*, the appropriate errors are taken into account. When more than one error appears within one transaction there is still just one transaction fault. Now the number of fault transactions can be calculated, so the formula for the reliability will be:

$$\text{Reliability} = \frac{\text{Number of Transactions - Number of Fault Transactions}}{\text{Number of Transactions}}$$

The results of this way of calculating are more congruent with the expected values, but there has to be done more research to investigate whether this is the right method.

4.6 Conclusion

At the beginning of this graduating assignment, the idea was to calculate the availability and reliability of a complete system. This was soon reduced to doing it first for one Transport conveyor, because it was clear that it was impossible to tackle a full system at once. Further, there is also thought about how to calculate

the availability and reliability for two or more conveyors when they are placed in series or parallel. Because there was noticed that the gap between the expected and the calculated values are a lot smaller for one single Transport Conveyor than for more Transport Conveyors placed in a system together. The conclusion is that the gap between the calculated and the expected value has its origin here.

During this graduating assignment some important steps were made. First, the VDI 3581 was analyzed, because in the information I found, this was the most important norm. Remarkable was that the previous version of VDI 3581 covers Availability AND Reliability and the most recent version of VDI 3581 covers Availability only.
In the latest version of VDI 3581 the way of expressing availability was not changed, but the experience of CSi was that the previous version of VDI 3581 did not satisfy the expectations. There was searched for the reason why VDI 3581 could not satisfy the expectations of CSi to express the availability.
The next step was made; logging more information about the errors in a database. The calculations of the Availability became more accurate compared to those in the starting situation, because the overlap-times of the errors could now be calculated, which have a great influence on the calculation of the Availability.

To get a representation of the reliability, the Markov model was used. This was necessary because there was no functional counter to see how many pallets the Transport conveyor has handled and the fraction of the faultily handled pallets could not be calculated. This Markov model uses a failure rate (λ) and repair rate (μ), which were time related, to calculate the reliability.
Previously, it was mentioned that the results of the Markov model were more similar to the values expected for the Availability, instead of the Reliability. Two different reasons were hypothesized for this fact. The first reason could be that the failure rates and the repair rates for the Reliability are time-related. Second, it should be taken into account that these calculations are based on a simulated situation in which failures have to be generated, because otherwise no failures would occur at all. It is hard to say how long one has to measure such a situation for the results of such a measurement to be reliable. These reasons should be seriously taken into account in subsequent studies.

In a later stadium of the project, during the pilot project, the functional counter of the Transport conveyor was introduced. This also was an important step, because now one can get a better notion of the number of pallets the Transport conveyor has handled, and also about the fraction of faultily handled pallets. Now this information is accessible, one is able to acquire a more accurate view of the system.
So, there are still possible aspects to be improved, in order to be able to calculate the Availability and Reliability more accurately, but this is material to be covered in subsequent studies.

Chapter 5

Communication in the Automation Industry

This part is about the communication of Material Handling Systems. As a reminder, the example of chapter 3 of how a physical architecture of a Material Handling System could look like will be given once again

5.1 The physical architecture

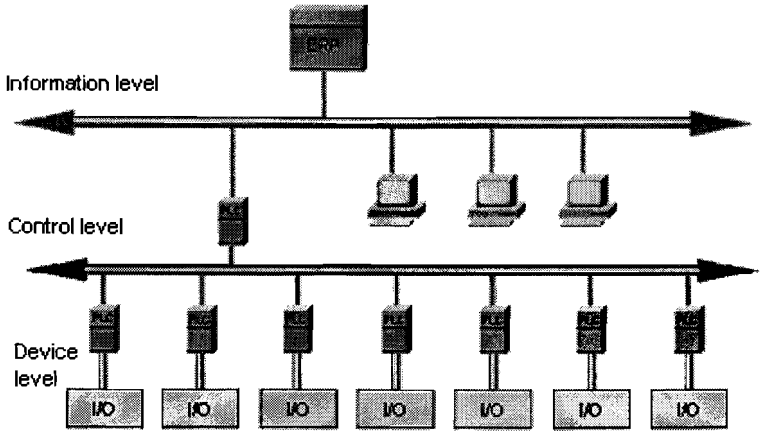


Figure 5-1 Architecture of Material Handling Systems

The figure above gives a possible physical architecture, which could be changed. The only certainty is that you have to handle the three levels. The number of devices or the kind of devices is variable, and therefore the different levels will be described functional.

5.2 The different levels of communication between the devices

CSi is a distributor of systems for Material Handling; this means that the separate devices have to be linked. So, through communication between the different devices, the devices become a system.
The different levels of communication will be explained now. In the text, some device-specific terms will be used; these will be explained further in Appendix G.

5.2.1 Device level

This is the lowest level of communication and it is characterized by the direct communication with the different I/O (sensors, actuators, etc.).

For example, when the I/O gives a signal about the status of a component, this status will be sent to a PLC. A program is running at this PLC that can react appropriately to different inputs, and engage in the appropriate reactions, for example turning on/off the motor.
Real-time communication is important In the automation industry and therefore an immediate reaction is required when there is a change in the status of an I/O component. Therefore, some requirements exist on the **response time** at this level of communication. The first requirement is that the response time is ***time-critical***. This means that the response time may not take too much time. This is a *relative* requirement, because it highly depends on the environment. In a Material Handling environment, a reaction time of second is satisfactory, but there are environments in which a reaction time of just milliseconds is required. So, the minimal demand of the communication in a Material Handling environment is for it to score *'middle'* at being time critical. Talking about real-time communication, **determinism** is also important. Determinism is the ability to reliably predict when data will be delivered, and repeatability ensures that transmit times are constant and unaffected by devices connecting to, or leaving, the network. So, determinism measures the consistency of the specified time interval between events. The protocol has to score at least *'middle'* at response time.

The environment can largely influence the communication. For example, turning on/off a motor could have its reflection on the communication. Therefore, the communication has to be **robust**. One of the possible ways to reach this is by ***maximizing the resistance to intervening variables.*** A way to enhance resistance to intervening variables is to provide the system with good wiring with enough isolation. The system can also be made more robust through ***error detection*** and ***error correction***.
The Material Handling environment, which has a lot of intervening variables, requires that the communication protocol scores *'high'* at robustness.

If nonetheless a communication failure appears, it is desirable to have some backup, so that when a communication failure occurs, another communication path is available. This is known as **availability** of the communication protocol. The now used protocol PROFIBUS does not have such alternative paths to 'jump in' when there is a communication failure. The only facility is that the network can be split, not failing the entire network, but only a part of it. Therefore, the communication protocol used to handle the communication of Material Handling systems has to score at least *'middle'* at availability.

Because more and more devices can generate information about their status, the load of the network and its complexity will increase. Therefore the **manageability of the network** becomes more and more important and the communication protocol has to score at least *'middle'* at this criteria. To guarantee a good manageability the following items have to be inspected: the ***access mechanism*** (How does the network discover the different addresses?), ***topology*** (Can the

topology easily be changed or not?) and ***complexity*** (Can it handle only low complexity or also higher complexity?).

Finally, one has to inspect the **ability to handle large amounts of data** (= data communication speed), which can be divided in:

- Low (< 1 Mbit/ sec)
- Middle (1 Mbit/ sec. - 10 Mbit/ sec.)
- High (> 10 Mbit/sec)

At the device level, the devices send more and more data packets, but these packets are but small. Therefore, it does not take much time to handle these packets, making a low ability to handle large amounts of data sufficient at this level of communication.

5.2.2 Control level

This is the level of communication that is characterized by two functions, which have to be fulfilled, namely:

1. Supervisory control[1] the devices situated below.
2. Data Acquisition[1] of the devices situated below;

The first function is just the same as was described in the section on the 'Device level' and has therefore the same criteria as mentioned above:

- *Response Time*
- *Robustness*
- *Availability*
- *Manageability*
- *Ability to handle large amounts of data*

Also, the scores that have to be minimally gotten by the communication protocol which handles the communication of the Material Handling system, are the same as described in the section on the 'Device level'.

The second function is used to collect data of the different devices situated below. Characteristics of Data Acquisition are that it is ***NOT*** time-critical and deterministic. So, robustness, availability, manageability and the ability to handle large amounts of data remain, which are also important criteria.
Earlier, it was mentioned that the systems are used in the automation industry. This means that there can be a lot of intervening variables, which have to be minimized. So here the conclusion can be made that the ***robustness has to be***

[1] A part of SCADA (Supervisory Control And Data Acquisition)

maximal, that the communication has got to have good ***resistance to intervening variables*** and that the communication scores has to score 'middle at least on ***availability***.

In the section on the device level, it was mentioned that more and more devices generate information about their status, which is interesting on this level of communication as well. Therefore it is important that the communication is able to handle all this information and that the devices could be re-configured or updated when needed. The criteria of ***manageability*** become increasingly important and the protocol has to score 'middle' here at least.

At this level of communication there is interchange of increasing quantities of data, so the communication protocol has to be able to ***handle large amounts of data***. So this is also an important criterion and the protocol has to score 'middle' at least on this.

5.2.3 Information level

In the example in figure 5.1 on the architecture of a Material Handling system, it is shown that there is communication between different components of the system, which all utilize the same communication channel.

The following functions have to be fulfilled here:

- Storing data of the devices, situated below.
- Visualization of the data, which was stored.
- Manage the data, so everyone can see which part of the data that is interesting for him.

Characteristics of this type of communication are that the response time is not that important and thus the communication does not have to be ***time-critical*** or ***deterministic.*** But still, the information delivery has to be ***available***.

This part of the communication process is not situated in a machine environment anymore. So there are no motors, etc. which can intervene with the communication process. Therefore, the susceptibility to intervening variables is no concern here and is ***robustness*** less important

The communication has to be ***available***, making sure that the information is accessible when it is needed for use. The network also has to be ***well manageable***, because of increasingly intelligent devices generating more and more available information. So the communication protocol has to be able to ***handle large amounts of data*** and has to score *'high'* at least.

At the moment, CSi is using PROFIBUS at the device and control level of the communication. This is a widely used and accepted standard in the automation industry. At the information level, CSi uses Ethernet with a bus topology for the communication.

But the target of this part of the graduating assignment is to analyze some other possibilities of communication protocols, so that a frame of reference is created.

5.3 Different types of communication protocols

There are two categories of data communication here, namely *single ended* (it uses a single wire for data transmission) and *multidrop* (one driver is connected to, and transmits on, a "bus" of up to the receivers). The different types of communication protocols will now be inspected and special attention will be paid to how they handle the criteria mentioned above.

5.3.1 Serial Communication

5.3.1.1 RS 232

RS 232 is an "Interface Between Data Terminal Equipment (DTE) and Data Circuit - Termination Equipment (DCE) Employing Serial Binary Data Interchange" and is a hardware specification. The standard is simply concerned with serial data communication between a host (DTE) and a device (DCE).
In short, the RS232 port was designed to communicate with local devices, and will support one driver and one receiver

Topology

RS - 232 is a single ended standard, what means that DTE and DCE have to communicate over a direct line. An example of this can be seen in figure 5-2

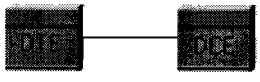


Figure 5-2 RS 232 connection

Addressing

Addressing is no problem here, because it is a one on one communication channel; so it always knows to whom it is communicating.

Protocol

The communication between these two components is full duplex; a component can send and receive simultaneously. This is realized by separating the send and receive channel.

For many applications, two data lines and two-handshake control lines are all that is necessary to establish and control communication between a host and a device. For example, a Material Handling System may need to interface with a photocell using a half–duplex communication scheme. At times the control systems may desire to read the status of the photocell. In this type of simple application, five signals may be all that is necessary (two for data, two for handshake control, and ground).

Time critical/ Deterministic

Because RS 232 is a single ended data communication, which means direct communication between two components with a fixed topology here, another device cannot communicate at the same time and interrupt the communication. Therefore it always takes the same time to communicate and therefore the communication is both *time-critical* and *deterministic*.

Error Detection

The RS 232 protocol only takes care for rough data transport. The signals necessary for serial communication are generated and received by the Universal Asynchronous Receiver/Transmitter (UART) and are not a part of the RS 232 communication, but they are necessary for correct communication.

The UART just mentioned performs the "overhead" tasks necessary for asynchronous serial communication. For example, the asynchronous nature of this type of communication usually requires that start and stop bits be initiated by the host system to indicate to the peripheral system when communication will start and stop. Parity bits are also employed frequently to ensure that the data sent has not been corrupted. The UART usually generates the start, stop, and parity bits when transmitting data and it can detect communication errors upon receiving data. This means that the influence of the intervening variables can be detected.

A note here is that there is no back up. If the communication line would fail, there is no communication possible at all. However, there **is** a control line, so it is possible to see whether the device on the other side is active or not.

5.3.1.2 RS 485

RS 485 is a hardware specification and is a specialized interface that is very common in the data acquisition world. Software protocol is not discussed in either specification. It is up to the designer to define a protocol suitable for the system.

Topology

RS 485 is defined as a multi-point bus (see figure 5-3) and will support 32 drivers and 32 receivers (bi-directional - half duplex - multi-drop communications over a single or dual twisted pair cable).

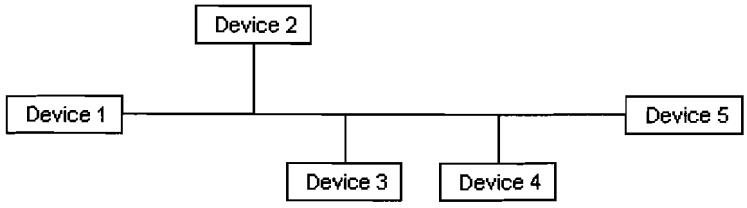


Figure 5-3 Bus topology

It uses differential signals on twisted pairs for receive and transmit. The typical use for RS485 is a single PC connected to several addressable devices that share the same cable. One can think of RS485 as a "party-lined" communication system (The addressing is handled by the Remote Computer unit)

Addressing

"How to address a RS 485 device?" The answer is, "It depends on the device." RS 485 specifies only that it is capable of connecting multiple devices on a single set of wires. It is the job of a particular device's software protocol to describe how to address a specific device.

Bus Protocol

RS 485 systems can be half duplex 2-wire systems (one twisted pair plus signal common/ground) or full duplex 4-wire systems. A RS 485 transmitter driver is activated to send data and is set to a high irnpedance tri-state at the end of transmission (which allows other drivers to transmit over the same pair of wires). Driver control can be automatic using a Send Data circuit, or manual by setting the RTS line or UART RTS control high for transmit, then low at the end of transmission. In a half duplex 2-wire system, the receiver is set to receive except when transmitting.

In a 2-wire system, all slaves and masters are normally in the receive mode. When one master transmits, all slaves and masters receive the signal and response, and all slaves must be able to ignore commands and responses to/from other slaves. Each slave must wait until transmit is finished plus a delay (for bus turn-around), before responding.

In a 4-wire system, all slaves are connected to the transmitter of the master(s). All slaves connect to the receiver of the master(s). Each slave must respond only to commands addressed to it, but no turn-around delay is needed. The slave can start responding immediately, even while receiving.

Each node in a multi-master type RS 485 system can initiate its own transmission creating the potential for data collisions. This type of system requires the designer to implement a more sophisticated method of error detection, including methods such as line contention detection, acknowledgement of transmission and a system for resending corrupted data.

Time critical/ Deterministic

The RS 485 protocol specifies the electromechanical structure for communication. It does not specify how the possible, different devices will respond, how it should react if an error should occur between the host and the devices, how the status of the different devices will be requested and in which order this would take place. Therefore it is difficult to say something about 'time critical' or 'determinism', because 'time-critical' and 'determinism' have to be ensured at a higher level of the communication.

Error Detection

The signals necessary for serial communication are generated and received by the Universal Asynchronous Receiver/Transmitter (UART) and are not a part of the RS 485 communication, although they are necessary for correct communication. (The working of the UART is explained in the paragraph concerning RS 232).

5.3.1.3 USB

The USB interconnect is the manner in which USB devices are connected to and communicate with the host.

Bus topology

The USB connects USB devices with the USB host. The USB physical interconnect is a tiered star topology. A hub is at the center of each star. Each wire segment is a point-to-point connection between the host and a hub or

function, or a hub connected to another hub or function. Figure 5-4 illustrates the topology of the USB.

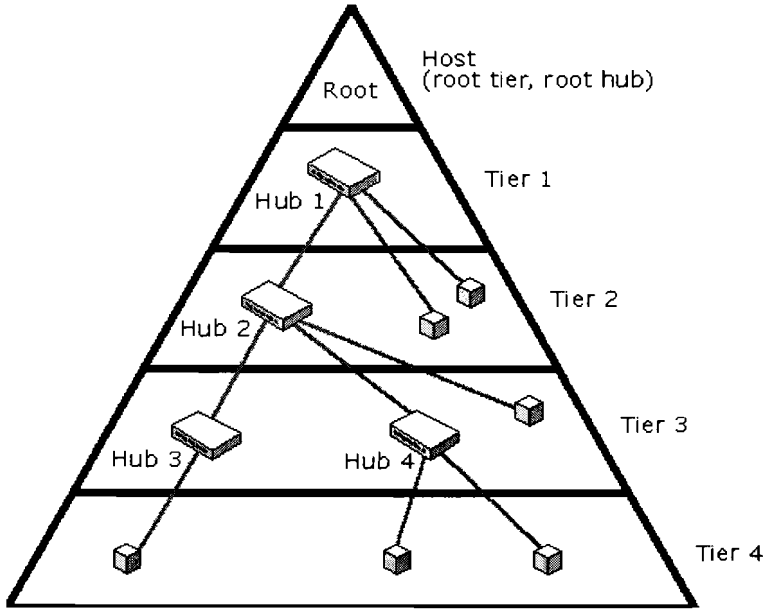


Figure 5-4 USB topology

Due to timing constraints allowed for hub and cable propagation times, the maximum number of tiers allowed is seven (including the root tier). Note that in seven tiers, five non-root hubs maximum can be supported in a communication path between the host and any device. A compound device (see Figure 5-4) occupies two tiers; therefore, it cannot be enabled if attached at tier level seven. Only functions can be enabled in tier seven.

Addressing

The Host Controller assigns a unique device address. During the configuration process, each device is assigned a unique address that it will respond to thereafter. No contention occurs since each device is assigned a unique address prior to enabling the next port. The standard Set Address request is used by software to assign the address.

Bus Protocol

The USB is a polled bus. The Host Controller initiates all data transfers. Most bus transactions involve the transmission of up to three packets. Each transaction begins when the Host Controller, on a scheduled basis, sends a USB packet describing the type and direction of transaction, the USB device address, and endpoint number. This packet is referred to as the "token packet." The USB device that is addressed selects itself by decoding the appropriate address fields. In a given transaction, data is transferred either from the host to a device or from

a device to the host. The direction of data transfer is specified in the token packet. The source of the transaction then sends a data packet or indicates it as data to transfer. The destination, in general, responds with a handshake packet indicating whether the transfer was successful.

The USB data transfer model between a source or a destination on the host and an endpoint on a device is referred to as a pipe. There are two types of pipes: stream and message. Stream data has no USB-defined structure, while message data does. Additionally, pipes have associations of data bandwidth, transfer service type, and endpoint characteristics like directionality and buffer sizes. Most pipes come into existence when a USB device is configured. One message pipe, the Default Control Pipe, always exists once a device is powered, in order to provide access to the device's configuration, status, and control information. The transaction schedule allows flow control for some stream pipes. At the hardware level, this prevents buffers from underrun or overrun situations by using a NAK handshake to throttle the data rate. When NAKed, a transaction is retried when bus time is available. The flow control mechanism permits the construction of flexible schedules that accommodate concurrent servicing of a heterogeneous mix of stream pipes. Thus, multiple stream pipes can be serviced at different intervals and with packets of different sizes.

Deterministic

USB is a polled bus, which means that the Host controller initiates all data transfers. This also means that this type of communication is deterministic, because no one else can start a data transfer than the Host controller and interrupt the present data transfer.

Robustness

There are several attributes of the USB that contribute to its robustness:

- CRC protection over control and data fields
- Detection of attach and detach and system-level configuration of resources
- Self-recovery in protocol, using timeouts for lost or corrupted packets
- Data and control pipe constructs for ensuring independence from adverse interactions between functions

Error Detection

The USB permits reliable end-to-end communication in the presence of errors on the physical signaling layer. This includes the ability to reliably detect the vast majority of possible errors and to recover from errors on a transaction-type basis. The core bit error rate of the USB medium is expected to be close to that of a backplane and any glitches will very likely be transient in nature. To provide

protection against such transients, each packet includes error protection fields. When data integrity is required, such as with lossless data devices, an error recovery procedure may be invoked in hardware or software.
The protocol includes separate CRCs for control and data fields of each packet. A failed CRC is considered to indicate a corrupted packet. The CRC gives 100% coverage on single- and double-bit errors.

Error Handling

The protocol allows for error handling in hardware or software. Hardware error handling includes reporting and retry of failed transfers. A USB Host Controller will try a transmission that encounters errors up to three times before informing the client software of the failure. The client software can recover in an implementation-specific way.

5.3.2 Bus structures with a master or multi master

Here, the most widely used protocols at the lower levels in the automation communication are discussed.

5.3.2.1 PROFIBUS

PROFIBUS is a consistent, open, digital communication system with a wide range of applications, particularly in the fields of factory and process automation. PROFIBUS is suitable for both fast, time-critical applications and complex communication tasks.

PROFIBUS *communication* is anchored in the international standards IEC 61158 and IEC 61784. The *application and engineering aspects* are specified in the general guidelines of the PROFIBUS User Organization. This fulfills user requirement for manufacturer independence and openness and ensures communication between devices of various manufacturers

Topology

PROFIBUS defines the technical characteristics of a serial field bus system with which distributed digital programmable controllers can be networked, from field level to cell level. PROFIBUS is a multi-master system and thus allows the joint operation of several automation, engineering or visualization systems with their distributed peripherals on one bus. PROFIBUS distinguishes between the following types of devices:

Master devices determine the data communication on the bus. A master can send messages without an external request when it holds the bus access rights (the token). Masters are also called active stations.

Slave devices are peripherals such as I/O devices, valves, drives and measuring transducers. They do not have bus access rights and they can only acknowledge received messages or send messages to the master when requested to do so. Slaves are also called passive stations. Since they only require a small portion of the bus protocol, their implementation is particularly economical.

Protocol Architecture

The protocol architecture is oriented to the OSI (Open System Interconnection) reference model in accordance with the international standard ISO/IEC 7498. In this model every transmission layer handles precisely defined tasks.

RS 485 transmission is the transmission technology most frequently used by PROFIBUS. The application area includes all areas in which high transmission speed and simple, inexpensive installation are required. Twisted pair, shielded copper cable with one conductor pair is used.
The bus structure permits addition and removal of stations or step-by-step commissioning of the system without influencing the other stations. Later expansions have no effect on stations that are already in operation.

An active bus terminator at the beginning and end of each segment terminates the bus. To ensure error-free operation, both bus terminators must always be powered. The bus terminator can usually be switched, in the devices or in the bus terrrıinator connectors.

Addressing

To address the data, PROFIBUS assumes that the slaves are built up as physical building blocks, or can be structured internally in logical function units, so-called modules. This model is also used in the basic DP functions for cyclic data transmission where each module has a constant number of input and/or output bytes, which are transmitted in a fixed position in the user data telegram. The addressing procedure is based on identifiers, which characterize the type of a module as input, output or a combination of both. All identifiers together give the configuration of a slave, which is also checked by the DPM1 when the system starts up. The new a-cyclic services are also based on this model. All data blocks enabled for read or write accesses are also considered as belonging to the modules. Slot number and index can address these blocks. The slot number addresses the module, and the index addresses data blocks belonging to a module. Each data block can have a size of up to 244 bytes. With modular devices, the slot number is assigned to the modules. Beginrıing with 1, the

modules are numbered consecutively in increasing order. Slot number 0 is for the device itself. Cornpact devices are treated as one unit of virtual modules. Addressing with slot number and index is also used here. Using the length specification in the read or write request, it is also possible to read or write parts of a data block. If access to the data block was successful, the slave responds with a positive read or write response. If it was not successful, the slave gives a negative response in which the problem is classified.

PROFIBUS Medium Access Protocol

The PROFIBUS Communication Profiles use a uniform medium access protocol. Layer 2 of the OSI reference model implements this protocol. This also includes data security and the handling of the transmission protocols and telegrams. In PROFIBUS, layer 2 is called Fieldbus Data Link (FDL). The Medium Access Control (MAC) specifies the procedure, which permits a station to transmit data. The MAC must ensure that only one station has the right to transmit data at a time. The PROFIBUS protocol has been designed to meet two primary requirements for the Medium Access Control:

- During communication between complex automation systems (masters), it must be ensured that each of these stations gets sufficient time to perform its communication tasks within a precisely defined time interval.
- On the other hand, for communication between a complex programmable controller and its assigned simple peripherals (slaves), cyclic, real-time data transmission needs to be implemented as fast and as simply as possible.

Therefore, the PROFIBUS medium access protocol includes the token passing procedure, which is used by complex bus stations (masters) to communicate with each other, and the master-slave procedure used by complex bus stations to communicate with the simple peripherals (slaves).

The **token passing procedure** ensures that the bus access right (the token) is assigned to each master within a precisely defined timeframe. The token message, a special telegram for passing the token from one master to the next master must be passed around the logical token ring once to all masters within a (configurable) maximum token rotation time. In PROFIBUS the token passing procedure is only used for communication between complex stations (masters).

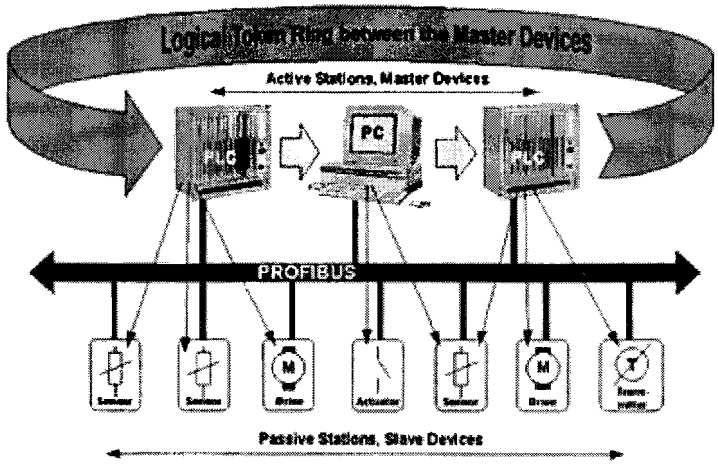


Figure 5-5

The **master-slave procedure** permits the master (the active station) that currently owns the token to access the assigned slaves (the passive stations). This enables the master to send messages to, or retrieve them from the slaves. This method of access allows implementation of the following system configurations:

- Pure master-slave system.
- Pure master-master system (token passing)
- A combination of the two

A token ring means the organizational lining up of active stations, which form a logical ring through their bus addresses. In this ring, the token, the bus access right, is passed on from one master to the next master in a predefined sequence (increasing addresses). When an active station receives the token telegram, it can perform the master role for a certain period of time and communicate with all slave stations in a master-slave communication relationship and all master stations in a master-master communication relationship.

The task of the bus access controller (MAC) of the active station is to detect this logical assignment in the startup phase of the bus system and to establish the token ring. During operation, defective or switched-off (active) stations must be removed from the ring and new active stations can be added to the ring. In addition, the bus access control ensures that the token is passed from one master to the next in order of increasing addresses.
The actual token hold time of a master depends on the configured token rotation time. In addition, the detection of defects on the transmission medium and on the line receiver, as well as the detection of errors in station addressing (e.g., multiple addresses assigned) or in token passing (e.g., multiple tokens or token loss) are characteristic features of the PROFIBUS medium access control.

Cyclic Process Data Transmission

The central controller (master) cyclically reads the input information from the slaves and cyclically writes the output information to the slaves. The bus cycle time should be shorter than the program cycle time of the central automation system, which for many applications is approximately 10 msec. In addition to cyclic user data transmission, DP provides powerful functions for diagnostics and commissioning. Data communication is monitored by monitoring functions on both the master and slave side.

Protection Mechanisms

Security and availability requirements make it necessary to provide DP with effective protection functions against parameterization errors or failure of the transmission equipment. To achieve this, monitoring mechanisms are implemented in the DP master and in the slaves in the form of time monitoring. The monitoring interval is defined during configuration.

Reliable operation is augmented by powerful error detection algorithms (CRC or Cyclic Redundancy Checking) and Watchdog timers.

The DP master 1 monitors data transmission of the slaves with the Data_Control_Timer. A separate control timer is used for each slave. The time monitoring is tripped when correct data transmission does not occur within the monitoring interval. The user is informed when this happens. If the automatic error reaction has been enabled, the DP master 1 exits its OPERATE state, switches the outputs of all assigned slaves to fail-safe status and changes to the CLEAR status.

The slave uses the watchdog control to detect failures of the master or the transmission line. If no data communication with the master occurs within the watchdog control interval, the slave automatically switches its outputs to the fail-safe status.

In addition, access protection is required for the inputs and outputs of the slaves operating in multi-master systems. This ensures that only the authorized master has direct access. For all other masters, the slaves offer an image of their inputs and outputs, which can be read from any master, even without access rights.

Acyclic Data Transmission on Demand

The extended DP functions make it possible to transmit a-cyclic read and write functions as well as alarms between master and slaves parallel and independent of cyclic user data communication. This allows the user to use for example an engineering tool (DPM2), to optimize the device parameters of the connected

field devices (slaves) or read out the device status without disturbing system operation. With these extended functions, DP meets the requirements of even complex devices, which often have to be parameterized during operation. The extended DP functions are mainly used for the online operation of the PA field devices by means of engineering tools. Transmission of the a-cyclic required data is performed with a lower priority parallel to the high-speed cyclic user data transfer. The master requires some additional time to carry out the a-cyclic communication services. This must be taken into account in the parameterization of the overall system. To achieve this, the parameterization tool usually increases the token circulation time somewhat in
order to give the master a chance to carry out not only cyclic data transmission, but also a-cyclic communication tasks.

Time critical and determinism

The **token passing procedure** ensures that the bus access right (the token) is assigned to each master within a precisely defined timeframe. The token message, a special telegram for passing the token from one master to the next master must be passed around the logical token ring once to all masters within a (configurable) maximum token rotation time. In PROFIBUS the token passing procedure is only used for communication between complex stations (masters) and therefore PROFIBUS is deterministic and time-critical.

5.3.2.2 AS-I

AS-I will usually be applied at the lowest level in a multi-level automation hierarchy. AS-I concentrates on the typical requirements to connect binary elements with a controller.
AS-I can be used as an interface physically integrated into field devices, for example actuators, sensors or other devices and elements themselves allowing the design of "intelligent" binary actuators, sensors or other devices and elements.

AS-I topology/ addressing

The AS-I system is a master-slave communication system composed of a single master and up to 31 slaves. Each slave has a unique address in the range of 1 to 31. This address is called the operation address. The operation address shall be non-volatile. Further there is a free choice of network structure.
The zero address is used during the change of a slave address. Normally, the zero address is non-volatile for factory new slaves.
A single transaction is composed of a master request and a slave response. If a slave responds to a master request it shall start its response within a period of 2.5 to 5.5 bit times after the end of a master request. The master will be able to accept the start of a slave response within a period of 2.5 to 10 bit times after the

end of its request to allow for propagation delay on the line and the possible use of repeaters.
A slave will not respond if it detects an erroneous master request. The slave will not give any negative response. The master will interpret the absence of a slave response as a negative response.

AS-I is an open, vendor-independent bus network that provides a low-cost solution for connecting binary products such as sensors, actuators, push buttons, valves, relays, etc., with higher level controllers such as PLCs or PCs. AS-I is used at the level directly below existing fieldbus and device-level network systems. AS-I transmits control and device data, configures the system architecture, powers the devices, and monitors the network, providing a complete system solution.

Protocol

A data exchange is realized by the processing of transactions. A transaction starts with a master request. The master will expect a slave response within certain time. If the master does not receive a valid response from the slave within this time, it may retransmit the master request once. After receiving a valid response and after the send pause has elapsed, the master will start the next transaction.

Time critical/ Deterministic

In a full configuration with 31 slaves, the scan cycle time is 5 ms. Scanning is deterministic and time-critical. Therefore, the AS-i network meets the real-time requirements of most control systems and processes.

Error detection

For the detection of errors AS-I uses a parity bit in the communication. The parity bit makes sure that the data received is composed of the same number of bits in the same order in which they were sent.

The AS-I network maintains a high degree of data integrity, provides non-stop monitoring of the network and peripheral devices, and provides excellent diagnostics.

High Availability

The repetition of a single telegram consumes only 150 us and is already taken into account in the specified cycle time.

5.3.2.3 InterBus

The open InterBus fieldbus system for modern automation seamlessly connects all the I/O and field devices commonly used in control systems. The serial bus cable can be used to network sensors and actuators, to control machine and system parts, to network production cells, and to connect higher-level systems such as control rooms.

Topology and Structure

In terms of topology, InterBus is a ring system, i.e., all devices are actively integrated in a closed transmission path. Each device amplifies the incoming signal and sends it on, allowing higher transmission rates at longer distances. Unlike other ring systems, the data forward and return lines in the InterBus system are led to all devices via a single cable. This means that the general physical appearance of the system is an "open" tree structure. A main line exits the bus master and can be used to form seamless sub networks up to 16 levels deep. This means that the bus system can be adapted quickly to changing applications.

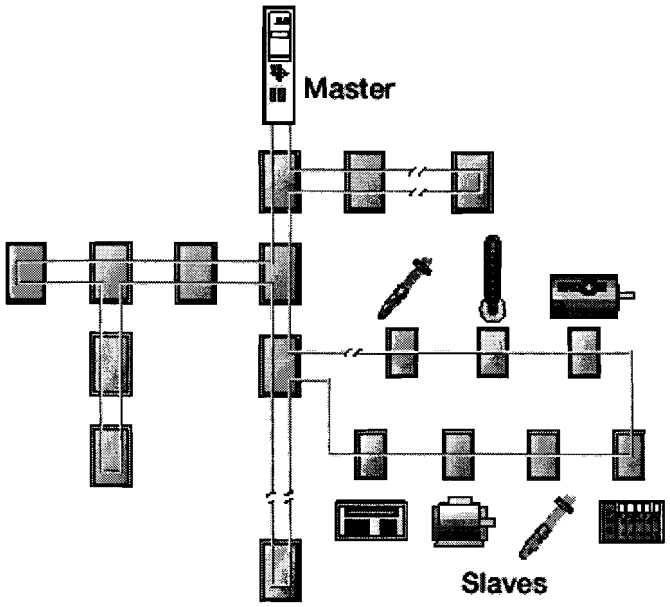


Figure 5-6 InterBus topology

Topology Flexibility

The InterBus master/slave system enables the connection of up to 512 devices, across 16 levels of networks. The last device automatically closes the ring.

Segmentation Flexibility

The point-to-point connection eliminates the need for termination resistors. The system can be adapted flexibly to meet the user's requirements by adding or removing devices.
Countless topologies can be created. Branch terminals create branches, which enable the connection and disconnection of devices. The coupling elements between the bus segments enable the connection and disconnection of a subsystem and thus make it possible to work on the subsystem without problems, e.g., in the event of an error or when expanding the system.

Addressing

Unlike other systems, in which data is assigned by entering a bus address using DIP or rotary switches on each individual device, the InterBus system assigns data automatically to devices using their physical location in the system. This plug and play function is a great advantage with regard to
the installation effort and service-friendliness of the system. The problems and errors, which may occur when manually setting device addresses during installation and servicing, are often underestimated. The ability to assign "easy to understand" software names to the physical addresses, allows devices to be added or removed without re-addressing existing devices.

Summation Frame Method – Master/Slave Structure

InterBus is the only bus system working according to the summation frame method that uses only one protocol frame for messages from all the devices. In this master/slave access method, the bus master acts as the coupling to the higher-level control or bus system.

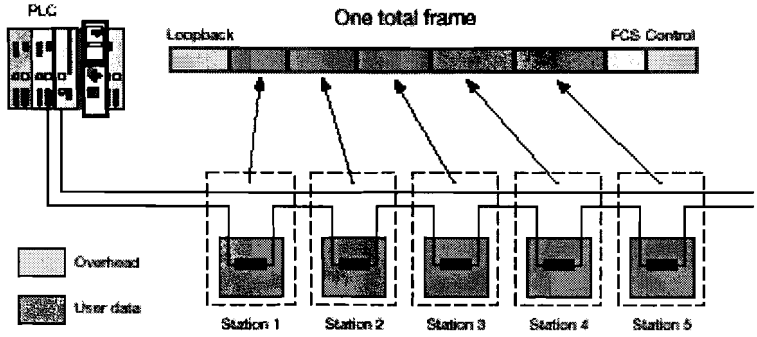


Figure 5-7 physical transmission method - summation frame method

The method provides a high level of efficiency during data transmission and enables data to be sent and received simultaneously (full duplex operation). With this data transmission method, InterBus ensures constant and predictable

sampling intervals for set points and real time control values. In summation frames, which consist of the header, the loop-back word, and data save and end information; data from all the connected I/O devices is grouped together in a block. The additional information that is required is transmitted only once per cycle. In practice, this method can be described as a register, which is formed by the devices that are connected in a ring system. In InterBus this consists of a number of binary memory cells, which push digital information from cell to cell to clock pulses. Each device has a certain number of buffers assigned to a preset number of cells for different tasks, e.g., data input and output for the process. Additional registers monitor the data transmission for errors. An InterBus device contains three registers that are connected in parallel. I/O data is transferred using the data register. The type of InterBus device is defined in the identification register. This enables the bus master to identify the devices and the bus topology, as well as to carry out addressing. Data is saved using the CRC16 register (cyclic redundancy check), where correct data transmission is checked.

Cycle Time and Calculation

The cycle time, i.e., the time required for I/O data to be exchanged once with all the connected modules, depends on the amount of user data in an InterBus system. The cycle time increases linearly with the number of I/O points, because it depends on the amount of information to be transmitted. A certain amount of time is needed for each bit. Because the summation frame has a set length, the cycle time also remains constant. In InterBus, the deterministic method of operation is provided by the summation frame method, which is essential for fast controllers.
Process data that is to be sent to the I/O devices is stored in the output buffer of the master in the physical order of the connected output stations. During data output, process information in the form of input data is simultaneously returned to the input buffer of the master. Once the entire summation frame has been sent and simultaneously read in again, all output data is correctly positioned in the individual devices. The data is made available to the host as defined by the user.

A network is established by connecting all the devices, whose length and structure corresponds exactly to the structure of the user data field in the summation frame telegram. The amount of user data for the summation frame method is over 60%. Bus access conflicts do not occur due to the master/slave structure. This means that potential error sources are avoided from the outset.

PCP Transmission

To transmit parameter data simultaneously as well as time-critical process data, a certain time slot must expand the data format. In several consecutive cycles, a different part of the data is inserted in the time slot provided for the addressed devices. The PCP software (Peripherals Communication Protocol) performs this task. It inserts one part of the telegram in each InterBus cycle and recombines it at its destination. The parameter channels are activated if necessary and do not

affect the transfer of I/O data. The longer transmission time for parameter data that is segmented into several bus cycles is sufficient for the low time requirements that are placed on the transmission of parameter information.

Determinism

An important feature of InterBus is determinism, i.e., the guaranteed time in which cyclic data transfer is carried out between spatially distributed devices. The summation frame method also ensures that the process image for all devices is consistent, because all the input data originates from the same point of scan time and the devices accept all the output data simultaneously.

Transmission Reliability

The bus master ensures transmission reliability by using the loop-back word. This unique bit combination is executed in a calculated number of bus system cycles. If it has returned to the master input buffer after this time, the ring is closed. Data is saved according to the CRC16 method. This information is attached to the data, and evaluated by the receiver.

5.3.3 Distributed mechanism

These protocols are normally used at the higher level of communication, for example for networks in offices. The protocols were developed for such environments, but we have seen that there are specific requirements in the automation industry, especially at the lower levels. Therefore, first the protocols will be described in a general way, and second the issue whether they can handle the requirements of the automation industry will be discussed. There are some possible solutions to solve the problems, and these solutions will also be mentioned.

5.3.3.1 Ethernet with CSMA/CD (IEEE 802.3)

The Ethernet protocol is by far the most widely used. Ethernet uses an access method called CSMA/CD (Carrier Sense Multiple Access/Collision Detection). This is a system in which each computer listens to the cable before sending anything through the network. If the network is clear, the computer will transmit. If some other node is already transmitting on the cable, the computer will wait and try again when the line is clear. Sometimes, two computers attempt to transmit at the same instant. When this happens a collision occurs. Each computer then backs off and waits a random amount of time before attempting to retransmit.

Topologies

The Ethernet protocol allows for linear bus, star, or tree topologies.

Bus, star and tree topology:

A linear bus topology consists of a main run of cable with a terminator at each end (See fig. 5-8). All devices are connected to the linear cable.

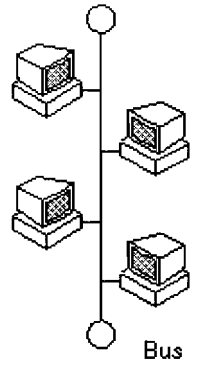


Figure 5-8 Bus topology

Bus topology requires no amplification or regeneration equipment, and all devices on the network have access to the bus. After waiting to determine an open line, a signal can be sent by any device along the network. It is expected that any signal be terminated at the end of the trunk.
One potential cost of this less expensive topology is that all devices are affected if the trunk cable fails. For this reason, more recent Ethernet networks are configured as stars.

A star topology is designed with each device connected directly to a central network concentrator (See fig. 5-9).

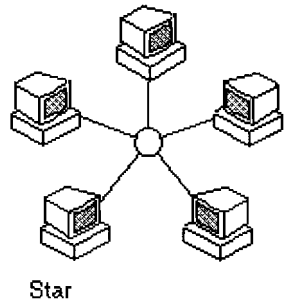


Figure 5-9 Star Topology

Data on a star network passes through the concentrator before continuing to its destination. The concentrator manages and controls all functions of the network. It also acts as a repeater for the data flow.
The advantage of this type of network is availability, for if one of these 'point-to-point' segments has a break; it will only affect the two nodes on that link. Other devices on the network continue to operate as if that segment were nonexistent.

A tree topology combines characteristics of linear bus and star topologies. It consists of groups of star-configured workstations connected to a linear bus (See fig. 5-10). Tree topologies allow for the expansion of an existing network.

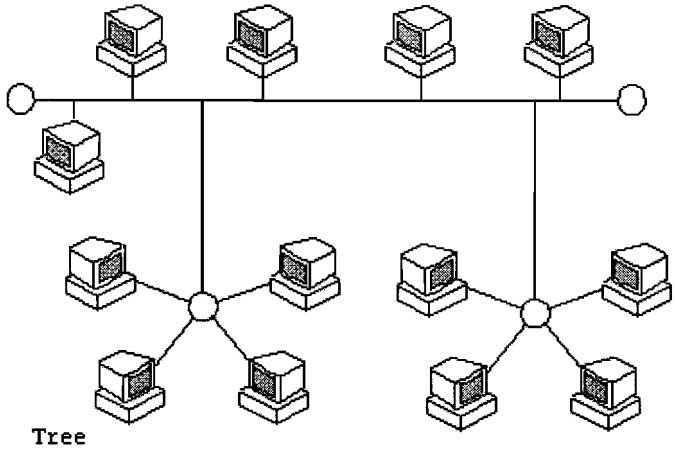


Figure 5-10 Tree topology

Ring topology:

The ring topology is also a possible option in combination with Ethernet (see figure 5-11), but this cannot function immediately; there has to be an adjustment. First there is explained why this will not function without this particular adjustment.

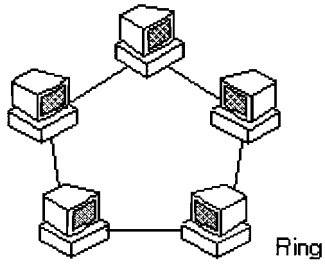


Figure 5-11 Ring topology

The access method that Ethernet uses here is the same as in Ethernet with a bus topology, CSMA/CD. In the case of Ethernet with a bus structure, when the message comes at the end of the bus the message will not be reflected through the use of a bus terminator. In Ethernet Ring, there is no end of beginning of the bus, so it is not known when the message arrived at the end, which is a kind of loop. Maybe the message will go around again and the station, which has to receive the message, sees the message again, while not knowing it has already seen this particular message. It becomes clear that Ethernet can not handle loops. Therefore, these loops have to be eliminated.
Therefore bridges or switches are introduced to couple the stations, so the Spanning Tree Protocol can be used.

Spanning Tree Protocol

Spanning Tree Protocol (STP) is a bridge-based mechanism for providing fault tolerance on networks. Spanning Tree Protocol is a protocol used by bridges (or switches) to remove redundant links from networks. As shown above, Ethernet with Ring topology could **not** function without the Spanning Tree Protocol! All spanning tree devices will send BPDU (Bridge Protocol Data Unit) frames out each of their ports and listen for BPDU frames from other devices. The protocol allows spanning tree devices to communicate path cost and identification information so that each device can block the highest cost path that is redundant. Therefore, STP allows you to implement parallel paths for network traffic and ensure the following functionality:

- Redundant paths are blocked (or disabled) while the main paths (the lowest cost path) are operational.
- Redundant paths are enabled if the main path fails.

The path cost factor is used to calculate the distance from each port of a bridge to the Root bridge.

Ports can be in one of the following states:

1. **Blocking** - The interface does not participate in frame forwarding.
2. **Listening** -The first transitional state after the blocking state when the spanning tree determines that the interface should participate in frame forwarding.
3. **Learning** -The interface prepares to participate in frame forwarding.
4. **Forwarding** - The interface forwards frames.
5. **Disabled** - The interface is not participating in spanning tree because of a shutdown port, no link on the port, or no spanning-tree instance running on the port.

In figure 5-12 the configuration of figure 5-11 is displayed abstractly as a graph, with the stations as nodes. A line connects each station by a bridge. The graph

could be reduced by a Spanning Tree, which eliminates the dotted line, because this is the redundant path, which is not normally used, except when the normal path does not function.

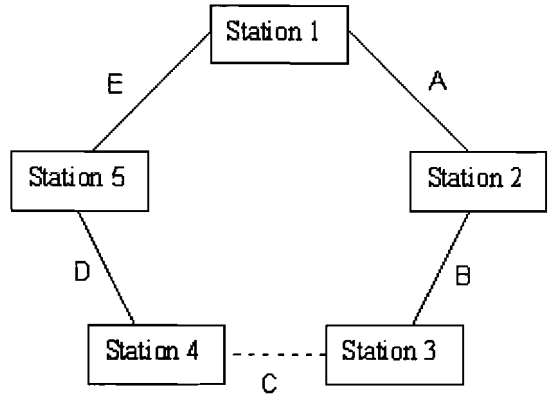


Figure 5-12

With this Spanning Tree there is just one path from one station to the other. When the bridges match, a Spanning Tree follows the communication in the matched route in the Spanning Tree. While there is one unique path from a station to another there are no redundant links possible.

To make a Spanning Tree the bridges have to decide which bridge has to be the root. All switches in the network participating in spanning tree gather information about other switches in the network through an exchange of BPDU data messages. This exchange of messages results in these actions:

- The election of a unique root switch for each spanning-tree instance
- The election of a designated switch for every switched LAN segment
- The removal of loops in the switched network by blocking Layer 2 interfaces connected to redundant links

For each LAN, the bridge with the highest bridge priority (the lowest numerical priority value) is elected as the root bridge. If all bridges are configured with the default priority (32768), the bridge with the lowest MAC address in the LAN becomes the root bridge. The bridge priority value occupies the most significant bits of the bridge ID.
When the bridge priority value is changed, the probability of the bridge being elected as the root bridge is changed; configuring a higher value decreases the probability; a lower value increases the probability.
The root bridge is the logical center of the spanning-tree topology in a bridged network. All paths that are not needed to reach the root bridge from anywhere in the bridged network are placed in the spanning-tree blocking mode.
BPDUs contain information about the sending bridge and its ports, including bridge and MAC addresses, bridge priority, port priority, and path cost. Spanning

tree uses this information to elect the root bridge and root port for the bridged network and the root port and designated port for each bridged segment.
Now the root bridge is known the tree will be made of the shortest paths from the root to each of the other bridges. This tree is the Spanning Tree. If a station or a bridge fails, another Spanning Tree has to be made. The result is that a unique path is formed from each station to the Root.
The packets, which will be transported from one station to another station, will follow this unique path, so the packets are routed directly to the right station, which is a property of Ethernet Switching Technology. Therefore the technique of Ethernet Switching will be explained below.

Ethernet Switching Technology

Originally Ethernet is a broadcast based protocol. In shared Ethernet, packets are flooded to all stations on the segment whether the packet is intended for them or not.

Switched Ethernet works differently. If a bridge or switch does not know the location of a station, the packets are flooded to all ports, until the station is found. From then on, packets are routed to that station and no other stations have access to those frames. However, packets that have a broadcast address for the destination will be sent to all stations. So a switch or bridge learns where the stations are through "Backward Learning", what means that it sees where the frames come from and save that information in a forwarding table.

Disadvantages of the Spanning Tree Protocol

There are several disadvantages to transparent bridging. First, the spanning tree protocol must be fairly conservative about activating new links, otherwise loops can develop. Also, all the forwarding tables must be cleared every time the spanning tree reconfigures, which triggers a *broadcast storm* as the tables are reconstructed.
To understand the *broadcast storm* the actions after a topology change will be described. When an STP bridge detects a topology change, it first notifies the root bridge, using a reliable mechanism as shown in the diagram below.

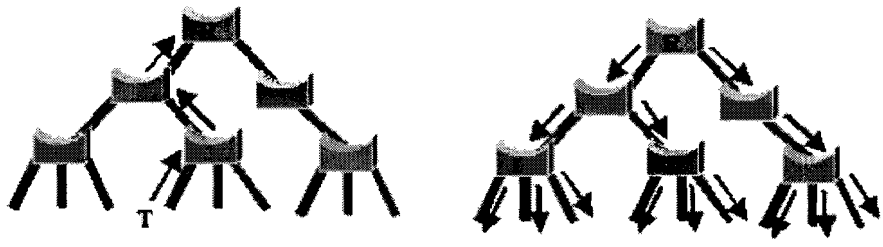


Figure 5-13 A topology change is generated on point T

Once the root bridge is aware of a change in the topology of the network, it sets the TC flag on the BPDUs and sends it out, which are then relayed to all bridges in the network. When a bridge receives a BPDU with the TC flag bit set, it reduces its bridging-table aging time to forward delay seconds, ensuring a relatively quick flushing of stale information.

This limits the usefulness of transparent bridging in environments with fluid topologies. Redundant links can sit unused, unless careful attention is given to root bridge selection. In such a network (with loops), some bridges will always sit idle anyway. Also, spanning tree table building after network failures takes considerable time and introduces long user delays. This is not desirable; therefore Rapid Spanning Tree Protocol (RSTP) has been developed, which is explained in the next part.
Finally, like all bridging schemes, the unnecessary broadcasting can affect overall performance. Its use is not recommended in conjunction with low-speed serial links.

So Spanning Tree Protocol is inefficient and disadvantageous when used in large networking protocol, but necessary for functioning of a Ring topology with Ethernet. Spanning Tree Protocol is better utilized when the network is made up of many point-to-point circuits.

Rapid Spanning Tree Protocol (RSTP)

The RSTP takes advantage of point-to-point wiring and provides rapid convergence of the spanning tree. Reconfiguration of the spanning tree can occur in less than 1 second (in contrast to 50 seconds with the default settings in the 802.1D spanning tree), which is critical for networks carrying delay-sensitive traffic.

The RSTP provides rapid convergence of the spanning tree by assigning port roles and by determining the active topology. The RSTP builds upon the IEEE 802.1D STP to select the switch with the highest switch priority (lowest numerical priority value) as the root switch. Then the RSTP assigns one of these port roles to individual ports:

- **Root port** - provides the best path (lowest cost) when the switch forwards packets to the root switch.
- **Designated port** -connects to the designated switch, which incurs the lowest path cost when forwarding packets from that LAN to the root switch. The port through which the designated switch is attached to the LAN is called the designated port.
- **Alternate port** - offers an alternate path toward the root switch to that provided by the current root port.

- **Backup port** - acts as a backup for the path provided by a designated port toward the leaves of the spanning tree. A backup port can exist only when two ports are connected together in a loop back by a point-to-point link or when a switch has two or more connections to a shared LAN segment.
- **Disabled port** - has no role within the operation of the spanning tree.

A port with the root or a designated port role is included in the active topology. A port with the alternate or backup port role is excluded from the active topology. In a stable topology with consistent port roles throughout the network, the RSTP ensures that every root port and designated port immediately transition to the forwarding state while all alternate and backup ports are always in the discarding state (equivalent to blocking in 802.1D). The port state controls the operation of the forwarding and learning processes.

Rapid Convergence

The RSTP provides for rapid recovery of connectivity following the failure of bridge, or a LAN. It provides rapid convergence for edge ports, new root ports, and ports connected through point-to-point links as follows:

- Edge ports – If you configure a port as an edge port on an RSTP switch by using the **spanning-tree portfast** interface configuration command, the edge port immediately transitions to the forwarding state. An edge port is the same as a Port Fast-enabled port, and you should enable it only on ports that connect to a single end station.
- Root ports – If the RSTP selects a new root port, it blocks the old root port and immediately transitions the new root port to the forwarding state.
- Point-to-point links – If you connect a port to another port through a point-to-point link and the local port becomes a designated port, it negotiates a rapid transition with the other port by using the proposal-agreement handshake to ensure a loop-free topology.

The bridge determines the link type from the port duplex mode: a full-duplex port is considered to have a point-to-point connection; a half-duplex port is considered to have a shared connection. You can override the default setting that is determined by the duplex setting by using the **spanning-tree link-type** interface configuration command.

Topology Changes Mechanism

When an STP bridge detects a topology change, it first notifies the root bridge, using a reliable mechanism as mentioned earlier in this paper.

The topology change mechanism has been deeply remodeled in RSTP. Both the detection of a topology change and its propagation through the network have evolved.
In RSTP, only non-edge ports moving to the forwarding state cause a topology changes. This means that a loss of connectivity is not considered as a topology change any more, contrarily to STP. When a RSTP bridge detects a topology change, the following happens:

- It starts the TC while Timer with a value equal to twice the hello timer for all its non-edge designated ports and its root port if necessary.
- It flushes the MAC addresses associated with all the ports.
- As long as the TC while timer is running on a port, the BPDUs sent out of that port the TC Bit set. BPDUs are also sent on the root port while the timer is active.

When a bridge receives a BPDU with the TC bit set from a neighbor, the following happens:

- It clears the MAC addresses learnt on all it ports except the one that received the topology change.
- It starts the TC While timer and sends BPDUs with TC set on all its designated ports and root port

This way, the TC is flooded very quickly across the whole network. The TC propagation is now one step process. In fact, the initiator of the topology change is flooding this information throughout the network. This mechanism is much faster than the STP equivalent. There is no need to wait for the root bridge to be notified and then maintain the topology change state for the whole network.

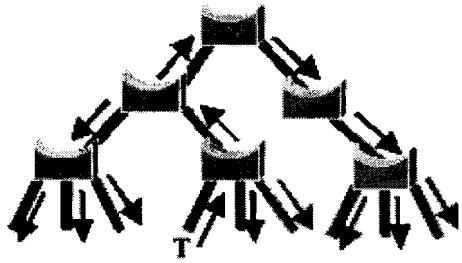


Figure 5-14 The originator of the TC directly floods this information through the network

So by using the Rapid Spanning Tree Protocol Ethernet can be used with a Ring Topology, which makes Ethernet more available, because some redundancy is build in. The convergence time to make the Spanning Tree becomes much smaller than when using the normal Spanning Tree Protocol.

Addressing (ARP)

On a single physical network, individual hosts are known on the network by their physical hardware address. Over the physical network, workstations communicate with each other with their Media Access Control (MAC) address. Higher-level protocols address destination hosts in the form of a symbolic address (IP address in this case). When such a protocol wants to send a datagram to destination IP address w.x.y.z, the device driver does not understand this address.
Therefore, Address Resolution Protocol (ARP) is provided that will translate the IP address to the physical address of the destination host. It uses a lookup table, sometimes referred to as the *ARP cache*, to perform this translation.
When the address is not found in the ARP cache, a broadcast is sent out on the network, with a special format called the *ARP request.* If one of the machines on the network recognizes its own IP address in the request, it will send an *ARP reply* back to the requesting host. The reply will contain the physical hardware address of the host and source route information (if the packet has crossed bridges on its path). Both this address and the source route information are stored in the ARP cache of the requesting host.

To reduce the number of address resolution requests, a client normally **caches** resolved addresses for a (short) period of time. The ARP cache is of a finite size, and would become full of incomplete and obsolete entries for computers that are not in use if it was allowed to grow without check. The ARP cache is therefore periodically flushed of all entries. This deletes unused entries and frees space in the cache. It also removes any unsuccessful attempts to contact computers that are not currently running.

All subsequent datagrams to this destination IP address can now be translated to a physical address, which is used by the device driver to send out the datagram on the network.

Flow Control

Inspecting Ethernet with CSMA/CD it becomes clear that this protocol prevents collisions through listening if the network is busy or not. This is not enough to make the network a reliable one. For example, it is possible that the messages do not arrive at the receiver correctly. Therefore, the Transmission Control Protocol will be used to get a more reliable end-to-end byte stream. The Transmission Control Protocol (TCP) is probably the most widely used protocol; it is also the most carefully tuned.
TCP guarantees the reliable, in-order delivery of stream of bytes. It is a full-duplex protocol, meaning that each TCP connection supports a pair of byte streams, one following in each direction. Therefore it includes a flow-control mechanism for each of these byte streams that allows the receiver to limit how much data the sender can transmit at a given time. Finally TCP supports a de-

multiplexing mechanism that allows multiple application programs on any given host to simultaneously carry over conversation with their peers. In addition to the above features, TCP also implements a highly tuned congestion-control mechanism. The idea of this mechanism is to throttle how fast TCP sends data, not for the sake of keeping the sender from overrunning the receiver, but so as to keep the sender from overloading the network.

So summarizing all TCP algorithm expects that timeouts be caused through congestion. There are two types of problems, which cause congestion, namely receiver capacity and network capacity.
Flow control with the advertised window, preventing senders from overrunning the capacity of receivers. Congestion control with congestion window, involves preventing too much data from being injected into the network, thereby causing switches or links to become overloaded.
In TCP/IP there is also a CRC check to see if the message was received correctly. So also this kind of errors could be detected.

Determinism/ time-critical

Formally, Ethernet is not known as a deterministic protocol, because all devices have the same right to send a message. So if a device has to send a message it has to listen if no on else device is sending. If someone else is sending then the device has to wait a random time, otherwise it could send directly.
Therefore it is hard to say how long it will take to communicate between devices, especially when the load of network will increase. This is hard to predict, which means that Ethernet is not deterministic and not very suitable for time critical communication.

There are some solutions to make Ethernet more deterministic. One of the methods is already used in field buses, namely polling. There is one master or there are multi masters that check the status of the devices cyclically. Doing this, it becomes easier to predict the communication time, because it is directed by the master(s).
Another option is priorization, which means splitting the different kinds of messages and giving them each a different priority. Important messages (for example real-time messages) have a higher priority and therefore, go before messages with a lower priority, making the system more deterministic.

5.3.3.2 Token Ring (IEEE 802.5)

IBM developed the Token Ring protocol in the mid-1980s. The access method used involves token passing. In Token Ring, the computers are connected in such a way that the signal travels around the network from one computer to another in a logical ring. A single electronic token moves around the ring from one computer to the next. If a computer does not have information to transmit, it

simply passes the token on to the next workstation. If a computer wishes to transmit and receives an empty token, it attaches data to the token. The token then proceeds around the ring until it comes to the computer for which the data is meant. At this point, the receiving computer captures the data. The Token Ring protocol requires a star-wired ring using twisted pair or fiber optic cable. It can operate at transmission speeds of 4 Mbps or 16 Mbps.
Unlike the bus, Token Ring uses a deterministic, rather than a contention-based, access method. In the Token Ring access method, an electronic signal called a token is passed from station to station on the ring, with each station regenerating the token as it passes by.
When a station wishes to transmit data over the network, it must wait until its neighboring station passes the token to it. It takes control of the station and then places a data packet on the network. Only after the data packet has made a full circuit of the ring, returning to its originator, the station releases the token for the next workstation.

Due to the increasing popularity of Ethernet, the use of Token Ring has decreased.

5.4 Evaluation of different techniques

Now the different techniques of communication have been analyzed, they have to be evaluated. Inspecting how each technique scores on the criteria mentioned in paragraph 5.1.

5.4.1 RS 232

Criteria	*Low*	*Middle*	*High*
Response Time			✓
Robustness	✓		
Availability	✓		
Manageability	✓		
Ability to handle large amounts of data	✓		

Response Time

While analyzing the determinism of a protocol, it is important to inspect whether it is predictable when the data will be delivered, whether there is low variance in the delivery or not, and whether the response is rapid or not.
RS 232 is a direct communication channel with some control lines, which makes it possible for each side to see whether the channel is in use or not. This prevents collisions and therefore these collisions can have no influence on the time of delivering the data.

Therefore, the conclusion can be made that RS 232 satisfies the demands of determinism and being time-critical. Therefore, it scores high on response time

Robustness

By using well-shielded wires the resistance to intervening variables could be maximized, but the protocol itself does not specify anything about maximizing the resistance to intervening variables. The fact that the designer himself has to take care for error detection or maximizing resistance to intervening variables makes that this protocol scores low at robustness.

Availability

RS 232 specifies only the electromechanical structure for communication and it does not specify how to design a highly available network. So the conclusion can be that RS 232 scores low at availability.

Manageability

RS 232 involves a direct communication channel, so there are two handshake control lines for the bi-directional communication. Furthermore, no access mechanism is necessary because there are only one sender and one receiver. RS 232, therefore, scores high on access mechanism.

A disadvantage is that RS 232 is a one on one connection and the designer has to route each device when communication is desirable. You can conclude that it is not flexible for topology changes and that is not cheap to route a cable to each device. So at this point RS 232 scores low when the complexity of the network becomes quite large.

Summarizing this all, RS 232 scores low at manageability, because the one on one connection is a major disadvantage.

Ability to handle large amounts of data

The maximum of data transmission speed that can be reached using the RS 232 protocol is 256 Kbit/ Sec. Therefore, RS 232 cannot handle large amounts of data and scores low at the ability to handle large amounts of data.

5.4.2 RS 485

Criteria	*Low*	*Middle*	*High*
Response Time	✓		
Robustness	✓		
Availability	✓		
Manageability	✓		
Ability to handle large amounts of data		✓	

Response Time

Inspecting the possible topology of the RS 485 protocol it can be seen that it has a bus structure. Further RS 485 specifies the electromechanical structure for communication and the receivers and master(s) are normally in the receive mode.
Each master in a RS 485 system can initiate its own transmission creating the potential for data collisions. The system requires the designer to implement a more sophisticated method of error detection, including methods such as line contention detection. Therefore, it scores low on determinism, because it is hard to predict when the data will be delivered.

Collisions could influence the delivery time of the data and can be avoided using a method for line contention detection, but this has to be implemented by the designer of the network himself. There is also no service for control of the load of the network, so the possibility exists that the network becomes overloaded. Polling and priorization could rectify this problem. The conclusion can be made that RS 485 scores low at being time-critical.

Surnmarizing this, RS 485 scores low at response time.

Robustness

By using well-shielded wires the resistance to intervening variables can be maximized, but the protocol itself specifies nothing about wiring. Therefore, the RS 485 protocol scores low on resistance to intervening variables.
The user has to take care himself for error detection and maximizing the resistance to intervening variables. Therefore, the RS 485 protocol scores low on robustness.

Availability

A note here is that the topology of RS 485 is a bus topology. This means that when a failure occurs in the bus, all the devices are unreachable. This makes the

protocol an unreliable one. Therefore, it can be concluded that RS 485 scores low at Availability.

Manageability

When analyzing the access mechanisms, special attention is paid to the way in which the protocol discovers the network, so which devices are connected. RS 485 does not specify how the devices are addressed. These addresses have to be configured at the beginning, because the protocol does not discover the different devices by itself and thus RS 485 scores low at manageability.

Ability to handle large amounts of data

The maximum of data transmission speed that can be reached using the RS485 protocol is more than 10 Mbit/ Sec; sometimes even 30 Mbit/ Sec. But when the complexity grows the effective data transmission speed would be around the 10 Mbit/ Sec and therefore, we can conclude that RS 485 can handle large amounts of data and that it scores middle at the ability to handle large amounts of data.

5.4.3 USB

Criteria	*Low*	*Middle*	*High*
Response Time			✓
Robustness		✓	
Availability		✓	
Manageability		✓	
Ability to handle large amounts of data			✓

Response Time

USB is defined as a polled bus, which means that the Host Controller initiates all data transfer and polls the devices. Because there is one initiator of data transfer, the Host controller can communicate with the devices directly and does not have to wait for, for example, a token. This all makes it possible to predict when the data will be delivered, and this makes the USB protocol a deterministic one. The USB protocol therefore scores high at determinism and being time-critical

Robustness

In the section on USB was mentioned that there are several attributes that contribute to its robustness, namely:

- CRC protection over control and data fields

- Detection of attach and detach and system-level configuration of resources.
- Self-recovery in protocol, using timeouts for lost or corrupted packets

But nothing in USB was mentioned about how to protect against intervening variables. Therefore, it can be concluded that the USB protocol scores 'middle' on robustness.

Availability

The topology of USB protocol is a tree structure with hubs. This means that when there is a communication failure the part of the topology under the hub cannot be reached, but the part above the hub can function normally. There is no alternative path that can still reach the device. The conclusion can be made that USB scores 'middle' at availability.

Manageability

The topology of the system has to be configured during the configuration process. Here each device gets a unique address and when the topology changes the configuration, this process has to be redone. This is not the case when a single device should fail, but only when the topology changes permanently. Therefore, USB scores middle at 'access mechanism'.

Ability to handle large amounts of data

The maximum of data transmission speed that can be reached using the USB protocol is 480 Mbit/ Sec. Therefore can be said that USB can handle large amounts of data and that it scores high at the ability to handle large amounts of data.

5.4.4 PROFIBUS

Criteria	*Low*	*Middle*	*High*
Response Time			✓
Robustness			✓
Availability		✓	
Manageability		✓	
Ability to handle large amounts of data		✓	

Response Time

The token passing procedure ensures that the bus access right is assigned to each master within a precisely defined timeframe. The token message, a special telegram for passing the token from one master to the next must be passed around the logical token ring once to all masters within a maximum token rotation time. In PROFIBUS, the token passing procedure is used only for communication between masters and makes that PROFIBUS to score high at being time-critical.

The token passing ensures also that there are no collisions on the bus, which could influence the time that the data will be delivered. So the conclusion can be made that PROFIBUS scores high on determinism.

Summarizing these sub-criteria the conclusion can be made that PROFIBUS scores high at response time.

Robustness

Different cable types (type designation A - D) for different applications are available on the market for connecting devices either to each other or to network elements. When using RS 485 transmission technology, use of *cable type A* is recommended. Always use a shielded data line (type A is shielded) to ensure high interference immunity of the system against electromagnetic emissions. The shield should be grounded on both sides where possible and large-area shield clamps used for grounding to ensure good conductivity and then PROFIBUS scores high at resistance to intervening variables.

Analyzing the reliability of the data exchange of PROFIBUS it can be seen that PROFIBUS checks whether the receiver has gotten the message correctly or not. PROFIBUS therefore is reliable on this criterion. Powerful error detection algorithms and Watchdog timers augment error detection. Therefore, PROFIBUS scores high at robustness.

Availability

When the network aspect of PROFIBUS is analyzed, it can be seen that it has a bus structure, which could be coupled by couplers. So when a connection in the bus fails, the whole bus cannot be reached and that there is no alternative way. The only advantage is that it is possible to divide the network and couple these by the couplers, so that when a part of the network fails the other part can still be reached. Considering this all, PROFIBUS scores 'middle' at availability.

Manageability

The topology of PROFIBUS is static so when there is a topology change the master has to reconfigure the topology. It is not self-detecting on topology changes, which decreases the flexibility of PROFIBUS and makes that PROFIBUS scores low at manageability.

Ability to handle large amounts of data

The maximum data transmission speed of PROFIBUS is 12 Mbit/ sec. This means that it scores high at the ability to handle large amounts of data. But when the complexity grows, it is logical that this speed will decrease. Therefore, PROFIBUS can handle large amounts of data and scores middle on this.

5.4.5 AS-I

Criteria	*Low*	*Middle*	*High*
Response Time			✓
Robustness			✓
Availability		✓	
Manageability	✓		
Ability to handle large amounts of data	✓		

Response Time

In a full configuration of AS-I there is one master and 31 slaves, which have to be scanned. The scan cycle time in a full configuration is 5 ms, because there is one master and collisions will be excluded because of this. The conclusion can be made that AS-I scores high at being time-critical.
Therefore it is possible to predict the time in which the data is to be delivered, which makes AS-I score high at determinism.

Robustness

A note here is the lack of shielding and the obvious concerns about noise immunity. Digital signals are encoded on the cable in a sinusoidal signal, which has a very narrow frequency bandwidth. Filtering which is distributed through the network rejects all extraneous frequencies, and in this way AS-I can be operated in electrically noisy environments without experiencing transmission errors.
So AS-I maximizes the resistance to intervening variables in spite of the lack of shielding. This is done in a very specific way, but it is very functional.

The parity bit is used in AS-I to check whether the data is delivered correctly. With this parity bit reliable data delivery can be guaranteed and therefore AS-I scores high at robustness.

Availability

The choice of the topology of AS-I is free; the designer decides for himself whether he wants to use hubs in his network or design the network in a way in which there are alternative paths. Therefore, the designer could design the AS-I network as available as he likes.

Manageability

Each slave has a unique address, which can be programmed manually. Therefore, the topology has to be configured by the user himself and it cannot be done automatically. Therefore, AS-I scores low at manageability.

Ability to handle large amounts of data

Analyzing the ability to handle large amounts of data it can be seen that AS-I can handle 167 Kbits/ sec. This is not very much. Therefore, the conclusion can be made that AS-I score low at the ability to handle large amounts of data.

5.4.6 InterBus

Criteria	*Low*	*Middle*	*High*
Response Time			✓
Robustness			✓
Availability		✓	
Manageability			✓
Ability to handle large amounts of data	✓		

Response Time

Data transfer is carried out cyclically between spatially distributed devices. The summation frame method ensures that the process image for all devices is consistent. This makes that InterBus scores high at determinism and being time-critical.

Robustness

In many cases, data from an InterBus system is transmitted through a copper cable using differential signal transmission according to RS 485. This means that a separate twisted-pair cable is required for the forward line and the return line.

When using optical fiber technology, the InterBus structure is the same as for copper technology because data is again transmitted by two fibers.
By using the twisted-pair cable or optical fiber, immunity for intervening variables is guaranteed and therefore, InterBus scores high on resistance to intervening variables.

Interbus uses the CRC16 method for detecting for possible errors. So when the receiver receives corrupted data, it can be detected.
There is also another mechanism that detects errors, namely during transmission pauses when no data is sent by the master, the operating data flow is filled by status telegrams. If there is a break in transmission of more than 25 ms, this is interpreted by all the devices as a system interrupt.
Therefore, the conclusion is that InterBus scores high at error detection.

Availability

When the network is considered, it can be seen that there is a ring topology, which is build through point-to-point connections from device to device. Each device node has two connectors, one that receives data, and one that passes data on to the next device, so in spite of the ring topology there is a way to communicate with the devices. There are no redundant paths to the devices. Another disadvantage is that one failed connection disables the entire network, so that the other devices cannot be reached as well. Therefore, InterBus scores middle at availability.

Manageability

Because of the unusual network topology, InterBus has the advantage that a master can configure itself because of the ring topology and without intervention from the user. This means that the only topology that Interbus supports is the ring topology and this is not very flexible. Therefore Interbus scores middle at manageability.

Ability to handle large amounts of data

Analyzing the ability to handle large amounts of data it is shown that InterBus can handle 500 Kbit/ sec. So it can be concluded that InterBus scores low at the ability to handle large amounts of data.

5.4.7 Ethernet

Criteria	*Low*	*Middle*	*High*
Response Time	✓		
Robustness	✓		
Availability		✓	
Manageability			✓
Ability to handle large amounts of data			✓

Response Time

This item encompasses one of the reasons why Ethernet is not used at the lower level of the automation industry. Originally, Ethernet was used in combination with bus topology. There are no masters in this topology and when someone has to send data it has to listen if someone else is sending. It is possible that someone else is doing the same thing at the same time, so collisions are possible.
This could be avoided by using, for example, a star topology where the devices are connected by a central switch or bridge.
The load of the network is also very difficult to manage while everyone has the same rights to send and there is no master to poll the devices.

Because it is difficult to manage the load of the network it is impossible to predict when a response will come. When the load is high it takes some time to communicate with a device, when the load is low the response time is short. Here it becomes clear that the response time depends on the load of the system and therefore it is difficult to use this protocol in a real-time environment.

So, originally Ethernet scores low at response time, because it is difficult to predict when the data will be delivered and how long it will take.

Robustness

Until now Ethernet has been used in Enterprise/ Office environments and characteristics of these environments are that there are not a lot of intervening variables. And even when an error occurs the data has to be resend.
In the automation communication, a lot of intervening variables are present, and the communication has to be real-time, so it is not desirable that these intervening variables have a great influence.
This could be reached by using the right cables. Originally these cables were not available, but nowadays these kinds of cables that can be used in environments with a lot of intervening variables become more and more available. Therefore, Ethernet scores middle at resistance to intervening variables.

TCP/IP could ensure reliable data delivery in an Ethernet environment, but this is at a higher layer of the communication and does not belong to Ethernet itself. TCP/IP calculates a checksum about the data and this all will be send to the receiver. The receiver also calculates a checksum and compares this with the received checksum. In this way, it can detect whether the data is corrupted or not. This makes that Ethernet scores low on error detection.

It was made clear that by using the right topology Ethernet became more reliable and that by using the right cables the resistance to intervening variables can be maximized. The network than becomes quite robust, but originally Ethernet scored low at robustness.

Availability

Originally, Ethernet was developed with a bus structure and one failed connection disabled the entire network, which is a great disadvantage. Over time, other topologies were developed, like star and ring topologies.
The first topology connects the device by a central switch or bridge, so when a connection to a device fails this device cannot be reached. But this has no influence on the accessibility of the other devices. The only disadvantage of this topology is that it does not have redundant paths.
The second topology builds some redundancy into the network, making it possible to reach a device in two ways. But for this to work, (R)STP has to be introduced. Now the redundant path can be used when a connection fails, which makes the network more available. The devices are connected using switches or bridges, which avoid collisions.

So it is clear that Ethernet can be made more available by using the right topology. Originally, Ethernet was developed with a bus topology and in this way Ethernet scores low at availability, but by using the other topologies Ethernet scores middle at availability.

Manageability

Above it was shown that ARP is used to detect addresses of the different devices. By using this protocol, the user becomes more flexible, because he does not have to configure the network before it can start.
This protocol requires the use of switches or bridges and these devices can store the addresses of the devices so they know how to route data to a device and therefore Ethernet scores high at access mechanism and thus at manageability.

Ability to handle large amounts of data

Ethernet was developed for an Enterprise/ Office environment, which is known as an environment where large amounts of data are transmitted. Ethernet started with a data transmission speed of 10 Mbit/ sec, but nowadays the speed is

increased to 100Mb/sec and even to 1 Gigabit/ sec. So it can be concluded that Ethernet is able to handle large amounts of data.

5.5 New Trend

Through new developments in the (automation) industry the devices become more and more intelligent; which means that they can deliver more and more information about their status. This information can be transmitted through the communication channel and this increases the load of the network.
Furthermore, the software in the devices, which takes care of more intelligence, has to be regularly updated or re-configured. Doing this, it would be useful to be able to do this at an 'arbitrary' place, instead of having to do this at each separate device. This only increases the load of the network more.
Therefore through a more intensive network load the manageability of the network becomes a more and more interesting point. With the fieldbuses like PROFIBUS and InterBus the data transmission speed and the flexibility of the network is *limited* and this could be a problem in the future when the load will increase. Therefore Industrial Ethernet becomes an option, which can handle large amounts of data.

Industrial Automation (IA) networks have historically been separate, both physically and conceptually, from Office/ Enterprise networks. This separation made sense, because of the widely divergent expectations and performance requirements for each type of network.

The questions are now whether Ethernet could be used at the lower levels, which problems would arise with this and whether these problems could be solved at this moment.

5.5.1 Industrial Ethernet

A key problem in Ethernet networks stems from their use of "shared access". Early networks, consisting of a coaxial cable "backbone" into which various devices were plugged, provided each device with equal, unlimited access to the network; therefore, a problem with any connection might disrupt the entire network. Moreover, in early shared Ethernet networks, all attached devices formed a single "collision domain;" traffic generated anywhere in the network could collide with messages sent from anywhere else. Therefore, network behavior under heavy loads was non-linear as traffic increased incrementally, the time for a message to reach its destination could increase exponentially. In addition, because of the lack of prioritization in early Ethernet, non-essential traffic (routine back-ups, for example) could degrade the capability of important messages to get through.

A series of Ethernet developments has provided designers with the tools necessary to overcome these problems.

The first development was the introduction of the "star topology". Star topology used hubs, linked directly to the backbone, to distribute messages to attached devices. In addition to enabling more flexible network design, star topology made networks more robust. A problem with a connection between a hub and one of its attached devices no longer brought down the entire network. However, since all devices still had full access to the network, they still formed a single collision domain.

From Switches to Topology

The use of switches has dramatically expanded the range of possible network topology and connection options. Switch ports can be connected directly to devices, or alternatively, to hubs or other switches. A special type of single-channel switch, known as a **bridge**, is often used to connect Office/Enterprise and IA networks; the network on each side of a bridge "sees" only the messages intended for its own nodes.

Switches have also made the use of decentralized network architectures far more feasible and attractive. Instead of using a mainframe to run a centralized network, each function within the enterprise (for example, process monitoring during a manufacturing run) can be managed by a smaller computer located on-site, isolating any potential problems and simplifying network installation, modification, and management. Only high-priority and/or selected messages are passed from the smaller computer through its switch connection to the broader network.

The newest generation of "smart" switches incorporates a protocol known as GMRP (for GARP Multicast Registration Protocol) that allows a switch and its associated group of nodes to "subscribe" to certain messages based on defined attributes. GMRP enables switches to send multicast frames only to members of a multicast group, instead of flooding the entire network with multicasts, helping to improve the performance of the network.

Changes in Ethernet so it becomes industrial

The development of Ethernet switches was a critical step that sidelined concerns about Ethernet determinacy and network-wide availability issues. Fueled by Ethernet's increased attractiveness and viability for IA network applications, the past decade has witnessed a series of developments that have transformed it from a potential network solution to an essential network solution.

Availability

IA networks must be **highly** available, they must continue to operate in the face of environmental assaults; accidental or deliberate network disruptions, and equipment failures. Network downtime can be enormously expensive, not only directly (for example, losing production capability), but also indirectly (for example, having pollution controls fail). Availability is the result of several factors, including manageability and supportability.

Availability is usually provided through equipment and path redundancy, coupled with firmware in the network devices that instructs the network to switch to alternate paths upon specific failures. Additionally, to provide availability even during loss of a power supply, higher quality products support dual power feeds, to ensure that no single power feed loss can bring the network down. In the ideal world, it would be possible to protect against, and recover from, any conceivable problem rapidly enough to prevent network failure. However, each increase in a network 's level of availability incurs increased expense.

A key problem for mission-critical Ethernet networks is the ability to recover from network path failure. Traditional commercial grade network hardware responds to a network path or device failure by switching to an alternate path, using the "Spanning Tree" standard in IEEE 802.3d. Spanning tree is a resiliency scheme that relies on a matrix of wiring between network switches that offers more than one wire path between all network infrastructure devices. Wiring a spanning tree "web" between network devices can increase the cost of physical network installation. In addition, since the definition of "high-speed" recovery is far different for Office/ Enterprise networks than it is for control networks, spanning tree was not designed to accommodate the time-limited recovery needed for many real-time process control applications. Spanning Tree may require as much as 60 seconds to identify and bypass the point of failure; during that time all attached devices are isolated.

A number of solutions have been developed to make recovery times acceptable for industrial control applications; however, because a faster response method has not been standardized for Ethernet, these solutions are proprietary and manufacturer specific. One example is "redundant ring" technology, which provides recovery from path failure within 500 ms. Whereas spanning tree requires a matrix wiring scheme between network devices, redundant ring topology requires only that all network switches are wired in a ring architecture.

Redundant ring topology has a couple of benefits, one obvious, and the other not as apparent. The obvious benefit is the speed with which redundant ring networks can respond to a network failure. When a fiber is broken, or a network switch fails, a Ring network will self-heal in less than 500ms, even in a network incorporating up to 50 switches.

The secondary benefit of a ring topology is the reduced cost of network cabling. When wiring Office/ Enterprise networks, fiber optic cable needs to be "home-runned" from a central switch to the switches or transceivers located in the field. The use of spanning tree redundancy requires even more cable. By not requiring every fiber run to travel back to the control room, wiring a ring topology requires far less fiber optic cabling, saving material and labor costs. Industrial users must weigh the combined cost of network components, cabling cost and installation labor before deciding whether they are truly saving money by using commercial network hardware in spanning tree 's cabling intensive network architecture.

Prioritization

The enormous bandwidth of Ethernet has become an invitation to expand not only the amount, but also the type of information transmitted through Ethernet networks. A wide (and steadily expanding) range of products is now being offered that permit video, voice, and other high-bandwidth data to be easily networked. It would seem to make little sense to lay multiple overlapping cabling systems for control data, surveillance video data and voice data, when all can be transmitted over a single robust, high-availability Ethernet network.
In most control networks, however, not all data on the network has the same importance to the user. For example, failing of the system is always more important than having an e-mail delivered immediately. To realize the cost advantages of using a single high-bandwidth network for multiple data streams **without** compromising the responsiveness required for the control network, one must ensure that the signal that the system is down isn't being slowed by the e-mail being transmitted at the same time.
The challenge of sorting out high priority traffic from low priority traffic is addressed by IEEE standards 802.1q and 802.1p,which provide for the addition of a 4-byte TAG field to the traditional Ethernet frame. Within the TAG field, 3 bits are reserved to allow the establishment of 8 levels of User Priority, or Quality of Service (QoS) (from Priority 0 to Priority 7). If two messages arrive at an Ethernet switch at or near the same time, the higher-priority message can "jump the queue" and be transmitted ahead of the lower- priority message(s).

In many cases, prioritization has been incorporated, by manufacturers of network devices, into the devices themselves; each Ethernet frame transmitted by the device carries the required level of prioritization. However, Ethernet-enabled devices in an existing control system may not support prioritization, and some end-device manufacturers don 't yet incorporate prioritization support. Therefore, **in-bound prioritization tagging** has been developed. In-bound prioritization tagging permits the Ethernet switch to be configured to add a prioritization tag to all Ethernet frames traffic in-bound on a port from a specific attached device. Moreover, this tag remains on the Ethernet frame, so that as the frame travels throughout the network, it is always handled at the appropriate priority level.

It is possible for switches to arbitrate the handling of message frames among the 8 priority levels using several different algorithms. "Raw prioritization" ensures that higher priority packets are always handled first, forcing lower priority packets to wait until all higher priority packets have been transmitted. However, in certain situations, this method may allow high priority packets to unfairly dominate network resources; therefore, the integrator and/or system administrator must define the priority level for each device with great care. In contrast, "Fair balancing" prioritization permits the allocation of switch processor time-share to each priority level, ensuring that no single priority stream monopolizes network resources, and guaranteeing some access to network resources for all traffic. The use of prioritization provides confidence for Ethernet network users that data from critical systems will always have an unencumbered path through the network.

Security

The use of a common Ethernet network to connect PLCs, Control PCs, administrative PCs and other disparate systems clearly provides a great economic benefit. However, with so many systems tied to a common network, concerns can arise regarding user security, as well as unwanted interaction between the different systems connected to the network. To address the need to create virtual "walls" within the common physical network, forward-looking Industrial Ethernet manufacturers incorporate virtual LAN (VLAN) support into their network devices.

A 12-bit segment within the same 4-byte TAG field used to define QoS (prioritization) can be used to assign message frames to a specific Virtual Local Area Network, or VLAN. VLANs allow a single physical network to be split into two or more virtual networks; by tagging a device as part of a specific VLAN; it will receive only messages specifically addressed to that virtual network. In practice, VLANs can make a single physical network act like many separate networks, with devices placed assigned to a specific VLAN only able to "see" other devices within the same VLAN.

In some cases, having completely distinct VLANs may create a challenge for sharing information or network assets such as printers. For example, two VLANs might be created, with all control devices in VLAN A, and all "front office" PCs in VLAN B. However, it may be desirable to provide access to an data historian and a printer from plant PLCs and PCs **and** front office PCs. Overlapping VLANs allow shared resources to be a member of both VLAN A and VLAN B.

VLANs can also be used to isolate any network device on a network that has been connected to the Internet via a broadband connection. For example, selected PCs connected to the network could be placed in a shared VLAN with a broadband connection to allow them Internet access, while PCs and control

equipment are rendered "invisible" to the broadband Internet connection (and the possibility of unwanted access).

Manageability

Modern control systems comprise three essential elements: the field controllers and devices, the network itself, and the interface. As the size of an Ethernet network grows, and the number of "plugged in" devices increases, it becomes increasingly important to be able to monitor and verify the health of the network in real-time. Simple Network Management Protocol (SNMP) was developed to allow a network administrator to be able to configure and monitor, from a central management station, the distributed hubs, switches, PCs and other network infrastructure equipment while they operate. The SNMP management station displays critical data about the health of the Ethernet system, such as port link status, network errors, and network traffic information.

The real-time data regarding network health is gathered by polling information from SNMP agents in the network infrastructure devices. If a problem occurs, the network administrator is able to see the it from within a network management software package, before the symptoms of the problem become evident, slowing the system or preventing it from performing correctly. Just as the temperature gauge or the fuel level gauge on a car 's dashboard help the operator avoid problems while operating a vehicle, SNMP can help an Ethernet network administrator head off problems in the network.
Not all network devices are manufactured with embedded SNMP agents. Full-range industrial Ethernet equipment manufacturers offer both SNMP-manageable versions of their devices and simple unmanaged devices. In addition, since many industrial users already have powerful data acquisition and control software packages that offer control of process equipment connected via the industrial Ethernet network, SNMP-OPC "gateway" software packages are available. These packages allow the existing process system to become the Ethernet network management interface. This eliminates the need to purchase and operate a distinct network management "dashboard;" the user can simply integrate network monitoring into the existing control system.

Physical

The final piece in providing Ethernet with true IA functionality is physical: ensuring that industrial Ethernet components and networks are capable of withstanding the environmental assaults common in manufacturing and other real-time applications.
Manufacturers now offer "hardened" Ethernet components, capable of providing full-spec performance in the face of shock, vibration, and extended temperature ranges (typically up to 60° C). In addition, components designed for the shop floor may be DIN-rail-mountable and utilize 24V power.

To fully exploit the advantages of distributed networks, it may be necessary to situate Ethernet switches in areas characterized by hazardous environmental conditions, or high levels of electromagnetic interference. For placement in such areas, Industrial Ethernet switches with the appropriate FM and UL approvals are available.

Electrical noise and interference are often problems for traditional office-grade unshielded Ethernet cabling in harsh environments. Industrial users have two options for use in network cabling. For short copper cabling runs in areas of moderate electrical interference, STP (Shielded twisted pair) copper cabling can provide greater noise immunity than UTP (Unshielded Twisted Pair). STP cabling has identical pin outs to UTP cabling, except that a shield is incorporated into the cable, and in turn fused to a metal sleeve at the connector. Higher-quality industrial Ethernet devices provide the necessary metallic shroud at their RJ45 connectors to support the STP cable shield connection, in addition to supporting UTP cabling.

For higher interference areas, or longer runs, fiber optic cable is the best choice. Fiber optic cable is immune to electrical interference, and its cost has steadily declined. Advances in manufacturing processes have led to the availability of industrially jacketed and strengthened fiber optic cabling systems that can withstand the environmental and physical abuse common in industrial installations.

Another potential solution, especially where running new cable or fiber is problematic, is wireless Ethernet. Wireless-enabled devices communicate with each other using a hub-like access center, which is usually integrated physically into the network.

5.5.2 Evaluation of Industrial Ethernet

Below, Industrial Ethernet will be evaluated on the same criteria that were used on the other protocols.

Criteria	*Low*	*Middle*	*High*
Response Time		✓	
Robustness		✓	
Availability			✓
Manageability			✓
Ability to handle large amounts of data			✓

Response Time

When two senders send data at the same time on the same channel, a collision occurs. Both senders detect the collision and wait for a random time. They than will try to send the data again. So, collisions influence the data delivery time in a powerful way and they make the prediction of the data delivery time difficult. Therefore, collisions have to be prevented, which is done by using switches. So, now it becomes a network with point-to-point connections and in such a network, collisions will be prevented, which makes the network more deterministic.

Another factor that can influence determinism is priorization. When no priorization is used, all the data in an Ethernet network has the same priority. It is possible that time-critical data cannot be sent because there is data with the same priority in the queue, but which is not time-critical. This data will be send before the time-critical data will be sent. This is undesirable, because determinism is important for time-critical data.
This could be solved by priorization. Using this, time-critical data is given a higher priority than non-critical data. The data with the highest priority will be send first over the network, which makes the waiting time for the delivery of time-critical data smaller. The data delivery time can be better predicted in this way, making the network more deterministic for time-critical data.

Above was mentioned that VLANs can be used to increase the security of the network. However, VLANs can also be used to increase the determinism of a network. VLANs can be used to isolate any network device on a network so not all of the data will pass the network. This means that the load of the network will be more manageable, preventing an overload. This makes the network data more deterministic.

Robustness

In the section on the physical part of Industrial Ethernet it was mentioned that manufacturers now offer "hardened" Ethernet components, capable of providing full-spec performance in the face of shock, vibration, and extended temperature ranges. This is important because there are a lot of intervening variables in the automation communication, especially at the lower levels (device and control level). By using the right components the resistance to intervening variables will be maximized.

Originally, Ethernet ensures reliable data delivery by the TCP/IP protocol. This protocol calculates a checksum of the data and sends this with the data to the receiver, but belongs to a higher layer of the communication. The receiver also calculates a checksum of the data and compares this with the received checksum and checks whether the data is corrupted or not. Therefore, reliable data delivery is ensured at a higher layer.

The robustness of the Industrial Ethernet network is increased, because, as earlier mentioned the manufactures offer "hardened" Ethernet components, causing the intervening variables to have less influence on the performance. When nonetheless a failure occurs, there is a redundant path to the device, so the device can still be reached. The conclusion can be made therefore, that Industrial Ethernet becomes much more robust.

Availability

A key problem for time-critical Ethernet networks is the ability to recover from network path failure. Traditional commercial grade network hardware responds to a network path or device failure by switching to an alternate path, using the "Spanning Tree" standard. This is necessary, because Ethernet with redundant paths does not function correctly as we have seen in Ethernet with CSMA/CD. But the disadvantage of a Spanning Tree is, that it requires as much as 60 seconds to identify and bypass the point of failure; during which time all attached devices are isolated.

A solution for a faster response is the "redundant ring" technology and the use of the Rapid Spanning Tree Protocol, which provides recovery from path failure within 500 milliseconds. Using this topology and the Spanning Tree, Industrial Ethernet becomes more available.

Manageability

Ethernet uses the ARP protocol to discover the network addresses. Nothing has changed in the way Industrial Ethernet discovers the addresses, therefore Industrial Ethernet scores equally high on access mechanism as Ethernet. Ethernet scored high on this criterion, therefore, Industrial Ethernet scores high too.

Ability to handle large amounts of data

Originally, Ethernet has a good ability to handle large amounts of data. Nothing has changed in Industrial Ethernet which influences this characteristic, so the conclusion remains that Industrial Ethernet has a good ability to handle large amounts of data.

5.6 Suitability of the different protocols at the different levels

Above the different protocols have been analyzed and how they score at the different criteria. Now the conclusions out of the available information can be composed.

In the table below, the different protocols will be mentioned and the table will show whether or not they can be applied at the different levels of communication. In the section called 'Implementation area' will be clarified why they can be applied or not.

	Device Level	*Control Level*	*Information Level*
RS 232	□	-	-
RS 485	-	-	-
USB	-	-	□
PROFIBUS	+	+	-
InterBus	+	□	-
AS-I	+	-	-
Ethernet	-	-	+
Industrial Ethernet	□	□	+

Table 5-1

- : Cannot be applied at this level
□ : Can be applied at this level with some restrictions
+ : Can be applied at this level without restrictions

Implementation area

RS 232

In the table, one can see that RS 232 could be applied at the device level, but the note here is that the complexity of the network cannot be too high. This, because RS 232 is a direct communication channel, with a cable that has to be routed from each device to which you are communicating. Therefore, the network is not so flexible and is expensive to route (this because it has to be routed to each device).
An advantage is that collisions are excluded, and have therefore no influence on the speed of the network. When the devices are fast enough this kind of protocol can be used in real-time environments and in normal environments, but with the restriction that the network has a low complexity.
Therefore it can be used at the device level with a low complexity, but not at the control and information level, because these levels are too complex most of the time.

RS 485

RS 485 specifies the electromechanical structure for communication and not the data exchange. Because it is a multi-point bus protocol collisions are possible and it is not defined when a device has the right to send.

Collisions have influence on the response time of the communication and therefore RS 485 is difficult to apply in a real-time environment. This can be solved, but this has to be done at a higher layer of the communication. Therefore, it is difficult to apply RS 485 at the device level and control level, because a constant and good response time is required for real-time communication.

RS 485 does not guarantee a highly available network. This has to be ensured at a higher communication layer or by the network designer himself. So when a highly available network is required, using only RS 485 is not enough. Therefore, RS 485 is not very suitable for the communication at the information level.

USB

The disadvantages, which were found in the USB protocol, were that no specification was found on how to maximize the resistance to intervening variables and the small maximum distance of 5 meter per cable segment.
In industrial communication there are a lot of intervening variables and the limited distances between the devices could cause problems. This makes USB not a very suitable protocol for the device and control level.

At the information level, however, USB can be used. The disadvantages remain that there are no redundant paths to a device when a connection fails, but this does not influence all of the devices, which is an advantage. Another disadvantage is the limited distance between the devices.
So when the distances in the network are not too large and no redundant paths are necessary USB is well suitable for communication at this level.

PROFIBUS

At the moment, PROFIBUS can be used at the device level and control level, because it satisfies the requirements for the response time (deterministic and time-critical), which is an important requirement for the automation communication at these levels.
The only disadvantage found in the protocol is that when there is a bus connection failure, the whole bus cannot be reached. This means that none of the devices can be reached. This could be partially solved by splitting up the network in a number of different parts, so that just one part of the bus cannot be reached during a failure. Till now, this is accepted in the automation industry.

At the information level, a more highly available network is desired, one that does not completely fail when there is a bus connection failure. Also, it is desirable to have a more flexible network (detecting dynamically the devices), which is not possible. So, these are the two disadvantages of PROFIBUS at the information level.

AS-I

The disadvantage of AS-I is that it cannot handle large amounts of data. This makes this communication protocol unsuitable at both the information level and the control level. At these levels there is quite a lot of data exchange, and is the 167 kBit/s not enough.

For the communication at the device level, AS-I can well be used. It satisfies all of the requirements for communication at this level, like being deterministic, time-critical and robustness. The only disadvantage of AS-I is that it cannot handle large amounts of data, but this should not be a problem at the device level, because there is not so much data exchange here.

InterBus

Inspecting how InterBus scores at the different criteria it can be seen that InterBus can be used at both device level and control level, because it satisfies the requirements for determinism, being time critical and robustness.
The disadvantage stays that when one bus connection fails (point-to-point connections, which form a ring) the whole network is disabled. This is the same as with PROFIBUS, but PROFIBUS has the advantage of having the possibility to split the network, therefore stopping the whole network from being disabled. Another disadvantage is that InterBus cannot handle large amounts of data, which becomes a requirement more and more in the future, especially at the control level.

The disadvantage mentioned above could be also a problem for the availability of a network, because the whole network is not reachable when there is a connection failure. Another disadvantage is that InterBus scores low at the ability to handle large amounts of data, so when the network load increases this could become a problem. These are two disadvantages that make InterBus not very suitable at the information level.

Ethernet

Earlier we mentioned that at Ethernet everyone has the same right to send data and that there is no master. So it is possible that a device is sending data while another device also has to send data. When this happens, one of the devices has to wait a random time.

It is also possible that two or more devices want to start sending at the same time and that there is a collision.
Therefore it is difficult to predict the response time and this makes Ethernet not very suitable for real-time communication, which is a demand at device and control level in automation communication.

Originally, Ethernet was developed for an Enterprise/ Office environment where the main task is to exchange data and therefore can conclude that Ethernet can handle large amounts of data.
Selecting the right topology could increase the availability of a network. So a ring topology is more available then a bus topology, because in a ring topology there is some redundancy. Therefore can be said that the designer can make the network highly available, which makes Ethernet suitable for the information level at automation communication.

Industrial Ethernet

By using switches, priorities and VLANs normal Ethernet could be made more deterministic and time-critical. Where Ethernet scores low at determinism and being time-critical, Industrial Ethernet scores middle at determinism and being time-critical. Restricting the network load could increase these scores. When the network load is not too high than you can say that the score becomes higher. Therefore you can say that Industrial Ethernet can be used at real-time communication if the network load is restricted.
The robustness and availability of the network could be increased by using a ring topology and "hardened" components. With a ring topology there is a redundant path to a device, so the availability is increased and by using "hardened components" the network is better resistant to intervening variables. So Industrial Ethernet is better suitable for the automation communication.
Therefore, the conclusion can be made that Industrial Ethernet can be used at the lower levels of automation communication with the restriction of the network load.

Ethernet can already be applied at the information level. The availability of the network is just better using Industrial Ethernet by redundancy and using "hardened" components. So, Industrial Ethernet is just as normal Ethernet suitable at the information level.

5.7 Conclusion

RS 232 is a good choice for a communication protocol, when it has to be implemented in a one on one connection, for example a PC-PLC connection, or in a very small network. It is able to handle the amounts of data that are common in these kinds of connections, and it is unnecessary to use a larger and more expensive protocol in these cases.

RS 485 has as advantage over RS 232 that it is possible to connect more than two devices at a time. It can well be used when the designer of a network is able to make his own protections against collisions, errors and intervening variables. It is difficult to use this protocol in most cases however, because there are no facilities at all with RS 485.

USB has as advantage over RS 485 that it has a better response time, better robustness, is better manageable and more highly available. But the distance between the devices can be no more than 5 meters, which is a serious disability when it has to be used in industrial automation settings. But when the network has to be implemented in one small room, for example a classroom, USB can be used using hubs.

PROFIBUS has as advantage over USB that it is able to bridge larger distances than 5 meters and it is a more robust protocol. Its disadvantage is that the speed with which data can be transmitted is lower than when using USB and only just as large as the speed of RS 485. Its speed is faster however than the speed of RS 232. It also is an expensive protocol.

AS-I does not score better than PROFIBUS at any of the criteria in the tables of the previous paragraphs. Major disadvantages are that it is not as well manageable as PROFIBUS and it is unable to handle large amounts of data. The robustness of this protocol is just as good as the robustness of PROFIBUS and they score just as high on the criterion of response time and availability.

InterBus has as an advantage over AS-I that it is better manageable, because it has a ring topology, so that addressing is automatic. The scores on the other criteria are the same as AS-I. A disadvantage of InterBus is that it does not score as well on the criterion of being able to handle large amounts of data as PROFIBUS.

Ethernet scores less than PROFIBUS, AS-I and InterBus on response time and robustness. Its scores on availability, however, are the same as the ones of these other three protocols, when the right topology (a tree- or ring topology) is chosen. Its advantages are that it is better manageable than PROFIBUS, AS-I and InterBus and that it is better able than these three to handle large amounts of data. These are major advantages because these criteria are becoming

increasingly important, because the devices are going to generate more and more information in the future.

Industrial Ethernet scores higher on the criteria of robustness, response time and availability than regular Ethernet, but it still performs worse than PROFIBUS, AS-I and InterBus. On the other criteria it scores just as high as Ethernet. The scores of Industrial Ethernet on all of the criteria are sufficient to support Material Handling systems.

With this all in mind the conclusion can be made that fieldbuses become a problem because of an increasing load of the network. Therefore an alternative has to be found that can handle the increasing data stream. Industrial Ethernet is a good option than, especially owing to the new developments, making it more deterministic and time-critical.
Therefore it is interesting for CSi to follow the new developments of the more intelligent devices and the trend of Industrial Ethernet at the lower levels of the automation communication.

Chapter 6

Conclusions and recommendations

At this point, some conclusions and suggestions for the future will be given.

The first following paragraph will do this for the research on availability and reliability of the Transport conveyor. In the paragraph thereafter, the communication of Material Handling Systems will be taken into account.

6.1 Conclusions on Availability and Reliability of the Transport conveyor

In the course of this project, a lot of new information about Material Handling Systems came up, especially about the Transport conveyor.

It was proven that VDI 3581 does not satisfy the expectations, because it is not able to calculate the availability of a system consisting of several Transport conveyors. The availability of more than one Transport conveyor can be better calculated when the start- en ending times of the failures of the devices are known.
Also, the view that the availability of a system can be calculated looking at the availability of the parts, of which it is constructed, as is the prevalent view of VDI 3581, is not tenable according to me, because this view fails to recognize the fact that there is considerable overlap between the failures of the conveyors.
The Markov model does take this overlap into account, but it was not until during the pilot project that the weakness of this model became clear; namely that the calculations of the reliability in this model resembled more the values of the availability in reality. This could be caused by the fact that the repair rates and failure rates are calculated in a time-related manner, which could be solved by using a functional counter, as was implemented in the pilot project. There are no available data on the outcome of this however.

A recommendation for the future is that a Markov model is a good way to express the availability and reliability, depending on the way in which the repair rate and failure rate are calculated.
A note here is that the time in which the failures are measured has to be large enough. When the time space is too narrow, the variance in the repair time could have a big influence on the result of the repair rate and accidence has not be excluded.

Another point for future research is to make more differentiations between the different errors. At this moment, it is not possible to make clear why an error occurs; because of the material, the operator, or the system. When this would be possible, one can express the Availability and the Reliability more accurately.

This point of improvement is especially of interest for CSi, because they are only responsible for the errors caused by the system.

6.2 Conclusions on the communication of Material Handling Systems.

The trend is that an increasing amount of devices become more and more intelligent; they can generate information about their status. For this change in intelligence, increasing amounts of software are implemented in the devices, which can be upgraded or reconfigured when needed.
Most of the time, the devices are connected through a network. Therefore it is desirable to be able to do this reconfiguring and upgrading from an 'arbitrary' place at the network, instead of having to do this on each device separately. When this reconfiguring and upgrading is done from an 'arbitrary' place on the network, the load of the network is increased. Here, the ability of the communication protocol to handle large amounts of data becomes more and more important. Because of this, several communication protocols are analyzed.

Up until now, PROFIBUS was used for the communication at the lower levels, but when analyzing it, it became clear that fieldbuses, like PROFIBUS, are able to handle the amounts of data that are common now, but when these amounts go up in the future, this could lead to problems.
Therefore, Industrial Ethernet becomes an option. Industrial Ethernet was no option in the past, because it was not deterministic and time critical enough. Nowadays, it become more and more deterministic and time-critical, especially for use in the Material Handling systems, because these systems are dealing with real-time communication, requiring a response time of seconds instead of milliseconds

Therefore, a strong recommendation for CSi is to follow the trend of Industrial Ethernet at the lower levels of the automation communication.

Appendix A

Literature at CSi

AVAILABILITY OF AUTOMATIC MATERIAL HANDLING SYSTEMS

In order to calculate availability and capacity it is necessary to consider each aspect in turn. This means that capacity must be ignored when considering availability, that is to say it is important to determine whether the system is capable of performing specified function not whether it can achieve specified throughput. Likewise, it is assumed that all capacity data applies to systems where availability is 100%.

This may appear to be an over simplification, but as a general rule capacity decreases considerably when the availability of a system falls. It is usually better to have simple, unambiguous rules and take their disadvantages into account rather than to have complex rules which can lead to misunderstandings.

DEFINITIONS

Availability:
The proportion of the total operating time for which a system is in working order.

Availability is expressed in percent and is 100% if no breakdowns occur.

To simplify calculations a relative availability of 1 is used instead of 100%, 98% availability thus being expressed as 0.98 etc.

Capacity:
The number of load units per hour, which the system can handle at 100% availability.

Working time:
Working time is the total time per measured period during which a plant is operational and can be in use for productional purposes.

Is the plant a full-continues production facility, than the working time per year is 365 days * 24 hours = 8760 hours.

Is the plant in operation during normal working days, i.e. closed during the weekends, in this case the working time is to be calculated as 365 * 5/7 * 24 hours = 6257 hours.

Breaks:
Breaks are to be understood as the total time per measured period within the working time during which the plant is shut down and not in use.

Examples for breaks can comprise lunch time, national holidays, etc.

Operating time:
Operating time is working time minus breaks.

Only such fixed breaks when the plant is shut down and not in use are counted.
Operating time is divided into available time and breakdown time.

Breakdown time:
Breakdown time covers all periods of interruption due to all *unplanned stoppages within operating time*, irrespective of whether they were caused by the customer, operating faults etc. or by the supplier.

Breakdown time is divided into that caused by the customer and by the supplier.
Breakdown time is always attributable to a component in the system and comprises:

* waiting time for specialist or service personnel
* troubleshooting time
* waiting time for spares
* repair time
* time for testing and restarting

Available time:
Available time includes *all planned stoppages within operating time* or such matters as maintenance, inventory, cleaning, machine adjustments, etc.

AVAILABILITY

In the case of a single machine, availability is easy to define and measure. It is also a simple matter to determine whether or not a stacker crane, a roller conveyor or an automatic AGV is operating. If these machines form part of a system, it does not necessarily mean that the entire system will be unserviceable if one part of it breaks down.
In practice, the availability of a system may still be 100% even when one or other of its components are not functioning. As an example consider an AGV system with 10 vehicles. If one vehicle breaks down it does not necessarily affect availability. There are still 9 vehicles availability to carry out the functions of the system. The capacity of the system is however reduced by 10%.
An exception to this rule is centralized computer control. If this is designed as a single-computer system without any facility for manual intervention, the entire system will come to a halt if it suffers from one vehicle breakdown.

SYSTEMS DESIGN AND CONTROL STRATEGIES

System design and control strategies have a decisive affection on the availability of a handling system. Consider as an example a high bay warehouse with four stacker cranes. If each crane is restricted to its own aisle, and each item is stored only in one aisle, then the breakdown of one crane results in serious consequences. One quarter of the items then becomes inaccessible. On the other hand, if each item is spread across more than one crane aisle, the breakdown of one or even two cranes will as a general rule have virtually no effect on the ability to reach all the items. The availability is still 100% but, as in the AGV system above, the capacity is reduced by 25% and 50% respectively.

If, in a second instance, four cranes with the aid of transfer car serve a larger number of crane aisles, availability is only marginally affected if one crane is taken out of service.

The question subsequently arises as to whether a project requires a study to determine how each constituent part will affect availability in the event of a breakdown. A temporary fall in the capacity of a system is generally acceptable, but

if the installation is not available to handle goods, then considerable problems usually occur.

EFFECT OF BREAKDOWN

The demand for availability is often set at a figure of between 90% and 99%, without any more detailed analysis of what this requirement involves. Instead, the customer should discuss and analyze with his supplier, the effect breakdowns will have on components, sub-systems and their control systems. The system must be designed to ensure the effect of breakdowns on availability is kept to a minimum. In particular, the control system should be split into sub-systems, since any breakdown here usually has a major effect on availability. It is usually a expensive, and as often a rule, impractical way of buying high availability by rasing operational reliability and reducing repair time for each component in a system. The correct solution is to choose a system with alternative transport routes. As will be shown later, a dramatic increase in availability will be obtained if parallel transport routes are present. The same applies to (computer) control systems. If instead of a single computer system where the computer has an availability of 98.5% a dual computer system is selected where each computer has an availability of 98.5%, than the total availability of the computer system will be 99.98%.

THE AVAILABILITY REQUIREMENT

An important factor in determining this requirement is that the definition of availability is not time-related. If an availability of say 98% is required, without defining the period of time, this requirement is generally understood as referring to a longer period, a year or so. 98% availability over one year is equivalent to a total downtime of 4.5 days. In the worst case an installation can therefore be out of order for 4.5 days in a row and still comply with this requirement.
For most installations such a long standstill is unacceptable. The availability requirement must therefore be specified with a maximum permissible downtime. This can be achieved by requiring a minimum permissible availability per month. A requirement of 95% availability per month allows a maximum downtime of one day.
For systems where many components (i.e. lifts, conveyors, transfers, etc) can perform the same handling, e.g. the above AGV system with 10 vehicles, one or

more units may be out of order for long periods without the availability being appreciably affected. A secondary requirement, that no component can have an availability of less than 80 %, will cover this aspect.
The availability performance may thus consist of one primary and two secondary requirements:
97% per annum for the entire system (primary requirement) but at least
95% per month (secondary requirement)
80% per month for each component (secondary requirement).

With present-day equipment no major problems are presented in reaching an availability of 97-98%, taken over a period of time as long as one year. By contrast, this requirement may prove more troublesome to meet in the short term. All equipment can break down and this may occur at any time. If a downtime of an entire day during a one-year period is completely unacceptable, then all important components and subsystems in the installation must contain alternative transport routes.

Rare faults

Rare faults (computer breakdowns etc.) leading to lengthy standstills and which statistically occur only once during a tenable period, may happen at the same time during one week, never to recur. Protection against this kind of breakdown at reasonable economic cost is not possible. It is especially problematic if such a breakdown should occur during an availability test. In an otherwise well-functioning plant such a chance occurrence should normally not be made the responsibility of the supplier.

RUNNING-IN TIME

Small transport systems can often attain the availability specified as soon as the installation is handed over. Medium and large systems generally require a running-in period during normal operation. This is to allow adjustment of the installation and its computer systems, and to also familiarize the operating and maintenance personnel.

The running-in time for medium systems is 6 months; for large ones, one year. A common error on the part of the purchase is to believe that system will give trouble free operation from hand-over. This not so.

It should be remembered that the contractual availability applies only to breakdowns caused by the supplier. The major source of such interruptions is however mostly due to the customer's own operating and maintenance personnel who have to work to new routines and product failures (i.e. broken pallets, badly taped cartons, etc.). It takes a considerable time to master all the functions of the system. Inexperienced staff make mistakes easily, require a long time to find the cause, decide on remedial action, and finally to deal with the fault.

MEASUREMENTS OF AVAILABILITY

Measuring the availability at hand-over is normally undertaken during a test period lasting 5-10 working days. The test days are jointly chosen by the customer and the supplier, and should be as representative as possible to the conditions specified in the contract.

During the test days, a record of operating and breakdown time is kept, and all remarks should be initialed by both partners.

In installations with a running-in time, the customer will continue to fill in operating records as a basis for calculating the availability over the agreed period, normally after six months or one year. An example of availability tests for a large-scale handling system is shown below.

During the hand-over test:
90% for the entire system, but at least 80% for each component.

After three months:
Measurement period: months 1, 2 and 3.
95% average for entire system, but at least
90% per month
85% per month for each component

After a half year:

Measurement period: months 4, 5 and 6.

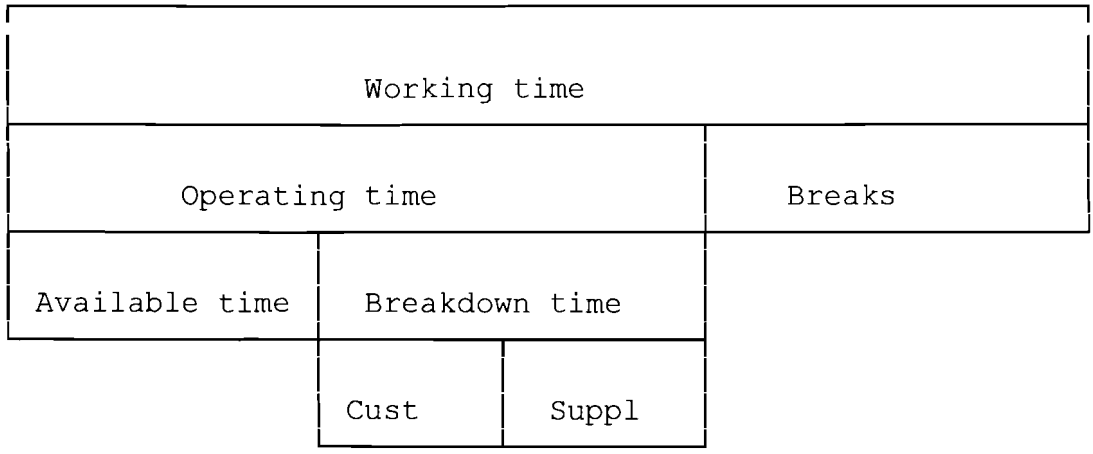
99% average for the entire system, but at least

98% per month

95% for each component

BREAKDOWN TIME

The following serves as a basis in calculating availability.



A clear definition of each item is given in the chapter "Definitions".

Example

A high bay warehouse has normal working hours between 07.00 and 16.00 and lunch between 12.00 and 13.00. Between 07.00 and 07.30 operation is shut down for planned maintenance. Between 11.30 and 12.00 there is an operational stoppage due to an operating fault and between 15.00 and 18.30 an operational stoppage due to a computer fault.

The allocation of time is:

Working time	07.00 - 16.00	= 9 hours
Pauses	12.00 - 13.00	= 1 hour
Operating time	9 hours - 1 hours	= 8 hours
Breakdown time cust.	11.30 - 12.00	= 0.5 hours
Breakdown time suppl.	15.00 - 16.00	= 1 hour
Total breakdown time	0.5 hour + 1 hour	= 1.5 hours
Available time	8 hours - 1.5 hours	= 6.5 hours

CALCULATION OF AVAILABILITY

The availability of component irrespective of who caused the breakdown is calculated as follows:

$$\frac{\text{Operating time - Breakdown time}}{\text{Operating time}}$$

In calculating the availability promised in a contract account should be taken only of breakdown for which the supplier is responsible. It is thus necessary to make a clear distinction between those caused by the supplier, and those caused by the customer.
The availability of a component with regard to breakdowns caused by the supplier is calculated as follows:

$$\frac{\text{Operating time - Breakdown time suppl.}}{\text{Operating time}}$$

Using the figures from the high bay warehouse example the total availability on that working day is:

$$= \frac{8 - 1.5}{8} = 0.81 = 81\%$$

The availability with regard solely to faults caused by the supplier:

$$= \frac{8 - 1}{8} = 0.875 = 87.5\%$$

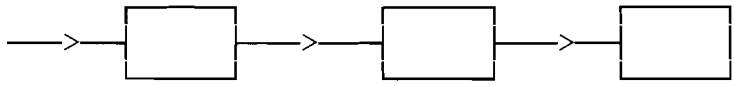
The availability with regard solely to faults caused by the customer:

$$= \frac{8 - 0.5}{8} = 0.935 = 93.5\%$$

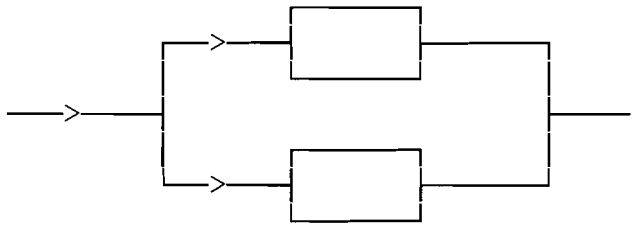
SYSTEM AVAILABILITY

There are two fundamental ways of connecting components in a pallet transport system:

* Series

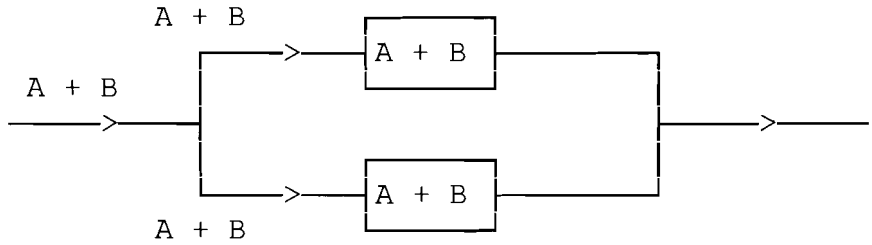


* Parallel

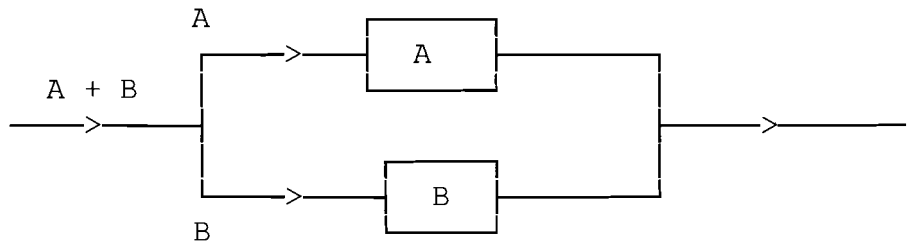


There are also two alternatives for the parallel connection:

* Transport can take place selectively across the parallel components



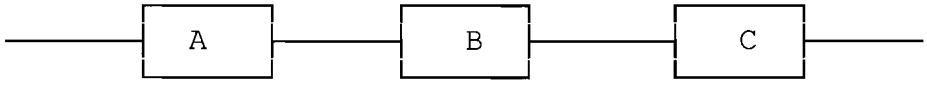
Transport of certain goods must take place across predetermined paths of the parallel system.



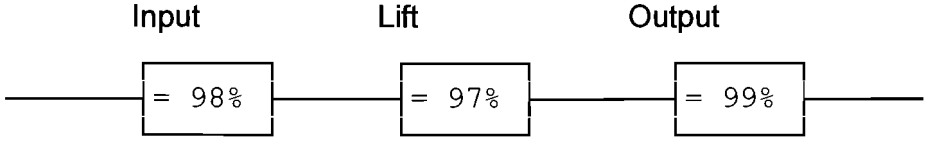
Series connection of components

The availability of a transport system where the components are connected in line is calculated as follows:

= A x B x C



Example: A pallet lift has an input conveyor and an output conveyor. The availability test has shown that the input conveyor has an availability of 98%, the lift 97% and the output conveyor 99%.



The availability of the system is then :

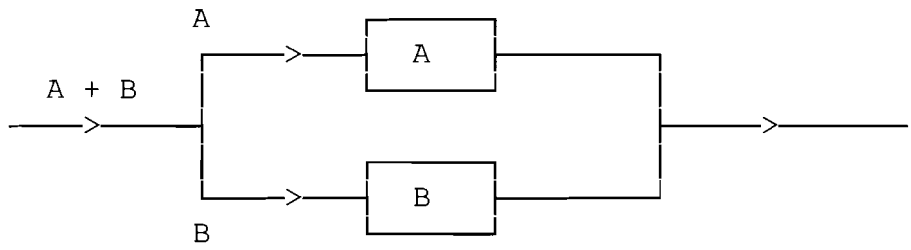
= 0.98 x 0.97 x 0.99

= 0.94 = 94%

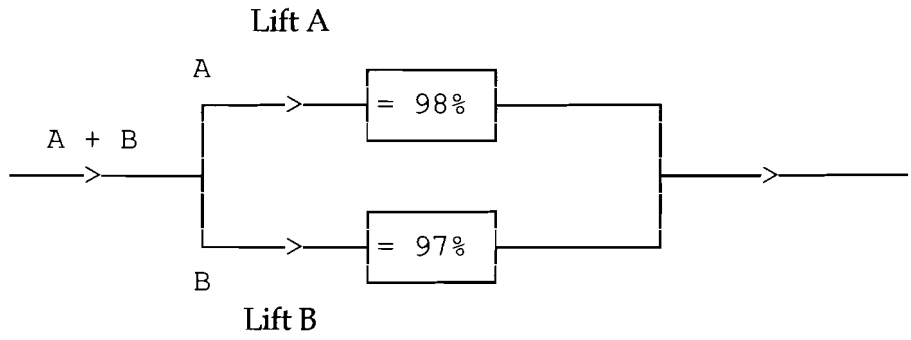
Parallel connection of components (undetermined paths)

The availability of a transport system where the components are connected parallel and the transport can take place selectively across the parallel components is calculated as follows:

= 1 - ((1 - A) x (1 - B))



Example: Two pallet lifts are placed parallel, pallets can be transported through either lift without any predetermination.
The availability test has shown that the pallet lift A has an availability of 98% and pallet lift B an availability of 97%.



The availability of the system is then:
= 1 - ((1 - 0,98) x (1 - 0,97))
= 1 - 0,02 x 0,03
= 0,9994 = 99,94 %

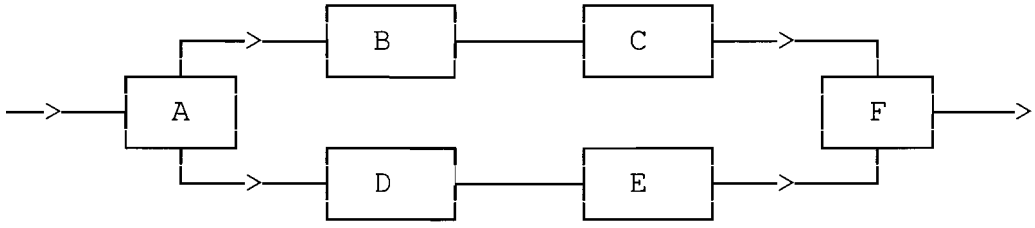
Parallel connection of components (predetermined paths)
In case the transport of goods in the above example must take place across predetermined paths, the calculation is as follows:

= 1 - ((1 - A) + (1 - B))
= 1 - ((1 - 0,98) + (1 - 0,97))
= 1 - (0,02 + 0,03)
= 1 - 0,05
= 0,95 = 95 %

Combination of serial with parallel connection of components (undetermined paths)

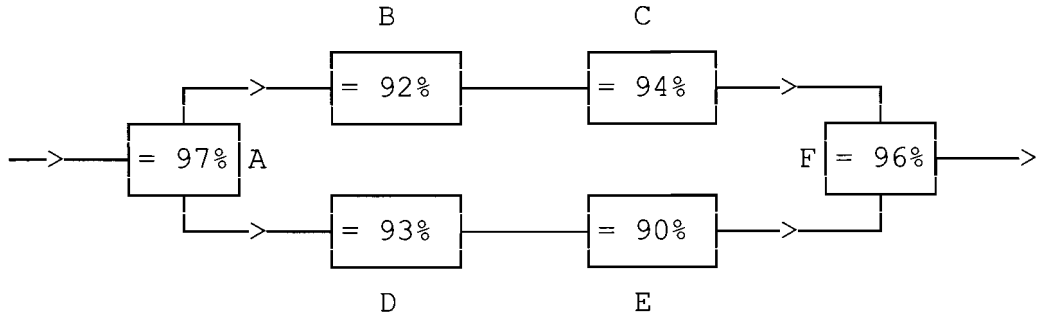
The availability of a transport system where the components are connected in a combination of serial and parallel components and the transport can take place selectively across the parallel components, is calculated as follows:

$$= A \times (1 - ((1 - B \times C) \times (1 - D \times E)) \times F$$



Example: Two pallet lifts (B and C) are placed in serie to elevate from a low level to a high level and to bring the pallets back to a low level. Two of these sets are placed parallel B/C and D/E) to increase the reliability. Pallets can be transported through either lift without any predetermination. The pallets are divided over the two sets of lifts via a transfer (A) and are again single lined via another transfer (F).

The availability test has shown the following availabilities:
- pallet transfer A 97 %
- pallet lift B 92 %
- pallet lift C 94 %
- pallet lift D 93 %
- pallet lift E 90 %
- pallet transfer F 96 %



The availability of the system is then :

= 0,97 x (1 - ((1 - 0,92 x 0,94) x (1 - 0,93 x 0,90)) x 0,96
= 0,97 x (1 - 0,1352 x 0,163) x 0,96
= 0,97 x 0,978 x 0,96
= 0,9107 = 91,07 %

Appendix B

Definitions of Book

of

M.J.P. van der Meulen

B.1 AVAILABILITY :

The ability of a product to be in a state to perform a required function under given conditions at a given instant of time or over a given time interval assuming that the required external resources are provided.
(CENELEC, ENV50129, 1998; CENELEC, prEN50126,1998)

The ability of a functional unit to be in a state to perform a required function under given conditions at a given instant of time or over a given time interval assuming that the required external resources are provided.
(ISO 2382-14,1997)

The ability of an item to be in a state to perform a required function under given conditions at a given instant of time or over a given time interval assuming that the required external resources are provided.
(ESA, ECSS-P-001A,1997; BSI, BS 4778-3.2,1991;IEC 50-191,1990)

The probability that the system or equipment used under stated conditions will be in an operable and committable state at any given time.
(NATO, ARMP-7,1996)

A Measure of the degree to which an item is in an operable and committable state at the start of a mission when the mission is called for at an unknown (random) time. The ability of an item to perform its designated function when required for use
(USA DoD, MIL-Std-109c, 1994)

The probability that a system will be able to perform its designated function when required for use.
(CCPS, 1993)

Dependability with respect to the readiness for usage. Measure of correct service delivery with respect to the alternation of correct and incorrect service.
(Laprie, 1992)

The fraction of time that the system is actually capable of performing its mission
(IEC 1131-1, 1992)

The degree to which a system or component is operational and accessible when required for use. Often expressed as a probability.
(IEEE Std-610.12, 1991)

The prevention of the unauthorized withholding of information or resources. (UK DTI, ITSEC, 1991)

Expected fraction of time during which a system or component is functioning acceptably
(Mus et al. 1987)

The probability that the system is operating satisfactorily at any point in time when operating under specified conditions, where the time categories include all but free time.
(Martz & Waller, 1982)

The ability of an item (under combined aspects of its reliability, maintainability and maintenance support) to perform its required function at a stated instant of time or over a stated period of time
(O' Conner, 1981)

B.2 OPERATIONAL AVAILABILITY :

The probability that an equipment/ system at any instant in the required operating time will operate satisfactorily under stated conditions where the time considered includes operating, corrective and preventive maintenance, administrative delay time and logistic delay time
(NATO, ARMP-7,1996)

The proportion of the defined operational period during which the equipment is available for the use without any performance limitations
(UK MoD, Def Stan 00-49,1996)

B.3 RELIABILITY :

The probability that an item can perform a required function under given conditions for a given time interval (t1, t2)
(CENELEC prEN50126, 1998; ESA, ECSS-P-001A, 1997; BSI, BS 4778-3.2, 1991; IEC 50-191,1990)

The ability of an item to perform a required function under stated conditions for a stated period of time.
(CENELEC, ENV50129, 1998; SSCP 85, 1995; BSI, BS 4778-9, 1991; Klinger et al., 1990; IEC 902, 1987; O' Conner, 1981)

The ability of a functional unit to perform a required function under given conditions for a given time interval
(ISO 2382-14,1997)

The ability of a system or component to perform its required function under stated conditions for a specified period of time.
(UK MoD, Def Stan 00-55,1997)

Probability that a system can perform a defined function under stated conditions for a given period of time.
(ISA, S84.01, 1996)

(1)The duration or probability of failure free performance under stated conditions.
(2) The probability that an item can perform its intended function for a specified interval under stated conditions
(USA DoD, MIL-Std-109C, 1994; USA, MIL-Std-721C,1992)

Probability that a component or system will function correctly under stated conditions for a stated period of time (CCPS, 1993)

The probability that an item is able to perform a required function under stated conditions for a stated period of time or for a stated demand
(Jones, 1992)

The ability of a functional unit to perform a required function under stated conditions for a stated period of time
(IEC 1131-1, 1992; IEEE Std-610.12, 1991)

Dependability with the respect to the continuity of service. Measure of continuous correct service delivery . Measure of the time to failure
(Laprie 1992)

That aspect of the safety integrity relating to random hardware failure in a dangerous mode of failure of the safety-related systems.
(HSE, 1991)

The probability that an item will perform a required function, under stated conditions, for a stated period of time. Since observed reliability is empirical it is defined as the ratio of times which perform their function for the stated period to the total number in the sample.
(Smith, 1981)

Appendix C

Standards and Norms

C.1 Nederlands Normalisatie-instituut

Availability :

Norm : NEN-EN-IEC 61703:2002 en;fr
Date : 2002-01-01
Title : Mathematical expressions for reliability, availability, maintainability and maintenance support terms
Summary : Provides mathematical expressions for reliability, availability, maintainability and maintenance support measures defined in IEC 60050-191.

Norm : NEN-ISO/IEC 2382-14:1998 en;fr
Date : 1998-01-01
Title : Informatietechniek;Woordenlijst;Deel 14: Betrouwbaarheid, onderhoud en beschikbaarheid / Information technology
Summary : Is intended to facilitate international communication in information technology. It presents, in two languages, terms and definitions of selected concepts relevant to the field of information technology and identifies relationships among the entries.

Norm : NEN-ISO 3977-9:2000 en
Date : 2000-01-01
Title : Gasturbines;Aanbesteding;Deel 9: Betrouwbaarheid, beschikbaarheid, onderhoudbaarheid en veiligheid / Gas turbines
Summary : Is to provide a basis for exchange of information about reliability, availability, maintainability and safety between gas turbine manufacturers, users, consultants, regulatory bodies, insurance companies and others. It defines terms and definitions used within this part of ISO 3977 and also describes component life expectancy, repairs and criteria for determining overhaul intervals.

Norm : NEN-EN 50126:1999 en
Date : 1999-11-01
Title : Spoorwegtoepassingen;De specificatie en het bewijs van de bruikbaarheid, beschikbaarheid, onderhoudbaarheid en veiligheid / Railway applications
Summary : This standard: - defines RAMS in terms of reliability, availability, maintainability and safety and their interaction; - defines a process, based on the system lifecycle and tasks within it, for managing RAMS; - enables conflicts between RAMS elements to be controlled and managed effectively; - defines a systematic process for specifying requirements for RAMS and demonstrating that these requirements are achieved; - addresses railway specifics; - does not define RAMS targets, quantities, requirements or solutions for

specific railway applications; - does not specify requirements for ensuring system security; - does not define rules or processes pertaining to the certification of railway products against the requirements of this standard; - does not define an approval process by the safety regulatory authority. This standard is applicable: - to the specification and demonstration of RAMS for all railway applications and at all levels of such an application, as appropriate, from complete railway routes to major systems within a railway route, and to individual and combined sub-systems and components within these major systems, including those containing software; in particular: to new systems; to new systems integrated into existing systems in operation prior to the creation of this standard, although it is not generally applicable to other aspects of the existing system; to modifications of existing systems in operation prior to the creation of this standard, although it not generally applicable to other aspects of the existing system. - at all relevant phases of the lifecycle of an application; - for use by Railway Authorities and the railway support industry.

Norm : NEN-EN 1699:1995 en
Date : 1995-09-01
Title : EDI;Bericht;Bericht over het verzoek om informatie over de transportschema's en -beschikbaarheid (IFTSAI) / EDI
Summary : The function of this message is to request transport schedule or availability information and to answer to such a request

Reliability :

Norm : NEN-IEC 60300-3-1:1999 en;fr
Date : 1999-06-01
Title : Beleid met betrekking tot betrouwbaarheid;Deel 3: Leidraad voor de toepassing;Sectie 1: Analysetechnieken voor betrouwbaarheid: Leidraad voor methodologie / Dependability management
Summary : Provides a standard method of test for determining the ability of a specimen of an electrotechnical product to withstand specified severities of impact. Describes the calibration of the test apparatus

Norm : NEN-IEC 60050-191:1991/A2:2002 en;es;fr;ru
Date : 2002-03-01
Title : Internationale elektrotechnische woordenlijst;Hoofdstuk 191: Betrouwbaarheid en kwaliteit van dienstverlening / International Electrotechnical Vocabulary
Summary : -

Norm : NEN-ISO 16269-7:2001 en
Date : 2001-04-01
Title : Statistische interpretaties van gegevens;Deel 7: Mediaan;Schatting en betrouwbaarheidsintervallen / Statistical interpretation of data
Summary : Specifies procedures for establishing a point estimate and confidence intervals for the median of any continues probability of a population, based on a random sample of size n from the population. These procedures are distribution-free, i.e. they do not require knowledge of the family of distributions to which the population distribution belongs. Similar products can be applied to estimate quartiles and/or percentities

Norm : NEN 11078:1994 en;fr
Date : 1994-04-01
Title : Analysetechnieken voor de betrouwbaarheid;Methode met betrouwbaarheidsblokdiagram / Analysis techniques for dependability
Summary : This standard describes procedures for modelling the dependability of a system and for using the model in order to calculate reliability and availability measures

C.2 Insurance Services Offices

Availability:

Norm : ISO/IEC 2382-14:1997
Date : 2002-01-15 (stage date)
Title : Information technology, Part 14: Reliability, maintainability and availability

Norm : ISO 3977-9:1999
Date : 1999-12-09 (stage date)
Title : Gas turbines, Part 9: Reliability, availability, maintainability and Safety

Norm : ISO 8927:1991
Date : 2001-09-25 (stage date)
Title : Earth-moving machinery : Machine availability

Norm : ISO 11994:1997
Date : 2001-11-15
Title : Cranes : Availability

Reliability :

Norm : ISO/IEC 2382-14:1997
Date : 2002-01-15 (stage date)
Title : Information technology, Part 14: Reliability, maintainability and availability

Norm : ISO 2394:1998
Date : 1998-06-11
Title : General principles on reliability for structures

Norm : ISO 3977-9:1999
Date : 1999-12-09 (stage date)
Title : Gas turbines, Part 9: Reliability, availability, maintainability and safety

Norm : ISO 5843-8:1988
Date : 1998-10-31
Title : Aerospace -- List of equivalent terms -- Part 8: Aircraft reliability

Norm : ISO 6527:1982
Date : 1999-11-26
Title : Nuclear power plants -- Reliability data exchange -- General guidelines

Norm : ISO 7385:1983
Date : 1999-02-05
Title : Nuclear power plants -- Guidelines to ensure quality of collected data on reliability

Norm : ISO 8930:1987
Date : 1994-11-25
Title : General principles on reliability for structures -- List of equivalent terms

Norm : ISO 14224:1999
Date : 2002-02-15
Title : Petroleum and natural gas industries -- Collection and exchange of reliability and maintenance data for equipment

C.3 Verein Deutsches Ingenieure

Availability :

Norm : VDI 3423
Date : 2002-01
Title : Technical availability of machines and production lines - Terms, definitions, determination of time periods and calculation

Norm : VDI 3542 Blatt 4
Date : 2000-10
Title : Safety terms for automation systems - Reliability and safety of complex systems (terms)

Norm : VDI 3581
Date : 2001-03
Title : Availability of transport/storage equipment as well as their subsystems and elements

Norm : VDI 4001 Blatt 1
Date : 1998-04
Title : General guide to the VDI-Handbook reliability engineering

Norm : VDI 4001 Blatt 2
Date : 1986-06
Title : Basic terms and definitions

Norm : VDI 4002 Blatt 2
Date : 1986-07
Title : Systems engineering considerations; introduction into the reliability problem of technical products and/or systems

C.4 DIN

Availability :

Norm : DIN IEC 56(Sec)402
Date : 1995-07
Title : Mathematical expressions for reliability, maintainability and availability terms

Norm : DIN IEC 60863
Date : 1988-08
Title : Electrical engineering; presentation of reliability, maintainability and availability predictions; identical with IEC 60863:1986

Reliability :

Norm : DIN 25424-2
Date : 1990-04
Title : Fault tree analysis; manual calculation procedures for the evaluation of a fault tree

Norm : DIN 40081-11
Date : 1976-11
Title : Guidance to reliability; electronic components, lot by lot and periodic inspection procedures

Norm : DIN IEC 56(CO)106
Date : 1985-11
Title : Electrical engineering; equipment reliability testing; part 2: guidance for the design of test cycles for equipment reliability testing; identical with IEC 56(Central Office)106

Norm : DIN IEC 56(Sec)283
Date : 1989-12
Title : Electrical engineering; equipment reliability testing; preferred test conditions; outdoor transportable equipment; low degree of simulation; identical with IEC 56(Secretariat)283

Norm : DIN IEC 60605-1
Date : 1986-03
Title : Electrical engineering; equipment reliability testing; general requirements; identical with IEC 60605-1, edition 1978

C.5 IEEE

Availability :

Norm : ANSI/IEEE Std 762-1987
Date : 05-1987
Title : IEEE standard definitions for use in reporting electric generating unit reliability, availability, and productivity

Reliability :

Norm : ANSI/IEEE Std 577-1976
Date : 11-1976
Title : IEEE standard requirements for reliability analysis in the design and operation of safety systems for nuclear power generating stations

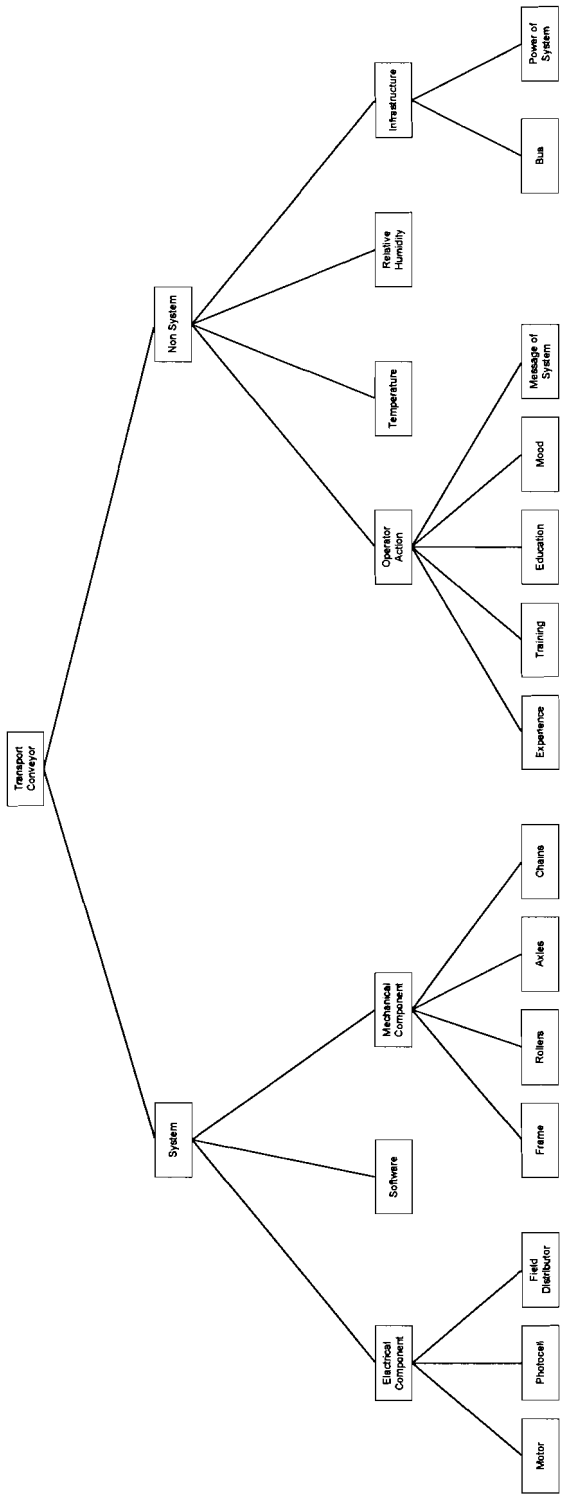
Norm : ANSI/IEEE Std 762-1987
Date : 05-1987
Title : IEEE standard definitions for use in reporting electric generating unit reliability, availability, and productivity

Norm : IEEE Std 1332-1998
Date : 10-1998
Title : IEEE standard reliability program for the development and production of electronic systems and equipment

Norm : IEEE Std 1413-1998
Date : 01-1999
Title : IEEE standard methodology for reliability prediction and assessment for electronic systems and equipment

Appendix D

Diagram of Transport Conveyor

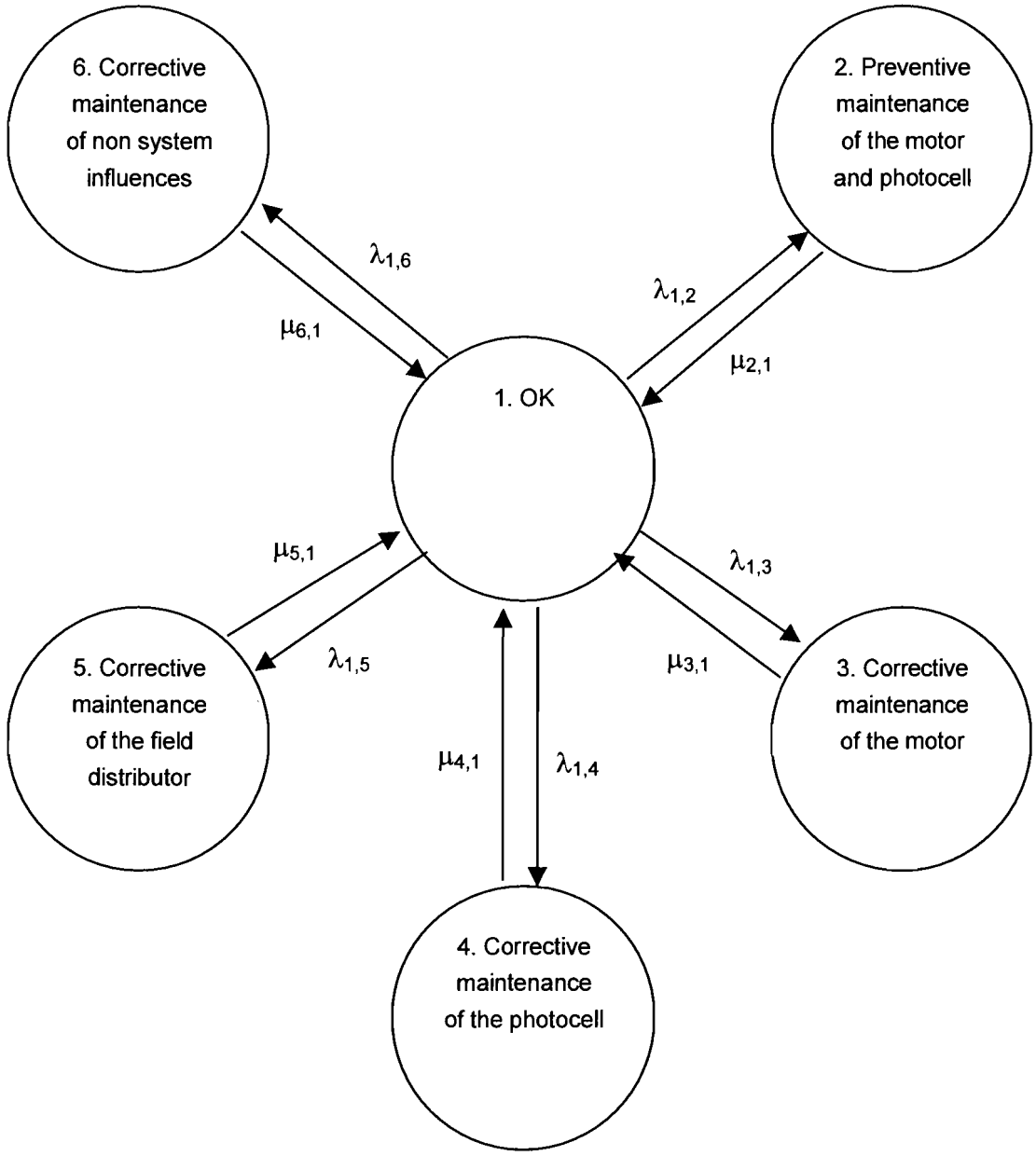

Transport Conveyor
Non System
Infrastructure
Power of System
Bus
Relative Humidity
Temperature
Operator Action
Message of System
Mood
Education
Training
Experience
System
Mechanical Component
Chains
Axles
Rollers
Frame
Software
Electrical Component
Field Distributor
Photocell
Motor

Appendix E

Markov Model

of

Transport Conveyor


6. Corrective maintenance of non system influences
2. Preventive maintenance of the motor and photocell
1. OK
5. Corrective maintenance of the field distributor
3. Corrective maintenance of the motor
4. Corrective maintenance of the photocell
λ1,6
μ6,1
λ1,2
μ2,1
μ5,1
λ1,5
λ1,3
μ3,1
μ4,1
λ1,4

Appendix F

Solving Markov models

F.1 Markov models: general appearance

In general a Markov model consist of a set of coupled differential equations for the state probabilities P_i and a starting condition for the state probabilities P_i at time t=0. If the assumption is made that the model has N states and that the (time-constant) transition rate from state I to state j is given by A_{ij} the set of differential equations would look like:

$$\frac{dP_1}{dt} = A_{11}P_1 + A_{12}P_2 + ... + A_{1N}P_N$$
$$\frac{dP_2}{dt} = A_{21}P_1 + A_{22}P_2 + ... + A_{2N}P_N$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$\frac{dP_N}{dt} = A_{N1}P_1 + A_{N2}P_2 + ... + A_{NN}P_N$$

In a shorter notation this can be written as:

$$\frac{dP}{dt} = AP$$

Where **P** is a vector containing all the state probabilities and **A** is matrix containing all the transition rates. The element of matrix **A** are defined by:

A_{ij} = transition rates from state i to state j. (for $i \neq j$)

$$A_{ii} = -\sum_{i \neq j} A_{ij}$$

The starting conditions can be written as:

$$P_1(t=0) = P_1^0$$
$$P_2(t=0) = P_2^0$$
$$\vdots \qquad \vdots$$
$$P_N(t=0) = P_N^0$$

In short this can be written as:

$$P(t=0) = P^0$$

Obviously at all times including t=0 the probability of the system being in any of the system state is 1:

$$\sum_{i=1}^{N} P_i(t = T) = 1 \qquad \text{for all } T$$

The effect of this last equation is that the set of differential equations is not independent. Taking only N-1 of the differential equations and combining these with the equation that the sum of all probabilities equals 1, can generally solve this problem. This results in the set of equations to be solved:

$$\frac{dP_1}{dt} = A_{11}P_1 + A_{12}P_2 + ... + A_{1N}P_N$$
$$\frac{dP_2}{dt} = A_{21}P_1 + A_{22}P_2 + ... + A_{2N}P_N$$
$$\vdots \qquad \vdots \qquad\qquad \vdots$$
$$\frac{dP_{N-1}}{dt} = A_{(N-1)1}P_1 + A_{(N-1)2}P_2 + ... + A_{(N-1)N}P_N$$

$$1 = P_1 + P_2 + ... + P_N$$

F.1.1 Find limiting state probabilities

The limiting state probabilities (the state probabilities at an infinite time) can be found by using the approach that at infinite time the changes in probabilities (derivatives) are 0. This means that the limiting state probability can be found by solving the set of linear equations.

$$AP = 0$$

Where **0** is the nullvector. In this set again the dependency of the equation can be removed by taking N-1 equations cornbined with the equation that the sum of all probabilities equals 1.

In Markov models a few different situations may occur. These situations are dependent on the existence of so-called absorbing states. These states, which only have transitions going in, but not going out.

- No absorbing states exist. The limiting state probabilities depend on the transition rates.
- Only one absorbing state exists. At any infinite time the system will always end up in this state, so the probability for being in this limiting states at infinite time is 1 and the probability for being in one of the other states at infinite time is 0.

- Two or more absorbing states exist. The system will always end up in one of these absorbing states, so the sum of the probabilities of being in either of these absorbing states at infinite time is 1. The probability of being in any of the other state at infinite time is 0. The subdivision of the probabilities of being in one of the absorbing states depends on the transition rates.

F.2 Solving Markov models using Laplace transforms

One way to solve the Markov models is using Laplace transforms. In this approach the set of differential equations is transformed into a set of linear equations for the Laplace transforms of the state probabilities. This set of linear equations can be solved, after which the Laplace transforms can be transformed back into the time domain.

Writing the Laplace transform of state probability P_i as p_i:

$$p_i(s) = \int_0^\infty e^{-st} P_i(t) dt$$

The set of differential equations after application of the Laplace transform can be written as:

$$\begin{aligned} sp_1 - P_1^0 &= A_{11}p_1 + A_{12}p_2 + \ldots\ldots + A_{1N}p_N \\ &\vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \\ sp_N - P_N^0 &= A_{N1}p_1 + A_{N2}p_2 + \ldots\ldots + A_{NN}p_N \end{aligned}$$

or in short:

$$sp - P^0 = Ap$$

or using **I** for the identity matrix

$$(A - Is)p = -P^0$$

This linear set of equations can be solved. The result is that each p_i can be expressed as a known function f_i of s:

$$p_i(s) = f_i(s)$$

Transforming back to the time domain assuming that the inverse Laplace transform of function $f_i(s)$ is given by $F_i(t)$ results in the solution for all the probabilities:

$$P_i(t) = F_i(t)$$

F.3 Solving Markov models using eigenvalues and eigenvectors

A different approach is using the assumption that a solution can be tried that looks like:

$$P(t) = Qe^{\vartheta t}$$

Inserting this into the set of differential equations leads to:

$$\vartheta Qe^{\vartheta t} = AQe^{\vartheta t}$$

or

$$AQ = \vartheta Q$$

This is an eigenvalue problem where θ is an eigenvalue and Q the corresponding eigenvector. If the eigenvectors are independent the general solution can be given by:

$$P(t) = c_1 Q_1 e^{\vartheta_1 t} + c_2 Q_2 e^{\vartheta_2 t} + ... + c_N Q_N e^{\vartheta_N t}$$

Where Q_i is the eigenvector corresponding with the eigenvalue θ_i and c_i are constants that can be calculated using the starting condition

$$P(t = 0) = P^0$$

F.4 Solving Markov models numerically using matrix multiplication

A different way of solving the Markov model is by using a non-analytical, numerical approach. In this approach the differential equation are approximated using a discrete time step Δt:

$$\frac{P(t+\Delta t)-P(t)}{\Delta t} = AP(t)$$

Solving for P(t+Δt) this can be rewritten as:

$$P(t+\Delta t) = (A\Delta A+I)P(t)$$

where I is the identity matrix.

Solving starting from t=0 now only requires matrix multiplication:

$$\begin{aligned} P(\Delta t) &= (A\Delta t + I)P(t=0) = (A\Delta t + I)P^0 \\ P(2\Delta t) &= (A\Delta t + I)P(\Delta t) = (A\Delta t + I)^2 P^0 \\ &\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ P(m\Delta t) &= (A\Delta t + I)P((m-1)\Delta t) = (A\Delta t + I)^m P^0 \end{aligned}$$

The time step Δt has to be chosen sufficiently small in order to prevent numerical errors due to instability problems. As a rule of thumb the time step should be chosen as a fraction of the smallest relevant time in the set of equations. In the systems discussed in this report this relevant time is generally the time associated with repair of detected failures (typically in minutes).

Another numerical problem that might arise is the fact that in the transition matrix two very different time scales play an important role: one time scale is that coupled to the failure rates which are typically in the order of 10^{-2} per hour. The second time scale is that repairs of (detected) failures are typically in the order of minutes. Especially when very reliable components (small failure rates) are used in combination with short repair times the numerical solution (the number of significant digits) used by the computer to evaluate the Markov model becomes a problem. In that situation a small time is required. A generally used representation for calculations using real numbers on a PC, the double types, has 15 significant digits. In the situation of a low failure rate and a small time step the difference between for example 1 and 1-λΔt may disappear in the computer representation. This problem can be solved by using a better representation (more digits) for real numbers on the PC, both for representing numbers as well as for the involved calculations, or by using dedicated solution algorithms. A set of equations showing this behavior is called a stiff set of differential equations.

Appendix G

Device Descriptions

Input/ Output (I/O)

This is the lowest level in the architecture. To one Programmable Logic Controller, one or more of the following components could be used for the Input / Output:

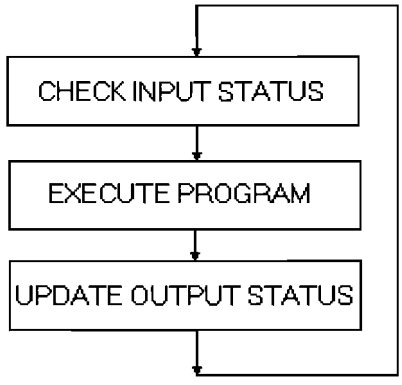
- Actuators (Motor, Alarm, Emergency Stop, etc.)
- Sensors (Photocell, etc.)
- Other Devices

These parts gave us the real status of the system. An example of Material Handling will be given to explain the function. The photocell gives us the signal that a pallet is detected, so the motor has to turn on and the pallet moves further. We see here two functions: detecting the state and reacting on that state.

PLC

A PLC (i.e. Programmable Logic Controller) is a device that was invented to replace the necessary sequential relay circuits for machine control. The PLC works by looking at its inputs and depending upon their state, turning on/off its outputs. The user enters a program, usually via software, that gives the desired results.

A PLC works by continually ***scanning*** a program. This is a scan cycle consisting of 3 important steps, namely:



Step 1-**CHECK INPUT STATUS**-First the PLC takes a look at each input to determine if it is on or off. In other words, is the sensor connected to the first input on? How about the second input? How about the third... It records this data into its memory to be used during the next step.

Step 2-**EXECUTE PROGRAM**-Next, the PLC executes the program one instruction at a time. Maybe the program said that if the first input was on then it should turn on the first output. Since it already knows which inputs are on/off

from the previous step it will be able to decide whether the first output should be turned on based on the state of the first input. It will store the execution results for use later during the next step.

Step 3-**UPDATE OUTPUT STATUS**-Finally the PLC updates the status of the outputs. It updates the outputs based on which inputs were on during the first step and the results of executing the program during the second step.

After the third step the PLC goes back to step one and repeats the steps continuously. One scan time is defined as the time it takes to execute the 3 steps listed above

SQL Server

SQL stands for Structured Query Language. SQL is used to communicate with a database. According to ANSI (American National Standards Institute), it is the standard language for relational database management systems. SQL statements are used to perform tasks such as update data on a database, or retrieve data from a database. Some common relational database management systems that use SQL are: Oracle, Sybase, Microsoft SQL Server, Access, etc. Although most database systems use SQL, most of them also have their own additional proprietary extensions that are usually only used on their system. However, the standard SQL commands such as "Select", "Insert", "Update", "Delete", "Create", and "Drop" can be used to accomplish almost everything that one needs to do with a database.

Clients

The clients are mostly used for the visualization of the system. Here the system can be analyzed and therefore the data coming from the SQL will be used.

ERP

An Enterprise Resource Planning (ERP) system is a generic term for an integrated enterprise computing system. It is a customized packaged software-based system that handles the majority of an enterprise's information systems requirements. It is a software architecture that facilitates the flow of information among all functions within an enterprise. For example; an engineer is only interested in the performance of the system and not in what the system has transported exactly. So he gets only information about the performance of the system from the ERP.
ERP sits on a common database and is supported by a single development environment. ERP systems are customized to support an organization' s business processes.

VLAN

A VLAN is a group of network resources that behave as if they were connected to a single network segment — even though they may not be. Personnel who are physically separated from each other are virtually connected on a VLAN. A VLAN is often easier to design, administer and manage than a physical LAN. Furthermore, a VLAN provides better quality control and security because the network can be segmented more effectively.

References

1. Berden T.P.J and A.C. Brombacher, P.C. Sander.
   The building bricks of product quality: an overview of some basic concepts and principles.
   International Journal of Production Economics. Vol. 67 (2000), nr.1, p. 3-15

2. Brombacher A.C. and H.A. de Boer, J. van 't Loo.
   Integration of reliability & tolerance effect analysis.
   In: Proc. Availability and Maintainability Symposium, Atlanta (GA), USA, 1989.

3. Brombacher A.C.
   Reliability by design.
   Stad : Wiley, 1992
   ISBN 0-471-93193-4

4. Brombacher A.C. and E. van de Geest, R. Arendsen , A. van Steenwijk, O. Herrmann.
   Simulation, a tool for designing in reliability.
   Quality and reliability engineering international, Vol. 9 (1993)

5. Elsayed Elsayed A.
   Reliability engineering.
   Stad : Addison Wesley Longman, 1996
   ISBN 0-201-63481-3

6. Johnson B.W.
   Design and analysis of fault-tolerant systems for industrial applications
   In: Proc. 4th International GI/ITG/GMA symposium on Fault-Tolerant Computing Systems.
   Baden-Baden (Germany)
   Stad : Springer, 1989

7. De Jongh S.
   Reliability and availability using simulation.
   Delft University of Technology, Department of Design, Engineering and Production, Chair Mechanical Engineering and Marine Technology, 2001
   Graduation Report, nr. 2001.LT.5443

8. Kales P.
   Reliability: for technology, engineering and management.
   Stad : Prentice Hall, 1998
   ISBN 0-13-485822-0

9. Lewis, E.E.
Introduction to reliability engineering.
Stad : Wiley, 1996

10. Meeker W.Q. and L.A. Escobar.
Statistical methods for reliability data.
Stad : Wiley, 1998

11. van der Meulen M.J.P.
Definitions for hardware and software safety engineers.
Berlin (Germany)
Stad : Springer, 2000
ISBN 1-85233-175-5

12. Military handbook Reliability Prediction of Electronic Equipment.
(MIL-HDBK-217E)
United States Department of Defense, 1987

13. Minderhoud S.
Quality and Reliability in Product Creation – extending the traditional.
Approach QREI November issue, 1999, Wiley & Sons

14. Petkova V.T. and P.C. Sander, A.C. Brombacher.
The use of quality metrics in service centers.
International Journal of Production Economics. Vol. 67 (2000), nr. 1,
p. 27-36

15. Rouvroye J.
Enhanced markov analysis as a method to assess safety in the process industry
Dissertation, Eindhoven University of Technology, 2001

16. Sander P.C. and A.C. Brombacher.
Analysis of quality information flows in the product creation process of high volume consumer products.
International Journal of Production Economics. Vol. 67 (2000), nr.1,
p. 37-52.

17. Yuan Lu and Han Tong Loh, A.C. Brombacher, E. den Ouden
Accelerated stress testing in a time driven product development.
International Journal of Production Economics. Vol. 67 (2000), nr. 1,
p. 17-26.