Analysis of Public Trouble Ticket Data*

Zsolt Pándi
Department of Telecommunications
Budapest University of Technology and Economics
Budapest, Magyar tudósok kõrútja 2., 1117, Hungary
e-mail: pandi@hit.bme.hu

Abstract

Trouble ticket databases keep record of registered network problems, their (likely) reasons, repair times and other relevant information. Even though this information is considered to be confidential, some research and education oriented networks publish it via the Internet. The present report contains the results of analyzing the content of two such databases and draws conclusions as to the properties of reasonable trouble ticket systems.

1 Introduction

Research related to survivable networks often faces the problem of getting reliability data from real systems. Vendors' interest is against publishing these data due to possible competitive issues, while network operators either do not keep track of this data or treat it as confidential information due to similar motives.

Some research and education oriented networks, however, do treat this kind of information differently and make their trouble ticket system publicly accessible through the Internet. Trouble ticket databases keep record of registered network problems, their (likely) reasons, the repair time and other relevant information [1]. Even though it is not possible to gain a complete picture by means of analysing these databases, it may very well yield valuable information and lead to conclusions regarding research and network operation practice.

The present report summarizes the results of the analysis of the data downloaded from two trouble ticket databases with read-only access to the public [2, 3]. Section 2 gives a detailed introduction of the sources of data and Section 3 introduces the applied methodology. Then the results are discussed in Section 4 and, finally, conclusions are drawn in Section 5.

2 Data Sources

The processed data is from two sources. The first one is the trouble ticket database of CA*Net4, the Canadian research and education oriented backbone network. From this source data of 1012 trouble tickets have been downloaded recording events in between January 1, 2002 and August 11, 2005. The topology of CA*Net4 is illustrated on Figure 1.

The second data source is the trouble ticket database of SWITCH, a network in Switzerland aiming at similar applications as those of CA*Net4. Events from between April 28, 1998 and June 16, 2005 were

*Technical report, Budapest University of Technology and Economics, Department of Telecommunications, December, 2005

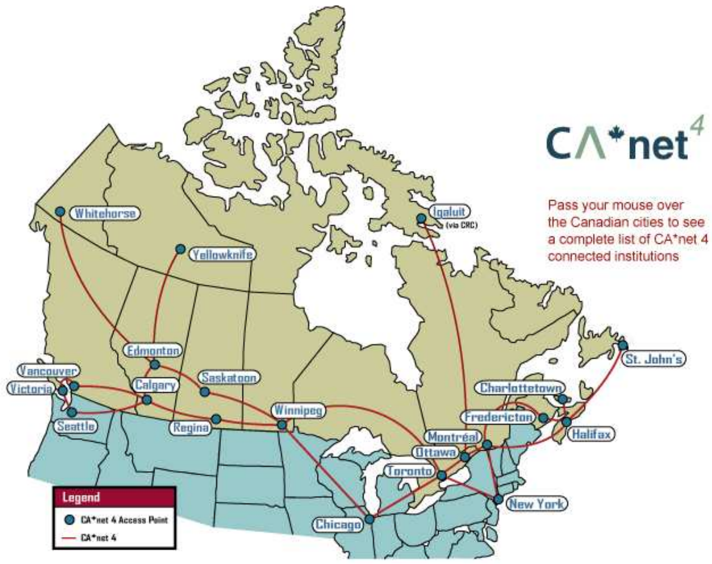


Figure 1: The topology of the CA*Net4 network [2]

recorded in 388 tickets, which have been downloaded. Figure 2 depicts the topology of the SWITCH network.

The downloaded data is available to the public by means of the respective websites [2, 3]. At the time of the processing no information was available as to the completeness or correctness of the records in the trouble ticket databases. Topology information included in this report is also public, and is repeated here in order to illustrate the size of the networks studied. However, detailed equipment information is not made available on any of the websites.

3 Methodology

3.1 General Structure and Lifecycle of Trouble Tickets

Records in the two examined trouble ticket databases share some common properties. They contain

- ticket IDs,
- the opening time of tickets,
- the closing time of tickets,
- ticket status: *open* or *closed*,
- event categories or short descriptions,
- ticket histories as textual descriptions of a series of related events, actions and — occasionally — dates.

In order to interpret the downloaded records it is necessary to overview the lifecycle of tickets, which also helps to establish some background based on which conclusions may be drawn.

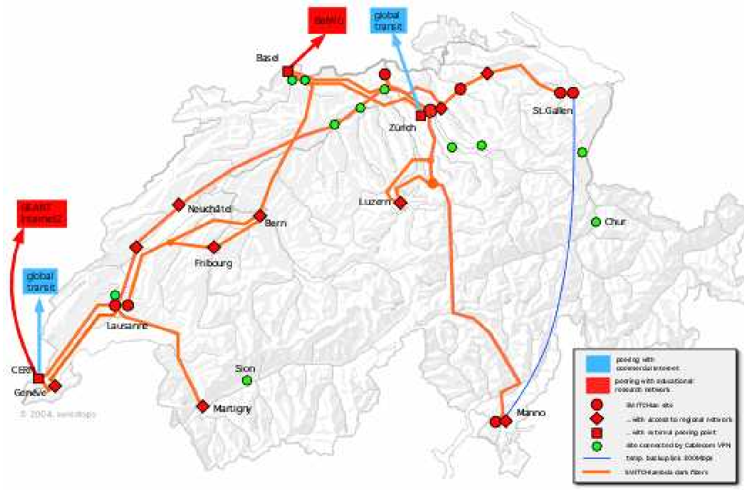


Figure 2: The topology of the SWITCH network [3]

A ticket is opened every time when either an equipment or service failure is discovered or information becomes available about an event in the future that may have an impact on the network[1]. The opening time allows of different interpretations. It may be either identical to the time of the discovery or the onset of the described event, or the creation time of the ticket record entry in the database. Likewise, the closing time has multiple possible meanings. It may be the time when the described failure was repaired, or the time when the ticket record became outdated, or the time when the next periodic check of the trouble ticket database was executed to collect open tickets, as well.

In general, when a ticket is created, its status is set to open. This signifies that the recorded event requires actions from the network operator. The necessary actions range from more attentive surveillance to direct intervention. The ticket remains open until its state is explicitly changed to closed. Practically, by means of filtering out the closed records of the trouble ticket database the remaining list contains current problems, i.e. , events to which special attention must be paid or which entail work that is still pending.

The purpose of the category field of tickets is to ease automated processing. Ticket category is chosen manually out of a predefined list of options. In contrast to this, the contents of the short description field is more suitable for displaying in a brief listing of event records, and is written as a free text in a human language. Thus, the rationale behind the two fields is somewhat different. Note that properly determining the contents of the category field is of utmost importance with respect to analyzing statistical data collected from the trouble ticket database.

The most ample and accurate information is given by the ticket history field — even if only suitable for human reading. Automatically interpreting the contents of this field represents a serious challenge, especially if formal rules are not adhered to and if the language may differ, as well.

The introduced ticket structure includes several options for network operation practice, which have a significant impact on interpreting statistical data, and thus have to be taken account of during the analysis. These issues are discussed in Section 4 along with the presentation of the results.

[1] Such as an announcement of scheduled excavations.

Category	Keywords
Scheduled event	maintenance, install, installation, relocate, relocation, relocations, upgrade, upgdrades, downgrade, downgrades, planned, test, tests, scheduled, work
Field event	cut, rodent, rodents, field
Link related	connectivity, connection, connections, link, links, circuit, circuits, fiber, fibers, fibre, fibres, interface, interfaces, card, cards, wave, waves, wavelength, wavelengths, cable, cables, patch, amplifier, amplifiers
Node related	router, routers, switch, switches, node, nodes, hardware, power, UPS, software, configuration, daemon, reboot, shutdown, performance, temperature, instability, equipment, generator, generators, junos, ios
Hardware related	cut, rodent, rodents, temperature, fiber, fibers, fibre, fibres, router, routers, switch, switches, hardware, wave, waves, wavelength, wavelengths, power, card, cards, infrastructure, cable, cables, patch, UPS, equipment, generator, generators, field, amplifier, amplifiers
Software related	crash, instability, performance, reboot, shutdown, software, configuration, misconfiguration, daemon, BGP, table, tables, junos, ios

Table 1: Keywords used for categorization of trouble ticket records

3.2 Processing Ticket Contents

Both of the trouble ticket systems had an initial trial period, the length of which may be tracked through the changes in the format and content of records. However, after this period records in both of the systems follow a unified structural pattern, thus ticket opening and closing times can be extracted relatively easily by means of text processing.

The category field in trouble ticket system of CA*Net4 only contain two different values and the tickets of the SWITCH trouble ticket database only contain a short description field instead. As a consequence, in order to categorize the records a keyword based approach is adapted here that relies on the full content of the records.

A record is considered to belong to a particular category if it contains at least one of the keywords associated with the category at least once. Analysis of downloaded trouble ticket data is based on the category definitions listed in Table 1.

Clearly, more sophisticated methods are available for the analysis, such as stem identification, synonym search, frequency based keyword selection, filtering of irrelevant words, vectorial characterization based on relevance or frequency dependent weights and cluster identification. Nevertheless, these methods are overly labor-intensive compared to the value of the expected outcome. Not even exact categorization would yield readily applicable data due to the properties of the input data (c.f. Section 5.2).

4 Results

Note that even though the two examined databases serve a similar purpose, they are operated according to somewhat different principles, which has to be remembered when evaluating data and drawing conclusions.

	CA*Net4	**SWITCH**
incomplete records	48.02%	2.06%
event intensity [1/h]	2.95526e-02	5.44175e-03
average ticket lifetime [h:m:s]	12:16:59	171:03:38
maximal ticket lifetime [h:m:s]	510:00:00	4243:49:00

Table 2: Overview of time statistics

	CA*Net4		**SWITCH**	
	test stat.	*p*-value	test stat.	*p*-value
time between successive tickets	13483.9	0	1261.61	0
ticket lifetime	62.5594	7.84356e-7	28.3741	0.012686

Table 3: Test statistics for chi-square goodness-of-fit tests of time data distributions

4.1 Time Data

Table 2 contains an overview of time data from the processed databases. Almost all of the records in the SWITCH database had their time fields comletely filled, whereas in the CA*Net4 database about half of the records did not contain any time data. This suggests that in case of the latter more reliable conclusions may be drawn based on time data. Even so, analysis of the SWITCH network data shows that average ticket lifetime is more than a week, and the ticket with the longest opening time was open for more than 176 days. Manual check of history fields of tickets with lifetime of several hundred hours altogether with the statistical data confirms the following observation. A large number of trouble tickets in the database of SWITCH recorded an *a priori* information on future events, and tickets were opened at the time of receiving the information and not at the time of the recorded event. Manual check of ticket history fields confirms also that ticket closing times often does not match the termination of the recorded event, but they show some later point in time, presumably that of a periodic check for expired tickets that remained open.

Table 3 gives a summary of statistical parameters in conjunction with the following sections. Chi-square goodness-of-fit tests are applied to determine wether the time data follows the exponential distribution. The tests had the null hypothesis that the respective distributions were exponential. During the analysis maximum likelihood estimations were used to determine the parameter of the exponential distribution[2]. The table shows the test statistics and the respective *p*-values. For further interpretation of the results the reader is referred to the Sections below.

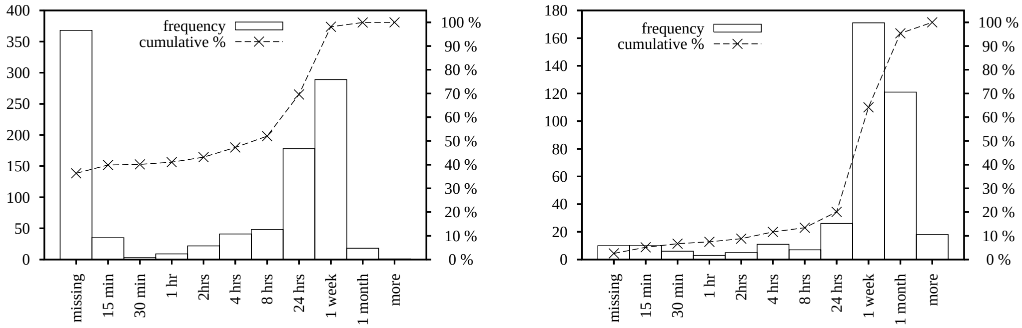
4.1.1 Disribution of interval lengths between successive tickets

The following observations may be made with respect to intervals between succesive ticket opening times, depicted on Figure 3. Event intensities[3] of the two networks differ by almost an order of magnitude. It may be explained by the larger extent of the Canadian network, as link failure statistics are directly related to link lengths. However, several factors have an influence on the event intensity, such as the number of failure-prone equipment or components, their reliability performance and ticket recording policies), and so further conjectures cannot be made.

Intervals between successive ticket opening times may be informally considered as intervals between failure events. However, analysis of the time data reveals that the exponential distribution, applied

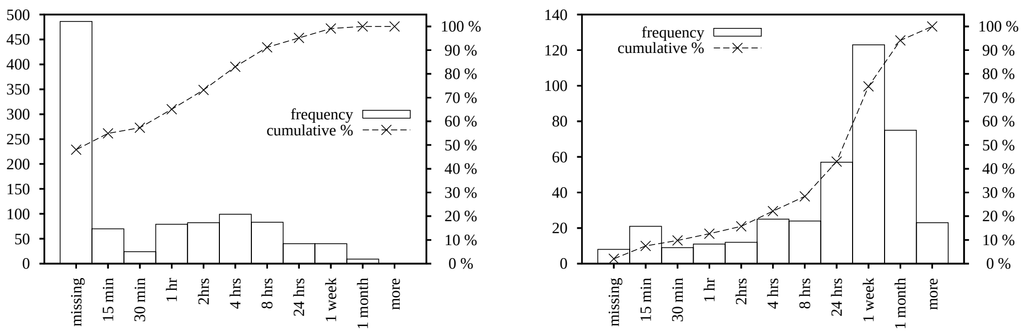
[2]In case of the exponential distribution the maximum likelihood estimation of the parameter of the distribution is simply the reciprocal of the average of the samples.

[3]Event intensity may be computed as the total number of events over the length of the obsrvation period.



(a) CA*Net4 (b) SWITCH

Figure 3: Distribution of interval lengths between successive ticket opening times



(a) CA*Net4 (b) SWITCH

Figure 4: Distribution of ticket lifetimes

widely in the literature to model such interval distributions, is unable to capture the sample distributions. This is clearly demonstrated by the p-values, which are practically zero, and therefore the condition of accepting the null hypothesis ($\alpha < p$) is not true even if the allowed maximal probability of type I error, i.e., the significance level, denoted by α, is set to a very small value[4].

4.1.2 Distribution of ticket lifetimes

Average ticket lifetime in the CA*Net4 network is more or less the MTTR value used in the reliability literature, while in the SWITCH network it is an order of magnitude higher (c.f. Section 4.1).

In case of the CA*Net4 network ticket lifetime typically corresponds to the beginning and termination of recorded events, as opposed to the SWITCH network, which is also demonstrated by the distributions on Figure 4.

The goodness-of-fit tests yield the conclusion that in case of the CA*Net4 network time data does not follow the exponential distribution, whereas in the SWITCH network depending on the selected significance level the null hypothesis may even be accepted. However, data from this network is rather descriptive of human behavior than equipment and cable feilure statistics.

[4] A type I error occurs in hypothesis testing when the null hypothesis is rejected even though it is, in fact, true.

	CA*Net4	**SWITCH**
scheduled events	55.04%	57.47%
field events	4.25%	6.44%

Table 4: Overview of some event categories

Due to the nature of the chi-square test it is possible to identify the places where the sample distribution significantly differs from the assumed distribution. The distribution of sample data from both sources has a heavier head than that of the exponential distribution, which is the most significant difference at the same time. One possible explanation is the following.

The length of pre-scheduled downtime follows a different distribution, as reaction time is missing from the total downtime. It means that pre-scheduled events are likely to entail shorter downtimes, which justifies the relatively high frequency of tickets that record short events[5].

4.2 Categorization

Before proceeding to ticket categorization, the reader is reminded that each record in the trouble ticket database documents an event that entails increased risk to the operation of the network. Only part of the records describe experienced decrease in the quality or outage of any service. In other words, opening a trouble ticket does not automatically mean that there is an outage in the network. This may be explained by the fact that in both networks application of resilience mechanisms, such as protection switching, prevented losses in some cases.

Table 4 shows that the ratio of field events is only about 5%. This is a warning sign that the assumption of perfect node equipment connected by failure-prone links, frequently applied in the literature for network modeling, may fail to capture all the relevant events during the operation of the network. Note, however, that different events have a significantly different impact on the network (in terms of overall performance measures). For example the amount of traffic affected by a cable cut is probably by orders of magnitude more than that affected by a line card failure, but even the latter may affect a whole subnetwork.

Table 4 also shows that more than half of the recorded events are scheduled. Obviously, these events should not catch the network operator off guard and so potential losses may be minimized. Another important property of these events is that since they are based on *a priori* information of the operator, the contents of the trouble tickets are written by the staff, i.e., networking professionals. If the trouble ticket system is open to the users of the network by means of e.g. a web interface, itmay not be the case. Therefore, in what follows the categorization results for scheduled and unexpected events are presented separately.

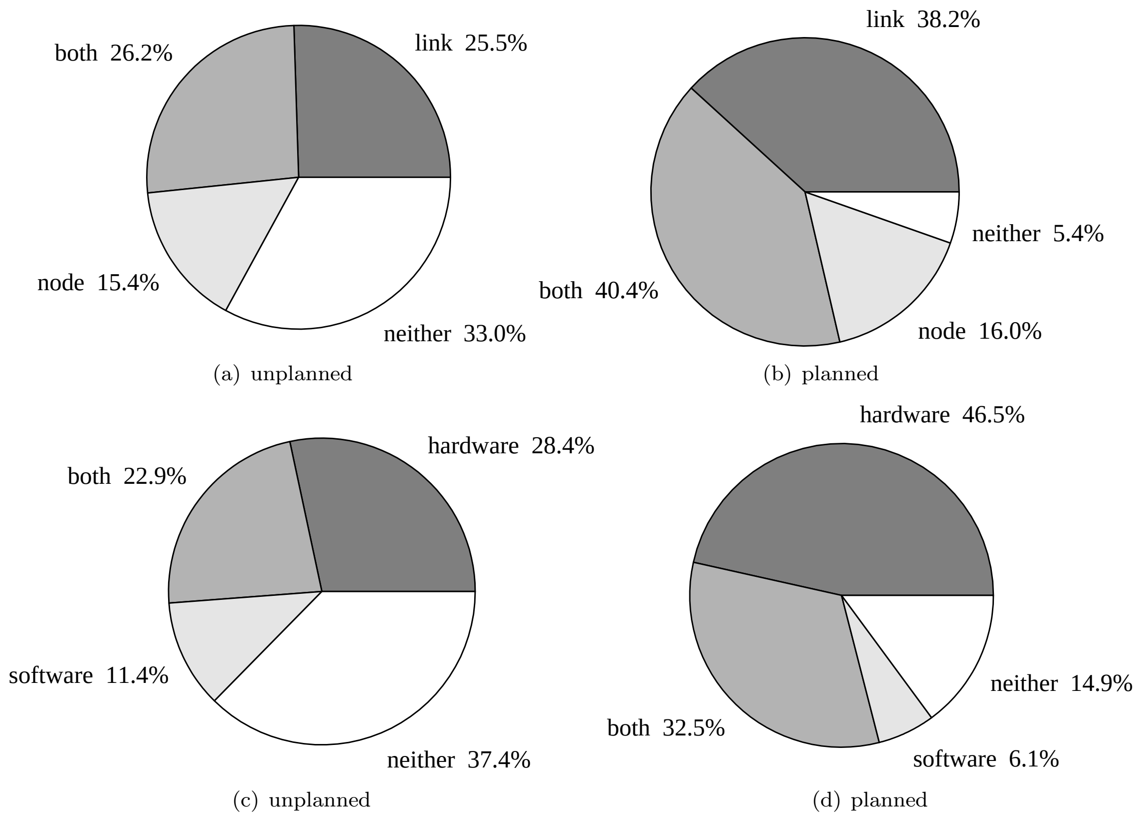
Figure 5 shows categorization results for the CA*Net4 network. Results for the SWITCH network are presented on Figure 6.

The most apparent phenomenon is that in both networks the ratio of tickets that cannot be categorized is significantly higher among unexpected events. This seems to support the earlier reasoning about ticket contents and opening of tickets.

The ratio of strictly node related events is more or less the same in both networks regardless of whether the records pertain to scheduled or unexpected events. In the SWITCH network the same is experienced with the ratios of strictly link related and both node and link related events. However, in the CA*Net4 network the latter two ratios significantly differ for unexpected and scheduled events.

Based on the existence of allusions to hardware and/or software components it is also interesting to examine the content of tickets. The ratio of strictly hardware related and both hardware and software

[5]With lifetime less than 15 minutes.



(a) unplanned (b) planned

(c) unplanned (d) planned

Figure 5: CA*Net4 trouble ticket categorization results

related events is significantly higher in case of scheduled events. It is probably due to the fact that the motivation for maintenance operations is very often the replacement or repair of some component. In contrast to this, strictly sofware related events are more frequent among unexpected events, which may in part be explained by the imperfect network operating practice and the hard-to-predict reliability performance of software components. Nevertheless, the data confirm that the frequency of software related events is not negligible at all, which further reinforces the need to model node failures, including software failures.

As to the data from the SWITCH network, in general, the same observations may be made (c.f. Figure 6). Note, however, that in this network the difference between the categorization results of unexpected and scheduled events is a lot smaller.

5 Conclusions

Based on the observations of the analysis some relevant conclusions may be drawn.

5.1 Properties of a Reasonable Trouble Ticket System

Some properties of the examined data sets suggest some basic principles of operating a reasonable trouble ticket system.

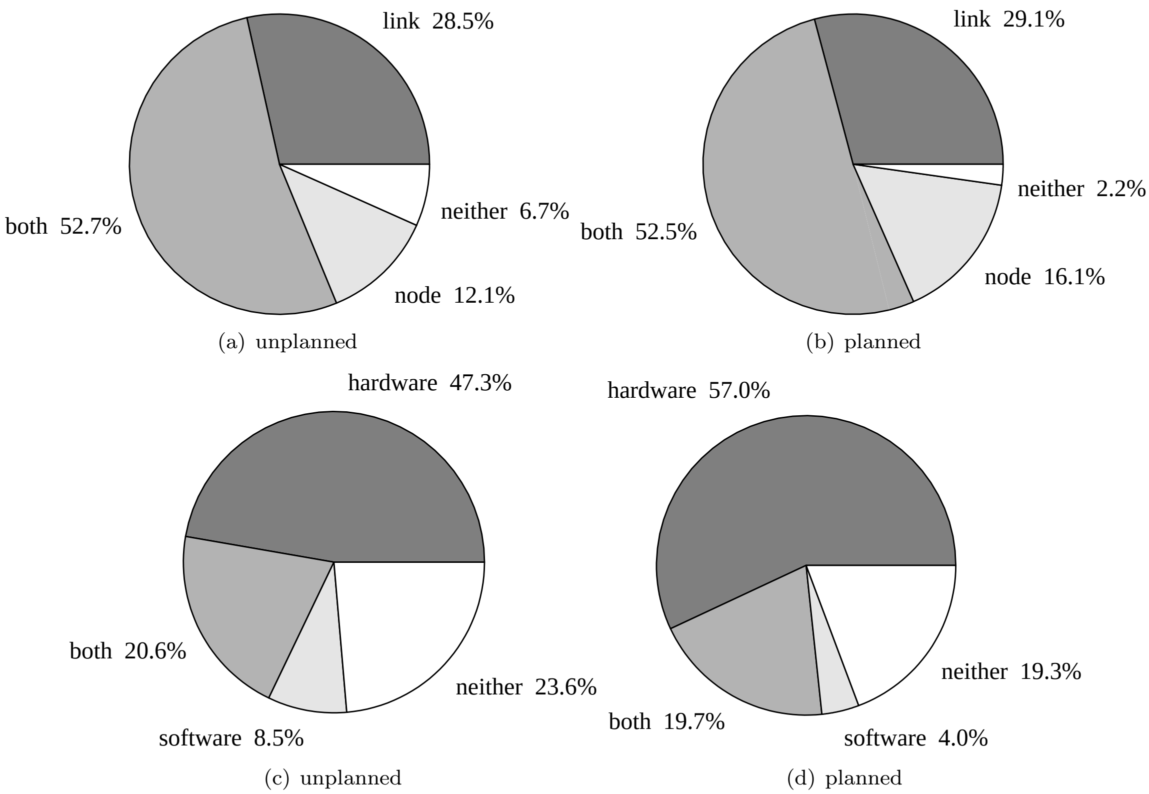


Figure 6: SWITCH trouble ticket categorization results

The sheer fact that analysis of text written by humans is necessary for the statistical analysis underlines the importance of appropriate categorization of records *at the time of creation* or, if it is not possible due to the lack of knowledge about the exact cause at the latest *when the tickets are closed*.

Moreover, it is advisable to separate ticket opening and closing times from beginning and termination times of the event recorded in the ticket. This way it is easier to treat differently administrative reaction times, including tickets left open for no particular reason. Another proposal to be discussed below also facilitates this.

It is very important that write access to ticket contents be restricted to members of the staff of the network operator, and that tickets may not be closed without the complete and accurate filling of their standard fields. Enforcing this policy would greatly improve the quality of ticket data, and the number of open tickets would characterize remaining work of operators better.

Exact, comprehensive and detailed category definitions are necessary for the categorization of equipment, services and users affected by events. However, it is inevitable to ensure that the system remains extensible by enabling the addition of categories in an on-demand manner, as well as to support attachment of log messages. The content of the ticket status field should provide a refined qualification instead of open/closed. The potential contents of this field may refer to both the status of the recorded event and that of the (re)actions. Timestamps should also be saved along with changes of the status field.

These observations are mostly in line with the recommendations of [1].

5.2 Research Aspects

As there are no guarantees concerning the completeness and the accuracy of records, and the exact equipment and cable list is not public, it is not possible to directly compare statistical data with values used in the reliability modeling literature.

The examined data, however, do support the observation that node equipment related and/or software failures also need to be modeled, and their effects and frequency need to be quantified by means of collecting statistics. The retionale behind this is that these events occur relatively often, and especially software related ones may cause significant losses. It is also observed that failure models using exponential distributions for approximation may not be accurate to a satisfactory extent to capture the behavior of real processes.

The bottom line is, however, that the toruble ticket systems of both CA*Net4 and SWITCH, for all their faults, should be considered as exemplary and important initiatives.

Acknowledgements

The author wishes to acknowledge the work of network operators of both CA*Net4 and SWITCH in the first place, who filled the trouble ticket systems with relevant data over the course of the past years and eventually shared it on the Internet. The author would also like to thank the devoted work of Ádám Ákos Mitterer and Gergely Bencsik, who downloaded and preprocessed the data, and Tivadar Jakab for showing the data sources and giving some initial ideas. The work was supported by grants from BUTE ETIK KKK.

References

[1] D. Johnson, *NOC Internal Integrated Trouble Ticket System Functional Specification Wishlist ("NOC TT REQUIREMENTS")*, RFC1297, January 1992

[2] `http://dooka.canet4.net`

[3] `http://www.switch.ch`