To appear in the *IEEE/IFIP 1996 Network Operations and Management Symposium (NOMS'96)*, Kyoto, Japan, April 1996.

TASA: Telecommunication Alarm Sequence Analyzer

or

How to enjoy faults in your network

Kimmo Hätönen, Mika Klemettinen, Heikki Mannila, Pirjo Ronkainen, and Hannu Toivonen

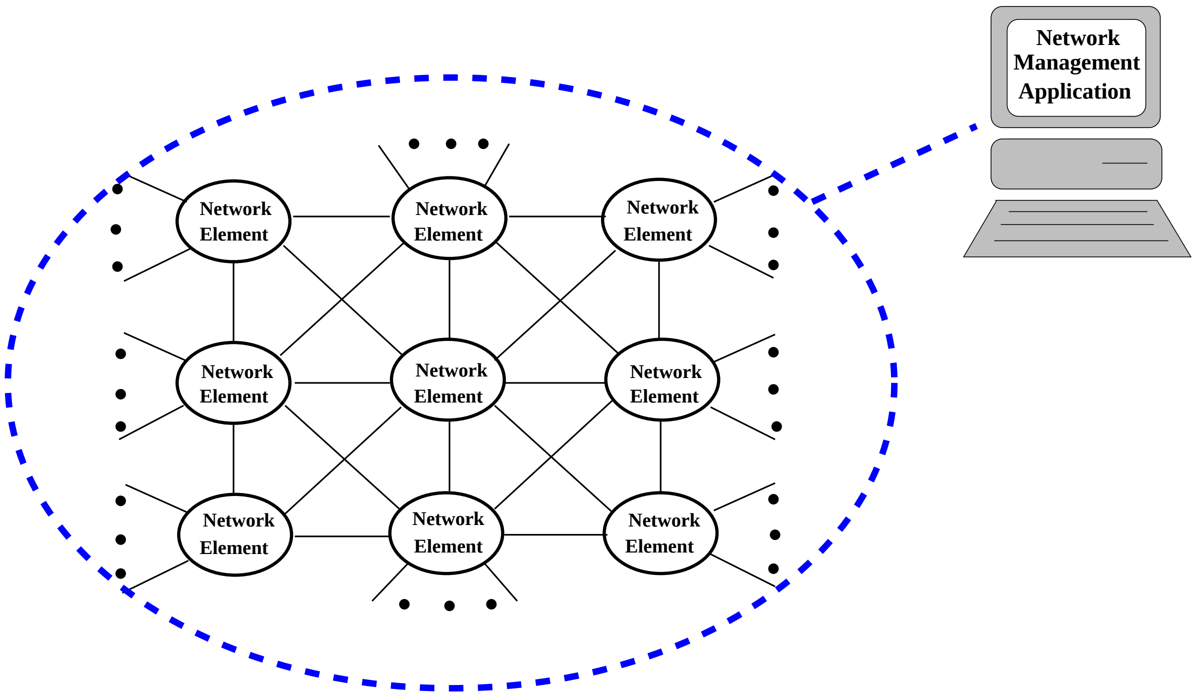
Department of Computer Science, P. O. Box 26
FIN-00014 University of Helsinki, Finland
e-mail: Heikki.Mannila@cs.Helsinki.FI

Abstract Today's large and complex telecommunication networks produce large amounts of alarms daily. The sequence of alarms contains valuable knowledge about the behavior of the network, but much of the knowledge is fragmented and hidden in the vast amount of data. Regularities in the alarms can be used in fault management applications, e.g., for filtering redundant alarms, locating problems in the network, and possibly in predicting severe faults.

In this paper we describe TASA (Telecommunication Alarm Sequence Analyzer), a novel system for discovering interesting regularities in the alarms. In the core of the system are algorithms for locating frequent alarm episodes from the alarm stream and presenting them as rules. Discovered rules can then be explored with flexible information retrieval tools that support iteration. The user interface is hypertext, based on HTML, and can be used with a standard WWW browser.

TASA is in experimental use and has already discovered rules that have been integrated into the alarm handling software of an operator.

Alarm diagnosis



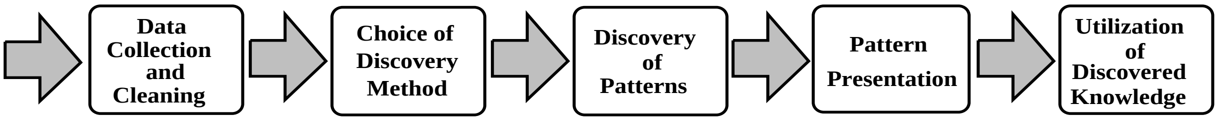
Today's large and complex telecommunication networks produce large amounts of alarms daily. An alarm is generated by a network element or its component to report an abnormal situation it has detected. Alarms are received by and handled by network management applications.

The flow of alarms from a telecommunication network contains a lot of detailed but very fragmented information about problems in the network. In fault management the alarm flow is examined, in order to isolate faults. However, the analysis is difficult, as alarms may be only remote implications of faults, and they can often be analyzed appropriately only in the context of other alarms and other knowledge. In addition, networks are large and alarms are very diverse, and alarms often occur in dense bursts. Also, networks and elements change and develop quickly.

Numerous expert systems and even specialized shells have been deviced to aid in the surveillance of alarms, aiming at alarm filtering, higher level description of network problems, and isolation of faults (see, e.g., [4] or recent articles in [9]). The development of such an expert systems is a very complex, tedious, and error-prone task. In addition to the knowledge that an expert can give, many unknown, interesting regularities may exist in the alarms.

In this paper we describe TASA, Telecommunication Alarm Sequence Analyzer, a system for taking advantage of the information potential hidden in the wealth of alarm data. (A more detailed description of TASA can be found in [3].) TASA semi-automatically discovers regularities in a sequence of alarms. The regularities can give more insight to the workings of the network, or indicate malfunctions or erroneous configurations. They can also be utilized in expert systems in alarm correlation as well as in prediction of faults. A related knowledge acquisition algorithm has been presented in [2].

Knowledge discovery



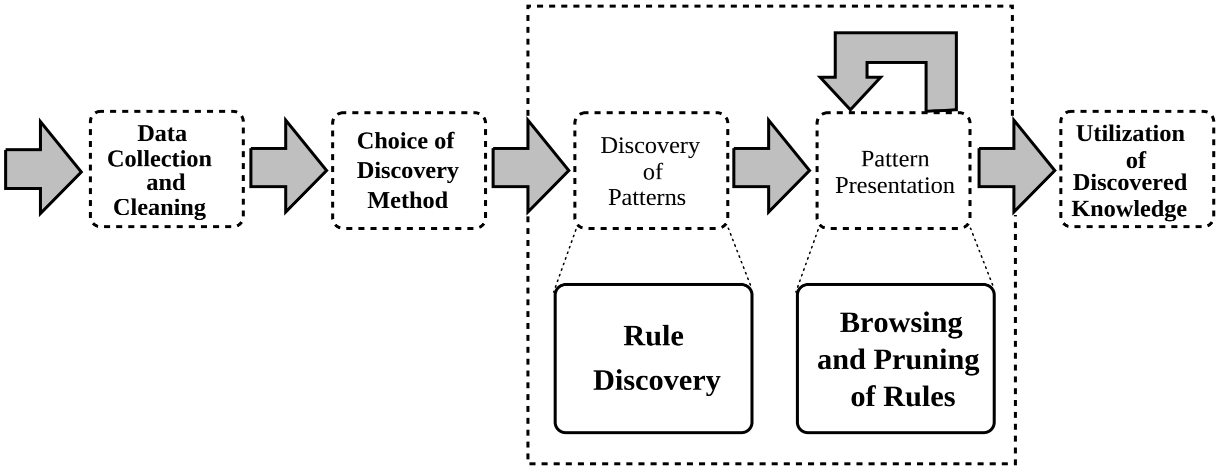
Automated discovery of regularities in large databases has recently become a field of intense interest, under the names knowledge discovery and data mining; see [1, 8] for overviews. We apply knowledge discovery to alarm databases, in order to find useful regularities in the sequence of alarms, for instance about how alarms occur together.

In the core of knowledge discovery are algorithms for discovering different types of patterns (rules, trends, etc.) from data. Knowledge discovery as a whole is an iterative and interactive process. It can, for example, be divided into the following tasks.

1. **Data collection and cleaning** (what types of data can be used, how are errors in the data corrected, what is to be done with missing data, etc.).
2. **Choice of pattern discovery methods** (what types of knowledge are to be discovered, parameter selection, etc.).
3. **Discovery of patterns.**
4. **Presentation of the discovered knowledge** (selection of potentially interesting patterns, organization and visualization of patterns, etc.).
5. **Putting the knowledge into use**, e.g., in an expert system.

No realistic KDD system can be expected to discover useful knowledge without interaction with the user: knowing the interests and using the background knowledge of users is vital for successful knowledge discovery. Also, iteration is essential: after receiving some knowledge the user is better able to focus the search to more interesting areas.

TASA: rule discovery from alarms



TASA is a novel system for knowledge discovery from telecommunication alarm databases. The aim of the system is to give operators useful, possibly new information about the behavior of the network. TASA supports the two central phases of the knowledge discovery process.

Pattern discovery TASA contains algorithms for discovering rules that describe associations of the alarms.

For instance, following types of rules can be discovered:

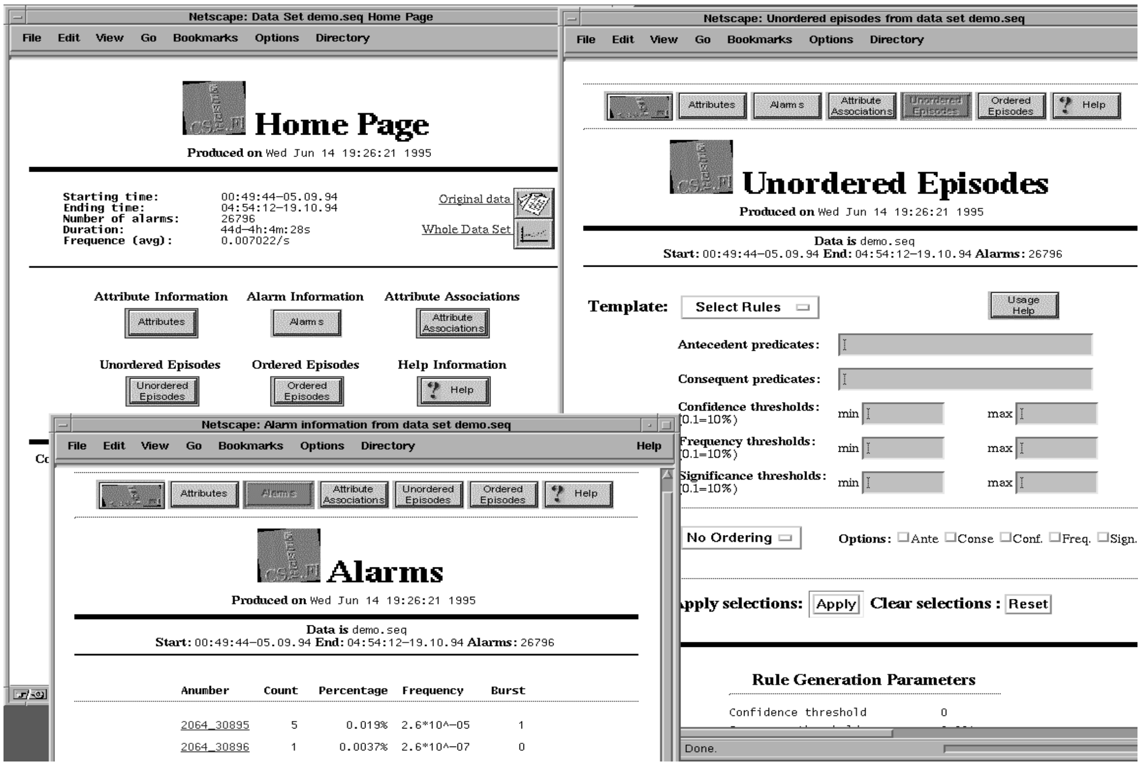
> If alarms of types *link alarm* and *link failure* occur within 5 seconds, then an alarm of type *high fault rate* occurs within 60 seconds with probability 0.7.

> If an alarm is sent during office hours by an element of type *base station*, then the alarm has severity level 1 with probability 0.9.

Pattern presentation TASA presents rules and other information in hypertext format, and offers tools for interactive browsing. The user can have different views to the rules, and iterate dynamically between the views.

We aim at a knowledge discovery process where very large collections of rules are discovered in the pattern discovery phase, and where the iteration and interaction are performed in the rule presentation phase. The advantage of this approach is that the response time is fast in the presentation phase where the rules are readily available, and that going back to a large alarm database for a new pattern discovery phase has to be done only seldom.

TASA system overview



TASA efficiently finds several kind of rules describing associations between alarms and associations between attributes of individual alarms. In the rule discovery phase, all potentially interesting rules that hold in an alarm sequence are discovered. The set of rules is typically very large (thousands of rules), and it includes also many uninteresting rules—most typically a rule fails to be interesting because, for the operator, it is trivial. At their best, discovered rules reveal unexpected and valuable information.

While the rule discovery phase is automatic and only relies on few parameters provided by the user, in the post-processing phase the role of an expert is vital. With the hypertext interface and information retrieval tools the user can have different views to the alarms, and he can select and order rules iteratively to obtain the most interesting and useful set of rules.

TASA is being developed in co-operation with telecommunication companies, and it is currently in prototype use. Rules discovered in initial tests have been integrated into the alarm diagnosis software of an operator.

Episode rules

$$A \xrightarrow{<30 \text{ sec}} B \quad (0.8)$$

$$\underbrace{A, B}_{5 \text{ sec}} \xrightarrow{<60 \text{ sec}} C \quad (0.7)$$

$$\underbrace{AB, CD}_{15 \text{ sec}} \xrightarrow{<4 \text{ min}} E \quad (0.6)$$

TASA discovers so called episode rules. Intuitively they state that when there occurs a certain combination of alarms there will soon occur a specified alarm [7]. For instance:

1. If alarm A occurs, then alarm B occurs within 30 seconds with probability 0.8.
2. If alarms A and B occur within 5 seconds, then alarm C occurs within 60 seconds with probability 0.7.
3. If alarm A precedes alarm B, and C precedes D, all within 15 seconds, then E will follow within 4 minutes with probability 0.6.

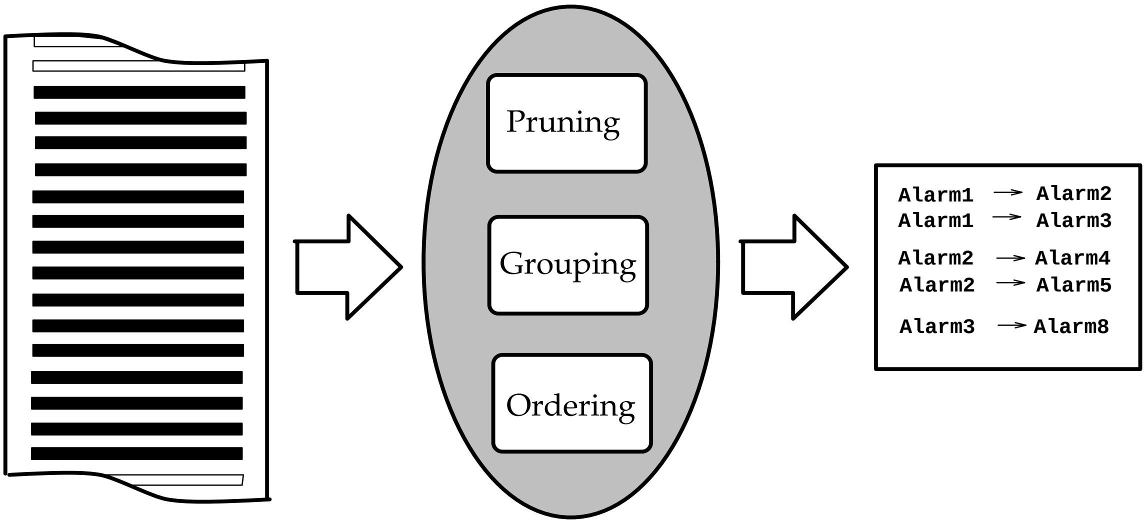
This type of knowledge was chosen for the following reasons.

1. Comprehensibility: such rules are easy to understand for humans.
2. Characteristics of the application domain: such rules can be representations of simple causal relationships.
3. Existence of efficient algorithms: rules of the above form can be discovered efficiently from tens of thousands of alarms.

In TASA, the user specifies the class of interesting episode rules by defining what types of partial orders are allowed in episodes, what types of alarm predicates are used, what are the time periods considered in the rules, and how frequent the rules must be in the sequence. Typical alarm predicates involve the type and severity of the alarm and the network element that sent the alarm. Time periods considered have ranged from 5 seconds to half an hour. The most common types of partial orders used are total orders, i.e., alarms are considered in a strict order; and trivial partial orders, where the order of alarms is not significant. The minimum frequency of rules has ranged from ten occurrences per day to two per week.

Note that TASA finds *all* rules that fulfill the given criteria. The idea is then to let an expert interactively explore the large set of rules. For the algorithms and experimental results see [7].

Rule presentation



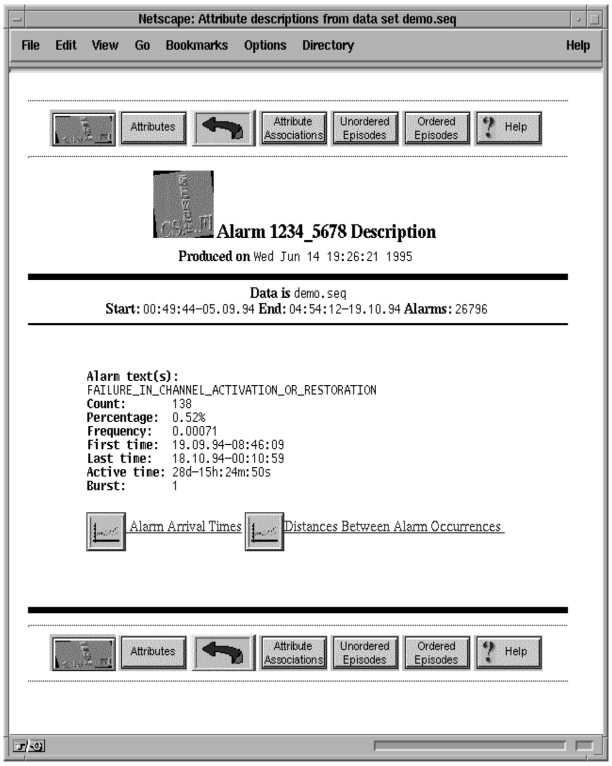
What kind of rules are useful or interesting varies from one situation to another, and it is often impossible for an expert to specify what is interesting before seeing what the results are like. TASA supports easy exploration of large sets of rules by powerful information retrieval tools. These tools can give different views to the discovered knowledge, and the views can be easily modified in an iterative and interactive fashion.

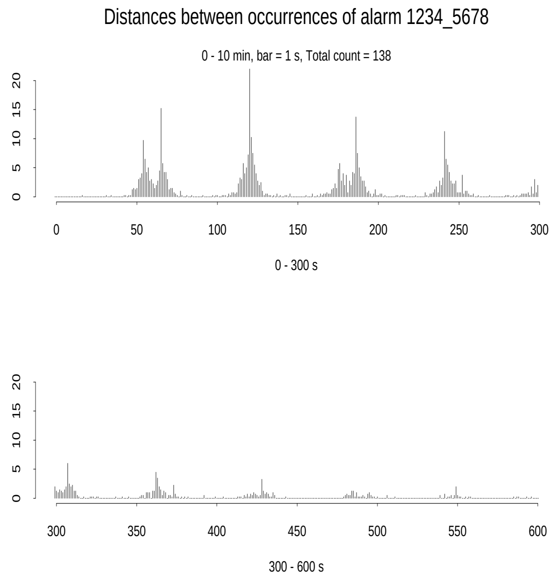
We have identified the following operations the users want to do with the rule sets:

1. **Pruning**: selecting or rejecting rules based on their properties, e.g., exclusion of rules explaining an uninteresting alarm, or inclusion only of rules that contain alarms from a given class.
2. **Ordering**: sorting of rules according to various criteria, such as rule confidence or statistical significance.
3. **Grouping**: clustering of rules into groups of rules that have similar effects in the analyzed alarm database.

Rules can be pruned and ordered according to various attributes of the rules, e.g., the alarm type, severity, rule confidence and frequency, and statistical significance of the rules. Additionally, rules can be pruned with templates [5], i.e., simple regular expressions that describe the form of rules that are to be selected or rejected. This technique is surprisingly powerful. For example, some background knowledge about the connections in the network can be taken into account by using templates that reject all rules that make no sense in the network topology.

Alarm information



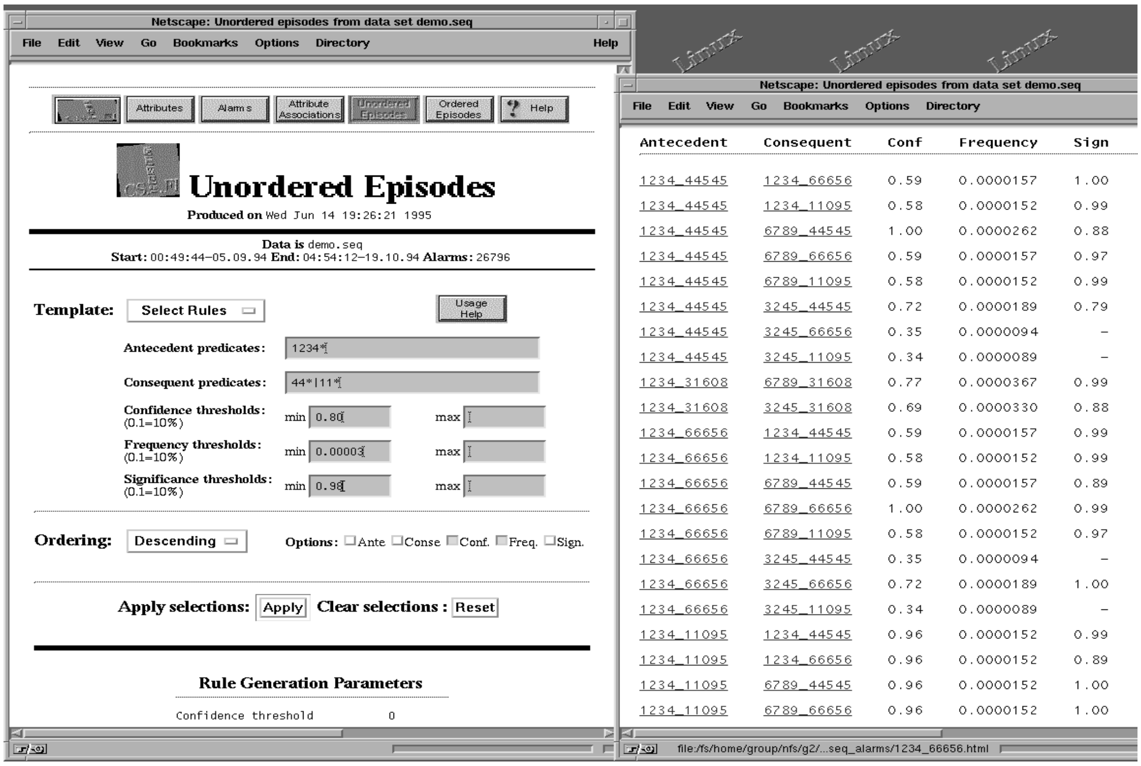


In addition to rules that describe the interconnections of alarms, fairly simple statistical measures are useful in giving an overview of alarms. TASA computes, for instance, the number and percentage of alarms of each type, their frequencies, and measures for whether alarms occur evenly or in bursts.

Information about associations between alarm predicates is also computed. Association rules give information about how properties occur together in alarms, such as "85 % of office hour alarms from network elements of type X occur during the rush hours." The combinations can be arbitrary large, as long as there is at least a user-given minimum number of alarms that match the combination. Just like episode rules, all association rules are found in one run, and then the rule presentation tools are used to browse the set of discovered associations. The algorithm for discovering association rules is described in [6].

Visualization of information is obviously an important part of knowledge discovery applications. Currently, the TASA system offers only simple facilities for this. For instance, a histogram of the distances between alarms of a certain type can be useful in understanding the distribution of alarms, e.g., their periodicities.

Hypertext interface



We have based the hypertext user interface of TASA on the HTML language. HTML (HyperText Markup Language) is a simple, general text based description language for creating hypertext documents and extensively used in the World Wide Web (WWW).

The advantages of using HTML are the following.

- Hypertext suits very well for the presentation of large amounts of data with inherent links but with no linear structure.
- Lists of rules, figures, text documents, and other diverse material can be easily mixed.
- It is easy to generate HTML automatically.
- Browsers and servers are available as public domain programs.
- The system can be used locally and also over a network.

With a hypertext interface, moving around in the discovered information is easy. For instance, by clicking on an alarm name in a rule the user arrives in the alarm description; from there the user can follow a hyperlink to have a look on a histogram of alarm arrival times, etc. Online user documentation is naturally also linked into the system.

The rule presentation tool is embedded to the HTML pages and is based on the fill-out form feature of the HTML language. The selections, orderings, groupings, etc. are given to the system by means of an HTML form. The operations are performed on the rule set, and a new document is created from the old one "on-the-fly" with the given criteria. The user can move freely in the history path and e.g. reformulate his selections.

Conclusion

We have described the TASA knowledge discovery system for analyzing large alarm databases from telecommunication networks. TASA aims at giving operators useful, possibly new information about the behavior of the network.

TASA supports two central phases of the knowledge discovery process. In pattern discovery, TASA finds automatically episode rules and association rules, and also computes several statistical measures. These types of knowledge are easy to understand. They give insight into the alarms, and can be used in network surveillance applications. In the rule presentation phase, exploration of large sets of rules is supported by simple but powerful pruning, ordering, and grouping tools. The hypertext interface of TASA is based on HTML, which offers facilities suitable for presenting linked information and simplifies programming.

The initial version of the TASA system has been tested using alarm data from telecommunication operators. Useful regularities have already been discovered by the system.

References

[1] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors. *Advances in Knowledge Discovery and Data Mining.* AAAI Press, Menlo Park, CA, 1996.

[2] R. M. Goodman and H. Latin. Automated knowledge acquisition from network management databases. In I. Krishnan and W. Zimmer, editors, *Integrated Network Management, II*, 541 – 549. Elsevier Science Publishers B.V (North-Holland), Amsterdam, 1991.

[3] K. Hätönen, M. Klemettinen, H. Mannila, P. Ronkainen, and H. Toivonen. Knowledge discovery from telecommunication network alarm databases. In *12th International Conference on Data Engineering (ICDE'96)*, New Orleans, Louisiana, Feb. 1996.

[4] G. Jakobson and M. D. Weissman. Alarm correlation. *IEEE Network*, 7(6):52 – 59, Nov. 1993.

[5] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo. Finding interesting rules from large sets of discovered association rules. In *3rd International Conf. on Information and Knowledge Management (CIKM'94)*, 401 – 407, Gaithersburg, Maryland, Nov. 1994.

[6] H. Mannila, H. Toivonen, and A. I. Verkamo. Efficient algorithms for discovering association rules. In U. M. Fayyad and R. Uthurusamy, editors, *Knowledge Discovery in Databases, 1994 AAAI Workshop (KDD'94)*, 181 – 192, Seattle, Washington, July 1994.

[7] H. Mannila, H. Toivonen, and A. I. Verkamo. Discovering frequent episodes in sequences. In *1st International Conference on Knowledge Discovery and Data Mining (KDD'95)*, 210 – 215, Montreal, Canada, Aug. 1995.

[8] G. Piatetsky-Shapiro and W. J. Frawley, editors. *Knowledge Discovery in Databases.* AAAI Press, Menlo Park, CA, 1991.

[9] A. S. Sethi, Y. Raynaud, and F. Faure-Vincent, editors. *Integrated Network Management IV.* Chapman & Hall, London, 1995.