Executive Summary

Trainmining

October 5, 2012

Chapter 2

Data analysis

2.1 Alarm database

The alarm database has is structured as follows:[1]

Table ER_ERRORS

Contains every alarm received by the Maintenance Station. Has the following fields:

- DVNI_ERRORNUMBER - Alarm identifier
- DVNS_ERRORTIME - Time-stamp for the alarm
- DVNI_INSTALLATIONCODE - Code of the installation in which the alarm was raised
- DVNI_SENDERINSTALLATIONCODE - Code of the installation from which the alarm was sent (might be different from the one which raised it)

Table IG_INSTALLATIONGENERAL

This table contains information on all the installations. Has the following fields:

- DVNI_INSTALLATIONCODE - Installation identifier
- DVNI_SYSTEMCODE - Type of system, as defined in the "SG_SYSTEMSGENERAL" table
- DVNI_VERSION - System version
- DVAC_SHORTNAME - Short name of the installation
- DVAC_INSTALLATIONNAME - Name of the installation
- DVAC_LOCATION - Location for the installation
- CHK_IS_NODE - Whether it is a node (doesn't directly send alarms, only raise them) or not

Table IG_NODO_INSTALLATION

This table gathers additional information on installations which are nodes. This is, installations that can raise alarms but need a Parent installation to send them. Has the following fields:

- IG_NODO_INSTALLATION - Identifier of the installation which is a node
- DVNI_FATHER_INSTALLATION - Identifier of the parent installation

Alarm information tables ERH_ERRORS_HSL1 or ERS_ERRORS_SAM_ENCE

Both these tables record information on the alarms. Either one or other table is filled depending on which version of the system is installed in the station. However, in terms of information, both contain the following fields:

- DVNI_ERRORNUMBER - Alarm identifier
- MESSAGE_ID - Unique alarm identifier
- MESSAGE_TYPE - Type of alarm, always set as "notification" (not relevant)
- INVOKE_TYPE - Tells whether the alarm has generated itself due to a connection or disconnection (if type is "node") or is generated by a diagnosis system ("saml") or energy system ("energy")
- INVOKE_NAME - Irrelevant, always set to "diagnosis"
- EVENT_TYPE - Defines the type of alarm which has been generated. Its possible values will be described afterwards.
- ADDITIONAL_TEXT - Alarm code
- ADDITIONAL_INFOS - Additional parameters to be shown in error message
- DVNI_ERRORCATEGORY - Alarm severity. Values from 1 to 5 indicating importance of the alarm, or -1 if the alarm indicates recovery from a previous failure.

The "ERH_ERRORS_HSL1" table, has one additional field:

- CLAZZ - Shows the type of system which has sent the alarm

The field "EVENT_TYPE" can have one of the following values:

- fieldElementAlarm - Alarm related to a field element
- fieldElementFailure - Failure in a field element
- operatorInformation - Information to the operator
- imCpuAndCommunications - Related to IM CPU or IM communications
- internalDiagnosis - Internal diagnosis of a system

- operationsDiagnosisCommunications - Communication error in Operation and Diagnosis systems
- ImFecVersions - IM or FEC version
- internalTraces - Internal traces of a system
- operatorCommandAnswer - Answer to an operator command
- CommProblem - Undefined communication problem
- Information - Information message: versions, etc.
- CommunicationsAlarm - Procedures and processes to carry information from one point to other
- QualityOfServiceAlarm - Loss of quality of service
- ProcessingErrorAlarm - SW or processing error
- EquipmentAlarm - Equipment failure
- EnvironmentAlarm - Related to the environment where the system is located
- other - Other

2.2 Statistic analysis

2.2.1 Alarm classification

In order to have a better insight of the provided databases and the mentioned descriptions, a preliminary insight was made, quantitatively analysing some of the parameters which seemed more relevant for alarm definition. Specifically, the chosen parameters are the following:

- EVENT_TYPE
- INVOKE_TYPE
- DVNI_ERRORCATEGORY (Error Category)

The proportion of each kind of alarms in each of the provided databases (Antequera, Camas, Segovia and Sevilla) is as follows:

In Sevilla, we observe an additional error category marked as “other”. If we make a deeper insight on those errors, we find that is a group formed by 77 alarms of the same type.

Hourly timeline

In order to make a first approach to data analysis, we decided to analyse the alarms on a hourly distribution, checking which types of alarms are more likely to happen in different hours during the day. The result is the following:

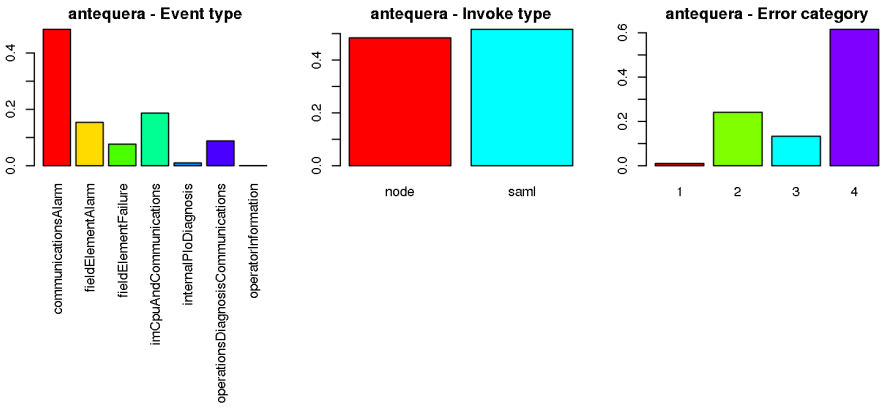

Figure 2.1: Alarm information for Antequera

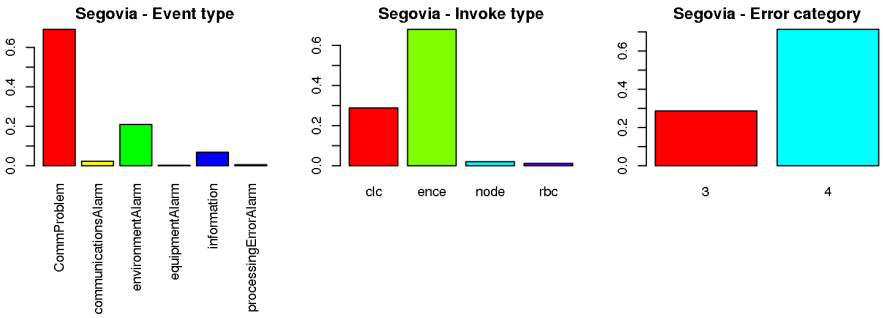

Figure 2.2: Alarm information for Segovia

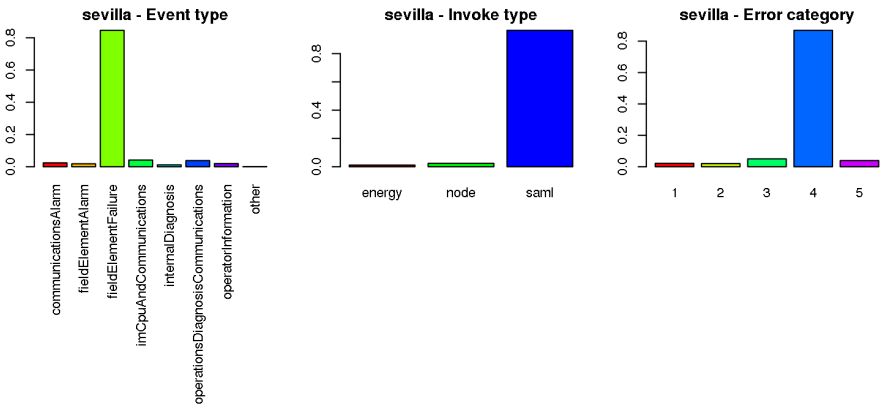

Figure 2.3: Alarm information for Sevilla

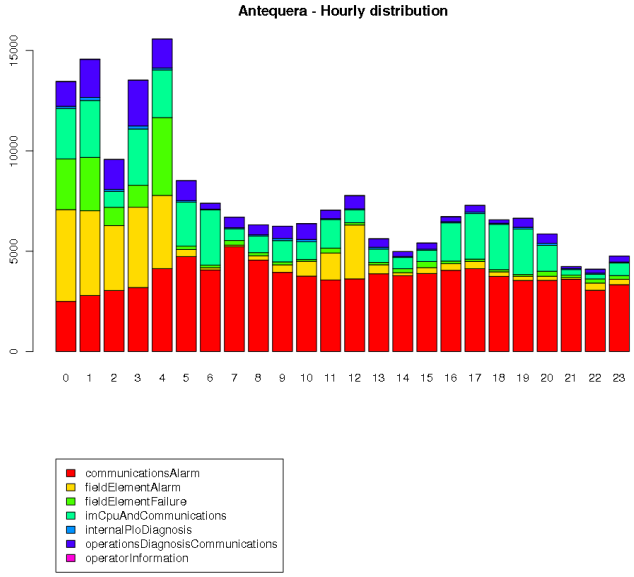

Figure 2.4: Hourly distribution for Antequera (stacked)

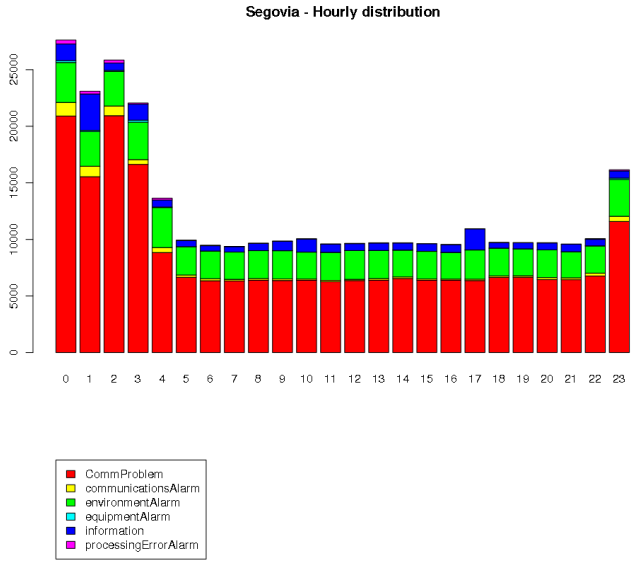

Figure 2.5: Hourly distribution for Segovia (stacked)

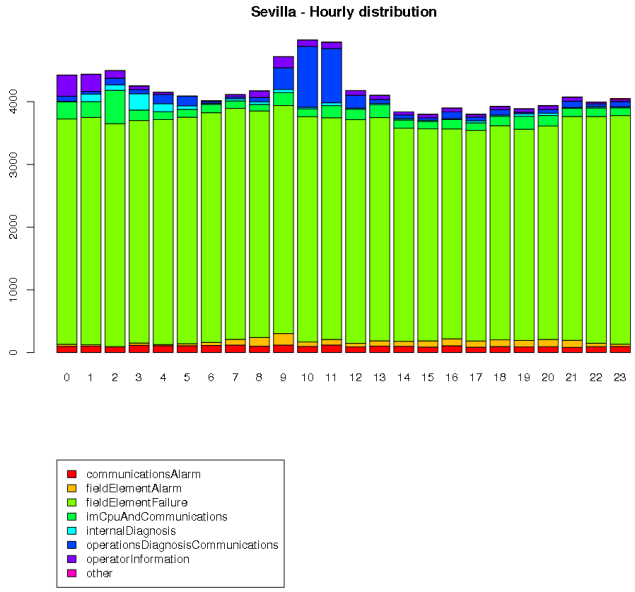

Figure 2.6: Hourly distribution for Sevilla (stacked)

Daily correlation

We have also generated graphics for correlation between number of alarms of each type during the day, and occurrences of other types of alarms. The result is as follows:

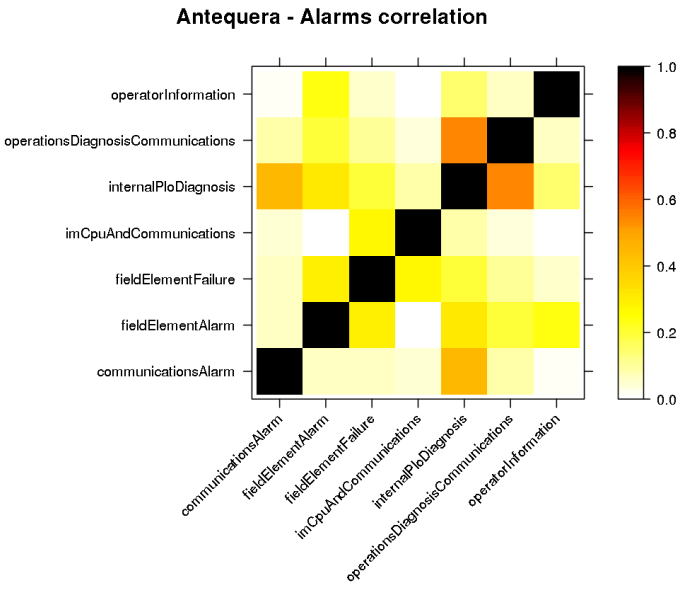

Figure 2.7: Daily correlation for Antequera

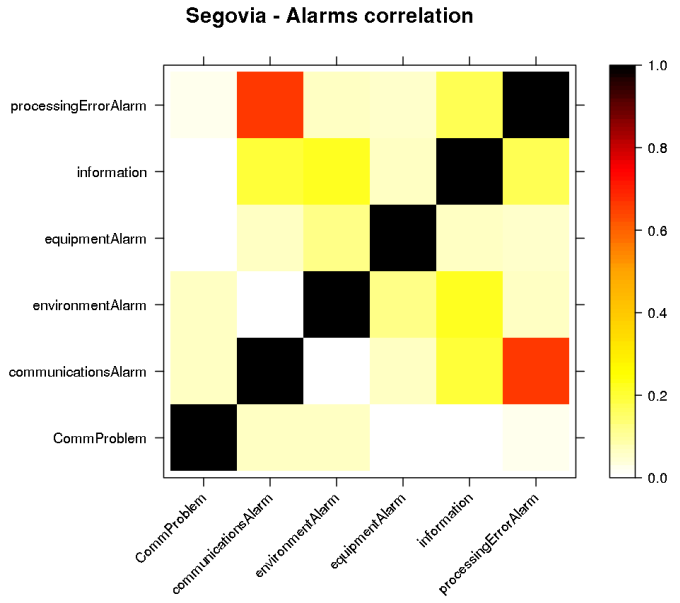

Figure 2.8: Daily correlation for Segovia

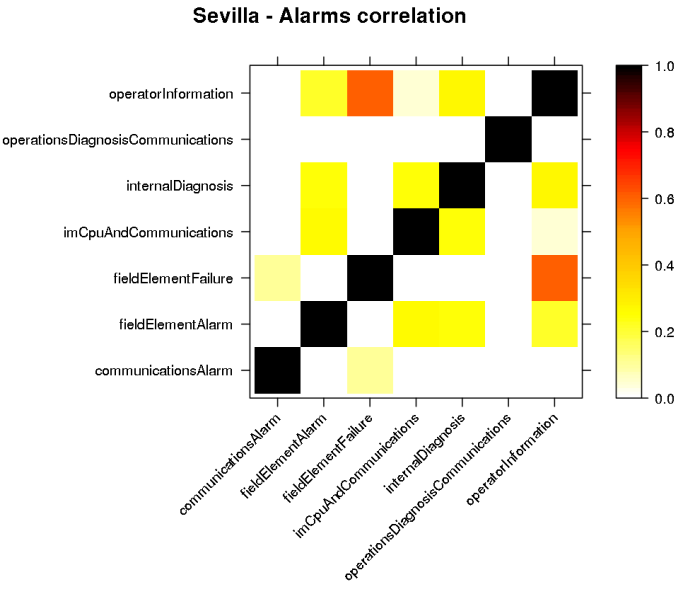

Figure 2.9: Daily correlation for Sevilla

Bibliography

[1] UM Feyyad. Data mining and knowledge discovery: Making sense out of data. *IEEE expert*, 1996.