HACKER: Human And Computer Knowledge discovered Event Rules for Telecommunications Fault Management

Roy Sterritt, Edwin P. Curran, Hongzhi Song

School of Information and Software Engineering, Faculty of Informatics, University of Ulster

Jordanstown, Newtownabbey, Northern Ireland, BT37 0QB, UK

{r.sterritt, ep.curran, gh.song}@ulster.ac.uk

Abstract--**Visualization integrated with data mining can offer 'human-assisted computer discovery' and 'computer-assisted human discovery'. Such a visual environment, reduces the time to understand complex data, thus enabling practical solutions to manyr eal world problems to be developed far more rapidly than either human or computer operating independently. In doing so the remarkable perceptual abilities that humans possess can be utilized, such as the capacity to recognize images quickly, and detect the subtlest changesi n size, color, shape, movement or texture. One such complex real world problem is fault management in global telecommunication systems. These systems have a large amount of built-in redundancy to ensure robustness and quality of service. Unfortunately, this means that when a fault does occur, it can trigger a cascade of alarm events as individual parts of the system discover and report failure making it difficult to locate the origin of the fault. Thisa larm behavior has been described as appearing to an operator as non-deterministic, yet itd oes resulti n a large data mountain thati s ideal for data mining. This paper presents a visualization data mining prototype that incorporates the principles of Human and Computer Discovery, the combination of computer-assisted human discovery with human-assisted computer discovery through a three-tier framework. Thep rototype is specifically designed to assisti n the semi-automatic discovery of previously unknown alarm rules that can then beu tilized in commercial rule based component solutions, "Business Rules", which are at the heart of many of todays fault management systems.**

Keywords:**Visual Data Mining, Knowledge Discovery, Data Visualization, Integration, Human -assisted Computer Discovery, Computer-assisted Human Discovery.**

I. INTRODUCTION

Visualization involves constructing graphical interfaces that enable humans to understand complex data sets. Visualization may be viewed as the link between the two most powerful information processing systems: Humans and the modern computer [6].

Uthurusamy (1996) proposed that a challenge of paramount importance for KDD was to have a more human-centered view. Pressing the need for highly interactive human-centered environments as outlined by the KDD process [2] would enable both *human-assisted computer discovery* and *computer-assisted human discovery*. Such tools would reduce the time to understand complex data sets would enable practical solutions to many real world problems far more rapidly than either human or computer operating independently [21].

Interest in the field of integrating KDD and visualization has been growing with these promises in mind and have been applied to various applications, in particular data exploration [7].

This paper presents at ool that integrates visualization and data mining incorporating the principles of Human and Computer Discovery, that is thec ombination of computer-assisted human discovery with human-assisted computer discovery through a three-tier framework. The tool is specifically designed to assist in the semi-automatic discovery of previously unknown alarm rules that can then be utilized in commercial rule based component solutions, "BusinessR ules", which are at the heart of many of todays fault management systems.

II. FAULT MANAGEMENT DOMIAN

A. Computer Networks

As the world becomes increasingly reliant on computer networks the complexity of these networks has grown along a number ofd imensions [16]. The phenomenal growth of the Interneth as shown a clear example of the extent to which the use of computer networks is becoming ubiquitous [1]. As users demands and expectations on networks become more varied and complex so do the networks themselves. As such, heterogeneity has become the rule rather than the exception [16]. Data, in any form, voice, movie, or actual information, may travel under the control of different protocols through numerous physical devices manufactured and operated by large numbers of different vendors. There is a general consensus, in dealing with such data, the trend towards increasing complexity willc ontinue rather than abate.

The complexity lies in the accumulation of several factors; the embedded increasing complexity of network elements, the need for sophisticated services and the heterogeneity challenges of customer networks [4].

B. Network Management

Network management encompasses a large number of tasks with various standards bodies specifying a formal organization of these tasks. The International Standards Organization (ISO) divides network management into six areas as part of the Open Systems Interconnection (OSI) model; configuration management, fault management, performance management, security management, accounting management and directory management which sit within a 7 layer network hierarchical structure.

Yet with the Internet revolution and the convergence of the Telcos and Data Comms other realities are forming. The trend is towards a flatter structure.

C. Faults, Fault Management and Alarm Correlation

Essentially, network faults can be classified intoh ardware and software faults, which cause elements to produce outputs, whichi nt urnc ause overall failure effects in the network such as congestion [22]. A single fault in a complex network can generate a cascade of events, potentially overloading a manager's console with information [15].

The fault management task can be characterised as detecting when network behavior deviates from the normal and in formulating a corrective course of action. Fault management can be decomposed into three tasks; fault identification, fault diagnosis and fault remediation [16]. A fourth task may be that off ault prediction, which may be a natural extension of faulti dentification [18].

Alarm correlation is a conceptual interpretation of multiple alarms such that a new meaning isa ssigned to these alarms [14] and potentially creating derived alarms [9].

D. The Domain Challenges

The principle aim behind alarm event correlation is the determination of the cause. The alarms represent the symptoms and as such, in the globals cheme of things, are not of significant interest once the failure in the network is determined [10].

The time to market and the R&D lifecycle of these products are continuously being squeezed while at the same time market demands for features and functionality increase with each release. It is also the nature of the domain that customers expect legacy support with the new systems as well as multi-vendor support.

This not only creates challenges for rule-based systems development but also creates a substantial rule-base maintenance burden [3]. As sucht echniques sucha s data mining to assist in discovery and development of rules in heterogeneous network environments is essential.

III. A THREE-TIER DISCOVERY PROCESS

It may be proposed that a flaw in data mining orK nowledge Discovery (KD) is that it is not user-centered [2] [8] [21]. A three-tier discovery process [17] is illustrated in Figure 1. Its aim ist he discovery of previously unknown and potentially useful association rules or in the exemplar domain alarm correlations directly from fault management data.

The objectives are to help avoid the traditional problems associated with RBS, namely the knowledge acquisition (KA) bottleneck and the maintenance burden [3].

The approach is more than a knowledged iscovery process since it is placing an equal weight against human participation. Obtaining a highly interactive human-computer environments hould facilitate practicals olutions to realw orld issues far more rapidlyt han either a human or computer operating independently could achieve [21]. It is hoped that combining the two into a process will assist in reaching the major goal of automation and understandability.

A. Computer-assisted Human Discovery

The aim is to discover hidden knowledge, unexpected patternsa nd new rules from data mountains. Visualization techniques of vast amounts of data allow the remarkable perceptual abilitiest hat humans possesst o be utilised, such as the capacity to recognise images quickly, and detect the subtlestc hanges in size, colour, shape, movemento r texture - and thus potentially discover new event correlations in the data. [17]

B. Human-assisted Computer Discovery

Data mining (discovery algorithms) may reveal hidden patterns and new rules yet these require human interpretation to transform them into knowledge. The human element attaches a more meaningful insight into the decisions allowing thed iscovered correlations to be coded as useful rules for fault identification and management. [17]

C. The Three-Tiers

Human-Computer Discovery, specifically computer-assisted human discovery and human-assisted computer discovery, can be incorporated together via a three tier process, providing an iterative processf or discovery and learning of rules for fault management.

The tiers are;

Tier 1 - Visualization Correlation

Tier 2 - Knowledge Acquisition or Rule Based Correlation

Tier 3 - Knowledge Discovery / Data Mining Correlation

The top tier (visualization correlation) has two distinct roles: (1) specifically to facilitate human discovery of correlations/potential rules (computer-assisted human discovery) and (2) visualizing the KD process - tier 3 (the human element of human-assisted computer discovery). The second tier (knowledge acquisition or rule-based correlation) like tier 1, has 2 distinct purposes: (1) discoveries of implicit orh idden knowledge from experts or documentation and (2) the validation of discoveries from tiers 1 and 3. That is, to move from discovered patterns(event correlations), be they through visualation or data mining, to knowledge (interruption, validation and coded rules) will require consultation with experts and/or documentation.

Thet hird tier (knowledge discovery correlation) mines the fault management data for more complex correlation rule candidates.

The application of the 3-tier framework is iterative and flexible in nature. As time goes on and the process is used andd evelopedi terativelyt he process moves from manual (human dependence) towards automation (computer dependence). For instance, the consultation of event consequence tables as specified by ITU by the expert to confirm that a discovery is in fact new knowledge may be coded in tier 2 as prior knowledge and utilsed when mining.

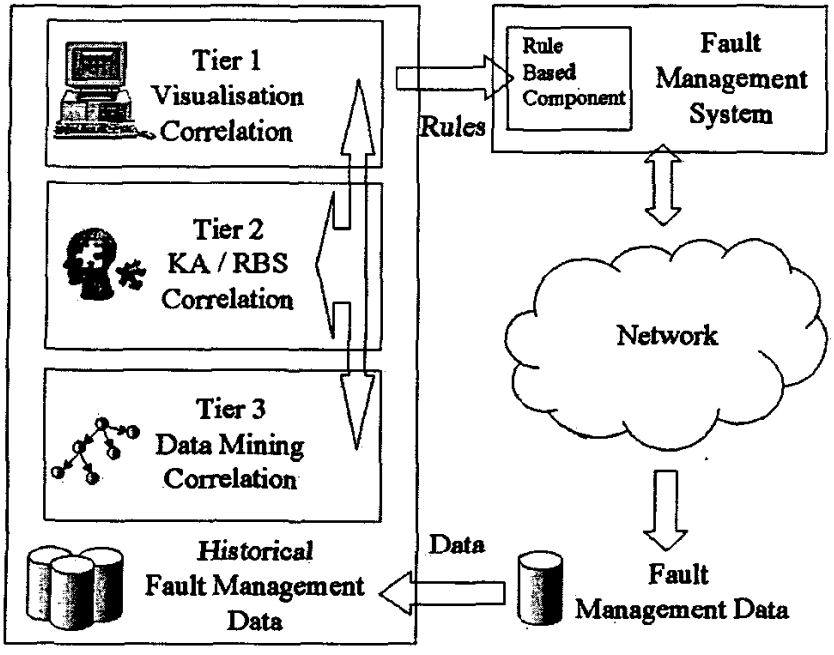

Figure 1 Three tier discovering process - Computer-Assisted Human Discovery and Human-Assisted Computer discovery

D. Business Rules

An increasing trend when developing FMSsi sf or component or APIs (Application Programming Interfaces) to providet he RBS capability. The market recently has seen many such 'BusinessR ules' component solutions and methodologiess uch as ILOG Rules, Blaze Advisor, USoft (now called Ness), Versata, Seec and Business Rules Solutions. Thel Log Rules approach is one of the most popular choices among Telco's with customers such as NortelN etworks [13].

As such, this move is more thanr educing development and programming time by utilizing COTS components. The 'Business Rules' approach is a move towards declarative programming from procedural [5]. To reduce the time-to-market further requires a refocus on automating as mucha s possible of the developmento f these declarative rules.

IV. The Tool

The starting point for the tool was an existing visualization tool [19], NxGantt - Stimuli-Event correlation Analyser (SEA) - Figure 2, that was utilised in tier 1. It was found to be a useful means for visualizing the large amounto f alarm and other event data found in the event logs from the element controller. It was used in the case study in [17] highlighting a computer-assisted human discovery. SEA applied the simple metaphor of a Gantt chart to visualise the life-span of an alarm and to facilitate easy placement with other alarms occuring on other network elements in that same space and time (i.e. visual correlation).

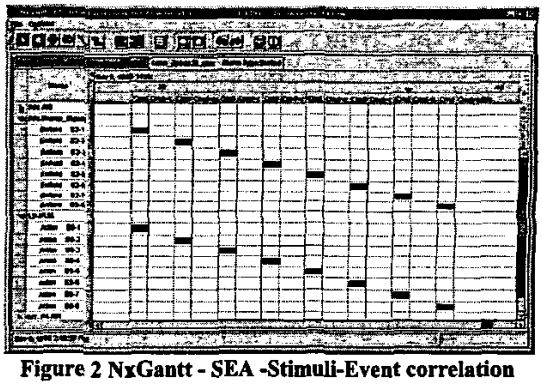

Figure 2 NxGantt - SEA -Stimuli-Event correlation Analyser

A. The Functions of the Tool

The main objective was to furtherd esign and develop NxGantt to capture the essence of the three-tier process, thus to facilitate both human and computer discovery while providing as much automation as possible. As such the extension of NxGantt from a visualization tool (tier 1)t o a system that incorporates a data mining approach (tier 3) with illustrated resultsv iewed in the visualization Gantt metaphor. The tool may also collect 'case' information from thed omain expert (tier 2) based on why a certain correlation should or should not be used to develop a rule. This last facility reflects the trouble ticket approach used in many faultm anagementa pplications.

The methodology behind the rule discovery has three tiers, visualization, knowledge acquisition (rule-base) and data mining (knowledge discovery). Although these tiers are

horizontal, in the sense that they follow a process, they should notb e considered exclusively so. The design takes into account vertical interaction - for instance it should be possible to visualise mined discoveries. To this end the design tries to capture additional information and develop a 'case' as well as a 'rule'. This would assist with knowledge acquisition/capture and store the reasoning behind the rules. It may also facilitate the introduction of other AI techniques in future such as Case-Based Reasoning. Firstly as part ofa fault management system utilizing the correlations and expert case knowledge. And secondly in automating more fully the rule discovery process by taking into account what lessons can be learnt from the historical information on correlations chosen or not chosen by the expert.

B. The Prototype of the Tool

Figure 3 highlights the 2 main threads through the tool - automated alarm correlation discovery (data mining) and manual alarm correlation discovery (visualization). Basically the process is discovery > case details > automated rule.

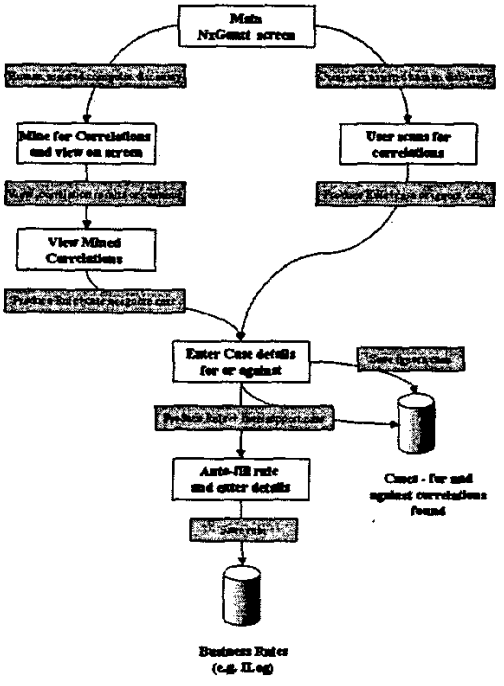

Figure 3 Computer and Human discovery options designed into the tool

It is important ina ny automated process not to assume the program knows best - the design of applicationt hus allows editing of the produced rules and cases as well as thev isual feedback on how these correlations have been discovered.

In Figure 4, screenshots (a) - (c) show the manual/human discovery approach which matches the lefth and side of the flow in Figure3 . Screenshots (d) - (f) show automated/computer discovery which matches the right hand flow in Figure 3.

Screenshot (a) depicts the visualization of data as was standard in theo riginal tool (Figure 2) with thea ddition of 2 alarms highlighted and manual rule option being chosen. Screenshot (b) requests (1) a rule name, (2) a higher order rule name that will be asserted in the system when these correlated alarms are retracted, (3)s ome explanatory reasoning from the expert as to why and (4) a diagnostic response that may be supplied to the user upon correlating these alarms. Screenshot (c) then automatically producest he rule in the correct syntax that may be edited or saved.

Screenshot (d) shows the (semi) automatic - find rules option being chosen with ther esults from them ining shown in (e). Upon choosing to develop a rule from a discovered correlation, will produce the same screen as in (b)a lthough note that in (f) the user has decided to ignore the correlation and fills in some reasons as to why.

The rule discovery is semi-automatic since not all information isa vailable to write a complete rule (thus the reliance on the case dialog). The mining algorithm basically searches for instances of alarms occurring together in time. This is problematic in that often the localc lock on a network elementi s off sync. Simply relying instead on the network managers time stamp doesn ot solve this either since it cannot be guaranteed that the alarms arrive to the manager in any certain order. This is handled by utilizing time windows, but obviously some uncertainty remains. Favored mined correlations for rule development then relies on frequency of that combinations occurrence although ultimately the decision lies with the expert user.

V. Future Development

At this stage of prototype the mininga lgorithm is not very sophisticated. Other mining algorithms that could handle the uncertainty in the data better could be 'plugged in'.

One challenge not addressed in this work is that of heterogeneous rule validation as rules are potentially being discovered from different sources. This would remove the need to manually confirm that they do not conflict with one another or the rules in the existing system.

You can never have enough integration and automation! As has been mentioned techniques such as CBR could be used once the 'case' database grows to assisti n the discovery task.

As the trend continues to derive and record specifications (tier 2) as soft-copy documents, often web enabled, this provides opportunities in itself for mining [12],[20] providing potential input for tier 3. Once captured, for example in a graphical casual model, the prior knowledge could then be used to refine future mining of the data [11].

VI. Summary and Conclusion

This paper presented the complexities and issues in fault management and identification namely that commercial systems tend to present a reduced set of symptoms of the fault not the actual fault itself. It then highlighted a three-tier process to assist in the semi-automatic discovery of new rulest hat could potentially reduce that set of symptoms further in ap ractical manner thata ddresses the issues.

The paper focused on the development of a specific tool that vertically slices the three tiers incorporating elements of visualization, knowledge capture and data mining reflecting both human-assisted computer discovery and computer-assisted human discovery.

The design of the prototype tool in the form of screen shots was then discussed with the purpose of highlighting features required from such a tool in terms of human andc omputer discovery, such as:

- Automation as far as possible
  - Correlations (computer discovery)
  - Correlation statistics (for human as well as computer discovery)
  - Rule coded from correlation (computer or human discovered)
  - Case coded from correlation (computer or human discovered)
- Transparency
  - Assumption expert knows best - therefore allow editing of automated bits
  - Visualization of discoveries
- Vertical hooks (to the other tiers)
  - Knowledge acquisition - capture why the correlations are relevant
  - Data mining - allow hooks in so mined discoveries can be visualised

The ultimate aim behind this work is that of integrating data visualization and knowledge discovery to create a human-centered process.

VII. ACKNOWLEDGMENTS

The authors are greatly indebted to our industrial collaborators Nortel Networks, Belfast Labs. We would also like to thank the Industrial Research and Technology Unit (IRTU) [Start-187 - The Jigsaw Programme 1999-2002] for funding this work jointly with NortelN etworks.

VIII. REFERENCES

[1] Bournellis C.,I nternet' 95.I nternetW orld,6 (11):47–52,1 995.

[2] Brachman R.J.,A nand T., "The Process of Knowledge Discovery in Databases: A Human-Centered Approach"., Advances in Knowledge Discovery & Data Mining,A AAI Press & The MIT Press: California,1 996, pp37-57.

[3] Bratko,M uggleton S.," Applications of Inductive Logic Programming",C ommunications of the ACM,V ol.3 8,n o.1 1, 1995,p p.6 5-70.

[4] Cheikhrouhou M., ContiP ., Labetoulle J.,M arcus K., IntelligentA gents for Network Management: FaultD etection Experiment.I n Sixth IFIP/IEEE International Symposium on Integrated Network Management,B oston,U SA, May 1999.

[5] Date C.J.,W hatN ot How - The Business Rules Approach to Application Development,f rom forth coming Book,2 001.

[6] Eick,S .G.i n Fayyad U. M., Grinstein G.G.,W ierse A.,(Eds.) Information Visualization in Data Mining and Knowledge Discovery,M organ Kaufmann Publishers,S an Francisco,U SA, 2002

[7] Fayyad U.M .,G rinstein G.G., Wierse A., (Eds.) Information Visualization in Data Mining and Knowledge Discovery, Morgan Kaufmann Publishers,S an Francisco,U SA,2 002

[8] Fayyad U.M., Piatetsky-Shapiro G.,S myth P.," From Data Mining to Knowledge Discovery: An Overview", Advances in Knowledge Discovery & Data Mining, AAAI Press & The MIT Press: California,1 996,p p1-34.

[9] Gruer D.,K han I.,O gier R., Keffer R.,A n Artificial Intelligence Approach to Network FaultM anagement, SRI International,M enlo Park,C alifornia,U SA.

[10] Harrison K.," A Novel Approach to EventC orrelation",H P, IntelligentN etworked Computing Lab, HP Labs, Bristol.H P-94-68, July,1 994,p p.1 -10.

[11] Heckerman,D .,B ayesian Networks for Data Mining, DM&KD 1, 79-119,1 997.

[12] IJCAI,1 999.W orkshop on TextM ining: Foundations, Techniques and Applications.1 997.

[13] ILog, www.ilog.com,2 000.

[14] Jakobson G., and Weissman M.,A larm correlation.I EEE Network,7 (6):52 -- 59,N ovember 1993.

[15] Oates T.,A utomatically Acquiring Rules for Event Correlation From Event Logs,T echnical Report9 7-14,U niversity of Massachusetts at Amherst,C omputer Science Department, 1997.

[16] Oates T.,F ault identification in computer networks: A review and a new approach.T echnical Report9 5-113,U niversity of Massachusetts atA mherst,C omputer Science Department,1 995.

[17] SterrittR .,D iscovering Rules for FaultM anagement, Proceedings of IEEE International Conference on the Engineering of Computer Based Systems (ECBS),W ashington DC,U SA, April 17-20,1 90--196,2 001

[18] SterrittR .,M arshall A.H.,S hapcott C.M., Mcclean S.I., Exploring Dynamic Bayesian Belief Networks For Intelligent FaultM anagementS ystems,I EEE Int.C onf. Systems,M an and Cybernetics, V,p p3646-3652,S ept.2 000.

[19] Sterritt,R .,C urran E.P.,A damson K.,S hapcottC .M.,. Visualization for Data Mining telecommunications network data, Data Mining II,(eds.) .E becken F.F.,B rebbia C.A.,W igend A., WIT Press,S outhampton UK,4 45-454,2 000.

[20] Tkach,e d.1998.T extM ining Technology: Turning Information into Knowledge.I BM White Paper.

[21] Uthurusamy R., "From Data Mining to Knowledge Discovery: Current Challenges and Future Directions", Advances in Knowledge Discovery & Data Mining,A AAI Press & The MIT Press: California,1 996,p p 561-569.

[22] Wang Z.,M odel of network faults,I n Meandzija B.,W estcott J.,(Eds.), Integrated Network ManagementI .,N orth Holland, Elsevier Science Pub.B .V.,1 989.

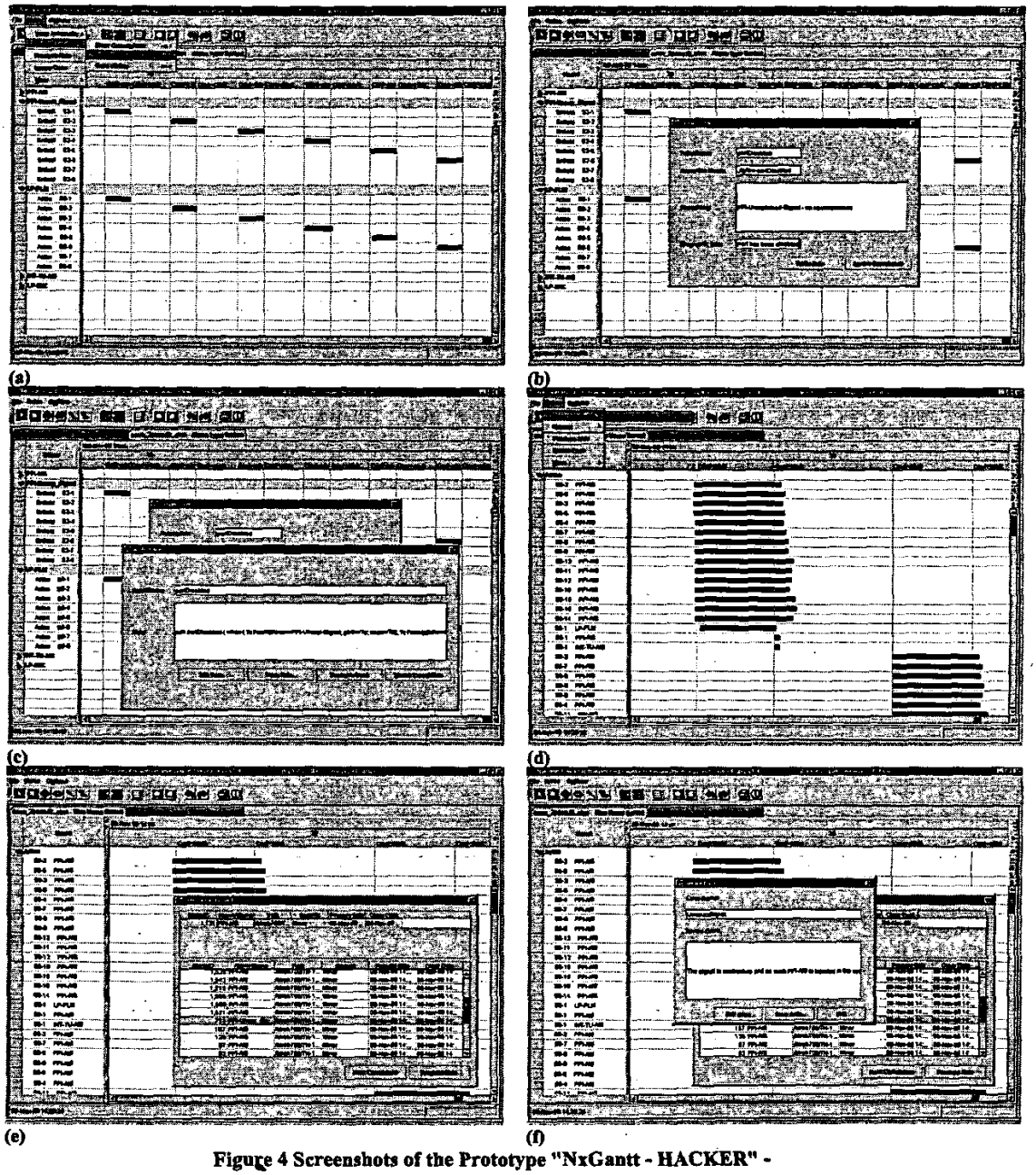

Figure 4 Screenshots of the Prototype "NxGantt - HACKER" - Human And Computer Knowledge discovered Event Rules