

A24544

EVENT CORRELATION

This invention relates to the operation of equipment management systems.

5           These systems are used for monitoring, and in some cases also controlling, the operating states of interacting equipments such as are found in communications networks or large industrial complexes. The present invention was devised with telecommunications networks in mind, but it can  
10 be applied to other systems having a large number of interacting equipments. In an equipment management system, monitoring devices sense the operating states of the various equipments and send signals to an equipment manager in response to significant events in the operation of their  
15 respective equipments. Signals may also be sent by human agency to report conditions in the equipment. The term "event" will be used in this specification to signify any state, or change in state, causing such a signal to be sent. In general these events will be faults and the monitoring  
20 devices will be fault detectors for sending alarm signals to the equipment management system in the event of a fault being detected.

For example, in a telecommunications network, an alarm signal caused by the failure of a switching centre would  
25 alert the system manager who would arrange for alternative routings to be made and who would also arrange for any necessary repair work to be done on the faulty switching centre. In some systems these responses may be automated, but more usually the faults will require human intervention,  
30 the system merely providing to the system manager details of any faults requiring attention. This allows the manager to organise the available resources efficiently, taking into account factors such as safety-criticality, priority, and the whereabouts of field staff.

35           This also allows non-significant alarms with known causes, such as those caused by equipment having been

---

This Sheet Has Been Replaced



disconnected for routine maintenance, to be disregarded by the system manager.

It is well established that effects of a network failure, such as the failure of a high bit rate line system, will propagate down through a hierarchy of dependent resources and initiate many nearly simultaneous alarm messages. Time and other resources can be wasted by investigating the sources of all the alarms if the underlying cause has not been identified.

10 A fault may affect all equipment directly connected to the source of the fault, or all equipment at one geographic location, (although they may be topologically remote from each other) or all equipment of a specific type. For example, external-radio interference could affect all  
15 radio links in the network operating on a specific frequency within radio range of the source of interference although, from the point of view of connectivity of the network, they may appear remote from each other. Based upon this understanding it is known that alarm messages that occur with  
20 close temporal proximity will tend to be associated or correlated.

The size of any temporal window within which alarm messages are considered as correlated has to be optimised - if it is set too large, the chance of alarm messages from  
25 unconnected resources arriving within the temporal window increases; if it is set too small, only a subset of the dependent alarm messages might arrive within the temporal window.

In certain circumstances, related faults may only be  
30 detected by the system at a later time. For example, this will be the case if an equipment not in operation at the time of the initial fault later attempts to establish contact with the failed equipment.

The operator of the network will be able to determine  
35 an appropriate window size according to the characteristics of the network. This will depend on the nature of the network and the faults being monitored within it. In

This Sheet Has Been Replaced



circumstances where faults are reported to the system by human agency rather than automatically, or in which their exact timing is difficult to measure, the window may be of the order of hours, or even days, whereas for systems continuously and automatically monitored a suitable time window may be measured in seconds.

Of course, the mere fact that two events have occurred together on one occasion is not necessarily indicative that there is a causal relationship between them. Whatever optimised temporal window is selected, there remains a finite probability that two or more independent alarm hierarchies will report within the same temporal window.

For convenience in the following discussion, events occurring in the same temporal window will be referred to as simultaneous.

Although temporal correlation within a single window is a simple technique to implement, it provides no great confidence that a dependency actually exists between alarm messages. Simultaneity cannot prove dependency and hence cannot identify a cause and effect relationship between alarm messages.

In known equipment management systems these problems have been approached by building up a computer model of the system, or relying on the knowledge of the operators. This requires detailed and up to date knowledge of the system and its accuracy depends on recognition on all of the factors which may be involved. Moreover, any such model would be specific to the individual equipment, and have no general applicability. The present invention addresses these disadvantages by employing an empirical approach to identification of related alarm events.

According to a first aspect of the present invention, there is provided a method of identifying likely relationships between events occurring in equipment managed by an equipment management system which includes: storing historical data concerning the times that specified events occur in the equipment during a reference time period;

This Sheet Has Been Replaced



identifying further occurrences of the specified events, analyzing the historical data to determine a measure of the relatedness between the further event occurrences, and providing the measure to an output means of the equipment management system.

This invention removes the need to build up a model of the system in advance. In one arrangement according to the invention, the historical data may be updated, allowing the method to "learn", allowing it to improve in accuracy as the database increases in size and to keep up with changes made to the system. It thus identifies events which are likely to be related, in the sense either that one is a direct consequence of the other, or that they have some common underlying cause. Although the events monitored will usually be fault alarms generated by devices monitoring the components of the equipment being managed, other events which could be related to such faults may also be input to the system, such as data from meteorological sensors, or faults reported to the management system by human agency, e.g. in response to a report from a remote user.

Significant power can be added to the prior art single window temporal correlation by using historical event data in this way to identify a probable relationship between event occurrences. Such dependencies can be identified through the use of a historical temporal correlation technique, according to the present invention, without prior knowledge of any relationship between resources.

Preferably a predetermined temporal window is defined and the further events are selected from the events occurring within the temporal window.

The measure of relatedness is preferably the statistical probability of the further events having occurred within the same temporal window by chance.

In one arrangement the relatedness of a selected event with two or more other events is measured, and the two or more other events are ranked in order of said statistical probabilities.



The output means is preferably a display, and events are selected for display only if said statistical probability is below a threshold value.

In such a display, displayed events may be highlighted  
5 if said statistical probability is below a threshold value, which may be a predetermined value, or may be selected according to criteria dependent on the range or pattern of said statistical probabilities.

The selected event is preferably selectable from  
10 amongst the different events identifiable by the system.

In a preferred embodiment the equipment management system monitors the equipment to be managed and detects the occurrence of the events.

However, at least some of the events may be identified  
15 to the system by human agency.

The historical data may be stored by

- selecting a reference time period, and
- dividing the reference time period into temporal windows of a predetermined size, and
- 20 - for each event to be analyzed, storing the identity of each temporal window within the reference time period during which that event occurs.

The analysis procedure may proceed as follows:-

- 25 - one of the temporal windows is selected to be analyzed;
- a first event occurring in the selected temporal window is selected;
- other events occurring in the selected temporal window  
30 are identified; and
- for each such other event, the stored identities of temporal windows in the reference time period in which the events occurred are used to calculate a statistical probability of such other event occurring by chance in the  
35 same temporal window as said first event.

An alternative analysis procedure is as follows:-

This Sheet Has Been Replaced



one of the temporal windows is selected to be analyzed; a first event occurring in the selected temporal window is selected;

5 other events occurring in a second temporal window related to the selected temporal window by a predetermined relationship are identified; and

10 for each such other event, the stored identities of temporal windows in the reference time period in which the events occurred are used to calculate a statistical probability of such other event occurring by chance in a temporal window having the same predetermined relationship to a temporal window in which the first event occurs.

The predetermined relationship is preferably a time difference.

15 The measure of probability of the events  $AR_m$  and  $AR_n$  occurring by chance in the same temporal window can be determined by: -

$$\frac{k!}{p! (k-p)!} r^p (1-r)^{(k-p)}$$

where

20 k = number of temporal windows in which an event of the first class occurs.  
r = number of temporal windows in which an event of the other class occurs.  
25 p = number of temporal windows in which events of both the first class and the other class occur together.

The reference time period may be fixed, or it may be continuously updated.

30 The events are, for example, fault conditions in a system.

Analysis may be done retrospectively, after the faults have been cleared, in order to identify the root cause of the fault in order to identify the unreliable component causing



the others to fail. Alternatively, analysis may be done in real time, during the fault condition, in order to locate and repair the cause of the failure, which may result in the clearance of the other fault reports without a technician's attendance being required.

The invention relies, in a preferred embodiment, on an analysis of the statistical probability of the simultaneous occurrence of the events. In essence, it relies on the fact that there is a very low probability of two independent events, both themselves rare, occurring simultaneously. It follows that if there is nevertheless a history of two rare events occurring simultaneously, it is probable that the events are not independent, and that there is a relationship between them. In this method the measure is the probability of the two events occurring by chance simultaneously. The method may be used to compare events selected pair-wise by the operator for possible matches, but in another arrangement the comparison is made between all events occurring simultaneously, which are then ranked in order of their calculated probability of having occurred together at one time by chance. These will generally fall into two groups, those with a high probability of having occurred together by chance, and those with a much lower probability, indicating that their simultaneity is unlikely to be a random coincidence, and that it is likely that there is a relationship between them. For a large system with many potential relationships to be assessed, some way of distinguishing such events is desirable.

The method may include the display of events in rank order, highlighting correlated events on a screen display, or displaying only events which are correlated. The threshold level below which a causal relationship is likely may be predetermined, or it may be varied depending on the pattern of probabilities determined on a case-by-case basis. This threshold level can determine which events are to be displayed or highlighted. Two or more thresholds may be



used, to identify different levels of certainty in the correlation.

Although the reference time period may be fixed it is preferable for the data to be continuously updated. This  
5 allows the system to learn and as the database becomes larger, it should become more reliable.

In some circumstances there may be a systematic delay between a report of a cause and a report of an effect, or between reports of two effects from a common cause. In  
10 systems where this is likely, the method may be repeated with the time windows staggered for the two events to be compared, for example such that an occurrence of a first event in a given time window is compared with an occurrence of a second event in the immediately previous window.

15 According to a second aspect of the invention, there is provided an equipment management system comprising monitoring means for identifying events occurring in the equipment to be managed by the system; storage means for storing historical data relating to such events; input means  
20 for selecting an event for correlation from among those identified by the monitoring means; correlation means for correlating the selected event with the other events identified by the monitoring means using the historical data stored in the storage means; and display means for displaying  
25 the correlations determined by the correlation means.

The system may include updating means for allowing the data held in the monitoring means to be added to that in the storage means, thereby allowing the data in the storage means to be kept up to date. The correlation means preferably  
30 performs a statistical correlation.

Means may be provided for allowing events to be reported to the monitoring means by human agency, eg. by a keyboard. This may be necessary, for example, if an alarm itself failed.

35 The display means preferably displays the events in such a way that correlated events can be visually

This Sheet Has Been Replaced



distinguished from uncorrelated events, for example by highlighting them.

The invention will now be described by way of example with reference to the accompanying drawings, in which: -

5        Figure 1 is a diagrammatic representation of a simple network to which an embodiment of the method of the invention is to be applied for illustrative purposes to determine the interdependency;

10        Figure 2 is a diagrammatic representation of illustrative historical alarm data for the network of figure 1;

      - Figure 3 is a diagrammatic representation showing the network of figure 1 with the interdependencies determined by the embodiment of the method of the invention;

5        Figures 4a 4b and 4c show dependency 'league tables' representative of the results obtainable by the method of the invention;

      - Figure 5 is a representation of a display produced by the method of the invention, performing a second exemplary correlation according to the invention; and

10        Figure 6 is a diagrammatic representation of an apparatus for performing an embodiment of the invention;

      To assist the understanding of this description of embodiments of the invention it will first be applied to the simple network of Figure 1 which has only six resources  
15        A, B, C, D, E and F and a reference time period divided into only ten temporal windows ( $t-9$  to  $t_0$ ).

      AR1, AR2, AR3, AR4, AR5, AR6 are the possible alarms generated by the resources.

20        Figure 2 shows an exemplary historical sequence of the alarm messages which are taken to have arrived from each resource over the reference time period.

      Following the method of the invention, an event occurrence is selected by the operator for analysis. In this example alarm AR1 occurring in temporal window  $t_0$  is  
25        selected. In this example all other events AR2 to AR6 occurred, simultaneously with AR1, in this window.



The probability (r) of each event occurring is obtained from the historical data by calculating the number of windows in which it occurred (X) divided by the total number of windows in the reference time period (n), i.e. 0.5 for AR1, AR2 and AR5; 0.1 for AR3 and AR6 and 0.9 for AR4.

Events AR2, AR4 and AR5 each occur simultaneously with AR1 five times, whilst AR3 and AR6 occur simultaneously with AR1 only once.

A dependency between two resources will tend to exist if their respective historical alarm sequences have a low probability of being similar by pure chance.

The probability of two historical alarm sequences,  $AR_m(t)$  and  $AR_n(t)$ , being similar by chance may be calculated from the Binomial distribution:

15

$$P(AR_m(t) \text{ and } AR_n(t) \text{ similar}) =$$

$$\frac{k!}{p! (k-p)!} r^p (1-r)^{(k-p)}$$

where:

k = Number of temporal windows where  $AR_n$  is active.

20 p = Number of temporal windows where  $AR_m$  and  $AR_n$  are active

r = Probability of  $AR_m$  being active

when testing dependency of  $AR_n$  upon  $AR_m$

r is determined empirically by the method, by calculating x/n, where n is the total number of temporal windows, and x is the number of windows in which  $AR_m$  is active. Having calculated the probability of each pair of alarms  $AR_m$ ,  $AR_n$  occurring together by chance, the pairings can be ranked in order of their probability to form a 'league table' of relatedness.

30

The probabilities of correlation as given by the formula above are shown in the tables of figure 4a to 4c, in increasing order of size. It will thus be readily seen from



figure 4a that the least likely alarms to have occurred simultaneously with AR1 by chance (and thus the most likely to be related), are AR2 and AR5.

A decision thresholding can be applied to determine where the break point between a dependent and non-dependent probability should lie. This threshold may be pre-determined, or it may be calculated on the basis of the clustering of results. For example, figure 4a shows the dependency league table produced by performing the historical temporal correlation of AR1(t) with all resources shown in figure 1.

AR2 and AR5 can be shown to fall within the same dependency hierarchy as AR1 itself.

The double line in figure 4a indicates where the dependent/non-dependent threshold should be applied. The probability of alarms being similar are clustered in two groups with a change by an order of magnitude across this threshold. In an operator's display, the area of the screen 2 above the threshold might be highlighted in some way. Alternatively, the correlated events may be the only ones displayed.

The method can be repeated selecting another alarm for comparison against the others.

For AR3 it can be determined that the probability of random correlation with AR6 is 0.1, whilst the probability of random correlation with any other resource is greater than 0.3 as shown in figure 4b. Similarly, using AR4, it can be determined that all correlations have probabilities of random occurrence of at least 0.24. It can thus be seen that event AR4 is not correlated with any other. These results are shown in figure 4c.

In the simple network of figure 1, one can see from visual inspection of figure 2 that these results are intuitively reasonable. A visual inspection of all historical alarm messages shows that the historical pattern of AR1(t) is similar to those of both AR2(t) and AR5(t) but is very different from those of AR3(t), AR4(t) and AR6(t).



AR3(t) has a pattern similar to that of AR6(t) but is very different to all other alarm sequences.

Although all resource alarms were temporally correlated in window t0, the historical temporal correlation of these alarm sequences would show that AR1 and AR2 were probably related to AR5 and that AR3 was probably related to AR6 (figure 3). No alarms appeared to be related to AR4.

However, in more complex systems having perhaps many hundreds of resources, such visual analysis is impossible.

Moreover, in real networks alarm signals occur at a much lower frequency than in the example and thus the number of time frames needed to provide a statistically useful database will be very large. Furthermore, although it would be possible for a skilled operator to manually sub-group many of these alarms, the data is not presented in prior art systems in a way that enables the most significant alarm (ie. the one upon which most other alarms are dependent upon) to be identified. The method of the invention is thus particularly suited for use in large systems where the number of possible correlations is too great to group intuitively.

When the probability of an event occurring (r) is small the Binomial distribution used in the first example approximates to the Poisson distribution:

$$P(\text{AR}_m(t) \text{ and } \text{AR}_n(t) \text{ similar}) = \frac{(kr)^P}{P!} e^{-(kr)}$$

As a second example, using real data, a one-off historical temporal correlation was performed between a selected resource in the BT Network Monitoring System (NETMON) which reported a fault at 10:53:43 on a certain date, and all other resources in the system reporting within the same window. The historical database comprises all December 1990 alarm data (fault reports), using a 150 second window size, thus giving 17,856 windows. The national NETMON database for December 1990 holds about  $2 \times 10^6$  alarms from about 40,000 resources. Clearly the identification of historical



patterns from such quantities of data cannot be done by simple inspection

In this example a threshold value will be determined in advance. From the figures above it can be seen that the  
5 average number of alarms per resource in the reference period is approx.  $2,000,000/40,000=50$ , and the average probability of occurrence in any given temporal window is thus  $50/17856=0.0028$ . Thus the probability of any randomly selected pair of resources reporting simultaneously in a  
10 given window is of the order of  $(0.0028)^2=7.8 \times 10^{-6}$ . There are  $(40,000)^2 = 1.6 \times 10^9$  possible pairings of resources, so there will be approximately 12,500 random correlations in any given temporal window.

To avoid being overwhelmed by these random  
15 correlations, a threshold value is chosen which only reports the most statistically significant correlations. In the following example a threshold probability of  $10^{-8}$  is used.

Figure 5 shows the top of the dependency "league table" resulting from performing the correlation method  
20 according to the invention on this data. This may be displayed on screen 1 (Figure 6). The top entry in the league tables is in respect of the fault reported by the selected resource. The threshold value lies off the bottom of this fragment of the table.

25 The table has five columns. The first column indicates the region in which the fault is located: NE = North East, S = Scotland, M = Midlands, L = London, NW = North West.

The second column specifies the actual origin of the  
30 alarm.

The third column indicates the nature of the fault prefixed by a two letter code indicating the location of the fault. (Note that the system may be alerted to a fault at a location remote from the fault itself.

35 The fourth column calculates the probability of the alarm fault occurring by chance.

The fifth column gives the correlation probability.



From their positions in this league table, it becomes easy to identify many associations between resources within an alarm dependency hierarchy, which would not otherwise be apparent.

5           Faults within this hierarchy are seen having the prefix codes (column 3) for Leicester (LE), Leeds (LS), Sheffield (SF), Edinburgh (EH), Cambridge (CB), London (L), Manchester (MR), etc. Some of the correlations revealed by this league table have obvious causes: for example the first  
10 thirteen all occur in the same location as the selected source. However, it is also possible to see that alarms emanate from three line systems all radiating from the same place coded LE/D (Leicester D): Leicester D - Derby F, Leicester D - Leeds G and Leicester D - Sheffield E. The  
15 apex of this dependency hierarchy appears to be a power related effect.

The results from this exemplary historical temporal correlation suggest that a problem with the power supply at the location coded LE/D is responsible for many alarms  
20 occurring over a wide geographical area. This conclusion could not have been drawn by looking at the results of a prior art single window temporal correlation because of the large number of unrelated alarms also present, but can be identified more easily from the ranking determined by the  
25 method of the invention.

Post analysis revealed that in this case a fault was indeed present in the power supply to part of the network in the Leicester area.

The principles of historical temporal correlation used  
30 in the present invention can be used by a network management system as a technique for self-learning of alarm dependencies. Artificial neural network principles provide one framework within which the self learning of alarm dependencies could function.

35           By recording correlations as they are identified, the system can build up a computer model of the dependencies so

This Sheet Has Been Replaced



that when particular patterns of alarms next occur it can more readily identify alarms as being related.

A network management system according to the invention is shown in figure 6. There is shown a network of resources  
5 R1 to R9 having a number of interconnections. Resources R1 to R8 have respective monitors AR1 to AR8 which report alarm conditions to the network monitoring means 2 for display to the operator 3 on screen 1. Resource R9 is not directly connected to the management system but field operative 4  
10 discovering a fault occurring in resource R9 can advise the operator 3, e.g. by telephone connection T, so that the fault condition can be reported to the monitoring means 2 by means of data input means 5 and input link 12.

Under the control of clock 9, the alarm conditions are  
15 periodically reported through updating link 11 to store 6, with their time of occurrence.

On observing from screen 1 a fault with which he or she wishes to correlate other faults, the operator 3 may use the input means 5 to select an alarm condition to be  
20 correlated. The information on the current alarm conditions is extracted from the monitoring means 2 by correlation means 7. Correlation means 7 extracts the historical data from store 6 and performs a statistical analysis (as described above) to calculate for each alarm currently reported to the  
25 monitoring means 2, the theoretical probability of it occurring at the same time as the selected alarm. The alarms are displayed in ascending order of probability on screen 1. Alarms having a probability of correlation below a predetermined value are identified by highlighting 8.

30 The monitoring means 2, screen 1, input link 12, clock 9, updating link 11, store 6 and correlating means 7 may be implemented as a computer provided with appropriate software.

---

This Sheet Has Been Replaced





CLAIMS

1. A method of identifying likely relationships between events occurring in equipment managed by an equipment management system, which method includes: storing historical  
5 data concerning the times that specified events occur in the equipment during a reference time-period; identifying further occurrences of the specified events; analyzing the historical data to determine a measure of the relatedness between the further event occurrences, and providing the  
10 measure to an output means of the equipment management system.
2. A method according to claim 1, wherein a predetermined temporal window is defined and the further events are selected from the events occurring within the temporal  
15 window.
3. A method according to claim 2, wherein the measure of relatedness is the statistical probability of the further events having occurred within the same temporal window by chance.
- 20 4. A method according to claim 3 wherein the relatedness of a selected event with two or more other events is measured, and the two or more other events are ranked in order of said statistical probabilities.
5. A method according to claim 3 or 4 wherein the output  
25 means is a display, and events are selected for display only if said statistical probability is below a threshold value.
6. A method according to claim 4 or 5, wherein the output means is a display and displayed events are highlighted if said statistical probability is below a threshold value.





7. A method according to claim 5 or 6, wherein the threshold is a predetermined value.
8. A method according to claim 5 or 6, wherein the threshold is selected according to criteria dependent on the  
5 range or pattern of said statistical probabilities.
9. A method according to any one of claims 4 to 8, wherein the selected event is selectable from amongst the different events identifiable by the system.
10. - A method according to any preceding claim, wherein the  
10 equipment management system monitors the equipment to be managed and detects the occurrence of the events.
11. - A method according to any preceding claim, wherein at least some of the events are identified to the system by human agency.
- 15 12. A method according to any of claims 3 to 10, wherein the historical data is stored by  
selecting a reference time period, and  
dividing the reference time period into temporal  
- windows of a predetermined size, and  
20 for each event to be analyzed, storing the identity of each temporal window within the reference time period during which that event occurs.
13. A method according to Claim 12, in which one of the  
25 temporal windows is selected to be analyzed;  
a first event occurring in the selected temporal window is selected;  
other events occurring in the selected temporal window are identified;  
30 and, for each such other event, the stored identities of temporal windows in the reference time period in which the events occurred are used to calculate a statistical



probability of such other event occurring by chance in the same temporal window as said first event.

14. A method according to Claim 12, in which one of the temporal windows is selected to be analyzed; a first event  
5 occurring in the selected temporal window is selected; other events occurring in a second temporal window related to the selected temporal window by a predetermined relationship are identified;  
and, for each such other event, the stored identities of  
10 temporal windows in the reference time period in which the events occurred are used to calculate a statistical probability of such other event occurring by chance in a temporal window having the same predetermined relationship to a temporal window in which the first event occurs.

15 15. A method according to Claim 14 in which the predetermined relationship is a time difference.

16. A method according to claim 13, 14 or 15, wherein the statistical probability of the events  $AR_m$  and  $AR_n$  occurring by chance in the same temporal window is given by

$$\frac{k!}{p! (k-p)!} r^p (1-r)^{(k-p)}$$

20 where

$k$  = number of temporal windows in which  $AR_m$  occurs.  
 $r$  = probability of event  $AR_m$  being active.  
 $p$  = number of temporal windows in which events  $AR_m$  and  
 $AR_n$  both occur when testing depending of  $AR_n$  upon  
25  $AR_m$

17. A method according to any of Claims 12 to 16 wherein the reference time period is fixed.





18. A method according to any of claims 12 to 16, wherein the reference time period is continuously updated.

19. A method according to any preceding claim, wherein the events are fault conditions in a system.

5 20. A method of operating an equipment management system in order to identify relationships between two or more events substantially as described.

21. An equipment management system comprising monitoring means(2) for identifying events occurring in the  
10 equipment(AR1-AR9) to be managed by the system; storage means(6) for storing historical data relating to such events; input means(5) for selecting an event for correlation from among those identified by the monitoring means(2); correlation means(7) for correlating the selected event with  
15 the other events identified by the monitoring means(7) using the historical data stored in the storage means(6); and display means(1) for displaying the correlations determined by the correlation means.

22. A system as claimed in claim 21, including updating  
20 means(11) for periodically supplying data held in the monitoring means(7) to the storage means(6).

23. A system as claimed in claim 21 or 22 wherein the correlation means performs a statistical correlation.

24. A system as claimed in claim 21, 22 or 23 further  
25 comprising input means(5,12) for reporting events to the monitoring means(2) by human agency.

25. A system as claimed in any one of claims 21 to 24 wherein the display means(1) includes means for visually distinguishing(8) events which are correlated from those  
30 which are not correlated.

