INTERNATIONAL DESIGN CONFERENCE - DESIGN 2006
Dubrovnik - Croatia, May 15 - 18, 2006.

DESIGN 2006

HEURISTICS FOR CHANGE PREDICTION

R. Keller, C. M. Eckert and P. J. Clarkson

Keywords: change prediction, design structure matrices, complexity measures

1. Introduction

Effective change management is a key to successful design development. As products and parts change, others can be affected, leading to further - often unexpected and costly - changes. These knock-on effects can jeopardise the timely delivery of projects and carry therefore a great risk for the entire design process. Assessing these potential knock-on effects is vital before selecting a possible change implementation. The Change Prediction Method (CPM) [Clarkson et al., 2004], captures the direct links between components in a systems and attaches impact and likelihood values to them. As change can only propagate through the direct links of components, the indirect risk of change propagating can be calculated, through what is essentially a brute force algorithm. The direct impact and likelihood values required for this computation are captured in a product connectivity model, that usually has the form of a Design Structure Matrix (DSM), but can also be interpreted as a directed digraph. This model is elicited in a two stage process in design meetings. A suitable product breakdown is developed and the likelihood and impact values of a change spreading from one component to an other are based on judgements of experienced designers (see [Jarratt et al., 2004]).
Depending on the size of model, the change prediction method can require a large amount of data (impact and likelihood values of direct change propagation), and be difficult to compute. To allow designers to easily assess direct and indirect change consequences, this paper proposes the use of simple, easily accessible heuristics that make use of complexity measures of graphs as employed in graph theory. These connectivity heuristics have certain advantages, such as simple computation, visibility and availability of the data, over the specific change values calculated in the current change prediction method. In this paper we will introduce these heuristics and discuss their merits by comparison with the values calculated by the CPM method.

2. Barriers to CPM

The change prediction method was used successfully in several industrial settings, however, it requires the commitment of internal champions to take up the method. This section explores which aspects of change propagation can act as barriers for the industrial use of the change prediction method.

2.1 Computational Cost

Computation of combined risk values used in the change prediction method as described in [Clarkson et al., 2004], requires high computational effort. The search for all propagation paths used to calculate the combined likelihood value of a change is essentially a “finding the k shortest paths” problem which can be solved in $O(m+n\log(n)+k)$ time. It can be shown that the total number of paths to be considered in a product linkage model with n components and a density of d (the number of links in the model divided by the possible number of links) can be estimated as growing exponentially with the number of components (see Equation 1).

$$k \approx E[p] = n\sum_{i=1}^{n-1} d^i \frac{(n-1)!}{(n-i-1)!} \quad (1)$$

For current product models described for example in [Jarratt et al., 2004], the number of all propagation paths is too large to be computed. For instance, for the model of a diesel engine with only 41 components and 254 direct links between them, the expected number of propagation paths between all component pairs is $8.47*10^{19}$ (the number of molecules in a cm^3 of air is in the order of $3*10^{19}$). Even when reducing the maximal considered path length to six, there are still $1.91*10^6$ expected paths for this relatively small product model. Especially in the light of interactive software, current computer hardware is not able to compute these numbers of propagation paths in an adequate time and it is impossible to compute change propagation, for instance, for products consisting of thousands of highly connected components. The heuristics described later in this paper have the advantage that they are less computational intensive and can even be calculated for very large product models.

2.2 Model building

The CPM method is applicable both to the generation of versions of an existing product and the design of a new generation of a product The data stored in the connectivity model used to predict change propagation [Clarkson et al., 2004] consists of quantitative values describing the direct change likelihood and impact of changes propagating directly from one component to the other one. These values are elicited in interviews and group meetings with experienced designers. However, these values are usually based on experience with past products and in very early stages of the design process, this data might not be available unless the new product is very similar. The heuristics described in this paper are purely qualitative measures used for instance in complexity theory to describe the complexity of a product (see [Summers and Shah, 2003]). While quantitative values might not yet be available or be difficult to assess for a different product, the product architecture might already be in place or remain constant, so that it can be the basis of a purely qualitative product connectivity model. The heuristics used in this paper can be applied to such a simple model.

Another problem is that it is very difficult to assess accurate likelihood and impact values for a change propagating between two components. As shown for instance in [Ayton and Pascoe, 1995], it is very difficult for human experts to assess likelihood values correctly. As these "uncertain" values are then used to compute change propagation effects with exact algorithms, it is questionable to what extent these values can be trusted.

A last argument in favour of simple heuristics is the time and therefore cost of building a product model. Establishing the product architecture and the qualitative links captured in a connectivity model alone requires valuable time and commitment of all designers involved. Capturing change likelihood and impact values increases the time necessary to build a model significantly and there is a trade-off between the benefits such a model provides and the effort needed to build and maintain it.

2.3 Visual Access

While there is a whole body of literature on how to visualise and represent qualitative graph structures, such as matrix-based representations like DSMs (e.g. [Browning, 2002]) and node edge diagrams (e.g. [Di Battista et al., 1994]), it can be difficult to incorporate quantitative likelihood and impact values into the visual displays. Attempts were made to show combined risk values in a DSM-like structure (see the combined risk plot introduced by [Clarkson et al., 2004] and Figure 1, left). However, these displays hide all direct links of the underlying model, and users are not able to identify whether a high change risk results from direct or indirect component connections (see [Keller et al., 2005]). There are, however, several possibilities to incorporate this information into node-link displays, such as using the edge width to indicate the likelihood or risk of a link. Spring layouts, where each link of a graph is modelled as a spring, offer the functionality to lay out a node-link diagram in such a way that the lengths of the edges can be approximately proportional to its weight (in this case: likelihood or risk values). However, these spring layouts are only an approximation and cannot visualise component connections when the change risk or likelihood in one direction is much larger

than in the other direction. [Keller et al., 2005] introduce a node-link display that shows such combined risk values for a particular change case (when one component is identified to cause a change). In this display (see Figure 2, right) the distance of one component to the initiating change component in the centre of the display is inversely proportional to the logarithmic change risk shown in a radial layout as calculated with the CPM method. The advantage of the heuristics described in the remainder of this paper is that they can be easily identified and assessed visually. Studies described for instance in [Ghoniem et al., 2004] use the number of common neighbours and the shortest path to assess how well DSMs and node-link diagrams are suited to show graph data. This shows that these measures have been proven to be visibly accessible.

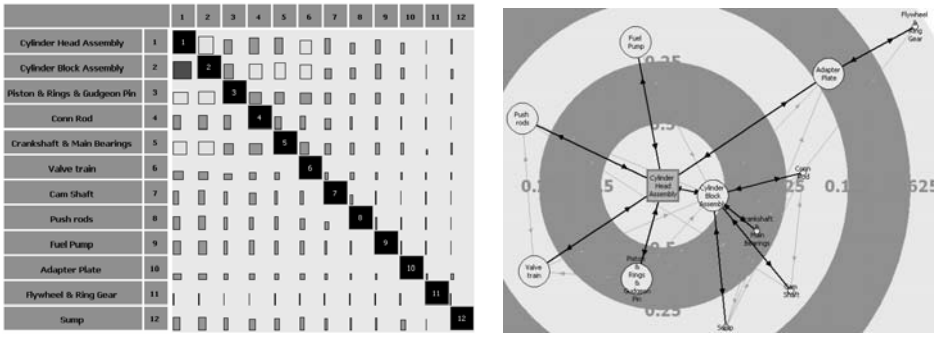

Figure 1. Indirect change risk visualised for a diesel engine: A combined risk plot (left) and a risk network (right) for the "Cylinder Head Assembly"

3. Simple Graph Measures as Heuristics

A set of well-established graph measures will be assessed in this section to investigate whether these measures can be used to indicate to how changes might propagate between components. The measures introduced here are simple graph theoretical measures, further descriptions can be found in standard text books on graph theory (e.g. [Boffey, 1982]). In a change context, similar measures (such as the shortest path for example) were used to define product modularity (see [Sosa et al., 2005]). The difference between this paper and [Sosa et al., 2005] is that the heuristics used here describe relations between two components, rather than providing centrality measures (that show how much a component is integrated in the product architecture) for one component. The heuristics (named H1-H3) will then be tested against the existing change prediction method that proved to produce valid results throughout different case studies. For this study, only change likelihood values will be considered as the heuristics do not incorporate any measures of impact and thus are not able to reflect change risks.

3.1 Shortest Path Length

The shortest path length between two nodes is the minimal number of links that connect these two nodes in the graph. A shortest path length of 1 for instance indicates that the two components are directly connected. Computing the shortest path between two nodes is one of the standard calculations performed on graphs and several algorithms exist that do this computation in little time. The standard algorithm to calculate the shortest path between two components is the Dijkstra algorithm that has a proven effort of $O(n^2)$ to compute the shortest path from one node to all other nodes (e.g. [Boffey, 1982]). See Figure 2 (a) for the shortest path between two nodes in a simple graph example that could be interpreted as a product connectivity model for a very small product.

The question is: “How does the length of the shortest path relate to the chance of change propagating between two components?” The hypothesis is that if the shortest path between two components is very small, the changes need not to travel a long distance so a propagation is more likely than for components that are very far away in the graph (H1).

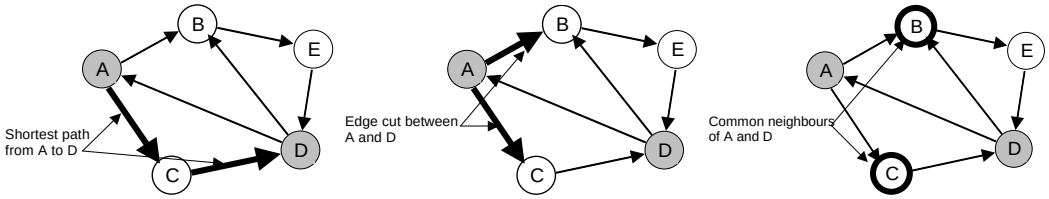

Figure 2. The shortest path from A to D has the length of 2 (a). Edge connectivity of 2 from A to D (b). Two common neighbours between A and D (c)

3.2 Edge Connectivity

The edge connectivity between two nodes in a graph describes how much material can flow between these two nodes under the assumption that each edge has a capacity of one. It can be shown that the edge connectivity is equivalent to solving the cut-problem in graphs [Boffey, 1982]. This means that the edge connectivity is equal to the number of edges that have to be minimally removed from the graph in order to remove the possibility to reach one node from the other (minimal edge cut). It can be shown that computing the edge connectivity between two nodes has an effort of $O(n^3)$ *[Boffey, 1982]*. For an example of the edge connectivity between two nodes, see Figure 2 (b).

If the edge connectivity between two nodes is high, it means that there are a high number of paths connecting these two nodes, allowing changes to propagate through different routes (H2). It can be argued that high edge connectivity means that there is a high change propagation likelihood. Low connectivity should indicate that there are not many possible change propagation routes, so changes are less likely to propagate between these two components.

3.3 Number of Common Neighbours

The number of common neighbours between two nodes is the number of components that share links with both nodes. The more common neighbours two nodes have, the more likely a change is to propagate between them as they have a high number of potential 2nd order connections between them (H3). In contrast to the heuristics described beforehand, this heuristic does not take directionality of the links into account, which might be a disadvantage.

See Figure 2 (c) for an example of how to determine the common neighbours between two components (note that in this example, the edge connectivity equals the number of common neighbours, in the general case, this does not hold). The number of common neighbours can be computed quite easily and has $O(n^2)$ effort from one to all components.

4. Assessment of the Heuristics

This section will discuss how well the heuristics established in the previous section can describe change propagation combined likelihood values. For this purpose, the values calculated for a well-studied product model will be compared with the values of the different heuristics. For other product models, a similar behaviour was observed (see Table 1). The diesel engine product connectivity model was elicited during an extensive case study at a diesel engine company described by [Jarratt et al., 2004]. This model was chosen as a well-studied example where the predictions using the CPM method clearly matched the expectations of experienced designers and historic change data. The model consists of 41 components connected through 254 direct linkages resulting in a density of 15.5%. Further analyses of this model are also described in [Keller et al., 2005].

For the analyses, a standard statistical modelling method (linear regression) is used as well as box plots as a visual method. A box plot is a well-known way of representing the density of data by showing important statistics (the box of one of the plots for instance shows 50% of the data, the central horizontal line represents the median).

A linear regression assumes a linear dependency between the influencing factors (independent variables *x*, here: the heuristic values) and the factor that is to be explained (the dependent variable *y*,

here: combined likelihood value of the change prediction method); and independence between the influencing factors. The resulting model of a linear regression is of the form $y=ax+b$ with the parameters a (the gradient) and b (the constant). In this paper we are not (in the first place) interested in the exact parameter values (that are likely to differ amongst different product models as a result of scaling issues), but in the significance of the parameter a. If this parameter is significantly different from 0, then there is an influence of the independent factors x on the dependent variable y. The R^2 value of the model, which signifies the model fit, is also of interest as it describes how much variability of the dependent variable is explained. However, as the heuristics are all qualitative we don't expect high values for R^2.
The resulting regression model could also be used to predict values for y given a value for x. However, the models shown later in this section should not be used for such predictions, as the factors will differ for each product model. For the sections 4.1., 4.2. and 4.3. such a simple one-dimensional regression will be used, in section 4.4. a combined model for the diesel engine model will be assessed that makes use of 3 independent variables (the three heuristics). As a third method, the most likely links using combined likelihood and the highest links identified by the heuristics were compared. A high match would indicate that the heuristic identifies most of the high-likelihood values of CPM.

4.1 Shortest Path Length

The regression model for the shortest path length between two components describing the combined likelihood value showed that the linear factor a of the linear regression is significantly less than 0 ($p<0.001$, the exact values for a can be found in Table 1) and signifies a negative trend, which means the smaller the shortest path length, the higher the combined change likelihood (on average) and thus proves H1. The model has an R^2 value of 0.332, which means that about 33% of all the variability of the combined likelihood value is explained through this model.
A visual representation of how the shortest path length describes the combined likelihood value is shown in Figure 3 (a), where this trend is immediately visible. It also holds that 47% of the 15% highest combined likelihood links are also amongst the 15% lowest shortest path links (47% match). This means the bigger the length of the shortest path between two components, the lower the change likelihood.

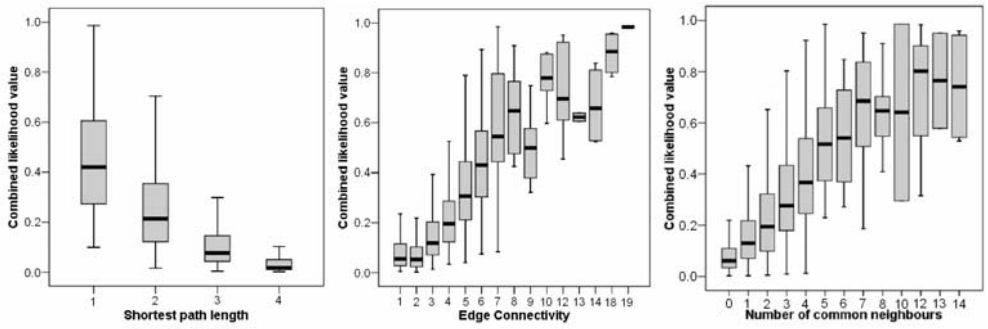

Figure 3. Box plots show the influences of the heuristics on the combined likelihood values computed by the change prediction method. a) Shortest path length, b) edge connectivity, c) number of common neighbours

4.2 Edge Connectivity

The regression model for the influence of the edge connectivity on the combined risk values gave the following results. Like in the model for the shortest path length, the parameter value for a is significantly different from 0 ($p<0.001$, however, in this case the parameter is significant larger than 0) which shows the existence of a positive influence of the edge connectivity on the combined likelihood

value and proves H2 (the exact values for the factor *a* can be found in Table 1, see also Figure 3 (b) for the visual representation). This means that the higher the edge connectivity between two nodes, the higher the expected combined likelihood of a change propagating between these two components. The R^2 value is calculated to be 45.4%, meaning that approximately 45% of the variability of the combined likelihood value is described through this model. Of the 30% highest link values there is a 64% match between the edge connectivity values and the combined likelihood values.

4.3 Number of Common Neighbours

The regression model for the number of common neighbours heuristic describing the combined likelihood value are as follows. Again, the parameter value *a* for the linear trend is significantly larger than 0 ($p<0.001$, see also Table), meaning that this factor has a significant positive effect on the combined likelihood values (H3). This trend can also be observed visually in Figure 3 (c). This model explains 37.9% of the variability of the combined likelihood ($R^2=0.379$) for the diesel engine example. Of the 30% highest link values (for both, combined likelihood and the number of common neighbours), this heuristic gives a 69% match. To summarise, the higher the number of common neighbours, the higher the combined likelihood values between a component pair.

4.4 Combination of Heuristics

Finally, a model trying to explain the combined likelihood value using all three heuristics was computed. One must mention that one of the assumptions of the regression, which is that the independent variables are independent, is violated in this example (as the shortest path length and the number of common neighbours for example depend on each other). However, the results are quite strong, giving a R^2 value for this model of 58.2%. Each of the factors in *a* (which is a vector of three values in this case) are significantly different from 0, supporting the values from the previous introduced models. This model gives quite a good approximation of the combined likelihood values given all three heuristics (see Figure 4 for a scatter plot that shows the value predicted by the model plotted against the combined likelihood value). Again, the most important finding is that all factors are significantly different from 0, so there always exists a trend. Using this combination of heuristics, there is a 77% match for the 46% biggest links. The main message of this model is that knowing the values of these three heuristics (that can even be visually seen from a graph representation), one would be able to predict a vast amount of the combined likelihood values without the need to use computational intensive algorithms.

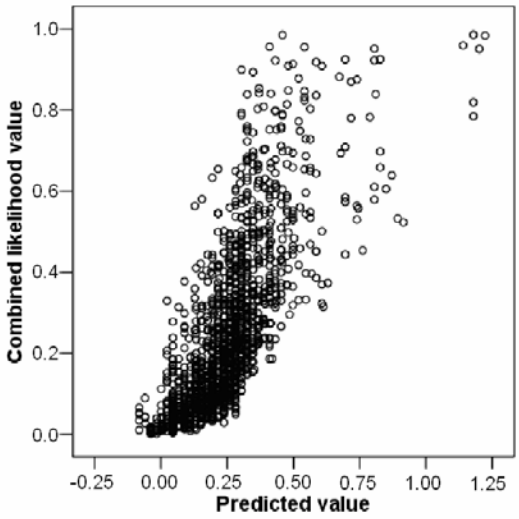

Figure 4. Scatter plot of the predicted model value and the combined likelihood value

4.5 Results for other product models

Table 1 summarises the results for the diesel engine model as well as for other product models that differ in size and density (dens.). All of these models are based on case studies conducted by other change researchers and are described elsewhere (see for example [Jarratt et al., 2004]). In all but one cases, the parameter value for *a* is significantly different from 0, which is the most important finding as it verifies that the trends observed for the diesel engine model hold in all other models. However, as one can see there are differences in how much of the variability of the combined likelihood values is explained by each model. The only model that seems to be difficult to assess with the heuristic values is the helicopter model (this is the consequence of mostly very small and some very large direct change probabilities that result in low R^2 values). Additionally, the common neighbour heuristic has no significant effects on the hairdryer model (probably due to the small size of the model). See the non-significant parameter for the common neighbours (indicated by *).

As one can see the individual parameter values for the factor *a* differ between the different models. For instance the factor for the shortest path for the model of the helicopter (see [Clarkson et al., 2004]) is in the order of *a=-0.01*. The match between the highest components is also quite high for all product models; the lowest value is 33% for common neighbour heuristic and the small hairdryer model. Even for the helicopter model, which has only very small R^2 values for the heuristics, the matches are always above 50% (the shortest path heuristic for the helicopter even has the highest match amongst all models and heuristics).

Table 1. Results of the heuristics for different product models

			Shortest path			Edge conn.			Common neighb.		
Model	size	dens.	*a*	R^2	match	*a*	R^2	match	*a*	R^2	match
Diesel engine	41	16%	-0.16	0.332	47%	0.068	0.454	64%	0.071	0.379	69%
Helicopter	19	30%	-0.01	0.111	87%	0.004	0.086	62%	0.001	0.018	51%
Jet engine	32	28%	-0.18	0.270	54%	0.052	0.443	62%	0.048	0.401	65%
Hairdryer	6	73%	-0.22	0.211	86%	0.194	0.425	77%	0.02*	0.003	33%
Injector	15	27%	-0.08	0.495	75%	0.043	0.194	75%	0.050	0.241	45%

5. Discussion

The analyses presented in this paper clearly show the value of simple heuristics. The heuristics are easily accessible and easily understandable standard graph theoretical measures and they overcome some of the barriers for change prediction presented in section 2. It was also mentioned that these heuristics can even be assessed visually given an adequate representation of the underlying product connectivity model. The heuristics are also based on qualitative connectivity models, which do not require any change likelihood, or impact assessments to be made by expert designers. As this kind of data could also be obtained from other models (i.e. CAD models that store information about spatial relations between components or circuit diagrams capturing electrical links), predicting change propagation based on heuristics could also be done automatically, without the need of manually building and maintaining product models.

However, the heuristics do not have the full predictive power of the change prediction method, which can be seen for instance in the relatively low R^2 values for the helicopter model and as they are qualitative rather than quantitative measures there will always be mismatches between these two methods. In that sense, practitioners should carefully decide which method is appropriate at which time in the design process. In later stages of the design, the standard change prediction method has many advantages over the proposed heuristics, as most of the data is available. In early stages of the design where only rough ideas of the design exist (and especially no quantitative data is available), the use of heuristics, however, can be advantageous. Also for assessments of the product model, when no computer assistance is at hand, the heuristics are clearly valuable.

6. Conclusions

This paper introduced several simple heuristics that allow the assessment of change propagation in complex products. The bases of these heuristics are simple graph theoretical measures. These were the length of the shortest path, the edge connectivity and the number of common neighbours of two components in a product connectivity model. It was shown that the heuristics do explain the combined likelihood values calculated by the currently used change prediction method to a great extend. The heuristics inferred from the analyses are the following:

- The longer the shortest path between two components, the less likely is change propagation;
- The higher the edge connectivity between two components, the more likely is change propagation between this pair of components;
- The more common neighbours two components share, the more likely is change propagation.

The validity of these heuristics was tested for a number of different product models against the results of the existing change prediction method and showed that they can be used to predict change propagation on the basis of change propagation likelihood. The lack of accuracy of these heuristics is outweighed by their simplicity, as they do not rely on any quantitative change propagation likelihood and impact values and can be computed easily. Future research will examine how combinations of these heuristics can be used to classify components of a complex product in respect to their behaviour in change propagation and how heuristics can be used in order to predict change risks rather than change likelihood as they do now.

Acknowledgements

This research is funded by the EPSRC.

References

Ayton, P. and Pascoe, E., "Bias in human judgement under uncertainty?" The Knowledge Engineering Review, Vol. 10, No. 1, 1995, pp. 21-41.

Boffey, T. B., "Graph theory in operations research", Macmillen, London, 1982.

Browning, T. R., "Process Integration Using the Design Structure Matrix", Systems Engineering, Vol. 5, No. 3, 2002, pp. 180-193.

Clarkson, P. J., Simons, C. and Eckert, C. M., "Predicting Change Propagation in Complex Design", ASME Journal of Mechanical Design, Vol. 126, No. 5, 2004, pp. 765-797.

Di Battista, G., Eades, P., Tamassia, R. and Tollis, I. G., "Algorithms for drawing graphs: an annotated bibliography", Computational Geometry: Theory and Applications, Vol. 4, No. 5, 1994, pp. 175-198.

Ghoniem, M., Fekete, J.-D. and Castagliola, P., "A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations", Proceedings of InfoVis 2004, Austin, Texas, USA, 2004, pp. 17-24.

Jarratt, T., Eckert, C. M. and Clarkson, P. J., "Development of a Product Model to Support Engineering Change Management", Proceedings of the TCME 2004, Lausanne, Switzerland, 2004, pp. 331-342.

Keller, R., Eger, T., Eckert, C. M. and Clarkson, P. J., "Visualising Change Propagation", Proceedings of ICED '05, Melbourne, Australia, 2005.

Sosa, M. E., Agrawal, A., Eppinder, S. D. and Rowles, S. D., "A Network Approach to Define Modularity of product Components", Proceedings of ASME-DETC2005, Long Beach, California, USA, 2005.

Summers, J. D. and Shah, J. J., "Developing Measures of Complexity for Engineering Design", Proceedings of ASME-DETC2003, Chicago, Illinois, USA, 2003.

René Keller
PhD Research Student
University of Cambridge, Engineering Design Centre
Trumpington Street, Cambridge, CB2 1PZ, United Kingdom
Tel.: +44 1223 332828
Fax.: +44 1223 766963
Email: rk313@cam.ac.uk
URL: http://www-edc.eng.cam.ac.uk