Causality and Sensitivity Analysis in Distributed Design Simulation

by

Jaehyun Kim

Bachelor of Science, Naval Architecture and Ocean Engineering,
Seoul National University, Seoul, Aug 1994

Master of Science, Mechanical Engineering
Massachusetts Institute of Technology, June 1998

Submitted to the Department of Mechanical Engineering
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February, 2002

Signature of Author____________________
Department of Mechanical Engineering
October 25, 2001

Certified by____________________
Professor David R. Wallace
Thesis Supervisor

Accepted by____________________
Professor Ain A Sonin
Chairman, Department Committee

This page is intentionally left blank

Thanks God,

You sent me to the place where I do not deserve to be

And gave me a chance to enjoy this long journey

Above all,

Thanks for giving me a family (my father, mother, brother in law and sister)

To whom I can dedicate this dissertation

This page is intentionally left blank

Causality and Sensitivity Analysis in Distributed Design Simulation

by

Jaehyun Kim

Submitted to the Department of Mechanical Engineering
on October, 2001 in Partial Fulfillment of the Requirements for the
Requirements for the Degree of Doctoral of Philosophy in
Mechanical Engineering

Abstract

Numerous collaborative design frameworks have been developed to accelerate the product development, and recently environments for building distributed simulations have been proposed. For example, a simulation framework called DOME (Distributed Object-oriented Modeling and Evaluation) has been developed in MIT CADlab.

DOME is unique in its decentralized structure that allows heterogeneous simulations to be stitched together while allowing proprietary information an simulation models to remain secure with each participant. While such an approach offers many advantages, it also hides causality and sensitivity information, making it difficult for designers to understand problem structure and verify solutions.The purpose of this research is to analyze the relationships between design parameters (causality) and the strength of the relationships (sensitivity) in decentralized web-based design simulation. Algorithms and implementations for the causality and sensitivity analysis are introduced.

Causality is determined using Granger's definition of causality, which is to distinguish causation from association using conditional variance of the suspected output variable. Sensitivity is estimated by linear regression analysis and a perturbation method, which transfers the problem into a frequency domain by generating periodic perturbations. Varying Internet latency and disturbances are issues with these methods. Thus, algorithms are developed and tested to overcome these problems

Thesis Supervisor: David R. Wallace

Title Professor of Mechanical Engineering

This page is intentionally left blank

Acknowledgments

I still remember the day when Prof. David gave me his offer, which was June 24th 1998. That was one of the happiest days of my life. I really appreciate his great guidance and generous funding. Working with him is a great privilege not only because of his ingenuity and creativity, but also his enormous capacity of joy and great sense of humor.

I am deeply grateful about the fact that I could have Prof. Seth Lloyd and Prof. Kamal Youcef-toumi as my committee members. They have given me an invaluable guideline for this thesis.

One of my greatest thanks goes to Prof. Sanjay Sarma. I learned a lot from him though the period of working with him.

I really appreciate Kathy Wang at CIPD for her help and encouragement for the final part of this job.

People at MIT CADlab have been excellent colleagues for me. Thanks for their support and friendship. Especially I want to give my special gratitude to Elaine Yang.

Last five years could have been more difficult without the friends here. I could enjoy my days at MIT thanks to Pyeongwon, Byeonghyuk, Sangbeom, Cheolmin, Joonil, Junho, Jeesung, Heejae, Yoonkyu, Kyujin, Jongyoon Jincheol, Sunwoong, Jungwook, Sangjun, Taesik, Yongsuk. Hyunkyu, Jinwoo, Sunjoon, Jinwook, Joon, Hyuksang, Jinpyeong, Seungkil, Hyungsoo, Jitae, Kihyun and Kiseog.

I also appreciate my friends back there in Seoul for their friendship and encouragements. Thanks Jinwon, Sangseo, Woochang, Seokwon, Sangkyun, Pilsang, Taejoon, Bumwoo, HyungKwan, Jaeyeol, Chanki, Yoojong, Jinwook, Yongjoon, Gene-Oh, Dohoon, Yangso, ByeonKyu, Keunwook, Dongbeom, Donggun, Sungsoo, Dongsik, Keunbae, Wonseok, Kwansoo, Heungjoo and Jaesung. Especially, I want to share this honor with my best friend Jaeha.

I am deeply grateful to Prof. Key-Pyo Rhee, Kwang June Bai, JongGae Shin, Hang S Choi, Hyochul Kim and JongNam Joo at Seoul National University.

YeDarm family, especially pastor Hosik Kim and ChangJoo Kim, deserves more than just thanks from me for their prayer.

Most precious thanks of mine goes to my family (Mom, Dad, Brother in law and Sister). They have given me love and faith I do not deserve. Maehyung, thanks for having been such a great mentor to me. I think I am privileged with your thoughtful advice. Noona, without your warm care and deep understanding on me, I couldn't come to this point. Mom, thanks for always putting some faith on me (even though the history does not allow you to do that). I can stand against all the hardships due to your belief in me. Dad, your sincere prayer and moving letters have given me the courage to finish this journey. I believe God granted the wish you made in your New Rochelle days.

This page is intentionally left blank

Table of Contents

List of Figures

1 Introduction

1.1 Motivation

As a result of rapid progress in communication technology, numerous Internet-based design frameworks have been developed to accelerate the product development and improve product quality. However many of these applications focus on a specific design stage or domain rather than the integration of simulations from a wide variety of design perspectives.

A design framework DOME (Distributed Object-oriented Modeling and Evaluation) has been developed by MIT CADlab to streamline designer-mediated integration of unstructured design simulations as illustrated by Figure 1-1. DOME is unique in its decentralized peer-to-peer simulation architectures, which allows heterogeneity of design simulations and protects proprietary information of participants. However, this decentralized approach also results in hidden simulation causality and sensitivity. Causality and sensitivity information is needed to support decision-making, to provide designers with a sense of the system structure and to ensure robust and efficient computation.

The purpose of this thesis is to analyze efficiently (in terms of time and memory) the relationships between design parameters (causality) and the strength of their relationships (sensitivity) when a model or simulation structure is not visible directly.

In this work, algorithms and implementations for the causality and sensitivity analysis are introduced. Token (message) passing, Bayesian network approach, regression analysis and perturbation method are considered and applied to analyze hidden causality, emphasizing time efficiency and precision. Sensitivity analysis is performed using the inferred causal structure. First, linear regression methods were tried, and issues in matching input changes with outputs are discovered due to the varying latency of the Internet. To overcome this problem, linear interpolation and cross-correlation were used to find a precise match. Second, a perturbation method combined with theories from signal processing is developed. The method takes advantage of the relatively large time-scale of change in Internet traffic to overcome varying latency and latent structure problems.

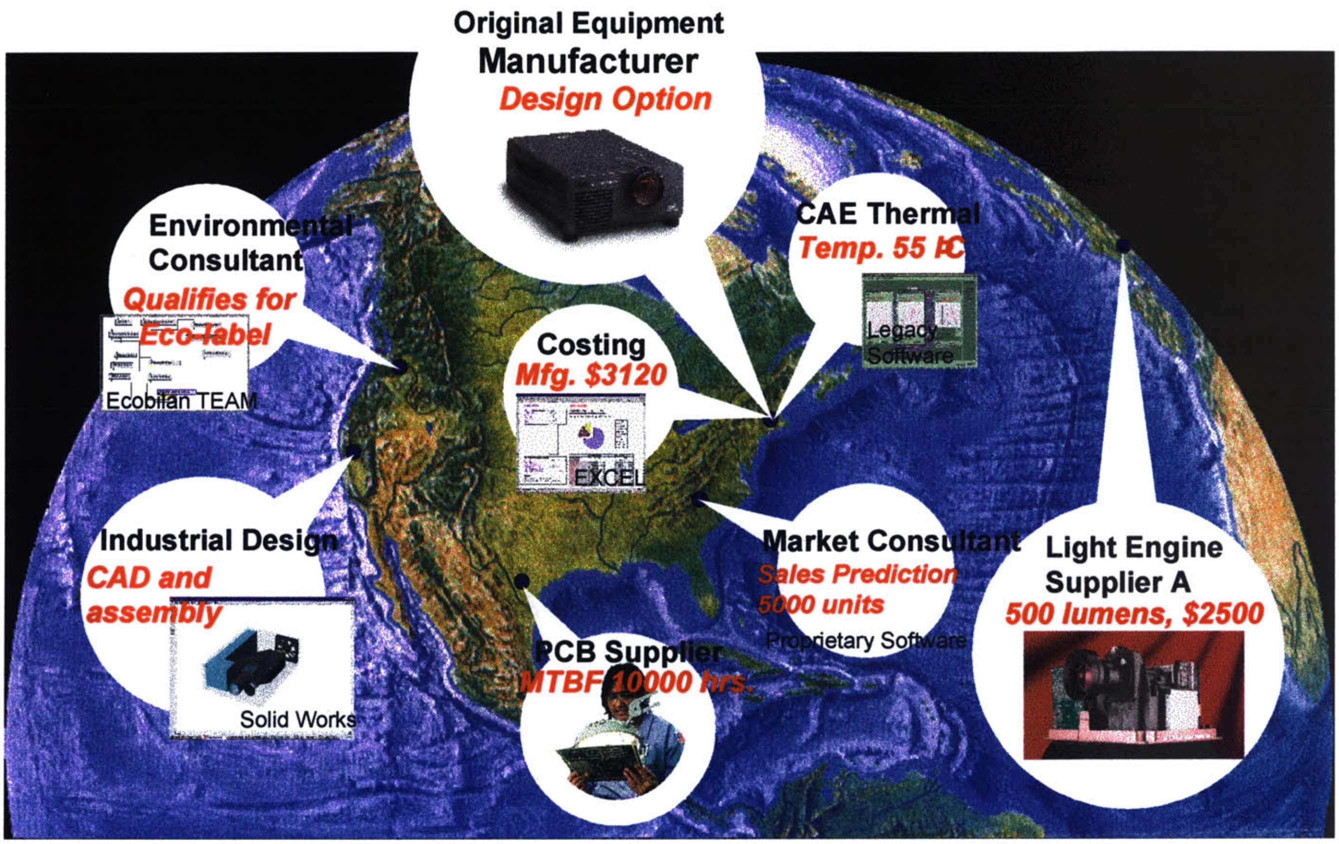


Figure 1-1 Design Simulation in a Distributed Design Framework

1.1.1 Regression analysis

Regression analysis is a methodology to determine the dependency of a variable (dependent variable) on another variable (explanatory variable) using distributions of the variables. Sensitivity can be directly determined by regression analysis, but a statistical measure[1] is needed to supplement the regression analysis to figure out causality.

In estimating regression coefficients, distribution of an explanatory variable, conditional upon a dependent variable, is needed to be known. Simulation modules in the DOME are MIMO (multiple input multiple output) systems. Thus, output values corresponding to the changes of an input must be selected to estimate the conditional distribution.[2]. However, there is an unknown delay between changes of an input and an output, because the DOME is using the Internet as a communication channel. As a result, the matching needs to be determined statistically.

In this thesis, using a periodic (frequency) perturbation is suggested as a solution for the matching problem. An interval of time between changes of an input and an output is estimated using the frequency information, and a series of corresponding output values are selected by its deviation from an expected interval[3]. In employing the periodic perturbation, maintaining the frequency is a primary issue, which is affected by the Internet latency. Thus, characteristics of the Internet latency depending on a time scale and protocols are studied in chapter 4.

1.1.2 Perturbation Method

Perturbation method is using periodic perturbation to transfer the problem from a time domain to a frequency domain. An input value is periodically perturbed and a response of an output is represented in the frequency domain. Sensitivity and causality is detected using an amplitude and existence of a frequency component respectively.

Maintaining an original perturbation frequency, during propagation through a design network is a primary issue. Thus, characteristics (especially a time-scale) of the Internet latency are studied to minimize the latency variation. Also, statistical methods are adopted to refine the results.

1.2 Goal of the research

The goal of this research is to develop and implement new algorithms for causality and sensitivity analysis in a distributed design simulation environment. Decentralized distributed design frameworks with a flat hierarchy have many advantages, such as scalability, heterogeneity between design objects and hiding proprietary information. However, it also has the problem of hidden causal information, which makes decision-making and system diagnostics difficult. The focus of the work is to overcome the hidden causal information problem with minimum computation time and a reasonable degree of precision. Internet latency

---

[1] Granger causality

[2] The selection of output values corresponding to an input will be referred as a matching problem.

[3] A series of output values arriving with intervals, that most closely approximate an expected interval, are selected.

characteristics will be addressed as a corner stone for resolving the unknown varying latency, which complicates cause-effect matching problems in regression analysis, and a non-uniform interval problem when using perturbation methods. Approximation algorithms will be presented to match regression analysis inputs to their corresponding outputs. To differentiate the effects when using perturbation methods, the problem is transferred to the frequency domain by generating periodic perturbation. Effects can thus be distinguished by the frequency of the perturbation. Algorithms to resolve the problem of non-uniform interval due to the varying latency are elaborated.

1.2.1 Importance of the importance of causal information in conceptual design stage

In the conceptual design stage, the knowledge on causality and sensitivity is essential for the designers to efficiently manage constraints and to achieve designs that are robust to the manufacturing tolerances or noise from aging or the working environment.

Constraint management

Design processes are constraint oriented, meaning that the goal is to satisfy relationships between various design parameters and performance goals. Constraints come from physical laws, performance, geometric and topological requirements or aesthetic considerations, etc. Constraint management includes maintaining consistency (no conflict or redundancy) and completeness (no unconstrained degrees of freedom) of the constraint set, achieving efficient constraint evaluation by selecting the relevant subset of constraints to be managed, and supporting decision-making through qualitative analysis (Serrano and Gossard 1992).

Robust design

The distribution of values associated with a given design parameters arises from manufacturing variations, product aging or noises from the environment. Such distributions can be reduced either by tightening manufacturing processes or by changing the design so that the parameters become less sensitive to the noise. Generally the former is expensive or sometimes even impossible to achieve. The latter is called robust design or *Taguchi's method.* The goal of Taguchi's method is to achieve the easy to manufacture design and best performance in operation (David 1992).

1.2.2 Properties of the decentralized distributed design environment

A decentralized distributed design environment is advantageous in its ability to accommodate heterogeneity (interoperability and accessibility), complexity (scalability and flexibility) and proprietary knowledge. However, it has the disadvantage of hiding causal information, which is needed for efficient constraint management and robust design.

1.2.3 Study on regression analysis and perturbation method

Regression analysis is a methodology to infer the dependency of one variable, the dependent variable on other variables, explanatory variables. The main purpose of the analysis is to estimate and predict the distribution of the dependent variable when that of explanatory variables is given. However, regression by itself cannot verify the causality, it can only show the strength of dependency. Statistical measures, such as Granger causality, are required to prove causality (Pearl 1988). In the context of distributed design framework, some conditions needed for the validity of the regression analysis, such as normality and independence between noise sources, are not satisfied and results in bias.

Thus, another approach, called the perturbation method, is proposed. The main issue in this approach is that the effects from multiple inputs cannot be differentiated simply from the perturbation values. To resolve this limitation, the problem is transferred into frequency domain by using a periodic perturbation. However, non-uniform delay of data delivery through Internet can still be an issue. Characteristics of Internet latency are studied and approximation algorithms are developed to resolve this problem.

1.2.4 Characterization of Internet latency depending on packet size and protocols

The RTT (round trip time) delay of packets delivery in the Internet will be characterized on the time scale of several milliseconds to minutes. RTT depends on packet size, involved protocols and traffic situation of the Internet (Bolot 1993). Properties of TCP (transmission control protocol) and UDP (user datagram protocol) will be explained and the effect of those properties on the RTT will be shown. Both protocols have pros and cons. The relatively long time scale of traffic situation variation of Internet relative to UDP communication will be used to overcome unknown and varying latency.

1.2.5 Implementation of algorithms

Based on the knowledge and techniques proposed in the thesis, an appropriate architecture for a computer-assisted design system, building upon the DOME framework is proposed and is implemented to explore the merits of the new algorithms. The framework and its implementation are also validated with an engineering design problem.

1.3 Problem statement

In this work, algorithms for causality sensitivity analysis are developed. Issues such as unknown varying Internet latency, nonlinearity and effect differentiation are considered and resolved.

1.3.1 Unknown varying Internet latency

Internet latency depends on protocols, the size of the packet, and network traffic. The unknown and varying latency of packet delivery through Internet causes difficulty in both regression analysis and perturbation methods. Several approximate algorithms and their implementation, based on statistical methods, will be presented.

1.3.2 Nonlinearity

In this work only quantitative constraints between design parameters are considered, which can be either linear or nonlinear. This thesis emphasizes regression analysis and perturbation methods, both of which have problems in approximating sensitivity for nonlinear cases. Algorithms to improve the robustness to nonlinearity are presented and the evaluated. Also the possibility of applying Bayesian network is introduced.

1.3.3 Effect differentiation

Each design object (relationship) is MIMO (multiple inputs and multiple outputs) system. As a result, the behavior of the output is likely to be a combination of the effects from multiple inputs. Especially when the connections between inputs and outputs are hidden effect differentiation becomes more difficult. This work will present methods to differentiate effect from specific input from mixture of effects.

1.4 Organization of thesis

In chapter 2, the background for this research is presented. First, general product design procedures including constraint management are studied. Second, the characteristics of DOME are reviewed. The unique properties of DOME and the pros and cons resulting from these properties are discussed. Third, the role and definition of causal information is presented. The difficulty in determining causality and sensitivity under the presence of proprietary information is elaborated. The problem of circular dependency is also addressed.

Chapter 3 reviews previous and on-going causality and sensitivity analysis research. General aspects of constraint management are illustrated, emphasizing the properties that should be satisfied for a constraint set to be valid. Previous causality work in CADlab implemented by Tom Almy is discussed. He suggested the perturbation method, and also pointed out the problem with effect differentiation (Almy 1998). Approaches used in econometrics are introduced, including regression analysis and Granger causality.

In chapter 4, Internet latency (the time taken for the packet to be delivered) is studied. Unknown and varying latency for packet delivery causes difficulty in implementing both the regression analysis and perturbation method. To overcome problems caused by latency variance, a different transport layer protocol is used, and filtering algorithms for approximation are devised.

Chapter 5 provides implementation results for regression analysis. Issues, such as matching and bias, are discussed and the solutions for these problems are suggested. The regression analysis cannot prove causality because it is based on the assumption that there is a causal relationship between explanatory variable and dependent variable. For this reason, Granger causality is introduced, which can statistically prove the causality between two variables. Experiments are performed and the results are presented, emphasizing robustness to nonlinearity, interruption from other inputs and latency variance.

In chapter 6, the perturbation method is explained and the implementation results are discussed. Periodic perturbation is used to differentiate the effects from multiple inputs by transferring the problem domain from time to frequency. The robustness of the method to nonlinearity, interruption from other inputs, and latency variation is compared with those of regression method.

Chapter 7 summarizes the work and suggests the future work.

2

Background

The vision of this research is to provide the causality and sensitivity information to the product designers in a computer network-centric distributed design environment.

In this chapter the vision of this research is described from the context of the DOME environment.

Existing models of the product design process are illustrated to emphasize the importance of constraint management. Characteristics of DOME, a decentralized network-centric design framework, are discussed. Finally the contribution of causality and sensitivity information[4] to the design process itself and the product quality is elaborated.

2.1 Product Design

Product design is the process to achieve an acceptable realization (or implementation) from the set of functional (behavioral), performance and external appearance specifications for an artifact (Subrahmanyam 1992). Functional and performance requirements come from customer needs. The basic design activities are the exploration of customer needs, generation of new ideas, evaluation of the generated ideas, and communication between designers (Nigel 2000).

---

[4] In the following chapters, the term causal information will be used to represent both causality and sensitivity information

Several designers such as French, Archer and Pahl and Beitz have developed their own models of design process (David 1992). Iteration between the generation of new designs and the evaluation of ideas is common factor in all of these models.

2.1.1 French's Model

According to the French's model, the design process consists of the following stages (Nigel 2000).

1. Problem analysis: The problem definition for designing the product is made. Also the limitations are placed on the product and the criteria for performance is established. The outputs of this stage are the goals, constraints and criteria.

2. Conceptual Design: In this stage, broad and abstract solutions are generated based on the problem definition and criteria that were generated in the previous stage. Conceptual design consists of the generation and evaluation of the concepts. Numerous concepts are generated as candidate solutions for the design problem (concept generation) and their feasibility are evaluated through the comparison with existing concepts for similar problems (concept evaluation). For the comparison, concepts must have the same level of abstraction. This is the most demanding stage of the design process because the most important decisions, such as functional decomposition, are made in this stage. The role of communication is highly important because many different kinds of domain knowledge relevant to the product development, such as, engineering science, practical knowledge, production method and also commercial aspects need to be integrated. This stage possesses the greatest potential for the product quality improvement. Also, the time to develop and the overall quality of the final embodiment design is very dependent upon the concepts. Effective design simulation environment can be valuable for this stage and following stages because the numerous trade-offs are made during the decision-makings process.

3. Embodiment of schemes: A concrete detailed design is the result of this stage. Choices are made if there are multiple candidates of schemes. Usually the outputs of this phase are a set of layout or assembly drawings.

4. Detailing: Large numbers of decisions for the individual details of every aspect of the design are made in this stage. Careful consideration must be given to prevent the errors and failure, because this is the final stage of the design process and the outcomes are given to manufacturers.

2.1.2 Pahl and Beitz Model

Pahl and Beitz's model of design process consists of four steps: task clarification, conceptual design, embodiment design and detail design (Nigel 2000).

1. Clarification of the task: Gathering the information about the requirement for the product. The product requirements are stated in terms of performance, cost and time. The collected information will be embodied by solution and design constraints.

2. Conceptual design: A function structure[5] is established in this stage. Principles relevant to the solution of the problem are sought and embedded in the concept variants.

3. Embodiment design: Layouts and forms for the product are determined based on the concepts generated in the previous stage. Technical and economical aspects of the design problem are considered.

4. Detail design: Attributes such as arrangement, form, dimension and surface properties of all individual parts are made. Material selection, drawing and production notes are the outputs of this stage.

2.2 Constraint Management

As is illustrated by the design process model by Pahl and Beitz, the initial step of product development is the identification of a task (functional requirement of the product), and the progression from task to product is punctuated by design decisions. Design decisions change the design state, which is a snapshot of the information about the product being designed. There are two different views on the progress of the design process. One view is that the design progresses by the comparison between design states and requirements for the product (David 1992). This view is based on the assumption that the requirements for the product are already known from the beginning stage, and that the difference between requirements and design states is recognizable. In this view, the difference between requirements and design states governs the process. However, the design problems are often ill-defined and the full requirements are hard to be provided at the onset. A second view is that the design progresses as requirements constrain solutions to a subset of all possible design solution candidates. As the design progresses, new constraints are added to further reduce the solution spaces until finally only one solution is left. "*Design is the successive development and application of constraints until unique solution is achieved.*" (David 1992).

[5] The function of the product and mathematical function are similar in that both are generating output from the given input

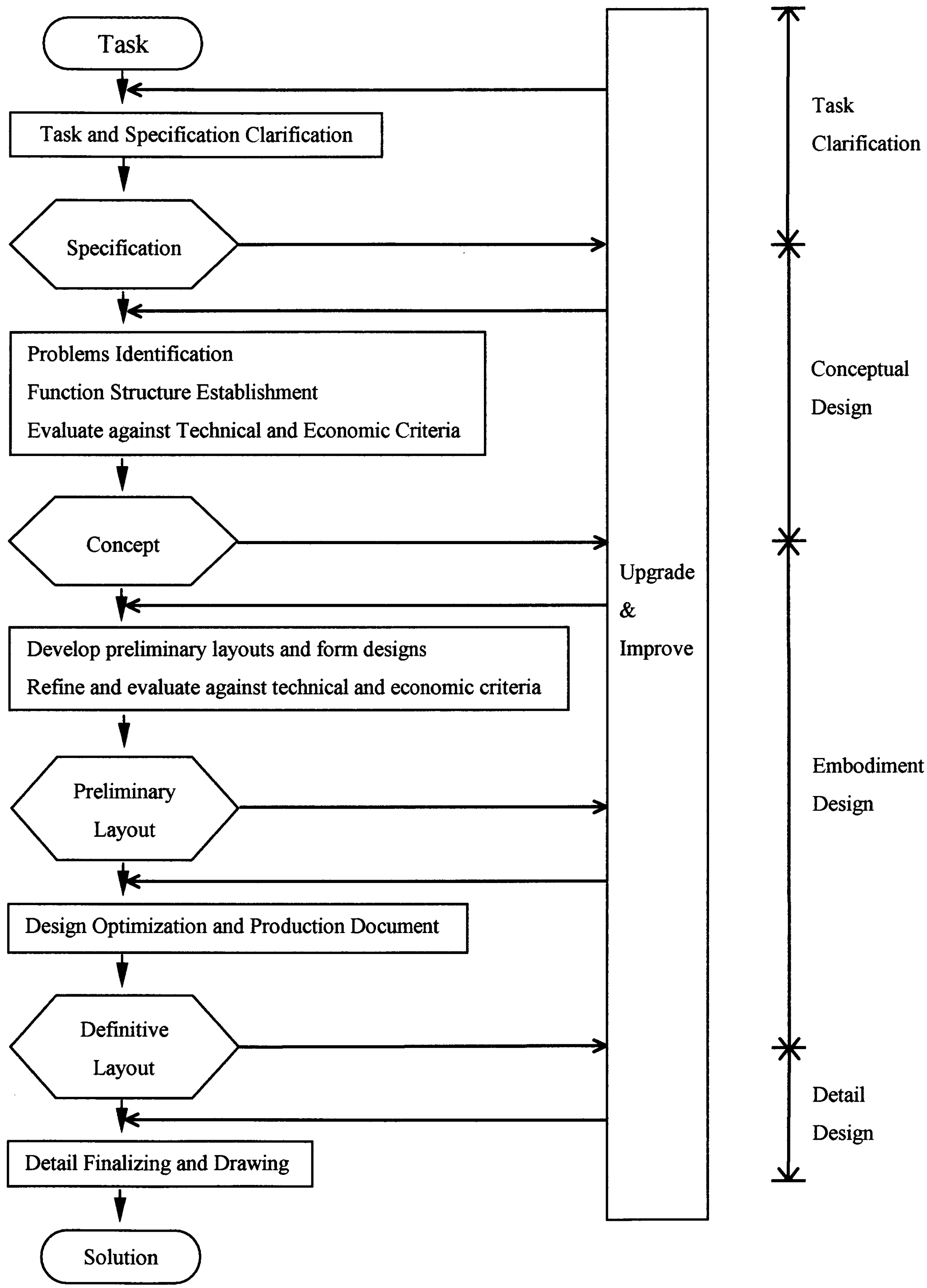


Figure 2-1 Pahl and Beitz's model of design process

The engineering design process- especially during conceptual design- is mostly focused on the recognition, formulation, and satisfaction of constraints. Constraints basically define the functional relations-, which need to be maintained between design parameters. Through the constraints, design parameters are related to desired performance, physical laws to be obeyed and geometric properties (Serrano and Gossard 1992). Also, constraints are used to represent domain knowledge relevant to the design problem (Baykan and Fox 1992).

The primary concern of this work is evaluation of constraints by tracking the causal chains. Causal information (causality and sensitivity between parameters) plays an important guiding role in the design decision-making process. When there are conflicts or under or over-constraint problems, causal information can be used to identify the specific aspects that need to be dealt with.

2.2.1 Constraint representation

The design task can be represented by constraints, with or without the objective function. Equation 2-1 is an example of constraints with an objective function.

$$
\begin{aligned}
&\min_{x,y,z} f(x,y,z) \\
&s.t. \\
&x^2 + y^2 = 25 \\
&x^3 + y^3 + z^3 < 289 \\
&x, y, z > 0
\end{aligned}
$$

Equation 2-1

Constraints with an objective function can be solved using general programming (optimization) methods. Tasks represented by constraints without an objective function can be solved using the constraint satisfaction methods[6].

Constraint structure (topology of the relation between variables) can be represented with a directed graph called constraint graph.

---

[6]Dechter, R., "Enhancement Schemes for Constraint Processing: Back-jumping, Learning, and Cutset Decomposition," Artificial Intelligence, Vol. 41, pp. 273-312, January 1990

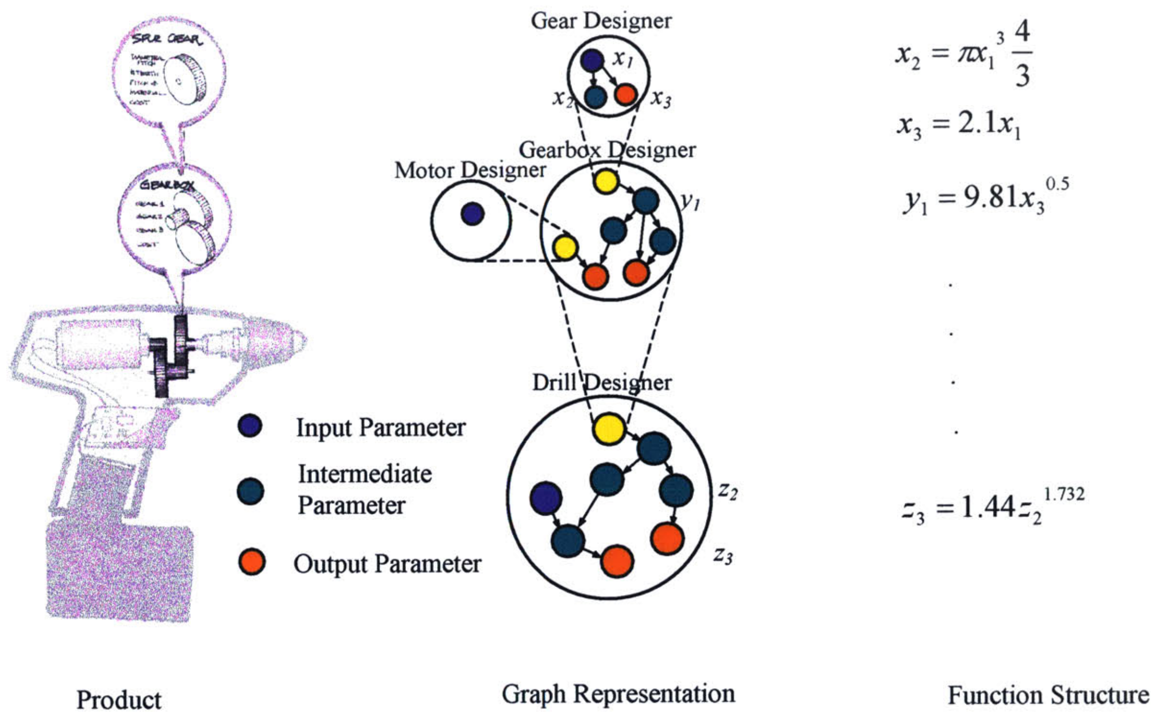


Figure 2-2. Graph Representation of structure constraints

As is illustrated in the Figure 2-2, the node pointed to by an arrow depends on the node on the other side of the arrow for its value. The direction of the arrow is reversed when the arrow indicates the dependency instead of causality.

2.2.2 Constraint generation

Design constraints are generated from two sources. First, constraints are formed by the designer's knowledge of mechanical devices or the specific problem being considered. In this case, the uniqueness of the solution comes from the uniqueness of the designer's knowledge. Second, constraints are added during the design process as a result of making design decisions. When adding such constraints, designers should be well informed about the characteristics of their solution.

2.2.3 Goal of constraint management

Consistency

There should not be any conflict or redundancy between constraints. This property guarantees the existence of the solution.

Completeness

There should not be unconstrained degrees of freedom left in a complete design. Uniqueness of the solution comes from this property of constraints.

Efficient evaluation of constraint network

In checking a validity (completeness and consistency) of a constraint set, relevant subset of constraints needs to be selected to reduce a problem complexity.

Providing guideline for the decision-making

Through qualitative analysis, such as causality analysis, a constraint set need to be characterized to support the decision-making (Serrano and Gossard 1992).

2.2.4 Detection of completeness and consistency

Table 2-1 Possible states[7] of a constraint set depending on the number of constraints

	Under-Constrained	Properly Constrained (Complete and Consistent)	Over-Constrained
Case 1: $\|N\|>\|F\|$	A	NA	NA
Case 2: $\|N\|=\|F\|$	A (Case 2-1)	A (Case 2-2)	NA
Case 3: $\|N\|<\|F\|$	A (Case 3-1)	A (Case 3-2)	A (Case 3-3)

$|N|$: number of unknowns, $|F|$: number of constraints, A: applicable, NA: non-applicable

As illustrated in Table 2-1, when the number of unknown variables is larger than that of constraints (Case 1), the constraint set is under-constrained. However, in Case 2 and Case 3, states of the constraint set cannot be distinguished by the numbers of the unknown variables and the constraints, because there can be redundant constraints. Case 2-1, 3-1 and 3-2 result from the redundant constraints. Redundant constraints are troublesome when they are in a strong component, because they cause singularity in Jacobean matrix. Case 2-1, 2-2, 3-1 and 3-2 can be distinguished by matching the unknown variables to the constraints. In Case 3-3, if unmatched constraints are consistent with the rest of successfully matched constraints, the unmatched constraints are redundant. Otherwise there is a confliction between the constraints, which requires a modification (Serrano and Gossard 1992).

[7] State indicates the completeness and consistency of a constraint set.

2.3 DOME

2.3.1 Communication and collaboration in product design problem

The product design problem can be regarded as a set of decisions to be made, which could be either qualitative or quantitative. Examples of qualitative decisions are selecting the color or material of a product, and examples of quantitative decisions include determining the dimension or the cost of the product.

Design decisions are made considering competing design objectives, which arise from the tasks or specifications that the product is expected to satisfy. As a result, decision-making has to deal with trade-offs in most cases, which necessitates efficient communication and collaboration between designers possessing different kinds of domain knowledge.

The DOME (Pahng, et al., 1998), which is an Internet based distributed design framework developed by MIT CADlab, provides a design environment enabling the real time collaboration between designers and simulations. DOME overcomes geographical barriers using the Internet as its communication channel[8]. As a result, wider variety of design simulations can interact in the design process, thus generating a larger solution search space.

Through real-time simulation of product characteristics, each designer can receive immediate feedback regarding his or her decision to facilitate the understanding of trade-offs involved with the decision. These advantages increase the chances for the various kinds of domain knowledge to be integrated into the design, which can be highly effective in improving the product quality. Additionally, the required period for the product development can be reduced dramatically[9].

Also DOME's object-oriented decentralized architecture[10] has the following characteristics.

1. Flexibility: Object-oriented decentralized characteristic of DOME allows both top-down and bottom-up approaches to the problem modeling.
2. Scalability: The DOME can adjust the size and complexity of the simulation depending on those of a design problem.
3. Heterogeneity: Each designer can participate in the design process focusing on their own competency, and exchange services with others in the design network. Services are exchanged as simulation inputs and outputs.

---

[8] However, using the Internet as communication channel causes the problems of asynchrony and difficulty for chronological tracking of the decisions. The asynchrony problem is caused by unknown and varying Internet latency, which is discussed in chapter 4.

[9] The advantage has been proved through the pilot project with Polaroid and Ford.

[10] This is the feature of service exchange network, the details of which are discussed in appendix.

4. Proprietary information: In DOME environment, designers do not have to reveal their own know-how because there is no central server managing the whole information involved with the design problem. Protecting proprietary information leads to the participation and sharing of services between a wider variety of designers. However, this feature causes difficulties in providing designers with knowledge of the problem structure. Problem structure can be well defined by the topology of the relations (constraints) between design parameters. For these reasons, providing an idea on problem structure through problem model will lead to a better set of decisions.

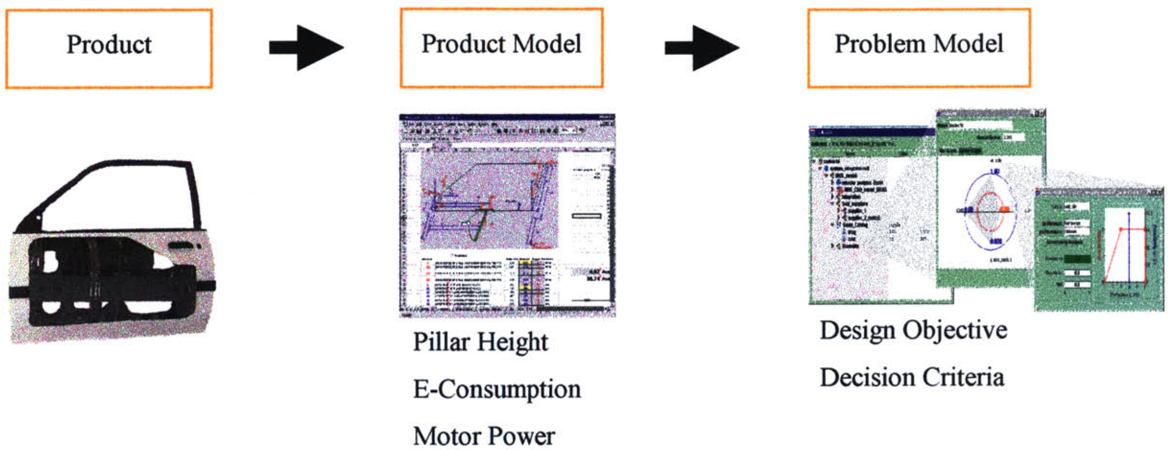


Figure 2-3 Design Problem Model

2.3.2 Object-oriented Modeling and Evaluation (OME) framework

Modeling and evaluating a design problem requires that a large number of aspects be considered. As a result, problems are often solved in a divide and conquer manner, dividing the problem into many sub-problems and solving each with designers from appropriate disciplines. Eventually solutions to the sub-problems are integrated and evaluated to form the complete solution for the whole problem. A divide and conquer approach can reduce the problem complexity and makes the problem easier to solve. However, it becomes more difficult for each designer to see the whole product, and understand global trade-offs. As a result, iteration is likely to be required which can be expensive in time and computation. The purpose of OME is to structure the decomposed sub-problems in a systematic way to identify interactions between sub-problems. Interactions provide designers involved in sub-problems with the knowledge needed to interact with the problem structure as a whole.

2.4 Causality

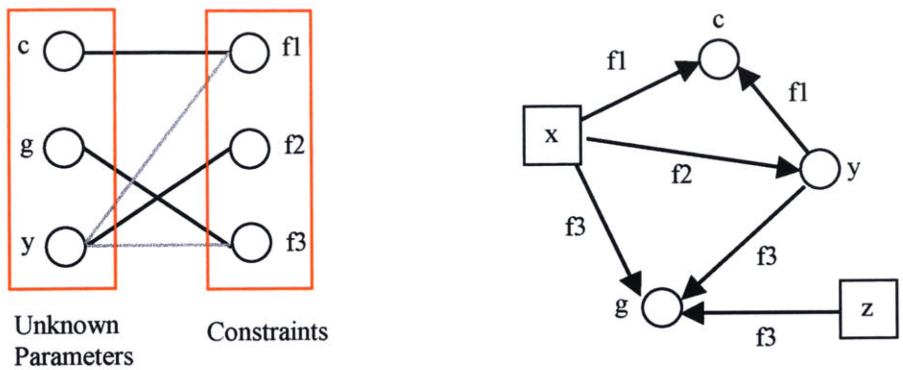
The direction and topology of the information flow in the design problem is called *causality*. The causal information defines the dependency between parameters by identifying the function structure of the design problem.

Causality information is important in estimating the expected performance of the product. Also, the feasibility of the design can be verified by tracking the causal chain. In terms of the constraint management problem, knowing causality information can narrow down the search space when trying to resolve conflicts.

2.4.1 Causality in local scope

Causality in a local scope can be determined manually by designers, or it can be automatically generated since complete knowledge of the sub-problem is available. Serrano and Gossard (Serrano and Gossard 1992) suggested a systematic method for automatic causal chain tracking.

1. Match unknown parameters (outputs) to the corresponding constraints.
2. Build bipartite graph. The nodes are divided into two sets: unknown parameters ($N = \{n_1,..,n_p\}$) and constraints ($F = \{f_1,..,f_r\}$). The edges ($E = \{e_1,..,e_k\}$) are connecting members of N to those of F only. No member of N is connected to another member of N and no member of F is connected to another member of F.
3. Therefore each edge in E indicates the unknowns and the constraints with which the unknowns are involved. As a result, there is no direct relation between constraints or unknown parameters: all the relations between unknown parameters are made through constraints and vice versa.
4. Find the maximum matching between N and F, which is to find the largest subset of E where no two pairs have the n or f in common. Maximum matching is complete when the number of matchings (cardinality) is same with $|N|$ or $|F|$. Here, the number of unknowns and that of constraints must be the same, which is one of the necessary conditions for the existence of the solution.
5. Once the matching is finished, the result can be applied to the original problem model network, which represents the topology of the constraints. In original problem model network, couple of parameters (nodes) are connected by an undirected edge representing the constraint in which both parameters are included. In this configuration, status of a parameter: whether it is an input or an output is not recognizable (Serrano and Gossard 1992). Steps 1 through 5 are based on the assumption that known parameters (inputs) and unknown parameters (outputs) are distinguishable and each constraint is MISO (multiple inputs and single outputs) system.



a) Bipartite Graph

b) Modified Original Problem Model

Figure 2-4 Graph Representation of Local Causality (Serrano and Gossard 1992)

In Figure 2-4, the squares are inputs and the circles are outputs. Input nodes do not have an incoming edge, indicating they do not depend on other variables. Gray links means that the corresponding unknown parameter is involved in a relation as an input. In Figure 2-4 the direction of the arrow indicates the causality. The direction of the arrow is reversed, when the arrow represents the dependency.

Constraint network evaluation

After matching is finished and the modified problem model is built, a constraint network can be evaluated. Properties such as consistency, completeness and existence of cyclic dependency are checked to ensure a valid constraint set. When there is a conflict between the constraints, backtracking is required, and the amount (number of steps) of backtracking depends on a tightness of coupling between the constraints (Sanjay and Agustin 1992). Cyclic dependencies are not allowable in some situations, like modeling manufacturing processes, because it means an inconsistency. However, in other situations, cyclic dependency means that the involved parameters need to be solved simultaneously (Serrano and Gossard 1992). The solution for the cyclic dependency problem will be suggested latter part of this chapter.

.1 Conflict resolution and feasibility checking

Decisions of design parameters need to be organized to satisfy a set of design goals as illustrated in Figure 2-5 (Sanjay and Agustin 1992). When there is a conflict between constraints, backtracking determines a design goal that requires modification. Also, a feasibility of a design goal can be verified by checking the feasibilities of design parameters that depend on the design goal.

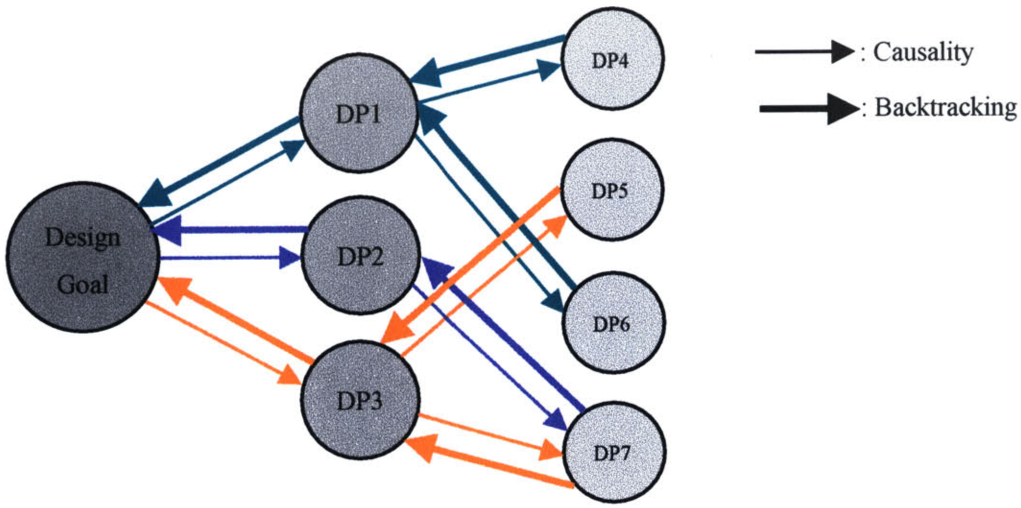
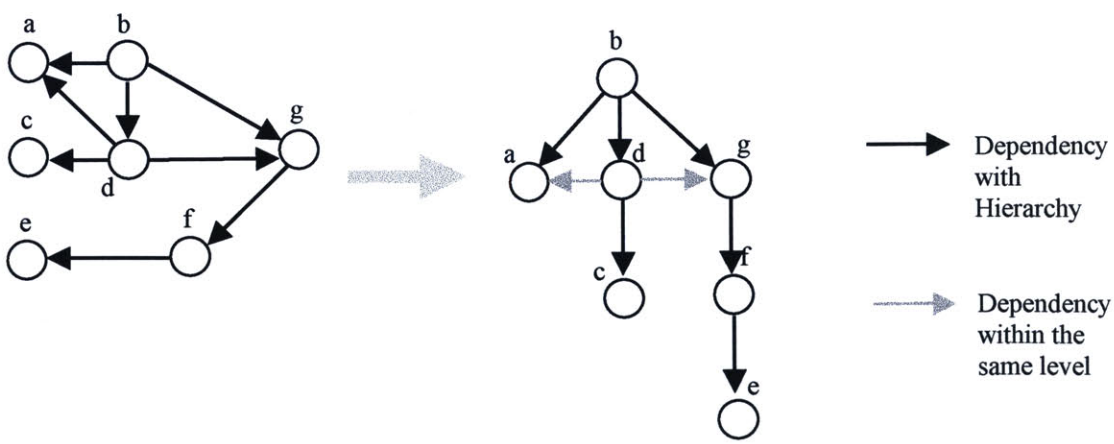


Figure 2-5 Design goal and parameters

Topological sorting

Under the assumption that the constraint network is a DAG (Directed Acyclic Graph), the constraint network behaves very similar to a tree. Topological sorting on a DAG is a process to sort the nodes of a graph so that no node is preceding the node from which the incoming edge is coming. In general, multiple solutions can exist for topological sorting, because many parameters could exist on a same layer. However, once the topological ordering is finished, the solution sequence and the causal chains are directly recognizable from the sorted graph.



a) Original Constraint Network b) Tree-resembling Sorted Graph

Figure 2-6 Causality (Dependency) shown by topologically sorted DAG (Serrano and Gossard 1992)

2.4.2 Causality in DOME framework

The object-oriented characteristic of DOME allows heterogeneity between design objects and protects embedded proprietary information. Thanks to the ability to include heterogeneous design objects, each designer can participate in design process with his own competency only by defining the service interface with the problem model. By protecting proprietary information, designers can hide core capabilities, including the structure of their models, during the integration process. However, this prevents other designers understanding the relationships between parameters, which is very important information in understanding the problem characteristics and trade-offs related to decisions. Causality information is valuable not only because it helps in decision-making, but also because it can be used to diagnose problems or solve problems efficiently.

2.5 Sensitivity

Quantitative aspects of causal information can be defined through the sensitivity, which is the strength of the relationships between parameters. With the sensitivity information, the designer can see the degree of impact of one parameter on another.

Sensitivity information is very important to product quality, because it directly indicates the robustness of the design to manufacturing variation, aging effects and environmental changes. In this work, sensitivity will be addressed from two different scopes: local and global.

Sensitivity information is also very important in specifying the tolerance of a feature. Rough tolerance can be given to a feature, to which the performance[11] is less sensitive.

2.5.1 Example-Motor Design

Suppose a designer participates in developing a motor, and his responsibility is to determine the torque and the angular velocity, given the predetermined user requirement for power.

The relation between the torque and the angular velocity is shown in Figure 2-7.

[11] The relation for the performance is assumed to known

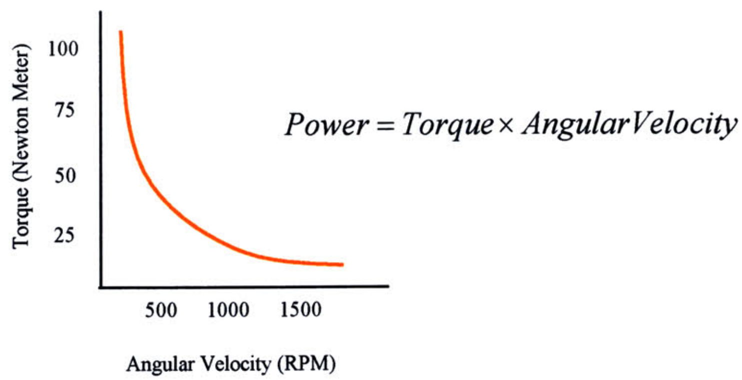


Figure 2-7 Torque and Angular Velocity

All values of torque and angular velocity on the line satisfy the power requirement. However, as illustrated in Figure 2-7, the sensitivity of power to torque and angular velocity varies along the line. For example, as the angular velocity increases the sensitivity of power to angular velocity ($\frac{\partial Power}{\partial Angular\,Velocity}$) becomes smaller and the opposite is true for torque.

In this situation if the motor manufacturer desires a motor to maintain a constant torque over angular velocity, determining the value of the torque in sensitive region and angular velocity in a robust (insensitive region) leads to a better design. Larger angular velocities will give a better design.

2.5.2 Sensitivity in a local scope

In this work, the local[12] scope of sensitivity is defined as the sensitivity between the parameters within the same design object. In local scope, there is no hidden causal information involved in determining sensitivity.

Single constraint case

Constraint: $f(x_1,..,x_n) = 0$

$$\frac{\partial f}{\partial x_1}dx_1 + \ldots + \frac{\partial f}{\partial x_n}dx_n = 0$$

Equation 2-2

---

[12] Being local means the information is within the reach of a designer.

Sensitivity of x_m to x_l: $\dfrac{dx_m}{dx_l} = -\dfrac{\dfrac{\partial f}{\partial x_l}}{\dfrac{\partial f}{\partial x_m}}$

Equation 2-3

Multiple constraint case

Constraint: $f_m(x_1,..,x_n) = 0, m = 0,..,j$

(Equation 2-4)

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_{21}}{\partial x_n} \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial f_j}{\partial x_1} & \cdots & \cdots & \frac{\partial f_j}{\partial x_n} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \\ \cdot \\ \cdot \\ \cdot \\ dx_n \end{bmatrix} = 0$$

(Equation 2-5)

Sensitivity of x_m to x_l: $\dfrac{dx_m}{dx_l} = -\dfrac{\sum_{m=1}^{j} \dfrac{\partial f_m}{\partial x_l}}{\sum_{m=1}^{j} \dfrac{\partial f_m}{\partial x_m}}, \; dx_k = 0, for\, k \neq l, m$

(Equation 2-6)

2.5.3 Sensitivity in the DOME framework

In the DOME framework, not only local but also global sensitivity information are required. In figuring out the global sensitivity, an analytic expression of constraints, such as (Equation 2-4), cannot be expected because of the existence of the proprietary information[13]. As a result, empirical or statistical methods may

[13] Proprietary information plays a role of black box in the constraint network.

provide a solution to the problem. However, in many cases, a designer cannot control all the parameters included in a constraint of interest. That is, a designer cannot make $dx_k = 0$ for all k except $k=l, m$ in (Equation 2-6). As a result, there is a difficulty with sensitivity analysis in effect differentiation.

2.6 Circular Dependency

In the design problem model, circular dependency between parameters represents the need for the simultaneous solving of constraints or relationships. When a perturbation is made to the set of relations within the circular dependency, the effect can either converge or diverge. Divergence means there is conflict between the requirements.

Granger's statistical method for detecting circular dependency from the behavior of the involved variables, and Serrano and Gossard's solution plan for the circular dependency are detailed in chapter 5.

3

Related Work

In this chapter, research relevant to this thesis is presented. Work on general constraint management in the area of computer network-centric design environment will be reviewed. Next, previous work in CADlab by Tom Almy (Almy 1998) will presented with his suggestion for perturbation and the effect differentiation method. Finally, causality analysis in econometrics is elaborated.

3.1 Constraint Management

Constraint management is intended to maintain

1. *Consistency*
2. *Completeness*
3. *Efficient evaluation.*

Various approaches have been developed to reduce complexity and computational effort by decomposing constraint sets in an appropriate way. Harary and Steward dealt with the consistent constraint set problem without consideration of inequality constraints (Harary 1962 and Steward 1962). Light suggested a way to check for completeness and consistency from Jacobian matrix of the constraint set, as illustrated in (Equation 2-2) (Light 1980). Chyz fabricated a two-dimensional geometric constraint manager to maintain consistency. Garret developed a methodology to select an applicable constraints set for a specific design problem. His approach used optimization to find a solution for the set of constraints. However, the approach did not deal with causality modification (Garret and Fenves 1986). Serrano developed a software package named MATHPAK, which has a reversing causality feature. MATHPAK can handle both geometric and non-geometric constraints (Serrano and Gossard 1986).

3.2 Design Complier

Daisie Boettner and Alen Ward developed a tool called the design compiler to convert the description of the product from a high-level (abstract) language to a detailed implementable (concrete) one.

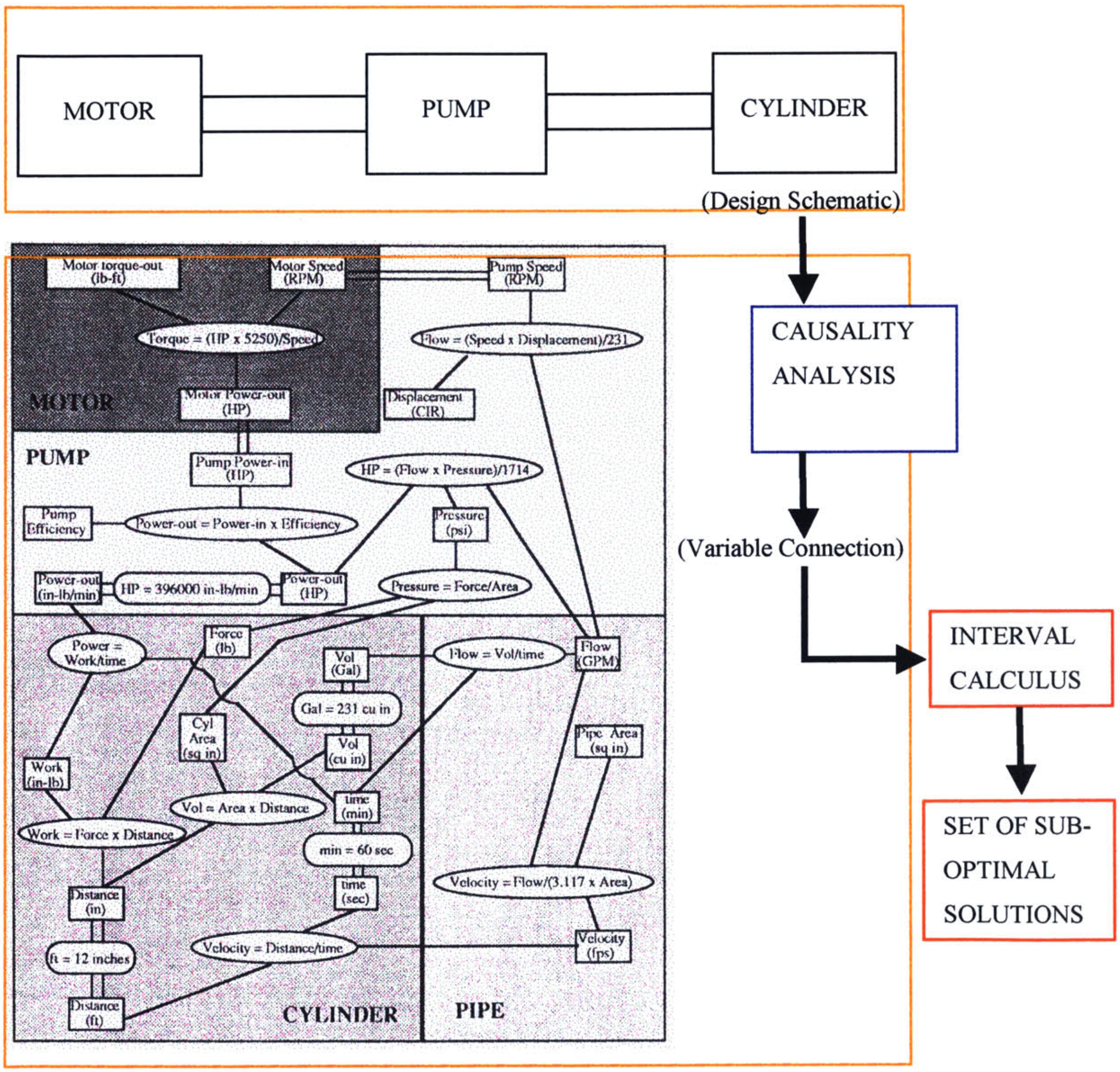


Figure 3-1 Design Compiler (Boettner and Ward 1992)

The idea was to propagate sets of values (intervals) instead of single values, for the purpose of providing a tool for elaborating solution spaces. Boettner and Ward focused on the fact that, in mechanical design, the relations between components are as important as formation of each component. The inputs for the compiler are specifications and utility functions, and the outputs are possible implementations. The compiler has an ability to find an optimal solution using a novel interval calculus mathematical scheme. Figure 3-1 is an example of the use of the design compiler.

3.3 Previous work on causality in CADlab

Previous research on causality and sensitivity analysis was done by Tom Almy in MIT CADlab (Almy 1998). The focus of his research was on ordering the variables and visualization of acquired causal information. He also suggested a perturbation method to find hidden causality and sensitivity information. The basic idea of perturbation method is to give a small variation to the input variable and observe the resulting change-in target (output) variables. The ratio of output change to the input variation is equivalent to the partial derivative of output to input, based on the assumption that the related constraint can be regarded as linear in the range of variation.

$\frac{\delta \text{output}}{\partial \text{input}_i}$

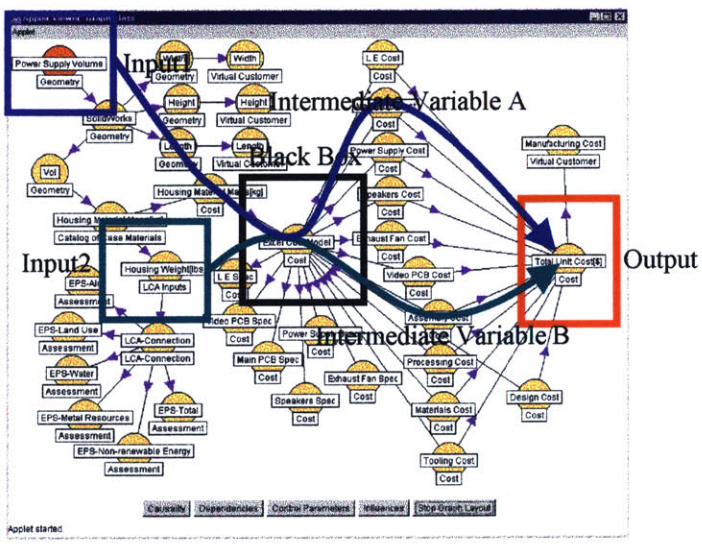


Figure 3-2 Visualization of the Causal Information (Almy 1998)

However it is explained in the work that the perturbation method had a problem with identifying the source of effects causing change in the output variable. As can be seen from Figure 3-2, it is not directly recognizable which input is connected to the output. Also, it is hard to track the intermediate variables through which the effect is propagated. Thus, the perturbation method by itself cannot directly determine the topology and the strength of relationships.

3.4 Causality and Sensitivity in Econometrics

Causality and sensitivity analysis has been studied widely in the field of econometrics. Many approaches have been based on regression analysis, which will be explained in detail in the later chapter. The focus of the research is whether the relation between two parameters is a mere correlation or causation. For example, if the sugar price of Mexico is observed to change after the price of milk in the United States changes, two different scenarios can arise from the same observation: 1) the sugar price of Mexico *causes* the change in milk price within the United States, or 2) there is a third factor, such as coffee price in Brazil, that affects both prices with a different time lag. Figure 3-3 shows two different configuration of causal chain with same phenomenon

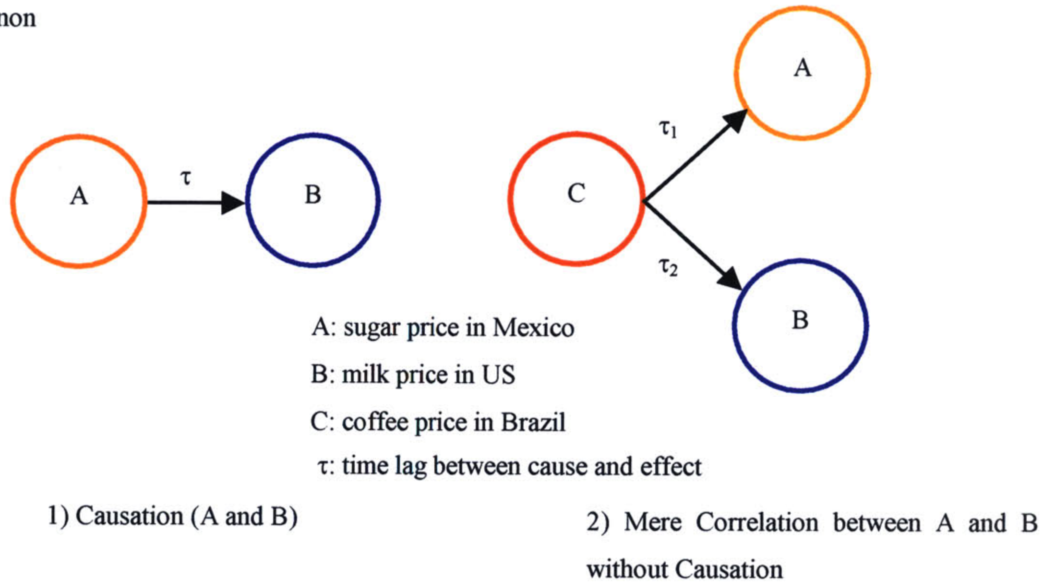


Figure 3-3 Causation and Correlation

When a time lag between C and A (τ_1) is smaller than that between C and B (τ_2), the difference between 1) causation and 2) mere correlation is difficult to distinguish. ok

3.4.1 Detection of causality

Distinguishing causation from correlation has been an important issue in econometrics. Regression analysis is based on an assumption that there is a causal relation between parameters, so regression cannot be used as verification for causality-it can only provide an approximation for sensitivity. Granger (Granger 1969) suggested a way to detect causality from the time series. The basic idea behind Granger causality is that if the variance of one variable, say Y, is larger than its variance conditional upon past value of the other variable, say X, then we can conclude that X causes Y.

Granger Causality: $\sigma^2(Y_t \mid \overline{A_t}) < \sigma^2(Y_t \mid \overline{A_t} - \overline{X_t})$

Equation 3-1

3.4.2 Circular dependency

Granger also suggested the method to detect circular dependency between variables by checking the following equations.

$$\sigma^2(Y_t \mid \overline{A_t}) < \sigma^2(Y_t \mid \overline{A_t} - \overline{X_t})$$

$$\sigma^2(X_t \mid \overline{A_t}) < \sigma^2(X_t \mid \overline{A_t} - \overline{Y_t})$$

Equation 3-2

3.4.3 Measuring sensitivity

Regression analysis is a methodology to determine the dependency of one variable, the dependency variable, on another variables, explanatory variables. The main purpose of the analysis is to estimate and predict the distribution, namely mean or average, of the dependent variable when those of the explanatory variables are given. Mainly, the dependent variables in regression analysis are stochastic or random variables. A stochastic variable is a variable that has a certain probability distribution. The explanatory variable does not enable one to perfectly predict the behavior of the dependent variable due to the presence of errors and unidentifiable explanatory variables. This explains the stochastic characteristic of the dependent variable. Generally it cannot be expected that all the explanatory variables are identified. The effects from unidentifiable variables are represented using disturbance terms. However, there are certain conditions between disturbances and explanatory variables-which must be satisfied for the analysis to be able to produce a valid result.

In this work, the regression analysis will be used as one of the tools to assess the sensitivity based upon the magnitude of regression coefficients.

3.4.4 Bias

Fundamental assumptions for the validity of the regression method are:

1. No correlation between disturbances
2. Zero mean of disturbances.
3. Equal disturbance of the disturbances.
4. Zero covariance between explanatory variable and disturbance.

There are several other conditions, but this list provides the conditions of relevance to this thesis. Disturbances can be understood as unknown or ignored causes. In the situation of DOME, the first two assumptions cannot be satisfied. As a result, bias is unavoidable in regression analysis. Several theories have been developed to resolve the bias problem.

Instrumental Variable Method

For a regression analysis to produce a valid result, the covariance between explanatory variable and the disturbances must be zero, which cannot be guaranteed in DOME due to the existence of proprietary information.

The instrumental variable (IV) method is a method to remove the correlation between explanatory variable and disturbances by using a proxy variable for the explanatory variable. A proxy variable is defined as the variable that shows a strong correlation with the explanatory variable but has zero-covariance with disturbances. Such a proxy variable is called instrumental variable. However, finding an appropriate proxy is not always possible. In case that a proxy cannot be found one may have to use maximum likelihood estimation techniques (Gujarati 1995).

3.4.5 Latent Variable Model

The goal of latent variable modeling is to make the relation between two observed variables explicable using a third variable that is unobservable. The relation between explanatory variables with latent variables and their indicators needs to be modeled (Ben-Akiva 2000). The latent variable modeling can be considered as a data reduction method because it reduces the full set of explanatory variables to a few latent variables. A latent variable is an unobservable variable, which affects the behavior of the observable variables. As it is difficult to observe, the behavior of latent variables are inferred by an indicator variable. The latent variable approach has many applications, including causal modeling. The main concern of the causal modeling is to find the parameter is a system equation model (SEM) that relates the explanatory variables with the dependent variables. (Everitt 1984)

3.5 Artificial neural network (ANN) approach

An ANN emulates a human intelligence through simplified replication of a brain's adaptive ability to learn from experiences. The ANN consists of many simple processors (units, nodes or neurons), which are equipped with a local memory. Function of the unit loosely imitates that of a biological neuron[14]. The ability of an ANN comes from inter-unit connection strength (weight), which is established by learning from training (Kevin 1997).

Previous work in MIT CADlab was done by Juan Deniz (Juan 2000). ANN was used as a surrogate model for a design simulation network (DOME) to predict an output for an untried set of inputs. Using the ANN could reduce the computation time, when the design simulation includes a sub-model with heavy computation time, such as FEA (Finite Element Analysis) module. For example, he applied the ANN to a

---

[14] ftp://ftp.sas.com/pub/neural/FAQ.html

genetic algorithm procedure and reduced the running time to 3 hours and 12 minutes, which original had taken 30 hours.

ANN can be used to estimate sensitivity. However, the most difficult parts of sensitivity analysis using neural networks are 1) collection of accurate data, 2) selection of good input factor and 3) interpretation of output data.

4

Internet Latency, Protocols and Node Structure-Statistics and Experiments

In this chapter, characteristics of Internet latency will be illustrated with statistics and experiments[15]. Latency depends on the Internet protocol, the interval between packet departures, the packet size and network traffic volume. Each protocol has its own pros and cons, and is suited for different applications.

The dependency of latency on the protocols, packet size, and time scales will be used to resolve the problem of unknown varying latency, which causes difficulties in implementing both the regression analysis and the perturbation method.

4.1 Latency Problem with Regression Analysis and Perturbation Method

In later chapters, two different methods will be suggested as a solution for the causality and sensitivity analysis: regression analysis and perturbation methods.

One of the most challenging parts in both approaches is to distinguish an effect of a specific input from the behavior of the output value, which is a combination of the effects from multiple inputs. For both of the methods, periodic signals are used as a mean to differentiate the effect. In regression analysis, frequency is

---

[15] As illustrated in Chapters 1.1.1 and 1.1.2, the Internet latency characteristic is important in implementing both regression analysis and perturbation methods.

used as a clue to match an input to an output. In perturbation methods, frequency information is used to transfer the problem from time domain to frequency domain for effect differentiation. Thus, maintaining an initial periodicity in the perturbation against the Internet latency variation is the most important issue. For this reason, sources and characteristics of the Internet latency are studied in this chapter to improve the precision of both methods.

4.2 Packet Switched Network

Generally, computer networks are packet switched networks, where the information is exchanged through discrete blocks of data called packets. In contrast with circuit switched network, packet switched network does not establish a dedicated connection between the source and the destination of the data blocks. Instead, many hosts share the network through packet multiplexing by the switches. The basic idea of multiplexing is similar to that of CPU timesharing (Roshan, et al., 1982).

4.2.1 Multiplexing: STDM, FDM and SM

The main role of the switch is to store and forward the packets to the next node in the network. For multiple hosts to share the network, packets from hosts are multiplexed at the switches. There are three different methods for multiplexing the multiple flows into one physical link.

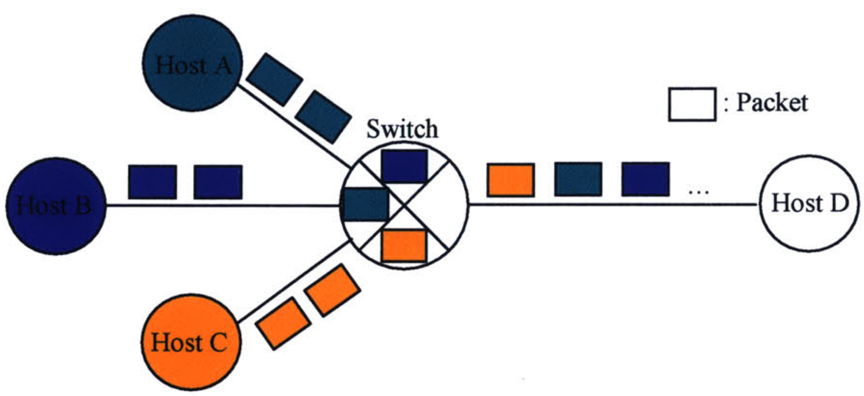


Figure 4-1 Multiplexing

STDM (Synchronous Time-Division Multiplexing)

This method allocates the same amount of time (quanta) for all the hosts, regardless of their demand. As a result, every host takes its turn regularly, but the link becomes idle when the host does not have a packet to send during its quanta.

FDM (Frequency-Division Multiplexing)

In this method, each host transmits packets through the same channel with different frequencies. This method also has the problem of idle frequencies when the host that uses specific frequency is idle.

In addition to the problem of idle time, both FDM and STDM have difficulty in scaling, when a new host is added to the network.

SM (Statistical Multiplexing)

The idea of SM is similar to that of STDM in that the links are shared over time. However the amount of time allocated for each host is not predetermined but it is determined on-demand. The links are used by the hosts in the order of demand. However, the size of a packet needs to be limited to guarantee that every host eventually takes its turn. This limit on size may require the host to fragment its message into several packets, and the receiver has to reassemble the packets into an original message (Peterson and Davie 1996).

The transmission of data through packet switched networks using SM experiences unknown varying latency (delay), but it is the most cost-effective way of sharing the network, because the resources are used in the most fine-grained manner. The advantages and disadvantages of the packet switched network are:

1. Advantages
    a. Efficient location and retransmission of the corrupted data block.
    b. Higher reliability by routing around the failed path
2. Disadvantages
    a. Unpredictable latency

4.2.2 Packet

A packet is a block of data that is used as a unit for information transmission. There are two reasons for using packets. First, the sender and receiver need to coordinate transmission in order to exchange data correctly. Dividing the data into blocks enables both sending and receiving sides to identify and locate the corruption of a data block. When there is an error (packet loss), the sender has only to re-send the lost packets, which is more efficient than resending all of the information. Second, it ensures fair access by each computer to the communication facilities. Infrastructure sharing between computers is unavoidable because such facilities as router and switch are too expensive for each computer to have one to itself (Peterson and Davie 1996).

4.3 Bandwidth and Latency

Bandwidth and latency are two main criteria used to measure link performance. If one compares data transmission with water flowing in a pipe, the diameter of pipe is equivalent to bandwidth and length of the pipe is latency.

4.3.1 Bandwidth

Bandwidth is represented by the number of bits transmitted in a unit time.

$$Bandwidth = Size\,of\;Packet\,/\,Transmission\,Time$$

Equation 4-1

In Equation 4-1, the transmission time is the time taken for a unit of data to be transmitted. The unit of bandwidth is bps (bit per second). For example, the channel of bandwidth 10 Mbps can transmit million bits per second. The performance can be dominated either by bandwidth or by latency, depending on the application.

4.3.2 Latency

Latency is a time taken for a data to be delivered from one end to the other. Latency consists of three components: propagation time, transmission time and queue time. Propagation time is a time taken by a physical layer, such as a wire or cable. For example, the propagation time for cable is $2.3\times10^{8}\,meters/\sec$ and that of fiber is $2.0\times10^{8}\,meters/\sec$. Transmission time is the time required for a unit of data to be transmitted. Queue time is time spent at the buffers of network nodes, such as routers and switches etc (Peterson and Davie 1996). In Equation 4-2, propagation time and transmission time are fixed components, and queue time is a variable component (Bolot 1993).

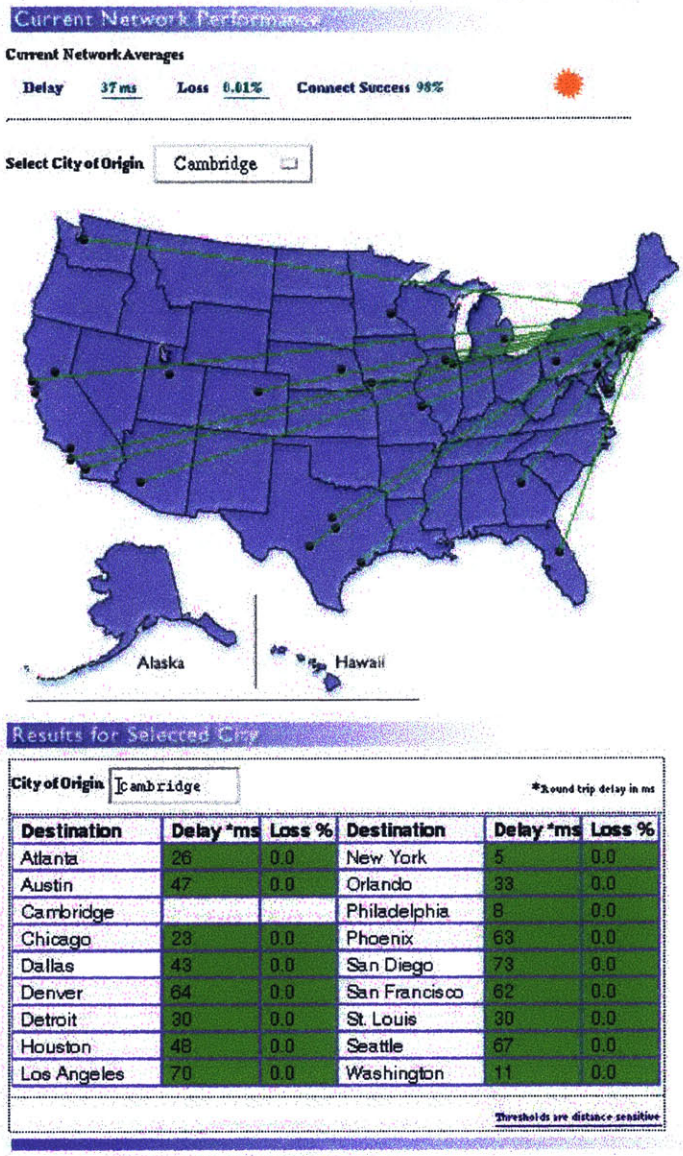
Merit Network Inc. surveys the Internet latency every 15 minutes. Shows an example.

$$Latency = Propagation\,Time + Transmission\,Time + Queue\,Time$$

Equation 4-2

$$Propagation = Distance\,/\,Speed\;of\;Light$$

Equation 4-3



Destination	Delay *ms	Loss %	Destination	Delay *ms	Loss %
Atlanta	26	0.0	New York	5	0.0
Austin	47	0.0	Orlando	33	0.0
Cambridge			Philadelphia	8	0.0
Chicago	23	0.0	Phoenix	63	0.0
Dallas	43	0.0	San Diego	73	0.0
Denver	64	0.0	San Francisco	62	0.0
Detroit	30	0.0	St. Louis	30	0.0
Houston	48	0.0	Seattle	67	0.0
Los Angeles	70	0.0	Washington	11	0.0

Figure 4-2 Latency statistics (http://ipnetwork.bgtmo.ip.att.net /current_network_performance.shtml)

RTT (round trip time) is the time taken for a bit to travel from one host to the other and back to the original host again. In some situation where the transmission requires acknowledgement from the receiving side, RTT is more meaningful than one-way latency. The size of a packet is one of the factors that dominate RTT and the probability of packet loss. One of the problems with packet switching networks is that they do not have control on delay (latency) of packet delivery.

4.4 The Internet Layers an Protocols

There are several different points-of-view on the architecture of connections between computers. The most commonly adopted model is the one with 7-layers defined by ISO (International Standards Organization) called OSI (Open Systems Interconnection) in Figure 4-3. According to the OSI, the Internet consists of an application layer, presentation layer, session layer, transport layer, network layer, link layer and physical layer (Peterson and Davie 1996).

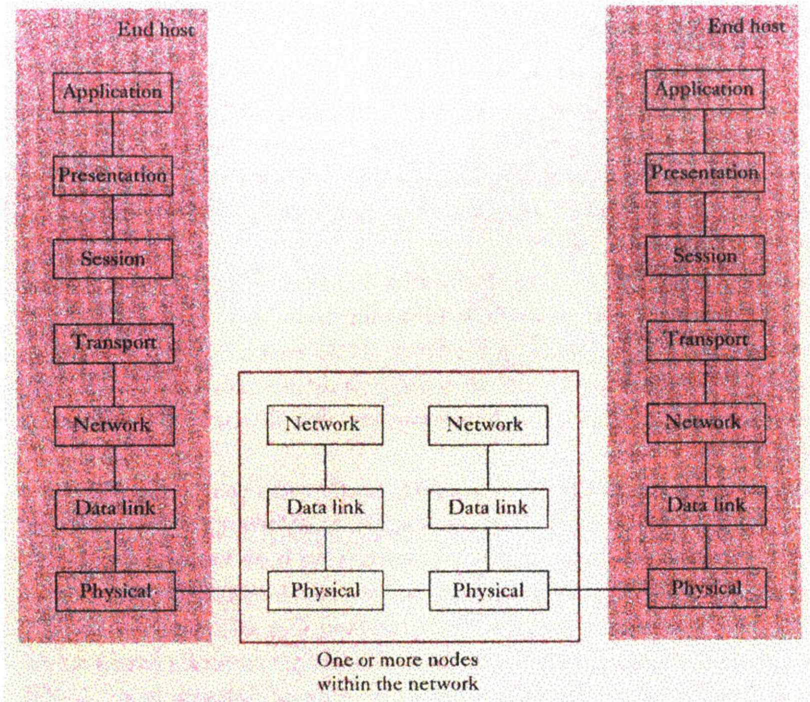


Figure 4-3 OSI Definition of Internet Layers

4.4.1 Application layer

Application layer is supporting the interoperability of an application that is working on the Internet. This layer is a closest layer to the users. Protocols working in this layer are ftp and http, etc.

4.4.2 Presentation layer

This layer deals with the format of data exchange. For example, the number of bits assigned to a character or an integer, and the position of a most significant bit.

4.4.3 Session layer

A name space is provided in this layer, to integrate different types of data stream used for a same application.

4.4.4 Transport layer

This layer implements the communication between end processes. At this level, the data exchange unit becomes a message rather than a packet. TCP and UDP are the working protocols in this layer.

4.4.5 Network layer

Datagram routing is done in this layer. The transmission between switches and the routers are executed, and a unit of data exchange is a packet. IP (Internet protocol) and routing protocols are used.

4.4.6 Link layer

Data transmission between network elements is done in this layer. Byte-stream is collected into a larger chunk of data called frame. Network adaptors typically implement this level. Ethernet is the most popular protocol.

4.4.7 Physical layer

Physical transfer of bits is executed through the wire or cable. The network, link and physical layer are implemented on all network nodes including switches (Peterson and Davie 1996).

4.5 TCP (Transmission Control Protocol)

TCP is one of the protocols used in transport layer. TCP is connection oriented. As a result, TCP is very reliable and packets arrive in the same order in which they are transmitted. However, its acknowledgement and retransmission mechanism can be time consuming. Moreover, high frequency packet transmission causes fluctuation in RTT, because of interactions between data and acknowledgement packets (Bolot 1993). Applications where TCP is used include HTTP, FTP and SMTP (email).

4.5.1 Acknowledgement and retransmission

TCP is connection-oriented in that it requires explicit connection establishment phase before starting communication, which is called handshaking. To transmit data through TCP, the sending side must always receive acknowledgement before the transmission of the next packet. Bolot (Bolot 1993) performed an experiment sending TCP packets in various time intervals to evaluate the structure of Internet load. He found that in realistic situations where the two-way traffic is allowed, the interaction between data and acknowledgement packets generate clustering of the acknowledgement packets, which causes rapid fluctuation of the variable component of latency (queue time in Equation 1-2) (Bolot 1993). Packets with a small time scale will arrive at the receiver with large variation of latency.

4.5.2 Flow control

TCP has flow control so that the sender does not overload the receiver. Flow control supports the communication between ends.

4.5.3 Congestion control

TCP also controls the rate of transmission to avoid overloading the switches or links. Congestion control supports the interaction between network and host. (Peterson and Davie 1996)

4.6 UDP (User Datagram Protocol)

UDP is the simplest possible protocol used in transport layer. It does not have flow control or congestion control. Also, it is not connection oriented and thus there is no handshaking, acknowledgement, or retransmission. As a result, it is faster in transmission than TCP. Applications using UDP are streaming media and Internet telephony, where old data does not need to be retransmitted in the event of failure. UDP does not exhibit RTT fluctuation for the high frequency transmission of packets.

4.6.1 End to end delay of packet using UDP

Bolot performed an experiment of sending UDP probe packets with a different time intervals to measure the structure of the Internet load over different time scales[16]. The time scale ranges from a few milliseconds to several minutes. He found that the packet size has a close relation with delay and packet loss[17].

[16] Analysis of latency and packet loss is important in designing a router, flow control algorithms or the size of a playback buffer.

[17] Packet losses are random unless the packet traffic uses large fraction of a bandwidth (Bolot 1993).

5

Causality and Sensitivity Analysis using Regression Method

5.1 Regression Analysis

Regression analysis is a methodology to determine the dependency of one variable, the dependent variable on other explanatory variables.

5.1.1 Linear regression analysis

Linear regression analysis describes the variation of response variable (dependent variable) using a linear combination of predictor effects (explanatory variables) and a random error term (disturbance). In these models, a causal relationship is implicitly assumed between the explanatory and dependent variables (Damodar 1995).

5.1.2 Linear population regression

Population regression estimates the conditional mean or expectation of a dependent variable for a given value of explanatory variables. The type of function used to approximate the mean determines the type of regression, namely, linear or nonlinear.

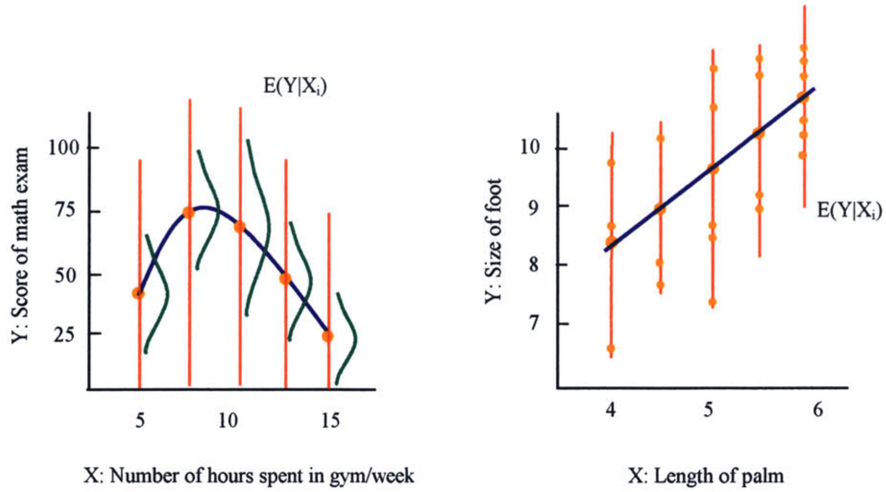


Figure 5-1 Regression Analysis

The functional form of linear population regression is

$$E(Y \mid X_i) = \beta_1 + \beta_2 X_i$$

Equation 5-1

Here, the β_1 and β_2 are regression coefficients, intercept and slope coefficient respectively. In approximating the individual value, Y_i given X, stochastic disturbance term is involved.

$$Y_i = E(Y \mid X_i) + u_i$$

Equation 5-2

The stochastic disturbance or error term u_i explains unobservable effects that influence the dependent variable but are not included in explanatory variables.

5.1.3 Linear sample regression

Sampled values of the dependent variable for a fixed value of explanatory variables is used to estimate the population regression model. The functional form of linear sample regression is in Equation 5-3.

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Equation 5-3

Here, Y_i, $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimator for $E(Y|X_i)$, β_1 and β_2 respectively. If an estimate of u_i is introduced into the equation, the resulting equation becomes

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

Equation 5-4

5.1.4 Linear in variable and linear in parameter

When dependent variable Y is a function of X, the regression is linear in variable. Another meaning of linearity in regression analysis is linear in parameter. For example, the following regression equation is linear in parameter but nonlinear in variable. (Damodar 1995)

$$E(Y \mid X_i) = \beta_1 + \beta_2 X_i^2$$

Equation 5-5

5.1.5 Types of data

In regression analysis, there are three different types of data.

1. Time series data: Time series is a data that is collected by regular sampling over a period of time.
2. Cross-section data: When a data is collected at one fixed point in time, the data is called cross-section data.
3. Pooled data: Both the characteristics of time series and cross-section data exist in pooled data. Such as the average scores of the test in three different classes last five years. A longitudinal or panel data is a special case of pooled data, which is a sampling the cross-sectional data of a same object at periodic interval (Damodar 1995). The set of data used in this work is classified into this category.

5.2 Estimation

Due to limitations in data availability, a population regression function (PRF) is estimated from sample regression function (SRF). The most common method of estimation is ordinary least squares (OLS).

5.2.1 Determining SRF

Based on the previously given definitions, the residual term in SRF is determined in Equation 5-6.

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$$

Equation 5-6

Here, e_i is the difference between the actual and estimated Y values. To make the SRF as close as possible to the actual Y value, the sum of errors, $\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$, needs to be minimized. That is least square criterion. Canceling effect of positive and negative errors can be eliminated by using least square criterion. $\sum e_i^2$ is a function of $\hat{\beta}_1$ and $\hat{\beta}_2$, which could be determined by

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}$$

Equation 5-7

And

$$\hat{\beta}_2 = \frac{\sum X_i y_i}{\sum X_i^2 - N\overline{X}^2}$$

Equation 5-8

In the equations above, $\overline{X}$ and $\overline{Y}$ are sample means of X and Y respectively and N is sample size. However, there are several assumptions underlying least square method.

The purpose of regression analysis is not only to get the estimated values, $\hat{\beta}_i$ and $\hat{\beta}_2$, but also to make an inference on a true value, β_1 and β_2. For this reason, the following assumptions are made on the generation of X_i and u_i. These assumptions underlie Gaussian, standard, classical regression models and could be changed. However, the method of fitting with SRF must also be changed.

1. $E(u_i \mid X_i) = 0$, For a given X_i, the expectation of error is zero. Basically this assumption means that the unobservable variables, of which the effects are embedded in u_i, do not affect the mean value of Y.

2. There is no autocorrelation between u's, namely $cov(u_i, u_j) = E[u_i - E(u_i)][u_j - E(u_j)] = E(u_i u_j) = 0$ when i and j are different.

5.2.2 Significance of regression coefficients

Under the assumption of normality of disturbances, the OLS estimators of β_1 and β_2 are normally distributed and the significance of β_2 can be estimated as follows.

$$\Pr[\beta_2^* - t_{\frac{\alpha}{2}} se(\hat{\beta}_2) \le \hat{\beta}_2 \le \beta_2^* + t_{\frac{\alpha}{2}} se(\hat{\beta}_2)] = 1 - \alpha$$

Equation 5-9

$$se(\hat{\beta}_2) = \frac{\hat{\sigma}}{\sqrt{\sum x_i^2}}$$

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}}$$

Equation 5-10

Here, α is the level of significance, which defines the region of acceptance. $se(\hat{\beta}_2)$ is an estimated standard error of estimator of $\hat{\beta}_2$, n is a sample size and $t_{\frac{\alpha}{2}}$ is the value of t-distribution with $\frac{\alpha}{2}$ level of significance.

5.2.3 Properties of SRF

1. It passes through sample means of Y and X.
2. $\bar{\hat{Y}} = \bar{Y}$ (From the fact that $\sum (X_i - \bar{X}) = 0$
3. The mean value of e_i is zero.
4. The error term e_i are uncorrelated with the predicted Y_i.
5. The error term e_i are uncorrelated with X_i.

5.2.4 Time lagged model

In regression analysis using time series data, the explanatory variable sometimes has lagged effects on the dependent variables. There are several different ways to model lagged effects, such as the distributed-lag model or dynamic model. The common idea underlying both methods is to distribute the lagged effects over the time series of dependent variable.

In this work, there are unknown delays associated with the propagation of values through the Internet. However, the current effect of one explanatory variable is not distributed over the upcoming time series of the dependent variable. Rather, there exists one specific corresponding value of the dependent variable for an explanatory variable. In such situation, the lagged effect is better modeled by estimating the value of delay rather than by distributing effects over time.

5.3 Causality

Regression analysis is based on the assumption that the dependent variable and the explanatory variables are causally related. The causality can be either a direct causation or a mere correlation through intermediate variables. Thus regression analysis by itself cannot be used as a tool to distinguish the causation from mere correlation[18].

However, the case of correlation also involves causal relations: the intermediate variables work as a common cause for the explanatory variable and the dependent variable. There can be no correlation without causation (Pearl 2000). Judea

5.3.1 Regression VS. Correlation and Causation

In 1961, Kendall and Stuart stated, "A statistical relationship however strong and however suggestive can never establish causal connection: our ideas of causation must come from outside statistics, ultimately from some theory or other." (Kendall and Stuart 1961). In other words, theoretical proof is required to assert a causal relationship.

Correlation analysis and regression analysis are based on very different concepts. The main focus of correlation analysis is to determine the strength of a relationship (correlation coefficient) while, the purpose of regression analysis is to predict or estimate the average value of a dependent variable for a fixed value of explanatory variables. In regression analysis, the explanatory variables are rendered deterministic but the dependent variables are assumed to be stochastic. Even when the explanatory variables are stochastic, they must be assumed nonrandom in repeated sampling. In correlation analysis, both variables are treated as random. (Damodar 1995)

[18] The distinction between direct causation and mere correlation was explained in detail in chapter3.

5.3.2 Granger causality

Granger suggested a theoretical way to verify the causality between two variables. Granger's definition of causality follows, where $\overline{U-X}$ represents set of past values of $U-X$, and U_t is all the information in the universe accumulated since *t-1*.

"*If* $\sigma^2(Y\,|\,\overline{U}) < \sigma^2(Y\,|\,\overline{U-X})$, *we say that X is causing Y, denoted by* $X_t \rightarrow Y_t$. *We say that* X_t *is causing* Y_t *if we are better able to predict* Y_t *using all available information than if the information apart from* X_t *had been used.*"

The underlying assumption of Granger's causality is that the causality is stationary (independent of time). In the case of causality in the distributed design framework, the stationary assumption need not be satisfied because the users need only a snapshot of causal information at the specified moment. However, one of the elements of the equation, U_t, is not realistic in that it needs a 'universe' of information to be collected,. When black box processes are involved, it will be impossible to observe the 'universe' of information.

Spurious causality

As discussed in the previous section, the 'universe' of variables is impossible to obtain. So a relevant subset of variables needs to be selected.

Let $X_t^D = \{X_t^i, i \in D\}$ denote a set of time series, where *D* is a set of integers. Denote *D(j)* as a set $\{i \in D, i \neq j\}$ and $X_t^{D(j)}$ as a set $\{X_t^i, i \in D(j)\}$, i.e., the full set of relevant time series except one, that is, X_t^j. The definition of Granger causality can be stated with new notation by replacing U_t with X_t and U_t-X_t by $X^{D(j)}$. For example, suppose that a relevant vector set (time series) consists only of two elements X_t and Y_t and the rest of the universe is assumed to be irrelevant, that is, *D* consists of X_t and Y_t. By definition, $\sigma^2(X\,|\,\overline{R})$ is a minimum predictive error variance of X_t using past values of R_t. Using this definition, if $\sigma^2(Y\,|\,\overline{Y}) > \sigma^2(Y\,|\,\overline{X},\overline{Y})$, then X_t can be regarded as causing Y_t. However, if there exist causes that are not included in *D*, then spurious causality will arise. For instance, suppose there is a third variable Z_t, which affects both *X* and *Y*, and there is only correlation between X and Y. Under this circumstance, if we define relevant set of time series D as $\{X_t, Y_t\}$, then the causality relevant to D will be spurious (Granger 1969).

Two variable model

Suppose that X_t and Y_t are time series, which are stationary and have zero means. The simple causal model is

$$X_t = \sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t,$$

Equation 5-11

$$Y_t = \sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + \eta_t, \quad t = m+1, \ldots, T$$

Equation 5-12

In Equation 5-12, ε_t and η_t are independent white noise series. If b_j is not zero, then Y_t is causing X_t, and if c_j is not zero then X_t is causing Y_t, by definition.

Various test methods for Granger causality

A null hypothesis[19] is tested using following causality tests. If a statistic, such as GS or GW, exceeds the critical F value at the selected level of significance the hypothesis is rejected (Damodar 1995), which means that there exists causality between the variables.

.1 Granger-Sargent test

Granger-Sargent test has substantial power (precision) and little upward bias in alpha, which is likeliness for the type-I-error[20]. It also has an advantage of being simple to implement.

$$GS = \frac{(PRSS - URSS)/m}{URSS/(T - 2m)}$$

Equation 5-13

The meanings of individual terms in the Equation 5-13 will be explained in the following section.

---

[19] Hypothesis that there is no causality between variables

[20] Type I error: An error of rejecting a true hypothesis; Type II error: An error of accepting a false hypothesis

.2 Granger-Wald test

In terms of overall precision, the Granger-Wald test shows the best performance. One of the great advantages that the test can offer is computational efficiency. Granger-Wald test is attractive in selecting a best subset of potential predictors[21](J.R. Bult, et al., 1997). The test loses the fewest degrees of freedom by the formation of the leads and lags, and also does not require any correction for serial correlation. However, it shows the greatest upward bias in alpha.

$$GW = T \cdot \frac{\hat{\sigma}_{\eta_t^*}^{\,2} - \hat{\sigma}_{\eta_t}^{\,2}}{\hat{\sigma}_{\eta_t}^{\,2}}$$

Equation 5-14

These two Granger tests are based on the model represented by **Equation 5-12**. For both of the tests, the distributions of ε_t and η_t are assumed to be NID(0,σ^2). In **Equation 5-12**, if X_t does not cause Y_t, then $c_j=0$ for all j or

$$Y_t = \sum_{j=1}^{m} d_j Y_{t-j} + \eta_t^*$$

Equation 5-15

where η_t^* a random disturbance term.

In Granger-Sargent test, PRSS is the residual sum of squares of **Equation 5-15**, and URSS is that of **Equation 5-12**.

The Granger-Wald test has an asymptotic χ_m^2 distribution under null hypothesis, which is asymptotically equivalent to the F-test [J.R. Bult, et al., 1997]. In **Equation 5-14**, $\hat{\sigma}_{\eta_t^*}^{\,2}$ is the estimate of the $var(\eta_t^*)$ and $\hat{\sigma}_{\eta_t}^{\,2}$ is the estimate of the $var(\eta_t)$.

[21] Reducing a set of potential predictors is one of the greatest purposes of causality test.

.3 Sims test

$$SI = \frac{(PRSS - URSS)/M}{URSS/(T - M - N - 1)}$$

M: maximum number of future value of dependent variable

N: maximum value of past value of dependent variable

Equation 5-16

In this method, the dependent variable is regressed on current, past and future values of explanatory variables. Under the hypothesis that there is no causality, the regression parameter for the future value of dependent variable is equal to zero.

.4 Modified Sims test

When the disturbances are serially correlated, the Sims test does not show the asymptotic distribution that it originally acclaims to exhibit if the null hypothesis is true. To overcome this problem, a Modified Sims (MS) test has been developed. In MS, lagged values of dependent and explanatory variables are included and the disturbance terms may be regarded as serially uncorrelated (J.R. Bult, et al., 1997)

$$MS = \frac{(PRSS - URSS)/M}{URSS/(T - 2P - M - N - 1)}$$

Equation 5-17

P is introduced into the equation to include the lagged values of dependent and explanatory variables.

.5 Haugh-Pierce (HP) test

The basic concept of Haugh-Pierce test is to determine the cross-correlation between the residuals of the explanatory and dependent variables. The residual of each variable comes from the univariate model fitting to each variable, and there is a causal relation between residuals as there is with the variables. Therefore causality can be detected by estimating the regression coefficient of residual corresponding to explanatory variable on past, current and future values of the residual corresponding to the dependent variable. Regression between the residuals is similar to that between variables except that the residuals are estimated using Box-Jenkins techniques. The statistics of HP are as follows.

$$HP = T\sum_{k=1}^{m} r_{uv}^{2}(k)$$

Equation 5-18

$$r_{uv}(k) = corr(\hat{u}_{t-k}, \hat{v}_t)\text{, } (K=1,..,m)$$

Equation 5-19

In Equation 5-19, m determines the degree of freedom of HP in Equation 5-18.

Performance comparison of the causality tests

.1 Definition of performance

Performance of a causality test is defined by its power and upward bias in alpha. Power is a probability of concluding causality, given that the causality exists, and alpha is a probability of concluding causality, given that causality does not exist. In addition to the absolute value of power and alpha, their sensitivity to influencing factors is an important issue. Factors that affect the performance include a sample size, disturbances, strength of a correlation, data generating process (autoregressive or white noise) and existence of a lagged variable [J.R. Bult, et al., 1997]. However, the existence of a lagged variable is not applicable in this work.

.2 Performance comparison

.2.1 Power

Granger-Wald test shows highest power and smallest sensitivity of power to the sample size. Granger-Sargent test and Sims test have substantial power, but Sims test is highly sensitive to a sample size. Modified Sims test and Haugh-Pierce test are worst in power but least sensitivity to a sample size.

.2.2 Alpha

Granger-Wald test has a greatest upward bias in alpha, and also highly sensitivity to the sample size. Modified Sims test and Haugh-Pierce test show a little bias in alpha and also least sensitive to a sample size.

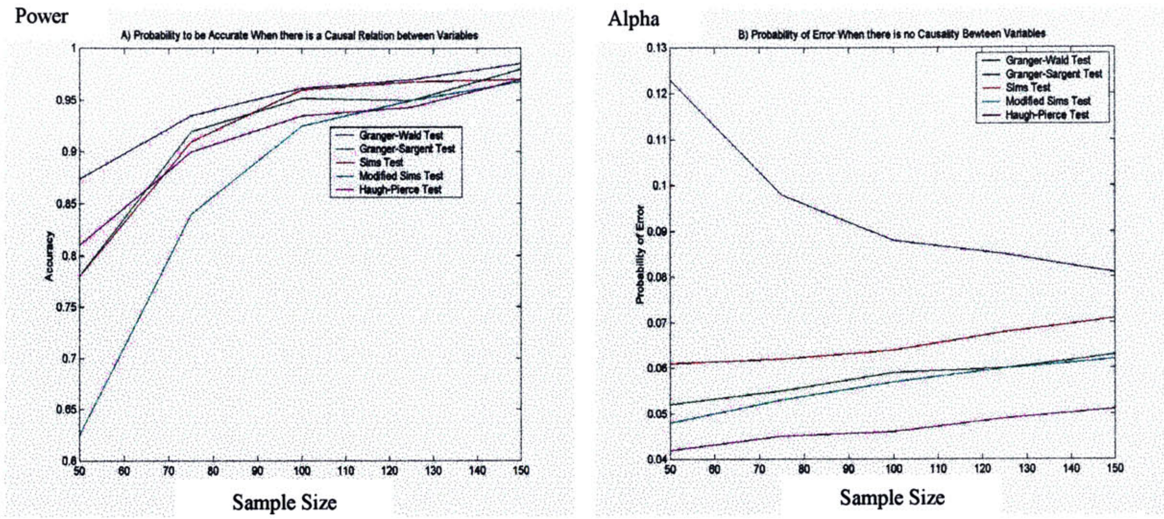


Figure 5-2 Influence of Sample size to Power and Alpha [J.R. Bult, et al., 1997]

In this work, Granger-Wald test is used, which is the most conservative way to select a good subset of input (explanatory or predictor) variables, in that it has highest power and largest upward bias in alpha.

Granger causality for the circular dependency (feedback)

Suppose there are two variables, X and Y. If X causes Y and Y causes X, there is a circular dependency or feedback between X and Y[22]. Applying the causality test for two variables with different direction can recognize the feedback.

[22] Details about detection and resolving the circular dependency problem are discussed in chapter 3.

5.4 Sensitivity

5.4.1 Matching

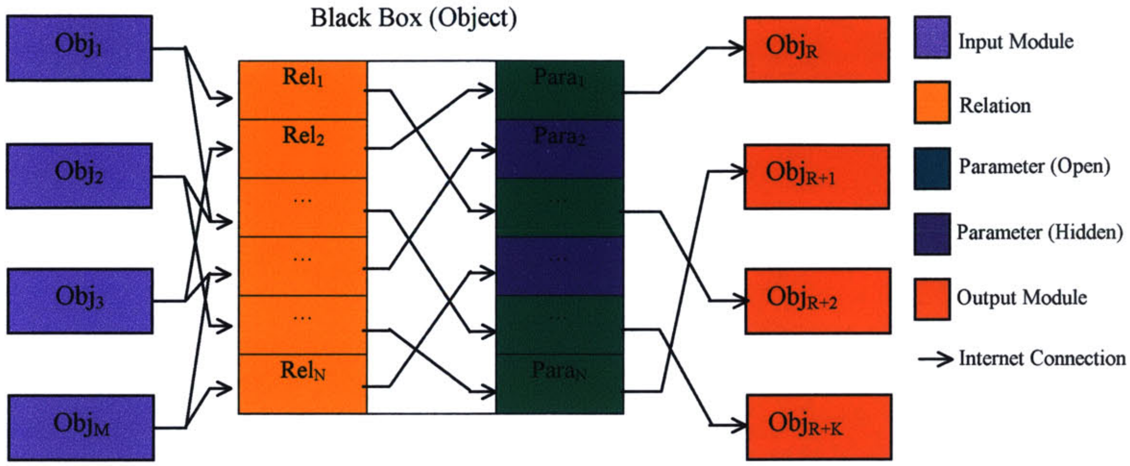


Figure 5-3 Black Box

Black box and varying latency

In the DOME framework, the relation plays a role of interface between the external variables (inputs) and internal variables (outputs). Each design object can contain multiple relations, and each of these relations is a MISO system (multiple input single output). This, each object can be regarded as MIMO system (multiple input multiple output) in terms of number of inputs and outputs.

In regression analysis the input must be matched with corresponding outputs. In the DOME framework, the data is transmitted through Internet, where the latency is not constant. In addition to this, existence of proprietary information makes it difficult to directly identify the effects from each input (cause) that compose the output behavior (effect).

Matching

In this work, regression analysis is used as a tool to find the relation between inputs and outpust in the presence of black boxes and varying latency. Considering the huge effect of data quality on the accuracy of analysis, matching is a very important issue.

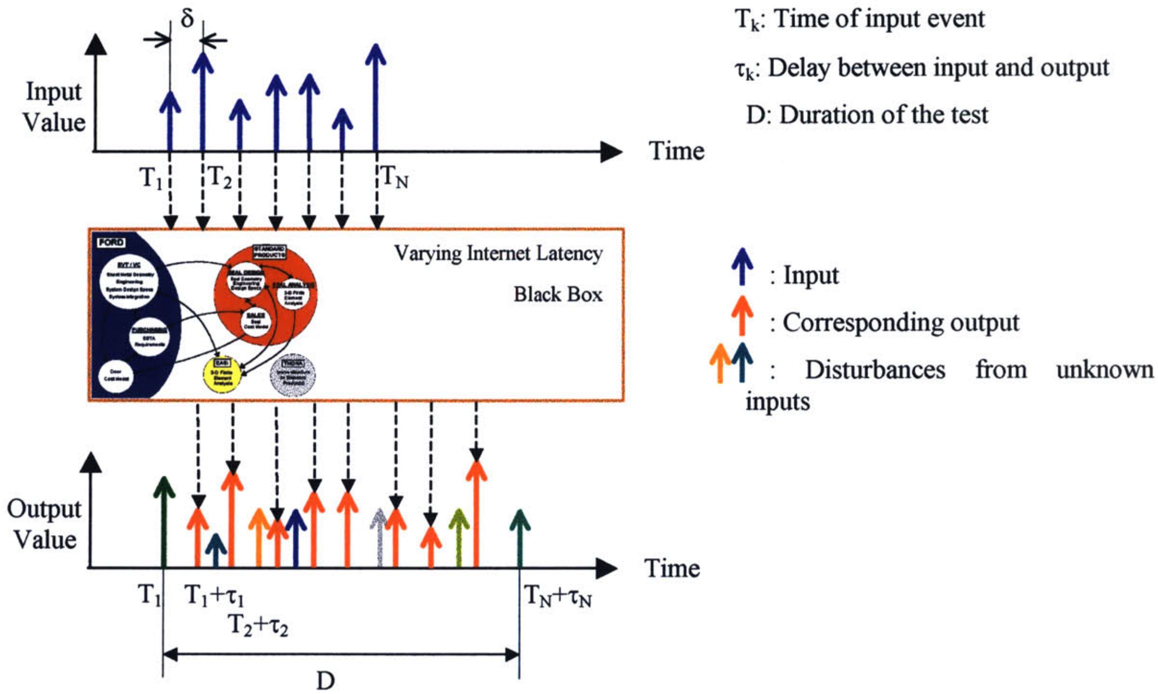


Figure 5-4 Matching

5.5 Periodic Inputs

The latency of Internet depends on both transport layer protocol and the network traffic condition. In addition, the shorter the length of interval between packet transmissions, the more likely the packets are sent through the same route in the network. If periodic inputs are generated with very small time scale compared with that of network traffic condition and transmitted through UDP, the variance of the latency can be expected to be minimum. The time scale of network traffic condition is relatively large compared with that of interval (period) between the transmissions of input values. According to the survey by Bolot, the time-scale of network traffic is several minutes. The frequency information will be used as a clue for matching in regression analysis. However, in the chapter for the perturbation method, the frequency information will have a different application.

5.5.1 Matching by interval

This method chooses the series of output values of which the arrival times shows the minimum deviation from the expected arrival time, that is $T_1+k\delta+\tau_1$, $(k=0,1,..,N)$. This method is implemented through the following steps.

1. Pick an arbitrary value at $T_1+\tau_1$ as starting point and build an output value time-series where the kth element is the one that arrives closest to $k\delta$ from the chosen starting point. See Figure 5-4.
2. Repeat this process with different starting points, that is, with different τ_1, until $T_1+N\delta+\tau_1$ becomes larger than D (duration of the test). The duration of the test is established by surveying the mean and standard deviation of the latency.
3. Select the output time series that shows a minimum deviation from $T_1+k\delta+\tau_1$.

5.5.2 Matching by value-cross-correlation

The focus of this method is to find the most probable lag between cause and effect by examining the cross-correlation between them. The amount of initial delay is determined by the cross-correlation of outputs and inputs. Granger's definition of causality lag is below. The original purpose of his research on causality and causality lag is to determine the relation between economic variables, where the speed of information propagation is unknown.

"If $X_t \rightarrow Y_t$, we define the (integer) causality lag m to be the least value of k such that $\sigma^2(Y|U\text{-}Y(k)) < \sigma^2(Y|U\text{-}Y(k+1))$. Thus, knowing the value Y_{t-j}, $j=0,1,..,m-1$, will be of no help in improving the prediction of X_t."

This definition is based on the assumption that the time-series is stationary. In this work, the analysis is performed after the complete set of output set is acquired that results from periodic inputs, so the time-series of outputs can be regarded as stationary.

5.5.3 Sample selection: statistic significance of sliding window method

Samples are automatically selected, once an initial delay, τ_1, is determined, because the overall time span is predetermined by the period of perturbations[23]. In selecting the initial delay, time window is used as illustrated in Figure 5-5. The statistic significance of the distance swept by the window is determined by the statistical distributiob of Internet latency (as illustrated in Figure 5-6) and the significance of the selected sample depends on the cross-correlation between inputs and outputs.

[23] Based on an assumption that the delay experienced by each packet follows a normal distribution, the overall span is most likely to be a multiplication of the number of perturbation and the interval between them.

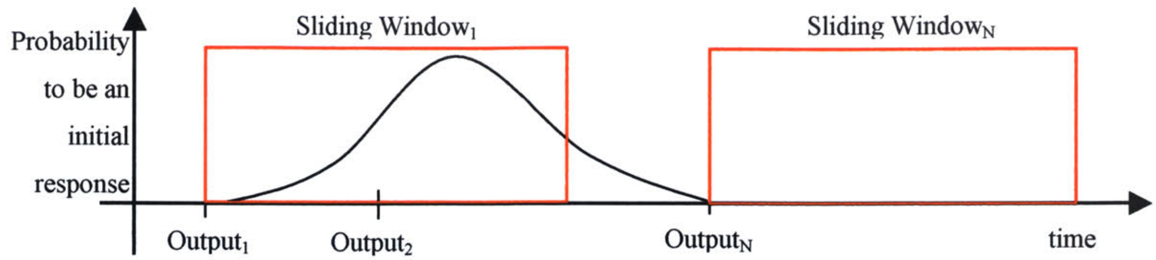


Figure 5-5 Significance of Sliding Window

5.5.4 Conditions for period selection

Currently, the minimum interval between perturbations is 10ms, which corresponds to a period of 20ms. This limitation comes from the capability of software[24] used for the experiments. Apart from software limitations, the interval between perturbations should be greater than the largest operation time of design objects. Considering the fact that most of design constraints are algebraic, the order-of-magnitude of operation time is negligible compared with the minimum interval between perturbations. If an operation time is greater than 10ms[25], the interval between perturbations should be larger than that operation time.

With respect to the maximum interval between perturbations, the total amount of time taken by the test should be smaller than the time scale of Internet traffic volume change. Matrix.net has been surveying the Internet latency every 15 minutes, and following is an example of their on-line report.

---

[24] Java was used to build the software for the experiments.

[25] Operation times of some applications, such as Cad systems, can be much greater than 10ms.

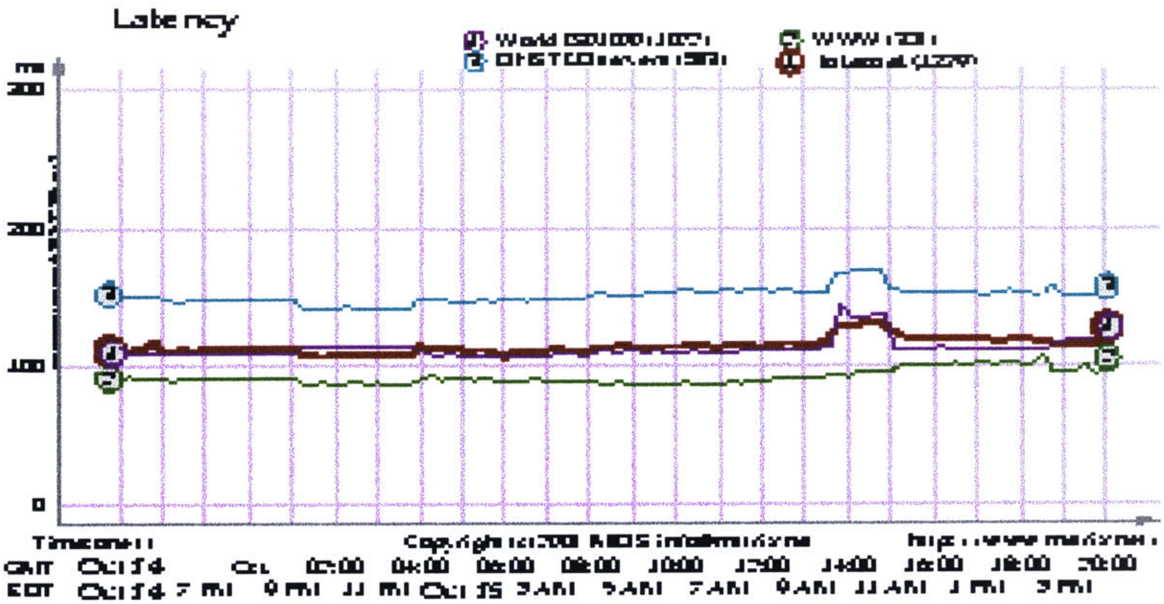


Figure 5-6 Internet Latency Statistic [http://average.matrix.net/]

As illustrated in Figure 5-6, latency does not change more than 30ms in 15 minutes. Thus, in this work, the maximum amount of time taken for the test is limited by 15 minutes.

5.6 Implementation

In this section, the regression analysis method is implemented and tested in the network centric framework. The result is presented with precision and robustness analysis.

5.6.1 Algorithm

In this section, the algorithm for causal analysis based on regression method is presented. The implementation of the algorithm is based on the node structure suggested in the previous chapter. The node structure was designed to obtain causal information without taking significant amount of time and resources originally allocated to the actual design process.

1. Observing the target variable behavior

First, the target object (variable) is selected and the observer (client) is attached to the variable. The target variable is a variable that is to be tested with its relation with an input. The duration of the observer is determined using latency statistics that record the history of value and arrival time of the target variable.

The observer is a Java client that listens to the value changes in the target variable through a UDP channel. In this work, it is suggested that the communication between design objects be made through two different kinds of channel depending on its purpose. TCP channel is used in actual design process where reliability is important. UDP is used as a separate channel for causality and sensitivity tests in which timing is more important than reliability. Communication through UDP channel maximizes the use of time by utilizing the idle time between design actions through TCP.

2. Generating the periodic input perturbation

Periodic perturbations are made to the input, and propagated through the UDP channel to the all related variables[26]. By setting the variation of the input to a sufficiently small, linear regression methods can produce a good approximation for the partial derivative of output with respect to the input.

3. Matching

The observer is attached to the target variable before the perturbations start to propagate. For the multiple perturbation propagation, the UDP channel in each object can multicast the value change. As a result, the behavior of the target variable is combination of effects from multiple inputs and it is likely that the stored history of target variable values outnumbers those of the input. Thus, matching (mapping) between input values and target variable values is necessary, as is explained in section 5.4.1.

4. Regression Analysis

Based on the matched history of input and output values, the regression parameter is estimated according to Equation 5-20.

$$\beta_2 = \frac{Cov(x, y)}{Var(x)}$$

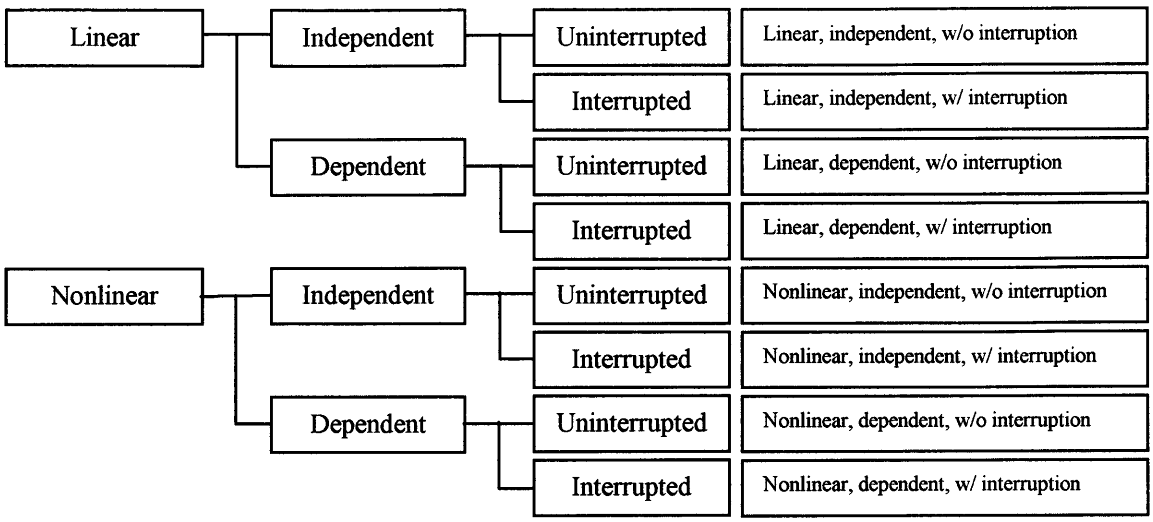
Equation 5-20

where x is input, y is target and $Cov(x, y) = E[(x - \mu_x)(y - \mu_y)]$.

5.7 Experimental Results

Experiments were performed for 8 different cases.

[26] The reason for giving periodicity to the perturbation was explained in the previous section.



Sensitivity is (non)linear when the actual relations existing in the black box are (non)linear.

(In)Dependent sensitivity means that the sensitivity of a specific input is (independent of) dependent on the value of another input variable that is unknown.

In addition to the linearity and dependency, which are determined by original relation, interruption also affects the precision of the test. The sources of interruption can be either causality test from other input or an value change in the actual design process.

5.7.1 Experiment-1: Linear and independent sensitivity

In the following experiments, the input, output and unknown input are Y, Z and X respectively.

1. Relation: $Z = 3X + 4Y$
2. Initial value: $X=2$, $Y=3$
3. Number of perturbations: *6*
4. Amplitude of perturbation: *0.05* for both
5. Interval between value generation: *X: 40ms, Y: 100ms*

Table 5-1 Linear Independent Case

Value (Z)	Arrival Time (ms)	Interval (ms)	Match 1	Match 2
18.4000	0	0	1 (18.4)	
18.7000	10	10		1
18.4000	50	20		
18.7000	90	30		
18.3000	100	20	2 (18.0)	
18.0000	120	30		2
18.3000	161	51		
18.7000	201	30	3 (18.4)	
18.4000	211	20		3
18.0000	301	80	4 (18.0)	4
18.4000	391	90	5 (18.4)	5
18.0000	501	110	6 (18.0)	6

• Values in parentheses are results without interruptions.

Within the time window there are two different plausible matches. As is shown in Table 5-1, Match 1 starts from 0ms and Match 2 starts from 20ms. The variances of Match 1 and Match 2 from the original interval (100ms) are 13.833ms^2 and 104ms^2 respectively. So Match 1 is chosen and the following is the result of regression analysis with the chosen time series.

Causality

$$\hat{\sigma}^2_{\eta_t} = 0.0482$$

$$\hat{\sigma}^2_{\eta_t^*} = 0.124$$

$$GW = 9.44$$

The statistic GW shows a χ^2 distribution with one degree of freedom. GW value of *9.44* corresponds to the probability of 0.9996 for the existence of a causal relationship.

Sensitivity

Table 5-2 Result-Linear Independent Case

Actual Sensitivity	Estimated Sensitivity	Percentage of Error
4	4.0832 (4.000)	2.08%

The value in a parenthesis is a result from 0th order interpolation.

Significance of regression coefficient

$se(sensitivity) = 1.10331$, $\alpha = 0.2$ and $t_{\frac{\alpha}{2}} = 1.372$. Thus, with 80% of significance, $sensitivity$ is between 2.486 and 5.513.

5.7.2 Experiment-2: Second Order independent case

1. Relation: $Z = 3X^2 + 4Y^2$
2. Initial value: *X=2, Y=3*
3. Number of perturbations: *6*
4. Amplitude of perturbation: *0.0005* for both
5. Interval between value generation: *X: 100ms, Y: 170ms.*

Table 5-3 Second Order Independent Case

Value (*Y*)	Arrival Time (ms)	Interval (ms)	Match 1	Match 2
48.0120	0	0	1	
48.0360	30	30		1 (48.024)
48.0240	100	20		
48.0360	190	30	2	2 (48.000)
48.0120	220	20		
48.0000	290	30		
48.0240	370	51	3	3 (48.024)
48.0360	390	30		

48.0240	501	20	4	
48.0000	531	80		4 (48.000)
48.0240	711	90	5	5 (48.024)
48.0000	881	110	6	6 (48.000)

Within the time window, there are two different plausible matches. As is shown in the table, Match 1 starts from 0ms and Match 2 starts from 30ms. The variances of Match 1 and Match 2 from the original interval (170ms) are 550.46ms and 30.5 ms respectively. So Match 2 is chosen and the following is the result of regression analysis with the chosen time series.

Causality

$\hat{\sigma}^2_{\eta_t} = 0.0003$

$\hat{\sigma}^2_{\eta_t^*} = 0.0004$

$GW = 2.000$

GW value of *2.00* corresponds to the probability of 0.8413 for the existence of a causal relationship.

Sensitivity

Table 5-4 Result-Second Order Independent Case

Actual Sensitivity	Estimated Sensitivity	Percentage of Error
24	18.4030 (16.0027)	23.32%

Significance of regression coefficient

$se(sensivity) = 6.26$, $\alpha = 0.2$ and $t_{\frac{\alpha}{2}} = 1.372$. Thus, with 80% of significance, $sensitivity$ is between 9.41 and 26.59.

5.7.3 Experiment-3: Exponential independent case

1. Relation: $Z = 4e^x + 7e^y$
2. Initial value: *X=2.5, Y=3.5*
3. Number of perturbations: *6*
4. Amplitude of perturbation: *0.0005* for both

5. Interval between value generation: *X: 150ms, Y: 200ms*

Table 5-5 Exponential Independent case

Value (Y)	Arrival Time (ms)	Interval (ms)	Match 1	Match 2
280.7701	0	0	1 (280.7701)	
280.8188	30	30		1
280.7701	190	160		
280.5381	200	10	2 (280.5381)	2
280.5869	340	140		
280.8188	401	61	3 (280.7701)	3
280.7701	481	80		
280.5381	591	110	4 (280.5381)	
280.5869	641	50		4
280.5381	791	150		
280.7701	801	10	5 (280.7701)	5
280.5381	1000	199	6 (280.5381)	6

Match 1 starts from 0ms and Match 2 starts from 30ms. The variances of Match 1 and Match 2 from the original interval (200ms) are 13.833ms and 600.5 ms respectively. So Match 1 is chosen and the following is the result of regression analysis with the chosen time series.

Causality

$$\hat{\sigma}^2_{\eta_t} = 0.00073$$

$$\hat{\sigma}^2_{\eta_t^*} = 0.00145$$

$$GW = 5.92$$

GW value of *5.92* corresponds to the probability of 0.9993 for the existence of a causal relationship.

Sensitivity

Table 5-6 Result-Exponential Independent Case

Actual Sensitivity	Estimated Sensitivity	Percentage of Error
231.8081	231.9241 (248.1756)	0.05 %

Significance of regression coefficient

$se(sensitivity) = 1.10331$, $\alpha = 0.2$ and $t_{\frac{\alpha}{2}} = 1.372$. Thus, with 80% of significance, $sensitivity$ is between 211.4 and 250.96.

5.7.4 Experiment-4: Nonlinear dependent case

1. Relation: $Z = XY$
2. Initial value: *X=2.5, Y=3.0*
3. Number of perturbations*: 6*
4. Amplitude of perturbation: *0.0005* for both
5. Interval between value generation: *X: 100ms, Y: 150ms.*

Table 5-7 Nonlinear Dependent Case

Value (*Y*)	Arrival Time (ms)	Interval (ms)	Match 1
7.502499	0	0	1 (7.502499)
7.5055	80	80	
7.503	150	70	2 (7.50)
7.50	180	30	
7.503	280	100	
7.5055	300	20	3 (7.502499)

7.5024999	380	80	
7.50	450	70	4 (7.50)
7.503	480	30	
7.50	580	100	
7.5024999	601	21	5 (7.502499)
7.50	751	150	6 (7.50)

In the nonlinear dependent case, the matching starting from 80ms is implausible and the variance of the Match 1 is 0.33ms. Following is the result of the regression analysis based on the given matching.

Causality

$$\hat{\sigma}^2_{\eta_t} = 0.000005299$$

$$\hat{\sigma}^2_{\eta_t^*} = 0.00000985$$

$$GW = 5.193$$

GW value of *5.193* corresponds to the probability of 0.9868 for the existence of a causal relationship.

Sensitivity

Table 5-8 Result-Nonlinear Dependent Case

Actual Sensitivity	Estimated Sensitivity	Percentage of Error
2.5	2.415 (2.5003)	3.4%

5.8 Result Analysis

5.8.1 Robustness to non-linearity

From the results of Experiment-2 and Experiment-3, it can be roughly observed that the non-linearity of the actual relation weakly affects the precision of the test. An error caused by non-linearity arises from the difference between an actual partial derivative and its linear approximation, which can be reduced by minimizing the perturbation. Table 5-9 illustrates a percentage of errors caused only by non-linearity.

Table 5-9 Error Caused by Non-Linearity

	Actual Sensitivity	Estimated Sensitivity without Latency Variation and Interruption	Percentage of Error*100
2^{nd} Order	24	24.004	1.66
Exponential	231.8081	231.9235	4.978

5.8.2 Robustness to interruption

Every experiment was completed with and without interruption from an unknown input. Table 5-10 is the result of each experiment without interruption.

Table 5-10 Error Caused by Interruption

	Actual Sensitivity	Estimated Sensitivity without Latency Variation and Interruption	Percentage of Error	Percentage of Error Increased by Interruption
Linear Independent	4	4	0%	0.25% (3.99)
Second Order Independent	24	24.004	0.0166%	4.15%(25.0)%
Exponential Independent	231.8081	231.9236	0.0497%	0.00034% (231.9241)
Nonlinear Dependent	2.5	2.5	0%	3.4% (2.415)

Interruption always degrades the precision of the analysis. To quantify the influence of the interruption on a precision of the analysis, an index DI (Degree of Interruption) in Equation 5-21 is defined.

$$DI = \frac{\Delta\, Output\ by\ Interruption}{\Delta\, Output\ by\ Known\ Input} * 100$$

Equation 5-21

Table 5-11 DI and Precision

	Degree of Interruption	Percentage of Error Caused by Interruption
Linear Independent	75	0.25%
Second Order Independent	50	4.15%
Exponential Independent	21.03	0.00034%
Nonlinear Dependent	120	3.4%

Degradation by interruption

Figure 5-7 shows that the degree of degradation from interruption is affected by the amplitude of the interruption value, δ, at the corresponding moment.

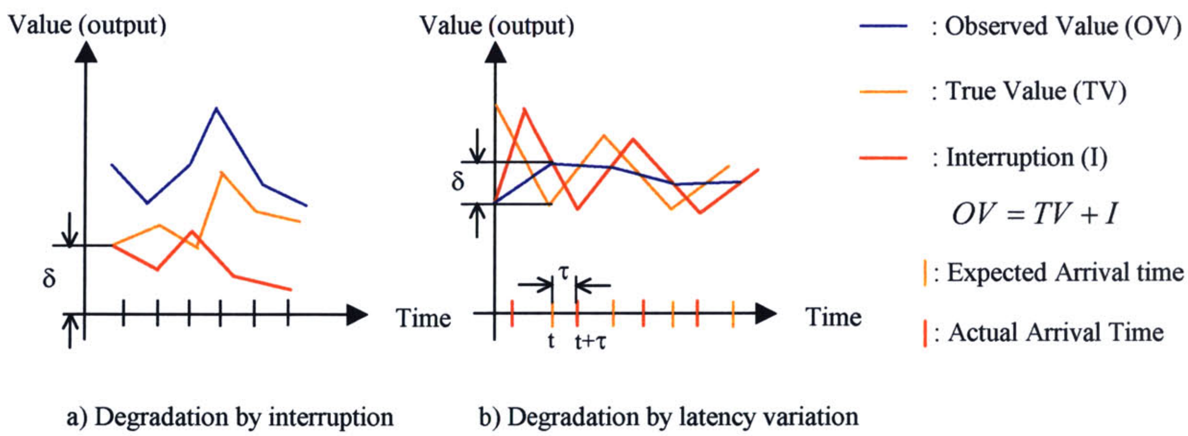


Figure 5-7 Degradation by interruption and latency variation

5.8.3 Robustness to arrival time variance

As illustrated in Table 5-12, the effect of arrival time variation on an error shows an abrupt change at a particular value. The reason for this can be either a coupled effect with other variables or the existence of a threshold. Thus, a multiple regression is performed between these three factors and the results are presented in Equation 5-22.

Table 5-12 Error by Arrival Time Variation

	Standard Deviation of Arrival Time	Percentage of Error caused by Arrival Time Variation
Linear Independent	3.74	0% (4.0)
Second Order Independent	8.97	33.32% (16.0026)
Exponential Independent	3.72	0% (231.9235)

Degradation by latency variation

The degradation by latency variation depends on the magnitude of τ in Figure 5-7. The degree of error depends on δ, which is a difference between values at t and $t+\tau$

5.8.4 Independent and Dependent Sensitivity

A dependent sensitivity[27] degrades the precision by deviating an observed value from a true value. Thus, a dependent sensitivity can be regarded as a special case of an interruption. This can be verified by applying the results of Experiment-4 to Equation 5-22.

5.8.5 Summary

Factors such as non-linearity, arrival time variation and interruption degrade the precision of the analysis. Section 5.8.1 through 5.8.3 present individual influences of these factors on the precision the analysis. Non-linearity alone shows minimum degree of degradation.

The influences of interruption and arrival time variation are stochastic: they can be either positive or negative. Thus, errors can compensate each other.

$$Percentage\ of\ Error = 0.000441 \times (DI)^2 + 0.0859 \times (AV) \times (DI) - 0.03269 \times (DI)$$

Equation 5-22

DN: Degree of Non-Linearity defined by *100*(percentage of deviation caused by non-linearity)*

AV: Arrival Time Variation estimated by standard deviation of arrival time

---

[27] Sensitivity of one parameter depends on other parameters.

PE: Percentage of Error

Equation 5-22 shows the result of multiple regression of the error on the factors [28]. When $DI > 1.01(AV) - 0.3845$, error is more sensitive to the AV than DI, which is true considering the order of magnitude of DI and AV.

[28] Non-linearity, arrival time variation and interruption.

6

Causality and Sensitivity Analysis using a Perturbation Method

In the previous chapter, causality and sensitivity analysis using regression methods was implemented.

In this chapter, a perturbation method is introduced to determine causality and sensitivity. Periodic perturbation is used to transfer the problem from time domain to frequency domain. The main idea of the method is to transmit frequency information along with perturbation values so that theories from signal processing can be applied for effect differentiation.

A mathematical background for transferring from time to frequency domains is presented, and methods to analyze frequency domain data are explained. Approaches to resolve the varying latency problem are discussed as varying latency causes difficulty in applying DFT (Discrete Fourier Transform). The perturbation method is implemented and the result is presented with analysis.

6.1 Frequency Domain

6.1.1 Fourier Transform

The Fourier transform is one of the most important techniques used in signal processing. It transforms a signal from the time domain ($y(t)$) to the frequency domain ($Y(f)$). Using Fourier transforms, the frequency content of a signal can be analyzed to determine the relative proportion of frequencies present in a signal (Karu 1995).

$$Y(f) = \int_{-\infty}^{\infty} y(t)e^{-j2\pi ft}dt$$

Equation 6-1 Continuous-time Fourier transform

6.1.2 Fourier Series

The Fourier transform of an aperiodic signal *y(t)* produces a continuous transform *Y(f)*. When a signal is periodic the Fourier transform of the input signal becomes Fourier series. In a Fourier series, the spectrum is represented by a discrete set of numbers (Fourier coefficients). Each of Fourier coefficient describes the corresponding frequency component of the signal. Fourier coefficients can be used to reconstruct the original signal from weighted harmonically related sinusoids (Karu 1995). There are several different ways to represent a Fourier series.

Sine-Cosine form

All periodic functions can be represented as follows.

$$y(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos\frac{2\pi nt}{T} + \sum_{n=1}^{\infty} b_n \sin\frac{2\pi nt}{T}$$

Equation 6-2

, where T is the fundamental period of original signal in seconds.

Magnitude phase form

$$y(t) = a_0 + \sum_{n=1}^{\infty} c_n \cos(\frac{2\pi nt}{T} + \theta_n)$$

Equation 6-3

Complex exponential form

$$y(t) = \sum_{n=-\infty}^{\infty} Y[n] e^{jn2\pi ft}, \ Y[n] = \frac{1}{T} \int_{period} y(t) e^{-jn2\pi ft} dt$$

Equation 6-4

, where in general, *Y[n]* is a complex number.

6.1.3 DFT (discrete Fourier transform)

Let y[n] be a signal with finite duration, which can be represented by:

$$y[n] = 0, \ for\, n < 0\, or\, n > N_1 - 1, N_1\, is\, an\, integer$$

Equation 6-5

Based on *y[n]* defined above, a periodic signal $\tilde{y}[n]$ can be defined as follows.

$$\tilde{y}[n] = y[n], \ for\, 0 \le n \le N - 1, \ N\, is\, an\, integer$$

Equation 6-6

The Fourier series for the $\tilde{y}[n]$ given in Equation 6-6 is

$$a_k = \frac{1}{N} \sum_{n=<N>} \tilde{y}[n] e^{-jk\frac{2\pi}{N}n}$$, which becomes

$$a_k = \frac{1}{N} \sum_{n=0}^{N-1} y[n] e^{-jk\frac{2\pi}{N}n}$$, when the summation is over the interval where $\tilde{y}[n] = y[n]$.

Equation 6-7

6.2 Causality and Sensitivity

In the perturbation method, the causality and sensitivity are detected using frequency information and the amplitude ratio respectively.

As in the regression analysis, hidden relations are assumed to be linear in the vicinity (operating range) of the original value. Latency variance causes difficulty in applying the theories from signal processing, so algorithms to resolve this problem are introduced.

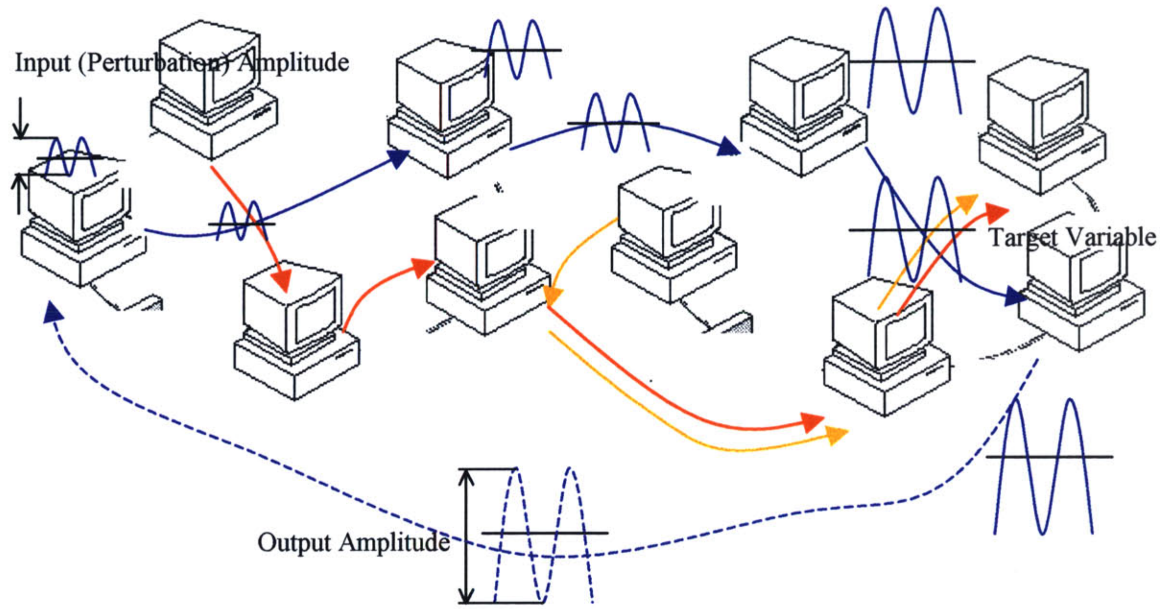


Figure 6-1 Causality and Sensitivity by the Perturbation Method

Figure 6-1, the solid line is a UDP perturbation propagation channel that is established in parallel with a TCP channel dedicated to the regular design process. The dashed line is a UDP observation channel, which is made only to observe the value change in target variable.

6.2.1 Causality

Causality is detected by checking the frequency of the value change in the target variable after propagating the periodic perturbation. The value change of target variable is recorded for a certain amount of time and then transferred to the frequency domain by DFT. The frequency components of the target variable are then analyzed to detect the input (perturbation) frequency.

Duration of Observation

The amount of time for recording is determined by Equation 6-8.

$$Duration = Initial\ Delay + Number\ of\ Perturbations \times Period + Clearance$$

Equation 6-8

In Equation 6-8, *Number of Perturbations* and *Period* are deterministic because they are known values. However, *Initial Delay* and *Clearance* are stochastic variables, of which the determination requires the Internet latency statistics and an approximation for the intermediate operation time. The Internet latency statistics are presented in 4.3.2[29].

In this work, the operation times of intermediate objects are assumed to be negligible[30].

6.2.2 Sensitivity

After causality is detected through the perturbation frequency, sensitivity can be analyzed. Based on the assumption of local linearity, the amplitude ratio is treated as the partial derivative of output to input.

$$Sensitivity = \frac{Output\ Amplitude}{Input\ Amplitude}$$

Equation 6-9

6.3 Implementation

6.3.1 Algorithm

1. Determine the frequency and the amplitude of the perturbation. The amplitude must be small enough for the linearity assumption to be valid. In this work, the set of solutions for the entire system of equations is assumed to be unique. Thus, the state of the system is restored after the perturbation vanishes. The uniqueness of the frequency needs to be guaranteed as it is used to distinguish the effect of a perturbation input. The possibility for the coincidence of frequencies is not considered. However, there are techniques to generate a distinguishable pulse or chirp that is used in radar technology. More details about chirp will be discussed in the latter part of this chapter, but implementation is left as future work.

2. Select the target variable and put the observer (client) of the target variable. The duration of observation depends on latency and the running time required for the each of the intermediate

[29] For example, the latency between Cambridge and Atlanta is between 70 and 90 ms. Within a subnet, the latency is less than 30 ms.

[30] An order of magnitude of the Internet latency is 10^{-2} second and that of the running time for an algebraic equation is 10^{-6} to 10^{-9} second. The latter is negligible comparing with the former. Considering the fact that the algebraic equation accounts for the largest portion of design constraints, neglecting an operation time is reasonable. However, in the DOME environment, a design object can be an application with huge running time, such as FEM module. In such cases, the operation time cannot be ignored.

objects[31]. Latency can be approximated by statistics, whereas running time cannot be predicted given the heterogeneity of the black box models.

3. Propagate the perturbation and record the output value behavior. The intervals between value arrivals will not be uniform.
4. Select the relevant set of output values from the recorded output values. The relevant values can be determined either by minimum variance from expected arrival time or by maximum cross-correlation with input.
5. Arrange the relevant values by interpolation.
6. Perform DFT and build the power spectrum of the target variable.
7. Determine causality by checking the frequency component and sensitivity using the amplitude ratio.

6.3.2 Frequency coincidence

Frequency coincidence happens when two or more design objects are trying to do causal analysis with same frequency at the same time. Using chirp is suggested to resolve this problem. Chirp is employed in radar to aid in distinguishing targets from surrounding clutter. It is a pulsed signal with a frequency that varies during the pulse time, usually from higher to lower frequency.

6.3.3 Interpolation

Due to varying latency, the interval between value arrivals from the perturbation signal is not constant. This causes difficulty in applying theories from signal processing. As stated in section 6.3.1, relevant values are selected by minimum variance of arrival time or maximum cross-correlation with the input signal. After the set of relevant values are chosen, the arrival times need to be rearranged.

---

[31] An intermediate object is a design module between input and the target variable, through which the perturbation is propagated.

Zero order interpolation

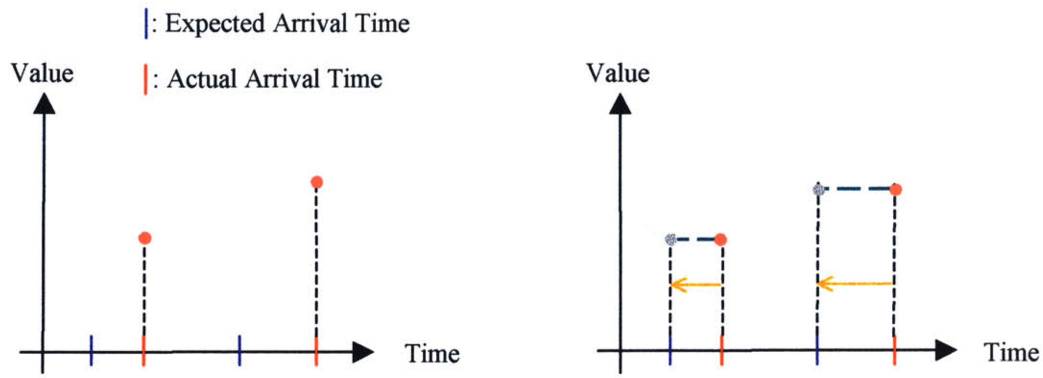


Figure 6-2 Zero Order Interpolation

First order interpolation

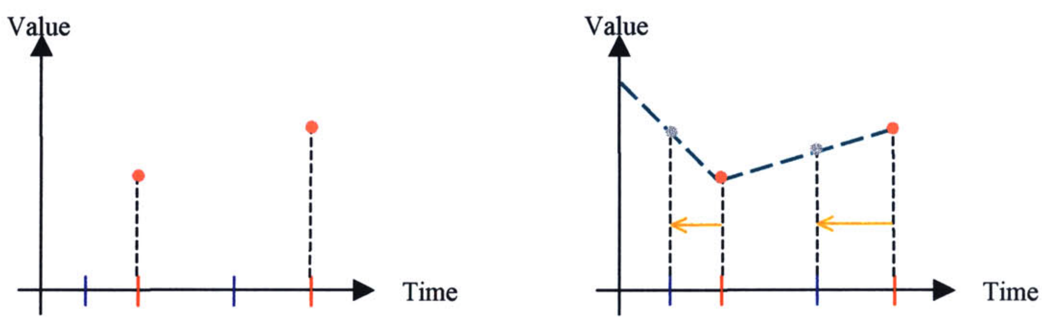


Figure 6-3 First Order Interpolation

6.4 Results

In the following, experiments using perturbation method are presented.

6.4.1 Experiment-1: linear and independent sensitivity

In the following experiments, the input, output and unknown input are *Y*, *Z* and *X* respectively.

1. Relation: $Z = 3X + 4Y$

2. Initial value: $X=2,\ Y=3$

3. Number of perturbations: *6*

4. Amplitude of perturbation: *0.05* for both

5. Interval between value generation: *X: 80ms, Y: 100ms (5Hz).*

Table 6-1 Linear Independent Case

Value (*Y*)	Arrival Time (ms)	Interval (ms)
18.4000	0	0
18.7000	10	20
18.4000	50	20
18.7000	90	30
18.3000	100	20
18.0000	120	30
18.3000	161	51
18.7000	201	30
18.4000	211	20
18.0000	301	80
18.4000	391	90
18.0000	501	110

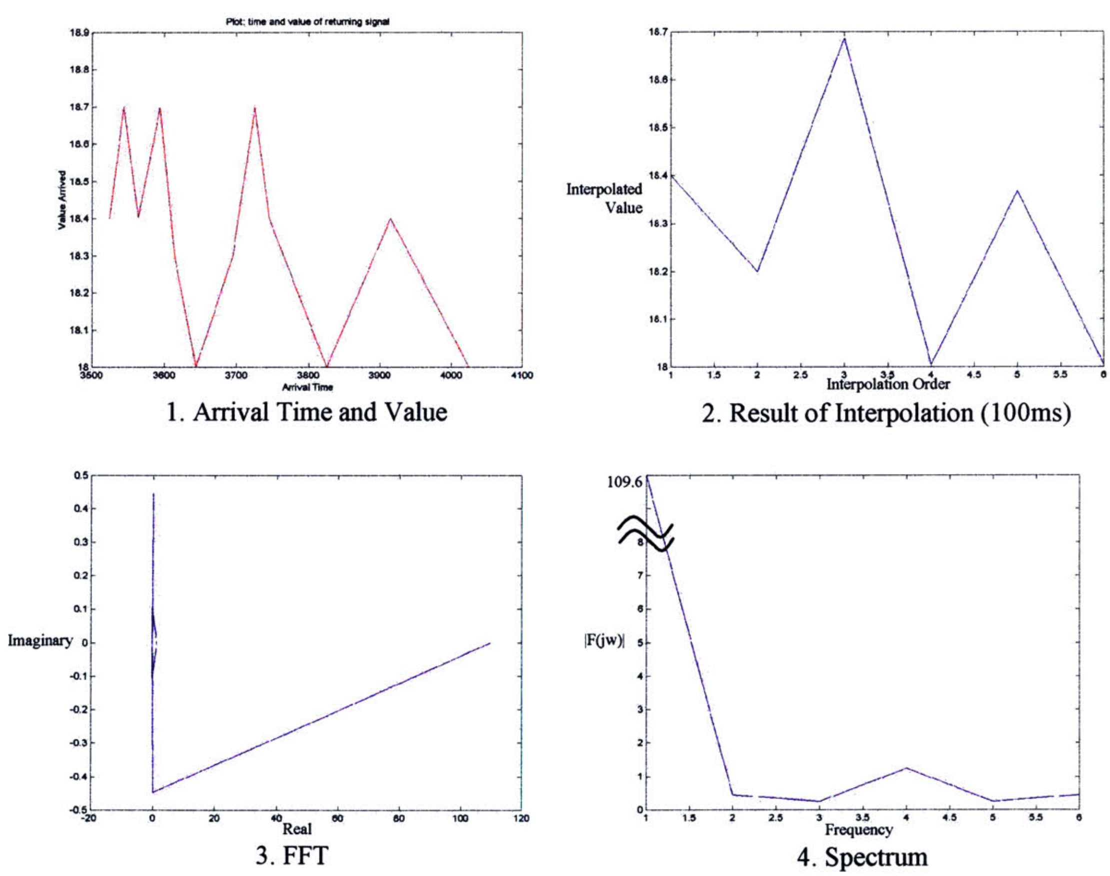
The values do not arrive in regular intervals. As explained in section 6.3.3, linear interpolation is applied to the time series, and Table 6-2 provides the results of an FFT (Fast Fourier Transform) and spectral analysis.

Table 6-2 FFT and Spectrum for Linear Independent Case

Interpolated Value (interval: 100ms)	FFT (1.0e+002 *)	Spectrum	Frequency (Hz)
18.40	1.0966	109.6626	DC
18.40	-0.0003 - 0.0045i	0.4477	1.666

18.6867	-0.0022 + 0.0011i	0.2479	3.333
18.0050	**0.0125 - 0.0000i**	**1.2453**	**5.0**
18.3673	-0.0022 - 0.0011i	0.2479	6.666
18.0036	-0.0003 + 0.0045i	0.4477	8.333

In Table 6-2, the first row corresponds to the direct current component of the signal of which the amplitude is 18.277 (109.6626/6).



1. Arrival Time and Value

2. Result of Interpolation (100ms)

3. FFT

4. Spectrum

Figure 6-4 Process to Generate Spectrum

Figure 6-4 shows the result of each step described in 6.3.1. The formulas to determine the spectrum line number corresponding to the frequency component of interest, and the amplitude of that component are provided in Equation 6-10.

$$1. \, Line\ Number = \frac{N}{R}$$

$$2. \, Amplitude = \frac{|F(jw)|}{N}$$

Equation 6-10

In Equation 6-10, N is the sample size and R is the number of samples in one period corresponding to the frequency component of interest. For example, the frequency component of $5Hz$ (original input frequency) corresponds to $R=2$ (=200/100) and the line number is 3, which is fourth line in the spectrum. The value of $|F(jw)|$ is 1.2453.and from second equation in Equation 6-10, the amplitude of 5Hz component is 0.20755 and the sensitivity can be derived by dividing the achieved amplitude by that of perturbation. In summary the sensitivity can be estimated using Equation 6-11.

$$Senstiivty = \frac{|F(jw)|}{N \cdot Perturbation\ Amplitude}$$

Equation 6-11

Table 6-3 Result-Linear Independent Sensitivity

Actual Sensitivity	Estimated Sensitivity	Percentage of Error
4	4.151	3%

In the linear independent case, the result acquired by perturbation method is identical with that of regression.

6.4.2 Experiment-2: second order and independent sensitivity

In the following experiment, the input, output and unknown input are Y, Z and X respectively.

1. Relation: $Z = 3X^2 + 4Y^2$
2. Initial value: $X=2,\ Y=3$
3. Number of perturbations: *6*
4. Amplitude of perturbation: *0.0005* for both
5. Interval between value generation: *X: 100ms, Y: 170ms (2.9412Hz).*

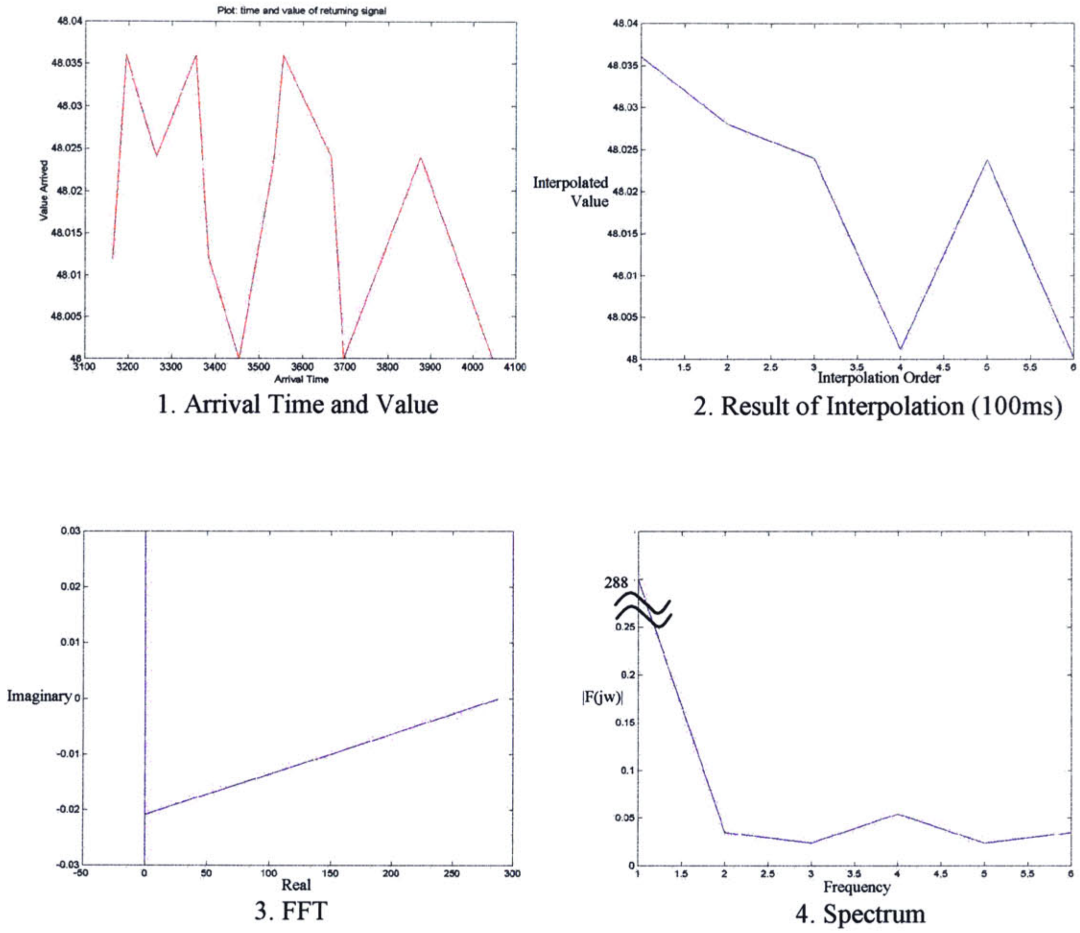
Table 6-4 Second order Independent Case

Value (*Y*)	Arrival Time (ms)	Interval (ms)
48.0120	0	0
48.0360	30	30
48.0240	100	70
48.0360	190	90
48.0120	220	30
48.0000	290	70
48.0240	370	80
48.0360	390	20
48.0240	501	111
48.0000	531	30
48.0240	711	180
48.0000	881	170

Table 6-5 FFT and Spectrum for Second Order Independent Case

Interpolated Value (interval: 100ms)	FFT (1.0e+002 *)	Spectrum	Frequency (Hz)
48.0360	2.8811	288.1132	DC
48.0283	0.0002 - 0.0002i	0.0348	0.9803
48.0240	-0.0000 - 0.0002i	0.0240	1.9607
48.0012	**0.0005 - 0.0000i**	**0.0545**	**2.9412**
48.0239	-0.0000 + 0.0002i	0.0240	3.9215
48.0001	0.0002 + 0.0002i	0.0348	4.9020

In Table 6-5, the frequency component corresponding to the original perturbation frequency is fourth one and the amplitude is 0.00908(= 0.0545/6)



1. Arrival Time and Value

2. Result of Interpolation (100ms)

3. FFT

4. Spectrum

Figure 6-5 Process to Generate Spectrum

Table 6-6 Result-Second Order Independent Sensitivity

Actual Sensitivity	Estimated Sensitivity	Percentage of Error
24	18.17 (16.0000)	24.29%

6.4.3 Experiment-3: exponential and independent sensitivity

In the following experiments, the input, output and unknown input are *Y*, *Z* and *X* respectively.

1. Relation: $Z = 4e^{X} + 7e^{Y}$

2. Initial value: *X=2.5, Y=3.5*

3. Number of perturbations: *6*

4. Amplitude of perturbation: *0.0005* for both

5. Interval between value generation: *X: 150ms, Y: 200ms (2.5Hz).*

Table 6-7 Exponential Independent Case

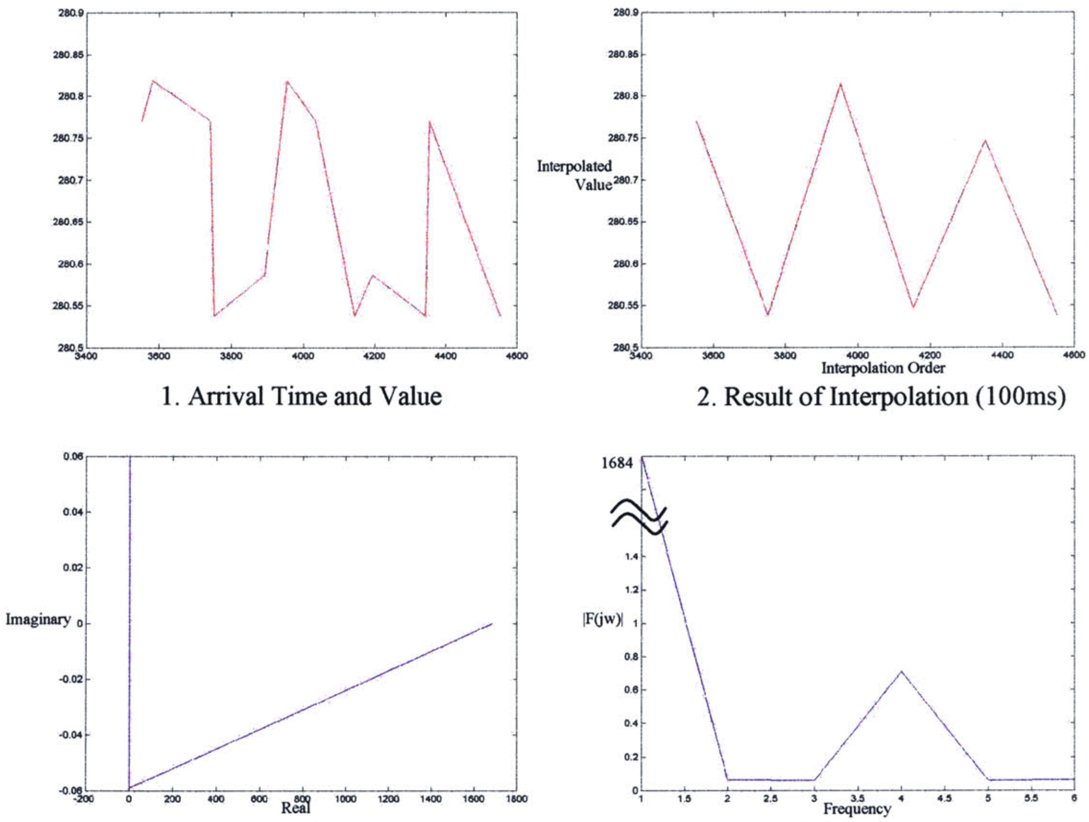
Value (*Y*)	Arrival Time (ms)	Interval (ms)
280.7701	0	0
280.8188	30	30
280.7701	190	160
280.5381	200	10
280.5869	340	140
280.8188	401	61
280.7701	481	80
280.5381	591	110
280.5869	641	50
280.5381	791	150
280.7701	801	10
280.5381	1000	199

Table 6-8 FFT and Spectrum for Exponential Independent Case

Interpolated Value (interval: 100ms)	FFT (1.0e+003 *)	Spectrum (1.0e+003 *)	Frequency (Hz)
280.7701	1.6840	1.6840	DC
280.5381	-0.0000 - 0.0001i	0.0001	0.8333
280.8150	-0.0000 + 0.0001i	0.0001	1.6666
280.5469	**0.0007 - 0.0000i**	**0.0007**	**2.5**

280.7469	-0.0000 - 0.0001i	0.0001	3.3333
280.5381	-0.0000 + 0.0001i	0.0001	4.1666

In Table 6-8, the frequency component corresponding to the original perturbation frequency is fourth one and the amplitude is 0.00012(= 0.0007/6)



1. Arrival Time and Value
2. Result of Interpolation (100ms)
3. FFT
4. Spectrum

Figure 6-6 Process to Generate Spectrum

Table 6-9 Result-Exponential Independent Sensitivity

Actual Sensitivity	Estimated Sensitivity	Percentage of Error
231.8081	236.808 (248.1666)	0.6%

6.4.4 Experiment-4: nonlinear and dependent sensitivity

In the following experiment, the input, output and unknown input are *Y*, *Z* and *X* respectively.

1. Relation: $Z = XY$

2. Initial value: *X=2.5, Y=3.0*

3. Number of perturbations: *6*

4. Amplitude of perturbation: *0.0005* for both

5. Interval between value generation: *X: 100ms, Y: 150ms (3.33Hz).*

Table 6-10 Nonlinear dependent Case

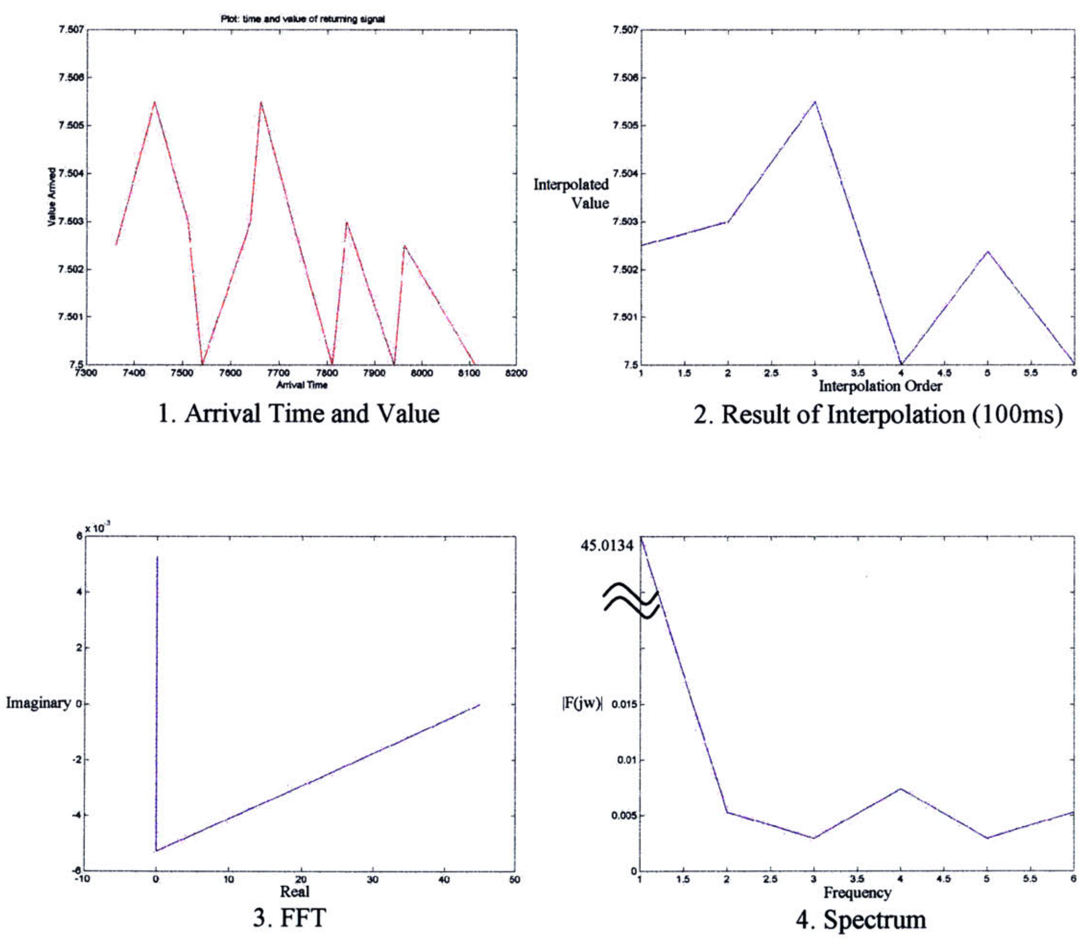
Value (*Y*)	Arrival Time (ms)	Interval (ms)
7.5025	0	0
7.5055	80	80
7.5030	150	70
7.5000	180	30
7.5030	280	100
7.5055	300	20
7.5025	380	80
7.5000	450	70
7.5030	480	30
7.5000	580	100
7.5025	601	21
7.5000	751	150

Table 6-11 FFT and Spectrum for Nonlinear dependent Case

Interpolated Value (interval: 100ms)	FFT	Spectrum	Frequency (Hz)
7.5025	45.0134	45.0134	DC

7.5030	0.0001 - 0.0053i	0.0053	1.1111
7.5055	-0.0029 + 0.0001i	0.0030	2.2222
7.5000	**0.0074 - 0.0000i**	**0.0074**	**3.3333**
7.5024	-0.0029 - 0.0001i	0.0030	4.4444
7.5000	0.0001 + 0.0053i	0.0053	5.5555

In Table 6-11, the frequency component corresponding to the original perturbation frequency is fourth one and the amplitude is 0.001233(= 0.0074/6).



1. Arrival Time and Value

2. Result of Interpolation (100ms)

3. FFT

4. Spectrum

Figure 6-7 Process to Generate Spectrum

Table 6-12 Result-Nonlinear Dependent Sensitivity

Actual Sensitivity	Estimated Sensitivity	Percentage of Error
2.5	2.467 (2.5)	1.3%

6.5 Results Analysis

6.5.1 Robustness to non-linearity

As illustrated in Table 6-13, non-linearity degrades the precision of the analysis only by an ignorable degree. The error caused by non-linearity comes from the difference between actual partial derivative and linear approximation.

Table 6-13 Error caused by non-linearity

	Actual Sensitivity	Estimated Sensitivity without Latency Variation and Interruption	Percentage of Error*100
2^{nd} Order	24.0	24.0	0
Exponential	231.8081	232.0000	8.27

6.5.2 Robustness to interruption

Every experiment has been done with and without interruption from the unknown input. Following is the result of each experiment without interruption.

Table 6-14 Error Caused by Interruption

	Actual Sensitivity	Estimated Sensitivity without Latency Variation and Interruption	Percentage of Error	Percentage of Error Increased by Interruption
Linear Independent	4.0	4.0	0%	0% (4.0)
Second Order Independent	24.0	24.0	0%	0% (24.0)

Exponential Independent	231.8081	232.000	0.083%	6.97% (248.1667)
Nonlinear Dependent	2.5	2.5	0%	0.012% (2.5003)

As shown from the results, interruption has a tendency to degrade the precision of the analysis. As defined in 5.8.2, a degree of interruption is measured by DI (Degree of Interruption).

Table 6-15 DI and Precision

	Degree of Interruption	Percentage of Error
Linear Independent	75	0%
Second Order Independent	50	0%
Exponential Independent	21.03	6.97%

As illustrated in Table 6-15, DI of linear independent case is *75* and that of exponential independent case is *21.03*. However, the percentage of error is higher with exponential case, which means that there are some other factors influencing the precision.

6.5.3 Robustness to latency variation

Table 6-16 present the error caused by arrival time variation. The latency variation strongly affects the selection of relevant variables, which affects the result of interpolation. Depending on the chosen value set, the result of interpolation can show an abrupt change.

Table 6-16 Error Caused by Arrival Time Variation

	Standard Deviation of Arrival Time (AV)	Percentage of Error caused by Arrival Time Variation
Linear Independent	3.74	6.67% (3.7333)
Second Order Independent	8.97	3.9% (23.0666)
Exponential Independent	3.72	1.67% (227.93)

6.5.4 Independent and dependent sensitivity

Dependent sensitivity can be regarded as special case of an interruption, in that it degrades the precision by directly deviating an observed value from a true value. In the perturbation method, a consistent relation cannot be established between the interruption and the error. Thus, in the later section, the contribution of interruption (along with that of dependent sensitivity) to the error will be measured considering coupling effects with other factors.

6.6 Comparison with Regression Analysis

Both the regression analysis and the perturbation method precision is influenced by non-linearity, latency variance and interruption. As is analyzed previously, the degree of degradation by one factor depends on the existence of other factors.

Regression analysis for the error factors in the perturbation method

To see the coupled influences of error-causing factors in perturbation method, regression analysis has been performed and is the result.

$$PE = 0.00018(DI)^2 + 0.0861(AV)(DI) - 0.2956(DI)$$

Equation 6-12

PE: Percentage of Error

AV: Arrival Time Variation (Standard Deviation of Arrival Time)

DI: $Degree\ of\ Interruption = \dfrac{\Delta(Output\ by\ Interruption)}{\Delta(Output\ by\ Known\ input)} \times 100$

DN: Degree of Non-linearity = (Error Caused by Non-linearity)**100*[32]

Suitability depending on range of error factors

.1 Arrival time variation

Sensitivity of an error to AV in regression analysis is

$$\left|\frac{\partial(PE)}{\partial(AV)}\right| = 0.0859(DI)$$

Equation 6-13

Sensitivity of an error to AV in perturbation method is

$$\left|\frac{\partial(PE)}{\partial(AV)}\right| = 0.0861(DI)$$

Equation 6-14

Regression analysis is slightly better than perturbation method in terms of robustness to the arrival time variation.

.2 Interruption

Sensitivity of an error to DI in regression analysis is

$$\frac{\partial(PE)}{\partial(DI)} = 0.000882(DI) + 0.0859(AV) - 0.03269$$

Equation 6-15

Sensitivity of an error to DI in perturbation method is

$$\left|\frac{\partial(PE)}{\partial(DI)}\right| = 0.00036(DI) + 0.0861(AV) - 0.2956$$

Equation 6-16

Subtracting Equation 6-16 from Equation 6-15 results in Equation 6-17.

$$\left|\left[\frac{\partial(PE)}{\partial(DI)}\right]_R\right| - \left|\left[\frac{\partial(PE)}{\partial(DI)}\right]_P\right| = 0.000522(DI) - 0.0002(AV) + 0.26291$$

Equation 6-17

When $DI > 0.0383(AV) - 503.659$ perturbation method is more robust to the interruption. Considering the order of magnitude of AV and DI, this will be the general case.

---

[32] The constant multiplied in DN and DI are set to match their order to other factors, which is required for a higher precision of the regression analysis.

6.7 Summary

There are several factors that affect the performance of the perturbation method, such as non-linearity, interruption and latency variation. These factors show coupled influence on the precision. Perturbation method shows better performance than the regression analysis, in terms of the robustness to the interruption.

7

Conclusions

7.1 Summary

In this thesis algorithms and implementations to determine hidden causal information in a network of distributed black-box design simulations are presented.

DOME (Distributed Object-oriented Modeling and Evaluation) is a design framework that enables the collaboration and communication between designers and vendors in the remote locations to streamline the product development process.

One of the unique properties of DOME is its decentralized architecture, which brings the advantages of scalability, flexibility, and most of all its capability to guarantee that proprietary information about different participants models (or services) is not exposed. As a result, the service providers can participate in design process without revealing their own know-how, which makes it possible for the wider variety of providers to join the product design process.

However, this capability of DOME hides causal information. Causal information consists of causality (topology of the relationship between design parameters) and sensitivity (strength of the relationship) information. Causality indicates the topology of the design network consisting of nodes and edges. Nodes are design objects (modules) that contain design parameters and edges are relationship between parameters. Sensitivity represents the strength of the relationship, which can be quantified by ratio of value change of output parameter to unit change of the input.

The hidden causal information prevents the participants from having the global understanding of the problem, or simply how they can affect the outputs of others. In addition, when there is a conflict between relations specified by designers, the hidden causality makes it impossible to select the design objects for negotiation. Also causal information is needed to efficiently solve a network of interacting simulations.

Two different approaches are developed to elicit hidden causal information: ***regression analysis*** and ***perturbation method***. Both methods are empirical in that they propagate a series of *periodic* test values through the simulation network, observe and analyze the behavior of the rest of the network. ***Periodic perturbations*** are propagated through a UDP channel in a very short period of time, on the order of *milliseconds*. In this time scale, which is small compared with the time scale of Internet traffic, latency can be expected to vary little. In addition, operating this higher frequency means that the causal tests are less likely to interrupt the actual design process executing through a TCP channel. Further, perturbations are generated as individual values, rather than an array of values, to avoid blocking the TCP traffic. When using TCP, high frequency packet transmission causes fluctuation in latency due to the connection procedures, such as handshaking, acknowledgement, and resending mechanisms. Thus UDP is used since it is a connectionless Internet transport layer protocol capable of enduring high frequency packet transmission without severe latency fluctuation. Uniform latency is very essential in maintaining the '*frequency*' information through the network that is needed for the causality and sensitivity analysis.

In regression analysis, series of output values are selected and matched to the corresponding input values, for which periodicity (frequency) information embedded in a perturbation is used. Causality is determined by comparing conditional variations of the output with and without regard to the input (Granger causality). Linear regression coefficient is estimated and sensitivity, which is a ratio of input and output covariance to the variance of an input.

In perturbation method, problem is transferred to a frequency domain using the periodicity. After switching the domain, the output is represented by frequency spectrum. Causality is verified by the existence of frequency component in the spectrum, and sensitivity is estimated by the ratio of amplitudes.

In implementing both methods, issues to be addressed included: non-linearity of the relations, variation in Internet latency, and differentiation of the effects under the presence of interruptions from other design objects. In terms of the robustness to the nonlinearity and latency variation, two methods do not show significant differences. As for the robustness to interruptions, perturbation method generally shows better endurance than the regression analysis.

As previous approaches, neural network and token (message) circulation (TC) are introduced. Neural network approach consists of data collection and the network training. However, when a design module with huge running time is involved, the data collection is prone to the same disadvantage with regression analysis and perturbation method. Also, the time required for the neural network training is not suitable for the fast evolution of a design assumed in this work. Another alternative TC is advantageous, when a design module with large running time exists in the network. However, a local causality needs to be determined to apply TC. In addition, sensitivity cannot be estimated using TC.

7.2 Future Work

One of the assumptions of this work is that the local constraint sets satisfy completeness and consistency criteria. However in practice, local constraint set management should precede the causal analysis to insure the uniqueness of a solution. Based on the inferred causal information, issues of global completeness and consistency can then be addressed. Another important assumption is that values are propagated through a single causal path. This assumption rules out the possibility of bias due to the violation of the independence condition in regression analysis, an echo in perturbation method. In the future the methodologies should be improved to eliminate the need for these assumptions.

7.2.1 Constraint management

One of the goals of causal analysis is the efficient maintenance of global completeness and consistency of the constraint set. Considering that design is highly constraint oriented, the constraints must be managed in a way that parallels the approach to solving the design problem. One of the most common approaches to constraint management is a divide-and-conquer method. Seeing that traditional design frameworks are centralized, the divide-and-conquer (constraints set decomposition) method is very natural.

DOME is unique for its decentralized architecture, that is, it does not have central server that keeps all the relative information. Through this property, DOME can offer great advantage of flexibility with its approaches to the design solution. Depending on the complexity, problem characteristics and amount of published information, design solution can be pursued either in a top-down or bottom-up manner, and the constraints set is build and managed in parallel with the problem solving method.

A systematic approach is required for the local constraints problem when the individual modules have greater complexity. The algorithms for global constraint management assume a locally satisfied constraint set. A standardized interface needs to be equipped for the module to provide minimum information required to manage constraints.

7.2.2 Multi-path causality

Only single path causal problems have been addressed. However, one parameter can also affect another from multiple paths. In such cases, the independence condition between explanatory variables is violated, which causes bias for regression coefficients. As a solution, latent variable modeling, instrumental variable methods and hidden Markov chain methods were suggested. In the perturbation method, effect propagation through multiple paths generates echo in output behavior, which causes difficulty in spectral analysis. Echo canceling filters are suggested to resolve this problem.

7.2.3 Circular dependency (Loop)

Circular dependency should be detected and handled. In chapter 5, Granger's algorithm to detect feedback was presented. Once the loop is detected it should be grouped into a super node. Super node can be represented a set of equations (when the constraints are all algebraic), and the solution should satisfy the equations simultaneously.

7.2.4 Avoid coincident frequency

When multiple design objects are propagating perturbations with the same frequency and at the same time, the coincidence of the frequency makes it difficult to differentiate the effect. To resolve this problem, the application of chirp was suggested as a future work.

7.2.5 Intermediate nodes with long execution times

Both regression analysis and the perturbation method approach the causal problem by applying a test signal to the network. However, when an intermediate design object with large execution time is involved, the time required for tests also grow proportionally. Running time depends on the amount of required computation, performance of the application and the values used in the test. In this circumstance, a messaging (token circulation) architecture through the predetermined causal path would be advantageous for determining causality. However, messaging cannot provide information on sensitivity.

References

Abul M. M. Masih and Rumi Masih., "Energy consumption, real income and temporal causality: results from multi-country study based on cointegration and error-correction modeling techniques", Energy Economics, Vol 18, No 3 pp 165-183, 1996.

Almy, Tom., "Context Dependent Model Visualization", M.S. Thesis, Dept. of Mechanical Eng., MIT., Cambridge, MA 02139, 1998.

B.S. Everitt., "An Introduction to Latent Variable Models", Chapman and Hall Ltd, 1984

Bolot, Jean-Chrysostome., "End-to-End Packet Delay and Loss Behavior in the Internet", INRIA, B.P. 93, 08902 Sophia-Antipolis Cedex, France, 1993.

Can A. Baykan and Mark S. Fox., "WRIGHT: A Constraint Based Spatial Layout System", Artificial Intelligence in Engineering Design, New Brunswick, New Jersey, vol. 1, pp.395-432, 1992.

Christopher A. Sims., "Money, Income and Causality", The American Economic Review, pp 540-552, September, 1972.

Cross, Nigel., *Engineering Design Methods*, New York. John Wiley & Sons Ltd, 2000.

C.W.J. Granger., "Investigating Causal Relations by Econometric Models and Cross-spectral Methods", Econometrica, Vol. 37, No. 3, July 1969.

Damodar N. Gujarati., *Basic Econometrics*, McGraw-Hill Book, 1995

Daisie D. Boettner and Allen C. Ward., "Design Compilers and the Labeled Interval Calculus: A Tutorial", Artificial Intelligence in Engineering Design, New Brunswick, New Jersey, vol. 1, pp.135-192, 1992.

David G. Ullman., *The Mechanical Design Process*, Hightstown, NJ: McGraw-Hill, Inc, 1992

Douglas E. Comer., *Computer Networks And Internets*, Department of Computer Science, Purdue University, West Lafayette, IN 47907, 1999.

D. Sanghi, A. K. Agrawala, O. Gudmundsson and B. N. Jain, "Experimental Assessment of End-to-end Behavior on Internet", Proceedings, Twelfth Annual Joint Conference of the IEEE Computer and Communications Societies. Networking: Foundation for the Future, IEEE Vol 2, PP 867-874, 1993.

Garret, J. and Fenves, S. J., *A knowledge-based Standards Processor for Structural Component Design,* R-86-157, Dept. Civil Engineering, CMU, Pittsburgh, PA 15213, September 1986.

G. Robert Malan and Farnam Jahanian., "An Extensible Probe Architecture for Network Protocol Performance Measurement", Proceeding, ACM SIGCOMM, 1998

Harary, F., *A graph Theoretic Approach to Matrix Inversion by Partitioning,* Numerische Mathematik, vol. 4, pp.127-135, 1982.

I. Abel., "Credit, imports and inventories in CPES-causality test for the Hungarian economy", Engineering Costs and Production Economics, Vol 19, pp 11-17, 1990.

Jan Roelf Bult, Peter S.H. Leeflang, Dick R. Wittink., "The Relative Performance of Bivariate Causality Test in Small Samples", European Journal of Operational Research (97), pp450-464, 1997.

Juan C. Deniz., "Learning Theory Applications to Product Design Modeling", M.S. Thesis, Dept. of Mechanical Eng., MIT., Cambridge, MA 02139, 2000.

Judea Pearl., *Causality: models, reasoning, and inference*, Cambridge University Press, Cambridge, UK, 2000.

Judea Pearl., *Probabilistic Reasoning in intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publisher, Inc, San Mateo, CA, 1988

Kathryn Blackmond Laskey., "Sensitivity Analysis for Probability Assessments in Bayesian Networks", IEEE Transaction on Systems, Man and Cybernetics, Vol 25, No 6, pp 901-909, 1995.

Kevin Gurney, *An Introduction to Neural Networks*, UCL Press, London, 1997.

Kimberly C. Claffy, George C. Polyzos and Hans-Werner Braun, "Traffic Characteristics of the T1 NSFNET Backbone", Proceedings, Twelfth Annual Joint Conference of the IEEE Computer and Communications Societies. Networking: Foundation for the Future, IEEE Vol 2, PP 885-892, 1993.

Koichi Yamada., "Reasoning for Causality Analysis with Possibility and Intensity Occurrence", IEEE International Conference on Systems, Man and Cybernetics, Vol 5, pp 848-853, 1999.

Larry L. Peterson and Bruce S. Davie., *Computer Networks A Systems Approach,* Morgan Kaufmann Publisher, Inc, San Francisco, CA, 1996.

Light, R., "Symbolic Dimensioning in Computer-Aided Design", M.S. Thesis, Dept. of Mechanical Eng., MIT., Cambridge, MA 02139, 1980.

M. D. Ryan and P. M. Sharkey., "The causal surface and its effect on Distribution Transparency in Distributed Virtual Environment", IEEE International Conference on Systems, Man and Cybernetics, vol 6, pp 75-80, 1999.

M. G. Kendall and A. Stuart., "The Advanced Theory of Statistics", Charles Griffin Publishers, New York, 1961, vol.2, cha.26, p. 279

Moshe Ben-Akiva., "Extension of the Random Utility Model", RU2000 at Duke University, Aug, 2000

P.A. Subrahmanyam, "Representation and Control for the Evolution of VLSI Designs: An Object-oriented Approach", Artificial Intelligence in Engineering Design, New Brunswick, New Jersey, vol. 1, pp.57-69, 1992.

Pahng, F., N. Senin, and D. Wallace, "Modeling and Evaluation of Product Design Problems," to appear in Computer-Aided Design, 1998.

Roshan Lal Sharma, Paulo J.T. de Sousa and Ashok D. Ingle., *Network Systems: Modeling, Analysis and Design*, Van Nostrand Reinhold Company, 1982.

Sanjay Mittal and Agustin Araya., "A Knowledge-Based Framework for Design", Artificial Intelligence in Engineering Design, New Brunswick, New Jersey, vol. 1, pp.273-293, 1992.

Steward, D.V., *On an Approach to Techniques for the Analysis of the Structure of Large Systems of Equations.* SIAM Review, vol. 4, No.4, 1962.

Serrano, David and Gossard, David, "Tools and Techniques for Conceptual Design", Artificial Intelligence in Engineering Design, New Brunswick, New Jersey, vol. 1, pp.71-116, 1992.

Serrano, David and Gossard, David, *Combining Mathematical Models with Geometric Models in CAE systems,* ASME, Proceedings of 1986 International Computers in Engineering Conference, July 1986.

Tiong-Hwee Goh., "Semantic Extraction Using Neural Network Modeling and Sensitivity Analysis", International Joint Conference on Neural Networks, Vol 1, pp1031-1034, 1993.

Zellner, Arnold., "Causality and Econometrics", Carnegie-Rochester conference series on public policy, pp 9-54, 1979.

Zoher Z. Karu., *Signals and Systems*, ZiZi Press, Cambridge, MA, 1995.