Design Basis for Achieving Information Integrity – A Feedback Control System Approach

by

*Mandke Vijay V.
Director and Professor,
School of Engineering and Technology,
Indira Gandhi National Open University,
(EMPC Premises), IGNOU Campus, Maidan Garhi
New Delhi-110 068 (India)*

and

*Nayar Madhavan K.
President,
Unitech Systems, Inc.
1240 E. Diehl Road, Suite 300,
Naperville, Illinois 60563, USA.
E-mail : mnayar@unitechsys.com*

Abstract

The paper studies errors in information system (IS) and loss of integrity in IS thereby. Specifically, the paper identifies accuracy, consistency and reliability as intrinsic integrity attributes that all information systems must satisfy. In networked computerized information systems, it is the inadequacy of application controls conceived at system design stage, and the possibilities of inherited errors and human errors that result in lapses in information integrity (II); in turn infesting IS with the possibility of massive risk of information pollution. The paper argues that these errors, which are made but not corrected, are due to factors drawn from system environment external to the application system and overlapping the user environment. This calls for automatic feedback control systems for on-line error detection (or estimation or prediction as the case may be) for integrity improvement in information systems.

1 INTRODUCTION

Networked computerized information systems of today see 'Data' as raw material, 'Data Product or Information' as processed data used to trigger certain management action, 'Processing' as the system function, and are characterized by (a) computing processes that include microcomputer and telecommunication and (b) pre and post-processing stage communication channels at various data/information processing nodes, that are people based and include data communication and transaction processing networks with world-wide reach. Such decentralized structure of IS has certainly facilitated organizations and individuals to work with shared data environments and with capture, use and control of growing, complex and diversified volumes of data and information; in turn affording business access to bigger markets. However, these networked computerized information systems make mistakes. It is this alarming reality that requires attention to questions of errors in information systems, of poor integrity of information systems and of finding methods, technologies and techniques for maintaining and improving integrity.

2 ERRORS IN INFORMATION SYSTEMS

Consider a conceptual representation of an Information System (IS) model as given in Figure 1 where $\langle e_i, a_i, v_i \rangle$ denotes a triple for data model (input to the information system) and $\langle e_o, a_o, v_o \rangle$ denotes a triple for information model (output from the information system); $\langle e, a, v \rangle$ representing datum a triple $\langle \text{entity, attribute, value} \rangle$ as developed by the database research community. This representation which permits treating data/information as formal organized collection allows to segment integrity issues into issues concerning entities, attributes and values thereby making it feasible to study IS integrity analytically.

where $\langle e_i, a_i, v_i \rangle$ denotes a triple for data model and $\langle e_o, a_o, v_o \rangle$ for information model.

Figure 1 Conceptual Presentation of an Information System Model.

Each stage in this networked computerized information system model contains errors. To elaborate, Data origin stage comprises designing and operating data collection procedures and codes, form filling (leading to data generation), data collection, and data preparation along with machine operation (where data is converted into machine-readable form). Various errors during operating data collection procedures and codes concern incorrect, incomplete and ambiguous manual, inappropriateness of manual language to user, non-availability of manual and carelessness, and they result into use of wrong procedure or code.

During filling in the data forms, errors encountered are in terms of ambiguous form filling direction, poor format, substitute or unauthorized person filling forms, poor motivation and once again carelessness, resulting into incorrect filling of forms. Coming to data collection stage, it is infected by errors caused by poorly designed forms and codes, poor handwriting and carelessness, resulting in omissions, inaccuracies, data in wrong place, loss of data, data manipulation, etc. Further, data preparation stage along with machine operation stage where data is prepared and converted into machine-readable form suffer from errors caused by poorly written keypunch instruction, poor procedures, hardware errors, poor maintenance, misrouting of documents, sabotage, carelessness; resulting in incorrect operations, fraudulent operations due to crime, data not processed on time, and machine breakdown.

Coming to the pre and post-processing communication channels, growth of online databases, distribution of information system resources across geographically separate locations, user community spanning entire enterprise, increased use of data communication functions, and information management networks combining terminal access to information processing and database resources, electronic mail, office automation, word-processing, facsimile, and graphics have understandably made 'communication channels' a very significant component of an information system. And it so turns out that these communication channels are extensively prone to errors.

To begin with unlike long distance voice communications and conventional radio broadcast, data communication content and performance are affected by inherent signal interference, usually referred to as communication noise thereby introducing errors in the data transmitted [Menkns, 1990]. Other causes affecting telecommunication and thereby introducing error in communication channel, relate to physical structure of telecommunication and to logical aspects of the data communication process itself. Specifically, causes relating physical structure of telecommunication (in addition to the factor of communication channel noise already mentioned) pertain to factors of electromagnetic signal radiation, circuit

switching deficiencies such as cross talk between circuits and the failure of mechanical or electronic components of the switch itself, and any interruption in the electrical power supply used by the switching facility.

As regards to causes relating logical aspects of data communication, they include failure of the software used in either data communication network management or in data communication itself, and the size and inherent complexity of the network being used. In addition to these, communication channel also suffers from problems of theft of service and circuit tapping, acts of sabotage, incidents of accidental destruction and of effects of adverse weather and water caused damage.

At the processing stage, this operation transforming 'data' into 'information' comprises machine operation, use of data files, use of systems and application programmes, and of processing operation itself. Errors during machine operation have been identified earlier. In respect of data files, errors caused are due to poor physical storage, lack of clearly defined responsibilities for data files, inadequate procedures, natural disaster and theft, fraud or sabotage; resulting in warped cards, dirty tape of disks, destruction of files, etc.

During the use of systems and application software, errors caused are due to out of sequence programming, wrong algorithms, wrong programming instructions, poor documentation, lax security; resulting in incorrect solutions and unauthorized changes. And coming to the processing itself, perhaps the most significant cause for errors is carelessness in data processing; resulting in records lost and use of incorrect file.

Finally, it is at the output stage that the user receives the 'information' which is the output of the information system. This 'information' constituting 'output' is used by the user as an aid in action or in management or in decision. This 'output' is the product of all input and processing under the information system model. Errors at this output stage are thus caused due to processing and operation errors; resulting in inaccurate and incomplete output.

Figure 2 gives a systems view of the errors in an information system identified above.

3 LOSS OF INTEGRITY IN INFORMATION SYSTEM

Chamber's dictionary gives the meaning of integrity as 'entireness, wholeness, unimpaired stage of anything, purity'. Errors as above at various stages of an information system certainly result in loss of integrity at each of these stages as also in the loss of overall system integrity. Specifically, errors at data origin stage,

Figure 2 System's View of the Information System Model in Figure 1 incorporating error implications

resulting into use of wrong procedures or codes, incorrect filling of forms, incorrect or fraudulent operations, data not processed on time, machine breakdown, etc. give rise to inaccurate, incomplete, backdated and insecured data, further threatened by loss of privacy in view of fraudulent operations.

Similarly, errors (during communication channel stage prior to processing) caused by communication channel noise, physical structure of telecommunication and failures in logical aspects of data communication, circuit tapping and theft of service, acts of sabotage, incidents of accidental destruction and the unpredictability of the complex networks used, give rise to inaccuracy, incompleteness, loss of confidentiality and loss of privacy in data. Further, during processing stage, errors in machine operation, errors in respect of data files, application and systems software and errors in processing itself, give rise to inaccurate, incomplete, insecured data further threatened by loss of privacy. As pointed above, communication channel to post-processing stage also contributes to the loss of accuracy, completeness, confidentiality and privacy – this time of processed data i.e., information.

Finally, errors at output stage also result in inaccurate and incomplete information.

4 DEFINING INTEGRITY ATTRIBUTES : A HEURISTIC TREATMENT

As can be seen, errors in information systems result in loss of integrity at each stage of the information systems, and thereby in the loss of overall system integrity. This integrity loss is in terms of the attributes (not to be confused with entity attributes referred to in this Section and paper) of : accuracy (purity), completeness (entireness, wholeness), data/information being up-to-date (i.e. timeliness implying accuracy inspite of time related changes in data/information), security and privacy (unimpaired meaning undamaged; purity).

Let us consider data/information modeled as a triple $\langle e, a, v \rangle$ as suggested in Section 2 above. As explained earlier, this affords a very meaningful approach whereby integrity attributes for an information system can be considered by studying them (integrity attributes) for the components of the triple, i.e., entity, attribute and value [Redman, 1992].

To elaborate, a universe for a company may comprise 'employees', 'products' and 'customer orders'. Employees and products represent 'entity classes or types' and customer order represents 'relationship' in the universe. For the purpose at

hand, an entity class, say employees class may be represented by attributes as follows:

EMPLOYEE = (Employee Number, Name, Department Number, Salary, Date of Birth, Sex)

Finally, a value provides information for specified attribute of a specified entity. For example, for employee entity, Albert, entity representation may be as follows:

Albert = (94256, Albert, 9, \$ 15000 p.a., 6.5.75, M)

In such case, Date of Birth attribute has the value 6.5.75 and Salary attribute has the value \$ 15,000 p.a. and so on.

4.1 Accuracy, completeness and timeliness attributes

With data/information model (in terms of triple) illustrated above, a clearer picture of integrity attributes could be obtained. For example, to study accuracy of the information on the company, one may study accuracy of the information on entity types, namely, employees and products and on relationship customer orders. Further, to study accuracy of information on entity type employees, one may study accuracy of information about attributes corresponding to entity type employees. Finally, to study accuracy of information about attributes, one may study accuracy of values for attributes, thereby making the exercise of studying accuracy of information on the company a viable exercise. For example, to study accuracy of information about salary attribute in the illustration above, one has to see if value of salary attribute \$ 15,000 p.a. is correct or not.

Similarly one can ascertain requirements of completeness and timeliness (i.e. accuracy inspite of time related changes) of information, both of which are 'necessary' for requirement of 'accuracy', though not sufficient.

4.2 Security (confidentiality) and privacy attributes

Finally, from the point of view of information model, requirement of security, meaning undamaged information, is analogous to accuracy, as any damage to information, that is to say, to the value of an attribute will only result in inaccuracy of the value.

Security also has an aspect of confidentiality. Further, security of information is also important from the point of view of privacy. However requirements of confidentiality and privacy, though they emerge as implications of errors in the information system, cannot be considered to be central requirements for all information systems, as there can be informations where confidentiality, and, for that purpose, security, and privacy may not be required.

Thus from the set of integrity attributes of accuracy, completeness, timeliness, security and privacy identified above, attribute of accuracy is central to an information system and attributes of completeness and timeliness are necessary for the attribute of accuracy. In other words, attributes of accuracy, completeness and timeliness are intrinsic to an information system irrespective of the use of information derived from the system. Against this, requirements of security in the sense of confidentiality and of privacy are optional to an information system and depend on the context and nature of use of information.

4.3 Consistency and reliability attributes

There are two other requirements that have not emerged in the integrity analysis so far and they are consistency and reliability of data/information. Information can be seen to meet consistency requirements if say attribute value satisfies the domain as well as the constraints on the value. Specifically, like completeness and timeliness requirements, consistency requirement is also a part of accuracy requirement, i.e., if data/information is accurate, then it is also consistent, but otherwise is not true.

Coming to the requirement of reliability its origin may be seen in the very choice of the information system model, wherein system output, i.e., information, is defined as what user receives as an aid in action or in management or in decision. Here no user is defined as such, but 'utility' or 'use' role of information is brought out. It is in this context, that the requirement emerges that information obtained be reliable. What is discussed here, is perceiving reliability as an accuracy with which the information obtained represents the data item in whatever respect the information system processed it.

Depending on the nature and type of information item, there can be methods and techniques for quantifying integrity attributes. For example, if one is talking about information item in respect of value of an attribute where, as in case of the example considered earlier, the attribute considered is 'Salary' which takes numerical value, then accuracy of information item on value of the Salary attribute at given month can be quantified in terms of say 'difference between value obtained and correct value' or in terms of 'error ratio of actual error and acceptable error' etc. But what if there is a problem of correct value of the identified source of the information item (also called standard) being undefined, or simply unknown. In situation an assumed standard itself may be incorrect as is often the case with data gathered some time in the past and with no corroborating evidence. In yet another situation there may be more than one correct value. Then there is a problem of how to quantify accuracy if the value does not lie on a real

line, i.e., it is not a numerical but say an alphanumeric or an alphabet or say even a picture or sound or multi-media presentation.

Whatever is said here about accuracy attribute is also valid for completeness, timeliness, consistency and reliability attributes. However, immediate effort here is not to offer methods for quantifying integrity attributes, but to state significance and centrality of requirement of integrity attributes of accuracy, completeness, timeliness, consistency and reliability in respect of an information system.

4.4 Intrinsic integrity attributes

Thus, taking entire discussion together, irrespective of the nature of use of the information obtained from the information system, attributes of accuracy, completeness, timeliness, consistency and reliability emerge as intrinsic attributes that an information system must meet, while attributes of security and privacy are optional depending on the context and nature of use [Mandke, 1996; Mandke and Nayar, 1997].

There is more to the intrinsic integrity attributes mentioned above. As pointed out earlier, attributes of completeness and timeliness are necessary for accuracy. That is to say, when checked for accuracy, the value of the information item also gets checked for its completeness and for it's being up-to-date (timeliness), as accurate value has to be complete and timely. In that sense, it is sufficient to check for accuracy only.

Similar is the situation in respect of consistency, too, as an accurate value also has to be consistent. However, difference is that, as mentioned earlier, consistency check is in terms of domain values and in terms of constraints without referring to real world objects and, therefore, a simpler and less expensive task offering first approximation on accuracy and, when checked in addition to accuracy, increasing overall reliability of integrity checking process itself.

It is within the above framework then accuracy (includes completeness and timeliness), consistency (satisfying constraints) and reliability (accuracy with which information item represents data item in whatever way information system processed it) emerge as intrinsic or basic or objective attributes of information integrity. As mentioned earlier, what one is considering here is an information system model which delivers Information for use. Therefore, depending on IS application area and industry standards, without reference to any specific user there may be different requirements/standards of information. For example, although accuracy of information is intrinsic and central to the information system performance, in many application areas irrespective of who is the individual user

it is not uncommon for four digits to the right of the decimal place to be rounded off without loss of much information. Thus, even though the intrinsic integrity attributes of accuracy, consistency and reliability emerge as basic integrity attributes all information systems must demonstrate, depending on the application area, each of these intrinsic integrity attributes may satisfy different application area specific industry standards [Mandke and Nayar, 1997].

4.5 Extrinsic integrity attributes

As observed earlier, depending on the context and nature of use, there would also be user specific integrity requirements, namely, security and privacy and these emerge as extrinsic or subjective attributes of Information Integrity. There are other subjective attributes of Information Integrity, too, as reported in the literature based on user survey in respect of data/information requirements [Delone and Mclean, 1992; Mandke, 1996]. These are : usability, independence, precision, relevance, sufficiency, understandability, freedom from bias, conciseness, brief, trustworthy, etc.

Just as in the case of intrinsic or objective attributes of Information Integrity, there is also a question of defining and, where possible, quantifying these extrinsic or subjective integrity attributes. But then this query is beyond the scope of the query at hand in terms of developing a design basis for achieving Information Integrity.

Figure 3 presents a system's view of the integrity implications for a networked computerized information system.

5 INFORMATION POLLUTION – A DIRECT IMPLICATION OF LOSS OF INFORMATION INTEGRITY

Traditionally, errors in information system (IS) and resulting integrity implications have been addressed as if they were preventable or correctable by building controls within application system. With time there have been efforts to put in greater inputs right at the system analysis and design stage, hoping that would ensure Information Integrity [Hussain, D. and Hussain, K.M., 1984].

5.1 Inadequacy of controls within computing system

However, reality is different. Main reason for this is ironically the inadequacy of controls designed to meet lapses in integrity at different IS stages. To elaborate, with reference to data collection procedures and codes, in view of shared data

Figure 3 System's View of the Information System Model in Figure 1 incorporating error implications and presenting emergent Integrity attributes

environments, in view of user community spanning entire enterprise and in view of human element of carelessness, it is extremely difficult to avoid situations like manual unavailable when needed, use of unauthorized manual, user inability to understand manual language, resulting in the use of wrong procedure or code and, therefore, in inaccurate and insecured data inspite of controls. Coming to the form filling stage, situations of poor motivation, carelessness and substitute or unauthorized persons filling out forms continue, resulting in incorrect filling of forms. As regards to data collection stage, errors caused due to carelessness are very difficult to eliminate completely. Similarly, controls at data preparation stage are not sufficient to eliminate error implications of carelessness in data preparation. Further, it is also extremely difficult to ensure controls like employee selection and training for foolproof results.

In respect of machine operation stage, errors caused particularly due to human behaviour like carelessness of operators, sabotage, desire for personal gain, documents lost or misrouted are difficult to eliminate. Further, once again, it is difficult to have controls like upgraded personnel selection or training that are foolproof. Similarly at communication channels at pre and post- processing stages, inspite of all communication controls, errors caused by failure of data communication software, channel noise, and unpredictability of data communication network due to its size and complexity, not to speak of adverse influence of people and weather, are difficult to eliminate. Also controls during machine operation at preprocessing stage and their inadequacy are same as that for machine operation following the data preparation stage discussed earlier.

Coming to data files, controls considered are controlled humidity storage, 'clean room' conditions, special cabinets, periodic cleaning, centralized storage under librarian, upgrading of storage procedures, backup data and controlled access to files. However, these controls cannot completely remove causes of theft, fraud or sabotage which stem from human behaviour. All controls in respect of systems and application software do not ensure removal of all errors, particularly those caused during programming and those due to poor documentation and lax security. Also all controls at processing stage cannot remove errors that are caused by carelessness. Finally, controls at output stage cannot always eliminate all operation or processing errors [Mandke, 1997].

It is these errors at different IS stages that (due to inadequacy of application controls) are made but not corrected, that then render an IS infested with information having poor integrity.

5.2 Inherited errors

Rapidly evolving technologies producing distributed information networks and shared data environments that are common place today pose yet another issue. Once an error occurs it can be considered an 'inherited error' [Banks and Weimer, 1992] if it is passed along to another computer, network, database, file, or the like. Inherited errors occur when an error is propagated beyond the system in which it originated. For example, if a personal computer on a local area network is used to prepare a report – and erroneous data is incorporated into that report – when the report is submitted to another computer system, an error is inherited [Wood and Banks, 1993]. Once again, various controls that are considered and implemented cannot remove these inherited error possibilities.

5.3 Human error: a significant integrity problem

As one critically considers the above discussion, amongst others, factors of carelessness, poor motivation, and other actions of people emerge as most significant factors contributing to errors in information system and, thereby, lapses in integrity. According to a study performed by the Executive Information Network [Bloombecker, 1989], 55% of the respondents involved in a survey considered human error as the most important integrity threat. Ironically, human error can also be one of the most serious problems causing system interruptions.

While the frequency of human error and the opportunity for human error are important considerations, the magnitude of the loss due to human error is also a major concern. Contrary to the general opinion of many information system practitioners, human error is not always a low-consequence threat. In fact, in terms of money lost, human error is the largest single cause of economic and productivity loss in the information system integrity arena. As an evidence of this, consider a study reported in Computerworld [Annon, May 1987], which attributes 52% of corporate information damage to human error. Single-incident losses can also be significant. And application controls cannot remove these human error possibilities, as can be seen from the feedback on error occurrences and their consequences as available from the field.

As computer systems become increasingly networked, as applications increasingly share data with one another, and as databases become increasingly distributed, the risk of data/information error – including risk of inherited error – increases, which amounts to the issue of data/information pollution.

6 FACTORS EXTERNAL TO COMPUTING (AND APPLICATION) SYSTEM RESPONSIBLE FOR LOSS OF INTEGRITY IN INFORMATION SYSTEM

As discussed through previous Section, errors in information systems are caused by factors not amenable to controls including application controls conceived at system design stage itself. Of course, literature reports research efforts in terms of identifying foolproof information requirements [Kliem, 1992; Mostert; 1994; Tompkins and Rice, 1996], but design experience shows this is something not easy to achieve.

This is because, due to factors detailed here, computerized information systems invariably have errors that are made but not corrected by the controls incorporated at system design stage. As can be seen, these factors, invariably have their presence mainly through the system environment which is external to computing (and hence the application) system and overlaps the user environment, though together they (the computing system and its external environment) constitute the Information System Model as in Figure 1. In spite of application controls, it is these external factors that then make information systems give rise to information which is inaccurate, inconsistent and unreliable [Mandke and Nayar, 1997; Nayar, 1996].

These external factors could be categorized into five major categories; namely, change, complexity, communication, conversion and corruption.

Change

Change may occur either in the content or in configuration of the system environment, resulting in a possibility of error introduction in the information system. Every hardware change, software release, and organizational change will come under this category, offering cause for error and, therefore, for a possibility of an inaccurate, inconsistent or unreliable information.

Complexity

Whenever one introduces complexity, there is a possibility of error introduction in the information system. Every new component, be it a programme, database or a network, adds new interfaces increasing the possibility of error introduction.

Communication

Communication stands for movement of data/information within or across enterprises and it also provides a chance for error introduction.

Conversion

Conversion, in this context, refers to the consolidation, decomposition or transformation of data. Whenever one converts data from one form to another, there exists a possibility of error introduction, resulting in information which may not be accurate.

Corruption

Finally, corruption pertains to human behaviour (poor motivation, desire for personal gain, carelessness, actions of people), to factors leading to inherited errors polluting the information systems, and to unpredictability (noise) of any kind leading to introduction of errors in computerized information systems.

Whether, in addition to controls discussed above, computerized information systems also incorporate human engineering design criteria at the system design stage itself or hardware and software vendors further incorporate error-checking filters into their products, it is these external error factors that then have to be addressed, if one were to resolve the question of errors in information system model so as to obtain information which is accurate, consistent and reliable. As observed by Svanks in her significant work entitled 'Integrity Analysis : A Methodology for EDP Audit Data Quality Assurance' published in 1984 [Svanks, 1984], an information system could be viewed as a production line in a manufacturing environment. Processing stage represents logic steps which utilize input transactions as raw material or parts to yield processed data base records, i.e. information, as the end product. A typical production line incorporates process control, but more importantly, also employs product control. The identification of faulty processes alerts product quality control to invoke special procedures such as tightened inspection, repair or discarding of finished goods. Conversely, the disclosure of substandard products suggests remedial action for specific processes.

Information System testing then becomes the equivalent of process quality control in that the errors revealed call for software revisions and maintenance. However, if the information system is already operative, a test result indicating error only suggests that such error might have occurred in the past. As test procedures do not access the production data base, no statement can be made as to

whether or not the information system error has occurred in real environment, nor which records have been affected by the erroneous process. And, therefore, the confirmation of potential or suspected anomalies on a live data base and subsequent integrity improvement becomes an essential facility (beyond application controls) within an information system.

7 NEED FOR AUTOMATIC FEEDBACK CONTROL SYSTEM FOR ON-LINE ERROR DETECTION AND INTEGRITY IMPROVEMENT IN INFORMATION SYSTEMS

In the information systems, the need is then to remove errors that are made but not corrected. When abstracted this implies, in the Information System Model in Figure 1, given that data/information is represented by a triple $\langle e, a, v \rangle$ and considering a particular example where say an output, i.e., processed data, i.e., Information is represented by entity class, namely, employees and where specific entity (e) under consideration is an employee by name Albert and where specific attribute (a) under consideration is Albert's Salary, then, by virtue of on-line errors present in the information system, at any time, there exists a possibility of information item on value (v) of Albert's salary being inaccurate, inconsistent or unreliable, i.e. it's being affected by error or say corrupted by noise, and, therefore, a more realistic representation of value (v) is $(v + \eta)$, where η represents noise or error component [Mandke and Nayar, 1997].

It is within this framework of error implications on data/information model wherein triple $\langle e, a, v \rangle$ is replaced by triple $\langle e, a, v + \eta \rangle$ and, as discussed in Section 2, considering that these error implications are present at each stage of an information system; namely, data origin stage, communication channel prior to processing stage, processing stage, communication channel at post-processing stage and output stage, a modified version of a conceptual schematic of an Information System Model in Figure 1 emerges, accounting for errors that are made but not corrected. The same is given in Figure 4.

It is these on-line errors that are then to be removed. For this purpose, one has to first detect errors and then correct them. In other words, what is required is to incorporate on-line learning and error correcting mechanisms in the Information System Model. Specifically, this calls for automatic feedback control systems with error detection and correcting technologies for improved information accuracy, consistency and reliability; technologies that maximize integrity of information systems – Information Integrity Technologies.

Figure 4 Modified Conceptual Presentation of an Information System Model in Figure 2 accounting for errors that are made but not corrected.

8 INFORMATION INTEGRITY TECHNOLOGY DESIGN

Rajaraman [Rajaraman, 1996] points out that integrity of the overall information system as in Figure 1 is ensured if the integrity of all parts of the system are ensured (see Figure 5).

Figure 5 Conceptual Presentation of Integrity of an Information System.

As a result, each stage and its intermediate stages in the information system as discussed in Section 2, along with the overall information system (with data as input and with processed data, i.e., information for use as output) considered as a black box, become the candidates for incorporating automatic feedback control systems or application and user specific software products constituting Information Integrity Technologies as above.

What Information Integrity Technologies will have to do is to follow data as it originates, moves over communication channel at pre-processing stage and gets processed, and follow processed data, i.e., information as it moves over communication channel at post-processing stage and gets used by user (as information system output), so as to detect where the error(s) occurs. Based on this, the Information Integrity Technologies would then need to take corrective action to remove error(s), so as to improve integrity of information system stage(s) under consideration and of the overall information system. In doing so, it would also be helpful if it is possible to measure integrity of a stage in the information system and of the overall information system, thereby offering a measure of integrity improvement achieved and a statement of improved level of integrity. Most importantly, such detection and correction of error(s) and improvement of integrity along with measurement of improvement and statement of integrity level achieved, will facilitate demonstration of integrity of computerized information systems rather than mere trusting them in the context [Mandke and Nayar, 1997].

However, apart from difficulties that are obvious in designing and developing such automatic feedback control systems leading to Information Integrity Technologies and in measuring integrity (i.e., accuracy, consistency and reliability) of information, the most important difficulty in perceiving such technologies is that it is impractical to follow, i.e., track and analyze every bit of data/information for all times as it flows through the information system stages.

Way out here is to consider Information Integrity Technology that takes a sample of input data at the output or at an intermediate point of an appropriately identified stage or sub-system of the information system and then follows or keeps track of the sampled records at output or intermediate points of subsequent stages (sub-systems), at a given point of time or at different points of time over a required time interval. Records so obtained at a given stage (sub-system) could then facilitate study of patterns of errors at that stage (sub-system) in the information system.

Utilizing the records available for different points of time over the time interval, error patterns can be studied even for a time variation. Based on patterns of errors

for a given stage (subsystem) so analyzed, causes of errors in the sampled data/information can be known so as to obtain corrective action (probabilistic mechanisms included) to eliminate the causes, remove the error and improve the integrity. In other words, Information Integrity Technology could comprise sampled data control system for the stage (subsystem) under the information system as also for the overall information system.

Figure 6 presents a block diagrammatic view of the Information Integrity Technology Design thus emerging.

Literature extensively refers to requirement of security in information systems. However, security meaning concern with controlling dissemination of information (confidentiality) is different from integrity of information having concern for accuracy, consistency and reliability of information in a computer system – requiring control over modifications made to information. As a result security and integrity require separate mechanisms. Further, the concept of ‘Trusted computing base and procedures’ implying an information system can be trusted over time without ability to provide the evidence that the trust is well placed is incompatible with internal control principles and therefore insufficient for the question of error free computerized information system. Also with networks becoming integral to information systems, there is need to recognize presence of ‘noise’ in data/information models; thereby conceiving probabilistic descriptions of information flow models and of on-line Integrity Improvement Mechanisms [Mandke and Nayar, 1997].

9 CONCLUSION

Errors in computerized information systems were relatively manageable as long as there was homogeneous system environment and centralized control over information. Emerging trends of globalization, changing organizational patterns, strategic partnering, electronic commerce and distributed computing have changed all this, posing risks to accuracy, consistency and reliability of information.

These intrinsic information integrity attributes are central to any information system in that in their absence the information systems will have massive amounts of polluted (error-filled) data and useless, even dangerous information.

These errors are essentially caused by on-line factors of change, complexity, communication, conversion and corruption which have their presence mainly through system environment which is external to computing (and hence the application) system and overlaps with the user environment. In spite of application

Figure 6 Conceptual schematic giving system's view for Information Integrity Technology Design

controls, it is these external factors that then introduce, in information systems, errors that are made but not corrected.

Need therefore is to design and develop automatic feedback control systems leading to application and user specific software products representing Information Integrity Technologies that (a) carry out on-line detection (filtering problem), estimation and prediction of errors contributing to loss of accuracy and consistency and of causes contributing to loss of reliability, and (b) implement integrity improvement action plan (probabilistic mechanism included) accordingly. It is such Information Integrity Technologies that could then also provide a measure (metric) of Information Integrity achieved, in turn facilitating demonstration of integrity of computerized information systems rather than merely trusting them in the context.

10 REFERENCES

- Annon (May 1987) Data view : loss leaders, *Computerworld*.
- Banks, W. and Weimer, J. (1992) *Human Factors in Computer Systems*, Prentice Hall Publishers, NY.
- Bloombecker, B. (1989) *Commitment to Security*, National Centre for Computer Crime Data, Santa Cruz, California.
- Delone, W.H. and McLean, E.R. (1992) Information Systems Success : The Quest for the Dependent Variable, *Information Systems Research* 3(1).
- Hussain Donna, and Hussain, K.M. (1984) *Information Resource Management*, Richard D. Irwin, Inc., Homewood, Illinois.
- Kliem, R.L. (October 1992) Back to Basics : Developing a Good Requirements Document, *Journal of Systems Management*, 16–19.
- Mandke Vijay, V. (1996) Research in Information Integrity : A Survey and Analysis, Proceedings of the JNCASR and SERC Discussion Meeting at IISc Campus, Bangalore on *Information Integrity – Issues and Approaches* (ed. Rajaraman V. and Mandke Vijay V.), published by Information Integrity Foundation, New Delhi, India.
- Mandke Vijay, V. and Nayar, M.K. (1997) *Information Integrity – A Structure for its Definition*, Proceedings of the 1997 Conference on Information Quality (ed. Strong Diane, M. and Kahn Beverly, K.), Cambridge, Massachusetts, USA.

- Menkns Belden, (1990) Understanding Data Communication Security Vulnerabilities, *Computer and Security*, **9**, 209 – 213.
- Mostert, D.N.J. and Solms, S.H. Von (1994) A Methodology to include Computer Security, Safety and Resilience Requirements as part of the user requirements, *Computers and Security*, **13**, 349 – 364.
- Nayar, M.K. (1996) A Framework for Achieving Information Integrity, Proceedings of the JNCASR and SERC Discussion Meeting at IISc Campus, Bangalore on *Information Integrity – Issues and Approaches* (ed. Rajaraman V. and Mandke Vijay V.), published by Information Integrity Foundation, New Delhi, India.
- Rajaraman V. (1996) Information Integrity – An Overview, Proceedings of the JNCASR and SERC Disucssion Meeting at IISc Campus, Bangalore on *Information Integrity – Issues and Approaches*, (ed. Rajaraman V. and Mandke Vijay V.), published by Information Integrity Foundation, New Delhi, India.
- Redman T.C. (1992) Data Quality : Management and Technology, Bantam Books, NY.
- Svanks Maija, I. (1984) Integrity Analysis : A Methodology for EDP Audit and Data Quality Assurance, EDP Auditors Foundation, Inc.
- Tompkins, F.G. and Rice, R. (1996) Integrating Security Activities into the Software Development Life Cycle and the Software Quality Assurance Process, *Computers and Security*, **5**, 218–242.
- Wood, C.C. and Banks, W.W. Jr (1993) Human Error : An Overlooked but Significant Information Security Problem, *Computer and Security*, **12**, 51–60.