FUTURE TASKING, PROCESSING, EXPLOITATION AND DISSEMINATION (TPED) DOMAIN, REDEFINING THE NEED

By

Richard L Landis

A MASTER OF ENGINEERING REPORT

Submitted to the College of Engineering Texas Tech University in partial fulfillment of the requirements for the degree of

MASTER OF ENGINEERING

Approved

____

Dr J Smith

____

Dr A Ertas

____

Dr T Maxwell

____

Dr M Tanik

June 29, 2001

ACKNOWLEDGMENTS

The author wishes to thank Dr. Tim Maxwell for providing an opportunity to exercise some of the key methods, tools and concepts presented in his creative thinking and generic design science course that sharpens the skills in a design engineers tool set; Dr. Ertas for coordinating and setting up the off campus Texas Tech Master's program that allowed so many fine engineers an opportunity to pursue a Master's of Engineering; Raytheon for giving their employees a way to better themselves; and finally, my family for understanding why I was not at home until late some evenings and not at home at all some weekends. Additional acknowledgement must be given to my fellow colleagues and friends who have provided help and advice at critical times during the development of this final report. Specific colleagues by name are Dr David Race for his help in hardware technology forecasting, Charles Sicking for his help in working out problems in the EXCEL spread sheet model used to develop some of the results, David and Quinta Seida, Richard Koshack and Steve Reynolds for their time spent in reading and providing questions and comments that helped edit and expand critical topics of this report.

TABLE OF CONTENTS

ABSTRACT

This paper is written as the final fulfillment of the requirements for a degree in Masters of Engineering a Texas Tech. The objective of this paper is to redefine the current Tasking, Processing, Exploitation, and Dissemination (TPED) concept, identify a new conceptual design and the performance needs to support this new concept, provide a technology forecast assessment, and then determine if projected technology will be available to meet this need by 2008. The paper describes the origins, background, and major parts of the current TPED domain, and the future changes facing the national intelligence community. Once the current TPED concept and future challenges are established, a suggested new TPED system concept to meet this change is introduced. From this new TPED concept, a conceptual design and performance needs are established. A sizing assessment of the new TPED design is conducted. Upon establishing the new TPED sizing need, a commercial technology trend forecast is performed and compared to the new TPED sizing and performance needed to determine feasibility of creating such a system within the 2008 timeframe..

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

ADCI - Assistant Director of Central Intelligence

Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) - A coordinated approach, called a framework, for Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance system architecture development, presentation, and integration.

Clusters – The set of dimensions that are grouped such that the relationship among dimensions are the design interdependencies.

COTS – Commercial-off-the-Shelf

Data object - Any piece of uniquely identifiable foundation data or mission specific data. Data objects are stored in electronic libraries and registered wherever possible to the Foundation Data grid (i.e., WGS-84). Other data associated with data objects includes minimum attributes of length, width, height, time and other information (e.g., when and where the object was created, first-observed, last-observed, etc).

Design Synthesis - A process that organizes functions or objects into a cohesive design and verifies that the results satisfy the requirements.

Dimensions - A set of thoughts, concepts, options, or other elements that are related, to which the relationship between the items in the sets supports the needs of a target design

Dissemination (Current TPED) - The process of getting the right information to the right place, at the right time.

Dissemination (Multi-Source TPED) - The sharing of information through collaborative communications

Distribution - The physical dissemination process that distributes data to where it belongs.

Exploitation (Current TPED) - All those value-adding activities that transform imagery into intelligence or, more generally, the link in the chain that transforms “information” into “knowledge”.

Exploitation (Multi-Source TPED) - the function of people seeing patterns in information that is normally associated with information analysis

Foundation data - Consists of controlled and orthorectified imagery, digital elevation, bathymetry, vector features including air and navigation safety, and other globally maintained data such as gravity and magnetics. This data is collected near worldwide, independent of missions, is relatively stable and tied to a common geometry.

Geospatial Information System (**GIS**)

Image Exploitation Support System (IESS) – A custom set of software that was created to manage and maintain exploitation activities and resources, and provide history of coverage, reporting, and work tasks.

Imagery Intelligence (IMINT) - Intelligence derived from visible or radar imagery. Examples include, RadarSAT, LandSAT, IKONOS, APF-70, etc.

Information generation support data - All data required by USIGS users to generate the imagery, imagery intelligence, and geospatial information of interest to USIGS customers. It includes, but is not limited to, collection sensor calibration data, models (e.g., Digital Elevation Models, surface runoff and erosion models, flooding and stream loading models), exploitation / mensuration support data, sensor-specific processing software modules, etc.

Information View - A presentation of foundation data, mission specific data, and intelligence using object data in a manner tailored to meet a customer's information need.

Information Need - A recognized gap in a decision-maker's / customer's / user's knowledge or information or data holdings.

Information Requirements (Current TPED) **-** A statement of all or a portion of an information need in a form that can be allocated to a component of an intelligence discipline for action. Information requirements include collection, processing, exploitation, search, retrieval, storage, and delivery requirements.

Information Requirements (Multi-Source TPED) - A statement of all information needs associated with a problem or question in a form that translates into intelligence source or knowledge discovery discipline for action. Information requirements includes questions to be answered, possible geospatial locations involved, type of activity being looked for, and any ontology relationships.

Interpretive Structural Modeling (ISM) – A generic design method that is used to document design options, organize the options into dimensions, and use these dimensions to identify clusters.

Joint Task Force (JTF)

Measurement and Signatures Intelligence (MASINT) - Technically derived intelligence that detects, locates, tracks, identifies, and describes the unique characteristics of fixed and dynamic target sources.

Mission Specific Data – Data consisting of intensified foundation data encompassing greater detail or additional features and / or attributes (information and / or intelligence) to meet mission requirements.

National Imagery and Mapping Agency (NIMA**)**

National Technical Means (NTM) – The intelligence collection capabilities available to the United States of America.

Options – The first level in the Triply Structure Quad and is the identified design alternatives to meet a target design.

Processing (Current TPED) The automated, rote application of algorithms that transform raw collection take into a product better suited for exploitation by a diverse set of analysts and for a diverse set of

Processing (Multi-Source TPED) Processing is the link in the chain that transforms "data" into "information" accessible to human analysts.

Requirements Analysis - The stage in the Systems Engineering Process by which system functional and performance requirements are identified and documented.

Requirements Engineering – The process that determines the requirements necessary to satisfy the need.

Systems Engineering (SE)

Secondary Image Dissemination (SID) – A subset derived image created from an image obtained from national technical means that is just a displayable softcopy non-manipulatable picture with annotations of items of interest that can be released to a lower classification level than which it was taken.

Signals Intelligence (SIGINT) – The interception of transmissions from broadcast communications systems such as radios, as well as radars and other electronic systems. SIGINT consists of several categories. Communications intelligence (COMINT) is directed at the analysis of the source and content of message traffic. While most military communications are protected by encryption techniques, computer processing can be used to decrypt some traffic, and additional intelligence can be derived from analysis of patterns of transmissions over time. Electronic intelligence (ELINT) is analysis of non-communications electronic transmissions. This would include telemetry from missile tests (TELINT) or radar transmitters (RADINT).

Space Reconnaissance (SR) Functional Reference Model (FRM) - A C4ISR representation of the top-level functional areas necessary to acquire, process, and provide access to reconnaissance information.

Tasking (Current TPED) - The value-adding process by which we try to ensure that the right data is gathered, at the right time. If collection capacity is a scarce resource, then tasking includes the optimization of that scarcity.

Tasking (future) – A process derived from a model based on abundance where information discovery is the issue.

Theory of Relations - The application of human reasoning applied through any of five relationships. The five relationships are: relations through inference (perceptual judgement, deductive, and ampliative), relations through modes of definitions (naming, extensions, intention, relationships), interpretive relations (definitive, comparative, influence, temporal, spatial, mathematical), mathematical relations (sets, tautologies), and structural relations (hierarchical, cycles).

United States Imagery and Geospatial Information System (USIGS)

Virtual Private Network (VPN)

World Geodetic System (**WGS)** the standard by which points on earth are measured in real space (the current standard is WGS-84)

Extensible hypertext markup language (XML**)**

XML query language (XQL) a proposed language by which queries can be made against material marked up by the tags specified in the XML standard.

CHAPTER I

INTRODUCTION

1.1 Project Description

There is no single systems design, in the strictest sense, for a Tasking, Process, Exploitation, and Dissemination (TPED) system [NIMA, 2000]. By some people's perspective, TPED is a "system of systems" but even that construct is misleading. TPED does embrace a concept of operations from which one may infer certain design concepts and one can substitute design concepts to modify the TPED system. The TPED is a complex (Class C) technological system as define by John Warfield, in "Science of Generic Design".

This paper describes the current TPED system as a technological system and reviews the current TPED domain and its generic functions and components. Using methods defined in "Science of Generic Design"[Warfield, 1994], a new multi-source TPED concept is discovered to support our country's need for information superiority in time of crisis. This process develops the significant new concepts that are used to define enablers and technologies needed for the newly defined multi-source TPED. Using domain specific design and analyses, the multi-source TPED functions and components are derived and the system conceptual design is established. This Multi-Source TPED conceptual design, is examined to define performance and sizing needs are defined. Finally, a commercial technology trend forecast is performed and compared to the new TPED sizing and performance needs to determine feasibility of implementation.

CHAPTER II

LITERATURE REVIEW

2.1 Background

As first explained in *Joint Vision 2010*, "today's military capabilities must transition to dominant maneuver, precision engagement, focused logistics, and full-dimensional protection". The evolution of these elements over the next two decades is strongly influence by the continued development and proliferation of information technologies. Information superiority—knowing more than enough about an adversary who knows much less than enough—is the key enabler for the practitioners of US diplomatic and economic policy. Geospatial information is nearly always the key to an international engagement, whether on the grand strategic level or at the "tactical" level.

With the advent of commercially available, high-resolution (less than 1-meter) satellite imagery, the United States has lost the exclusivity it once had. These images will be available, as never before, to any potential adversary. While it may be regrettable, it is not possible or desirable to turn back the clock. America's answer must be to use information faster and better. TPED, in all its dimensions, is the key to "faster and better". Our use of multi-source information and derived intelligence must put us "inside the adversary's decision cycle". The importance of TPED processes for information dominance cannot be overstated. Everyone agrees that multi source TPED processes are critical for information dominance; not everyone agrees on how.

2.2 Current Imagery TPED

NIMA has described TPED as a system of systems that will provide the tasking, processing, exploitation, and information dissemination service for all imagery. This includes imagery collected by (theater) airborne assets and by national technical means (NTM), as well as those services provided by commercial imagery entities. Commercial services can range from raw images to value-added products and fully exploited information. Programmatically, the imagery TPED includes all the people, hardware, software, communications and "O&M" for the entire Imagery and Geospatial Community (IGC) from the "national" level down to the theater joint task force (JTF)/component level.

2.2.1 Tasking

Tasking is the value-adding process by which the right imagery gets collected at the right time. Since collection capacity is a scarce resource, tasking includes optimization of the scarce collection resources. Today, technical insight into specific collection systems is necessary to accomplish good tasking. Consequently, a corps of trained intermediaries—who mediate between the information needs of intelligence consumers and the tasking of collection systems—are a critical part of the TPED process.

2.2.2 Processing

Processing is the automated rote applications of algorithms, that transform raw collection into a product better suited for exploitation by a diverse set of analysts and for a diverse set of purposes. There is a continuum between collection, processing, and exploitation. The collector can have embedded and/or "on-board" processing. On the other hand, processing can be at a "down-link" site. In any case, there usually are heavy computing demands and consequent economies of scale in processing, as well as, a requirement for intimate technical knowledge of the collector. For these reasons, processing is more closely tied to collection than to exploitation, both in systems design and in organizational responsibility. Because the processing "system" has as its input a well-defined collection system specification, and because it controls explicitly its output specifications, it is arguably the easiest function of TPED to architect. Said differently, it largely is isolated from the vagaries of human interaction—"free will" being the archenemy of system architecture.

2.2.3 Exploitation

Exploitation is the most abstract of the concepts and the easiest of the TPED functions to define. Exploitation comprises all those value-adding activities that transform imagery into intelligence, or the link in the chain that transforms "data" into "information". Because there are still an infinite number and variety of exploitation algorithms yet to be discovered, one is challenged to devise a meaningful exploitation architecture.

Figure 1 shows the seven step process today's image analyst uses for imagery exploitation and reporting. It must be noted that today's process requires two individuals, an Analyst and a Supervisor. Even though it is not a clean break between the analyst and the supervisor steps, one can separate the steps such that the analyst performs steps 1 through 5 and the supervisor performs steps 6 and 7. Currently, part of step 1 is being automated, however, the remaining steps are manual. An explanation of each step follows.

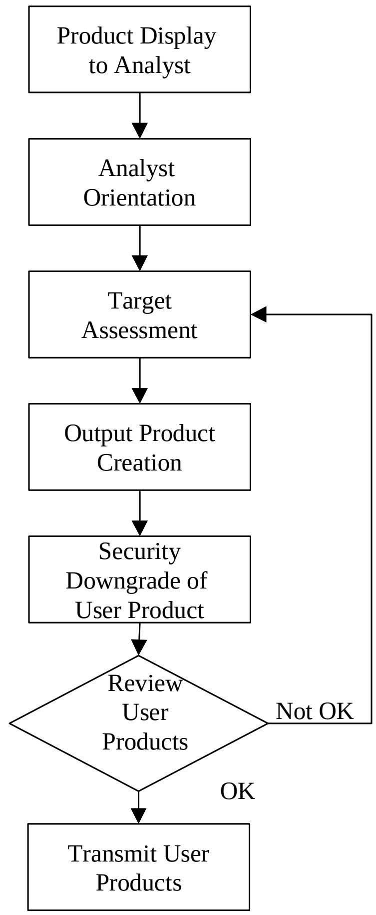


Figure 1 Current Image Analyst Process

1. Product Display to Analyst - Under current USIGS "pull concept", this step involves the operator selecting from a list of images needing to be worked, identifying the targets that are contained on the selected image from Image Exploitation Support System (IESS), recovering the target specifications and reporting requirements for the target that is being assessed, and transferring the mission image and any reference data to the workstation to be displayed.

2. Analyst Orientation - This step consists of the Analyst finding the target location on the mission image. Once the target location is found, the Analyst orients, adjusts and enhances the image at the target location. The Analyst reads the reporting requirements contained within IESS for the target being assessed. It must be noted that mission image orientation, adjustment and enhancement at the target location may fall either under Analyst Orientation or Target Assessment.

3. Target Assessment - This step consists of the Analyst assessing whether or not the user requirement can be satisfied and retrieving any usable intelligence against the target.

4. Output Product Creation - This involves the Analyst going into IESS Image Interpretation Report (IIR) Create and filling in all mandatory fields in the report item form. If it is requested

within the user request to create an associated Secondary Image Dissemination (SID) product for that target, then the analyst will create a SID product.

5. Security Downgrade - If a SID product is created, the analyst must check the classification of the mission image and if it meets certain downgrade criteria, then he can reclassify the SID product. Once a report item or a SID product is created by the analyst, then it is placed into a review queue for a supervisor. The Supervisor downgrades report items as a function of the review process.

6. Review User Product - This step consists of report item downgrade and review, and second person verification of SID products. Report item downgrade consist of the supervisor selecting the item for downgrade then having IESS reprocess that report item against the collateral project file. This processing removes the time over target and replaces other sensitive information fields on the image data line of the intelligence report item. The supervisor reviews the report item a second time to ensure that the image data line of the report item was modified correctly. SID product review consists of the supervisor reviewing the image and all normally non-displayable header fields of the SID product and verifying that no unauthorized data is contained within the image or the non-displayable fields.

7. Transmit User Product - This step consists of assembly of multiple report items into a report, then releasing of the report and SID product on the appropriate communications circuit depending on classification of products. Within IESS, the supervisor selects the report items that have been downgraded and reviewed, adds an addressee list for distribution, and then directs IESS to assemble the report items into a standard report format. IESS assembles the report items into a report, then places a standard message format around the report. The supervisor reviews the message and selects it for transmit.

2.2.4 Dissemination

Dissemination is thought of as getting the right information to the right place, at the right time. It is sometimes useful to decompose dissemination into two parts: the physical process of getting it there, “distribution”; and the logical process of deciding, “what goes where.” Of the two, the distribution historically appears to be the more expensive and difficult, and the most boring. The logical process of dissemination is by far the more intellectually challenging.

2.2 The Space Reconnaissance TPED

The Space Reconnaissance (SR) Functional Reference Model (FRM) is a representation of the top-level functional areas necessary to acquire, process, and provide access to reconnaissance information (currently IMINT, SIGINT, and MASINT) collected with space-borne assets. This model shows larger

TPED made up of multi-sensor collection information. The model accounts for functionality and interfaces associated with Satellite Operations, Space/Ground Communications, and Ground Operations. Table 1 through Table 3 lists the components for the space collectors, space and ground communications, and the ground parts of the space component of a larger TPED system.

Table 1 Space Collection Platform Components by Category

Sensor	Relay	Platform Infrastructure
Target Acquisition	Satellite Acquisition	Power Management
Relay Contact	Antenna Control	Momentum Management
Crosslink Communications	Frequency Management	Navigation and Guidance
Data Processing	Crosslink Communications	Attitude Control
Data Compression		Command Execution
Data Formatting		Health and Status Monitoring
Error Protection		Fault Management
Synchronization		Telemetry Generation
		Payload Interface

Table 2 Air and Ground Communications Components by Category

Satellite Narrowband Comm	Ground Narrowband Comm	Ground Wideband Comm
Frequency Management	Frequency Management	Frequency Management
Antenna Control	Antenna Control	Antenna Control
Data Rate Control	Data Rate Control	Data Rate Control
Data Transmit/Receive	Data Transmit/Receive	Data Receive
Modulation/Demodulation	Modulation/Demodulation	Demodulation
Tracking Support	Tracking Support	Data Formatting
Data Formatting	Data Formatting	Decryption
Encryption/Decryption	Encryption/Decryption	Data Alignment
Command Data Uplink	Command Data Uplink	Signal Conditioning
Telemetry Data Downlink	Telemetry Data Downlink	Synchronization
Range Data Uplink/ Downlink	Ranging Support	Error Detection & Correction
Satellite Wideband Comm		
Frequency Management		
Antenna Control		
Data Rate Control		
Data Transmit		
Modulation		
Data Formatting		
Encryption		

Table 3 Ground Components by Category

Satellite C & C	Mission Management	Product Dissemination
Command Generation	Resource Management	Final Processing
Telemetry Analysis	Status Collection	Formatting /Compression
Range Determination	Status Reporting	Electronic Dissemination
Ephemeris Generation	Mission Planning	Media Generation
Status Assessment	Performance Assessment	Physical Delivery
Other Collection Systems	Archive Maintenance	Quality Assurance
Airborne Collection	**Simulation & Forecasting**	Security Control
Ground-based Collection	Collection Window Simulation	**Analysis & Reporting**
Other Intelligence	Long Term Forecasting	Product Analysis
Commercial Collection	Relay Contact Determination	Information Extraction
Requirements Management	Collection Feasibility	Report Generation
Requirement Analysis	Target Weather Prediction	Report Distribution
Requirement Allocation	**Data Processing**	**Ground Infrastructure**
Requirements Tracking	Data Conditioning	1ntra-facility Data Routing
Resource Scheduling	Signal Reconstruction	Inter-facility Communications
Satisfaction Assessment	Quality Enhancement	Computing
	Support Data Generation	Data Storage
		Facility Services
		Human Interface
		Display Services
		Collaboration Tools
		Security Services
		Bandwidth Allocation

2.3 A New Multi-Source TPED – Any Source

From a macro view of information superiority, TPED is more than just imagery or multi-source space reconnaissance data. It is the integrated information collection and discovery of all critical knowledge provided to the practitioners of US diplomatic and economic policy at the right time to influence an outcome to a national or international issue. In reality, knowledge is all information about a topic, issue, or problem. What are the sources of knowledge? The sources of knowledge are:

1. Books, libraries, archives, data repositories, and local, national, and international demographic data stores.
2. Television, movies, archives, and Internet.
3. Geospatial intelligence reports and intelligence data from National Technical Means (NTM) and non-NTM (e.g. SIGINT, MASINT, and IMINT data from satellite, airborn and other collection means).

C H A P T E R I I I

CREATING A MULTI-SOURCE TPED CONCEPT

3.1 Applying Generic Design Methodology to the Problem

To assess multi-source TPED concepts and approaches, one can use selected Generic Design Science methods. The process and sequence of use are:

1. Identify the significant design options of the system.
2. Structure, categorize, and identify the design dimensions and clusters.
3. Provide sequencing of the dimensions within clusters and clusters within the system.

The result displays the design in an Option Field Representation and a TPED Delta Chart.

Using the following steps, one can define the new multi-source TPED concepts:

1. Establish the TPED system context, objectives, options, and criteria.
8. Use creative exercises to obtain an exhaustive list of multi-source TPED concept and approach options
9. Use Interpretive Structural Modeling (ISM) methodology to initially structure the Multi-Source TPED concept options into categories.
10. Name the categories, then distill out the multi-source TPED concept dimensions from the group of categories by syntheses, combining, and excluding options and categories.
11. Use the same ISM methodology used in step 3 above, discover and define the clusters by grouping and combining dimensions.
12. Establish a sequence flow of clusters using ISM.
13. Logically order the sequence of dimensions within clusters.
14. Display the completed design in a Option Field Representation.
15. Use Delta Chart representation to present the sequential flow of clusters that make up the multi-source TPED.

3.2 Establishing the Multi-Source TPED context, objectives, options, and criteria

To define a new multi-source TPED concept, generic design principles are applied to structured decision-making to assist in recognizing and analyzing the basic parts of decisions. The four basics of decision-making are:

1. **Context** - The context describes the situation surrounding the decision. The situation is the need for more encompassing TPED concepts to support rapid decision-making that draws knowledge from all potential sources of information.
2. **Objectives** - A clear understanding of problem centric outcomes that must guide decision-making and make it easier to deduce alternatives. The new Multi-Source TPED concept objective must establish a list of alternatives to support rapid administrative decisions based on knowledge discovery from any means.
3. **Options** - Significant effort must be spent uncovering all available options, studying how each may be implemented and what results they will produce. The concept options of a new Multi-Source TPED must be in addition to those options provided by the traditional TPED (e.g. imagery TPED, Space Reconnaissance (SR)Functional Reference Model (FRM)).
4. **Criteria** - The criteria used to select the best possible concept options are determined by the context and objectives. Hard criteria are conditions that must be satisfied in order to have a useful design decision, such as budget or time constraints. Soft criteria are conditions that require subjective assessment, and therefore can be more difficult to apply. The selection criteria for the new Multi-Source TPED are:

   - Rapid decision-making from the Intelligence Community (IC) in terms of minutes or hours vice days is available to support all US Government agencies for various types of resolution management activities.
   - The concepts needed to support a new Multi-Source TPED must be available by 2008 to allow implementation by 2010.

3.3 Applying Creative Thinking Methods to Identify Options

Typically, designers are asked to brainstorm to create a list of ideas. Recording of the creative ideas can be done by simply writing them on a piece of paper or transmitting them on a wireless, hand-held, numeric keypad so their individual responses can be recorded. Special effort to identify all options must be made. All options must be considered even if impossible, wild, or unrealistic, as these may lead to outstanding “out of the box” solutions. Finding ways to uncover new options that may not be obvious is a challenge. To encourage creativity, all options must be listed. Once all available options are listed, each option should be clarified and the most promising ones selected for further analysis.

Using imagination by word play [Plsek 1997,] as a guide, I used three manipulative verbs from Osborn’s (1953) checklist of manipulative verbs (minify, combine, and reverse) to create new conceptual ideas for a new Multi-Source TPED system concept. Table 4 through Table 6 contains the new conceptual ideas that I generated.

Table 4 Multi-Source Tasking Concepts through Creativity Tools

Stimulus Word	Conceptual Tasking Idea
minify	Make it easy for a policy or decision maker to ask for needed answers or obtain needs resources to obtain answers concerning national or international issues.
combine	Build on the idea of rapidly requesting and receiving timely answers, develop a method between the policy or decision making staff and the resources planners, and data collection means that seamlessly handle the generation and delivery of answers to the policy or decision maker's questions.
reverse	Think about how policy makers can answer their own questions. We can set up processes and tools to directly task knowledge discovery and collection resources in ways to provide graphical answers to policy or decision maker's questions. Then, facilitate results by providing process and tool experts to work with decision-making staffs.

Table 5 Multi-Source Processing Concepts through Creativity Tools

Stimulus Word	Conceptual Processing Ideas
minify	Make it easy for a policy or decision maker to obtain "what if" answers to an issue based on the decision-maker's selection alternative to an initial question through rapid alternative processing.
combine	Build on the idea of rapidly providing the needed knowledge just in time, develop a method between the policy or decision making staff and the processing system analyzing related data that could seamlessly handle the generation and delivery of answers to the policy or decision maker's questions.
reverse	Think about how policy makers can answer their own questions. We can set up processes and tools to provide, on demand, graphical answers to policy or decision maker's questions through data processing. Then, facilitate results by providing access to process and tool experts to work with decision-making staffs.

Table 6 Multi-Source Exploitation & Dissemination Concepts through Creativity Tools

Stimulus Word	Conceptual Exploitation and Dissemination Idea
minify	Simplify retrieval of needed information concerning an issue by policy makers directly from graphically displayable processed information on demand.
combine	Build on the idea of providing the required knowledge when it is needed, develop a means to seamlessly handle the delivery of the needed information to the policy maker when considered necessary.
reverse	Think about how policy makers can do their own analysis work. We can set up knowledge mining, discovery processes, and tools. Then, provide training programs to teach practical intelligence gathering and correlation. Finally, provide direction of these people. The policy maker would be helping them develop their policy staff.

Once the Multi-Source conceptual ideas were generated, each idea was analyzed and concepts needed to support these ideas were created. Table 7 contains the created concepts for a new Multi-Source TPED.

Table 7 Approaches and Technologies based on Multi-Source Conceptual Ideas

Conceptual Idea	Concept, Approach and Technology Options
Make it easy for a policy or decision maker to ask for needed answers or obtains needs resources to obtain answers concerning national or international issues.	Provide Multi source collection awareness capability
	Direct tasking of intelligence resources by policy or decision-maker (reduce cycle)
	Direct tasking of data mining and knowledge discovery resources
Build on the idea of rapidly requesting and receiving timely answers, develop a method between the policy or decision making staff and the resources planners, and data collection means that seamlessly handle the generation and delivery of answers to the policy or decision maker's questions.	Collaboration tools between policy/decision maker and resource planners
	Create common collection planning awareness that allows policy/decision makers to be automatically added for distribution to a currently planned resource collection in minutes
	Provide policy/decision-maker a multi-media shared area, direct interaction with data mining specialist and intelligence analysts supporting the policy issue or problem, and provide configurable user portals for rapid access to needed resources and information
	Provide virtual reality-meeting places where data mining specialist, intelligence analyst and policy/decision makers and staff can meet and work in near real time.
Think about how policy makers can answer their own questions. We can set up processes and tools to directly task knowledge discovery and collection resources in ways to provide graphical answers to policy or decision maker's questions. Then, facilitate results by providing process and tool experts to work with decision-making staffs.	Provide awareness of any source data collection from data mining specialist and traditional intelligence collection
	Provide facilitators to aid policy/decision makers in effectively using collaboration tools and resources
	Setup knowledge discovery through Enterprise Data Mining as an intelligence discipline
Make it easy for a policy or decision maker to obtain "what if" answers to an issue based on the decision-maker's selection alternative to an initial question through rapid alternative processing.	Provide for near real time reprocessing of tailored knowledge composite outputs based on alternative need selection by the policy/decision maker mining
	Provide for near real time reprocessing of tailored multi-source intelligence composite outputs based on alternative need selection by the policy/decision maker mining
	Provide for automatic or semi-automatic combining of tailored knowledge and multi-source intelligence composites outputs for the policy/decision maker

Table 7 Approaches and Technologies based on Multi-Source conceptual Ideas (continued)

Conceptual Idea	Concept, Approach and Technology Options
Build on the idea of rapidly providing the needed knowledge just in time, develop a method between the policy or decision making staff and the processing system analyzing related data that could seamlessly handle the generation and delivery of answers to the policy or decision maker's questions.	Provide for auto extraction of feature data, object detection, elevation, location, candidate identification list, foundation data or other to be determined data, based on profiling of historical need for similar situations
	Provide formatting of data to support automatic co-registration of multi-source data, auto construction of composite feature 3D model or other NIMA defined Information Views
	Product creation based on ontology of problem or issue construct
Think about how policy makers can answer their own questions. We can set up processes and tools to provide, on demand, graphical answers to policy or decision maker's questions through data processing. Then, facilitate results by providing access to process and tool experts to work with decision-making staffs.	Provide creation of event driven tailored product
	Provide policy/decision–maker with problem translation capability that translates their problem, issue and question into knowledge or intelligence resource collection requirements
Simplify retrieval of needed information concerning an issue by policy makers directly from graphically displayable processed information on demand.	Automatically create problem centric views of provided data based on policy/decision maker alternative selections and provide it to shared desktops
	Provide updated information and problem centric data on-demand
	Provide the capability for dynamic changes to problem centric situation to be automatically provided through smart push dissemination
Build on the idea of providing the required knowledge when it is needed, develop a means to seamlessly handle the delivery of the needed information to the policy maker when considered necessary.	Use trusted computer agents or similar technologies to find delta data updates pertinent to the policy/decision maker's question, or issue
	Automatically or semi-automatically generate knowledge and intelligence based composite products
	Create Policy Collaboration Centers with the communication, networking, processing and tools specialist support for problem and issue centric policy/decision-making operations
	Provide problem centric locations configurable on policy/decision-maker's portals
Think about how policy makers can do their own analysis work. We can set up knowledge mining, discovery processes, and tools. Then, provide training programs to teach practical intelligence gathering and correlation. Finally, provide direction of these people. The policy maker would be helping them develop their policy staff.	Use data mining experts to rapidly decompose the search parameters
	Provide profiling of historical need for similar search situations
	Provide collaboration tool specialist to operate the tools and display the results based on policy/decision maker's request for information

3.4 Using ISM for Categorizing Options

After a list of concepts had been generated, each concept was ranked or categorized to enhance focus on select capabilities. Using ISM techniques to organize the options into groupings, the following question was asked:

"In the context of a new Multi-Source TPED system, is concept option A similar to concept option B?"

And

"In the context of a new Multi-Source TPED system, does concept option A belong in the same category as concept option B?"

As part of this effort, each concept option within the relational groups was clarified and redundant concept options were removed. Table 8 shows the results of this allocation effort.

Table 8 Allocation of Multi-Source Concepts into Grouping by Relationships

Group	Concept
Group 1	Direct tasking of intelligence resources by policy or decision-makers
	Direct tasking of data mining and knowledge discovery resources by policy or decision-makers
	Provide updated information and problem centric data on-demand to the policy or decision-makers
	Provide the capability for dynamic changes to problem centric situations to be automatically provided through Smart push dissemination
	Use trusted computer agents or similar technologies to find delta data updates pertinent to the policy/decision maker's question, or issue
Group 2	Provide multi-source collection awareness to the policy or decision-makers
	Create common collection planning awareness that allows policy/decision makers to be automatically added for distribution to a currently planned resource collection in minutes
	Provide awareness of any source data collection from data mining specialist and traditional intelligence collection
Group 3	Provide collaboration tools between policy/decision maker and resource planners
	Provide policy/decision-maker a multi-media shared area, direct interaction with data mining specialist and intelligence analysts supporting the policy issue or problem, and provide configurable user portals for rapid access to needed resources and information
	Provide virtual reality-meeting places where data mining specialist, intelligence analyst and policy/decision makers and staff can meet and work in near real time.
	Create Policy Collaboration Centers with the communication, networking, processing and tools specialist support for problem and issue centric policy/decision-making operations
	Provide problem centric locations configurable on policy/decision-maker's portals
Group 4	Provide facilitators to aid policy/decision makers in effectively using collaboration tools and resources
	Set up knowledge discovery through Enterprise Data Mining as an intelligence disciple
	Provide collaboration tool specialist to operate the tools and display the results based on policy/decision maker's request for information
Group 5	Provide for near real time reprocessing of tailored knowledge composite outputs based on selection of alternative need selection by the policy/decision maker
	Provide for near real time reprocessing of tailored multi-source intelligence composite outputs based on selection of alternative need selection by the policy/decision maker

Table 8 Allocation of Multi-Source Concepts into Grouping by Relationships (continued)

Group	Concept
	Provide for automatic or semi-automatic combining of tailored knowledge and multi-source intelligence composites outputs for the policy/decision maker
	Provide for auto extraction of feature data, object detection, elevation, location, candidate identification list, foundation data or other to be determined data, based on profiling of historical need for similar situations
	Provide formatting of data to support automatic co-registration data multi-source data, auto-construction of composite feature 3D model or other NIMA defined Information Views
	Provide creation of event driven tailored product
	Automatically create problem centric views of provided data based on policy/decision maker alternative selections Shared desktops
	Automatically or semi-automatically generate knowledge and intelligence based composite products
Group 6	Create products based on ontology of problem or issue construct
	Provide policy/decision–maker with problem translation capability that translates their problem, issue and question into knowledge or intelligence resource collection requirements
	Use data mining experts to rapidly decompose the search parameters
	Provide profiling of historical need for similar search situations

3.5 Organizing Relationships through Option Field Representation

To develop profound knowledge about my new multi-source TPED concept, I examined, analyzed, and compared how every idea relates to and affects every other idea. Using the Theory of Relations to structure options, a detailed examination of relationships between essential concepts and elements was performed and a chart was created to insure that all relationships are identified. The next task is to name the categories, and then distill out the Multi-Source TPED concept dimensions from the group of categories by syntheses, combining, and excluding options and categories. Once I named the option categories, a close look at each category was made and a determination made as to if it should be a dimension. I felt that each category played a significant role in the new Multi-Source concept and, therefore, should be considered a dimension.

Figure 2 and Figure 3 show the results of this effort in an Option Field Representation.

Direct Tasking/Dissemination

- Direct tasking of intelligence resources
- Direct tasking of data mining and knowledge discovery resources
- Provide updated information and problem centric data on-demand
- Use trusted computer agents or similar technologies to find delta data updates pertinent to the policy/decision makers question, or issue

Information Awareness

- Provide multi-source collection awareness
- Create common collection planning awareness that allows policy/decision makers to be automatically added for distribution to a currently planned resource collection in minutes
- Provide awareness of any source data collection from data mining specialist and traditional intelligence collection

Collaboration

- Provide collaboration tools between policy/ decision maker and resource planners
- Provide Policy/decision maker a multi-media shared area, direct interaction with data mining specialist and intelligence analyst supporting the policy issue or problem, and provide configurable user portals for rapid access to need resources and information
- Provide virtual reality meeting places where data mining specialist, intelligence analyst and policy/decision makers and staff can meet and work in near real time
- Create Policy Collaboration Centers with the communication, networking, processing and tools specialist support for problem and issue centric policy/decision making operations
- Provide problem centric locations configurable on policy/decision maker's portals

TIE LINE

Figure 2 Option Field Representation of first three Demensions

Experts & Specialists

- **Provide facilitators to aid policy/decision makers in effectively using collaboration tools and resources**
- **Setup knowledge discovery through Enterprise Data Mining as an intelligence discipline**
- **Provide collaboration tool specialists to operate the tools and display the results based on policy/decision maker's request for information**
- **Use trusted computer agents or similar technologies to find delta data updates pertinent to the policy/decision maker's question, or issue**

Tailored Products

- **Provide for near real time reprocessing of tailored knowledge composite outputs based on selection of alternative need selection by the policy/decision maker mining**
- **Provide for near real time reprocessing of tailored multi-source intelligence composite outputs based on selection of alternative need selection by the policy/decision maker mining**
- **Provide for automatic or semi-automatic combining of tailored knowledge and multi-source intelligence composites outputs for the policy/decision maker**
- **Provide for auto extraction of feature data, object detection, elevation, location, candidate identification list, Foundation data or other to be determined data, based on profiling of historical need for similar situations**
- **Provide formatting of data to support automatic co-registration data multi-source data, auto construction of composite feature 3D model or other NIMA defined Information Views**
- **Provide creation of event driven tailored product to automatically create problem centric views of provided data based on policy/decision maker alternative selections Shared desktops**
- **Automatically or semi-automatically generate knowledge and intelligence based composite products**

Problem Auto-translation

- **Product creation based on ontology of problem or issue construct**
- **Provide Policy/decision maker a multi-media shared area, direct interaction with data mining specialist and intelligence analysts supporting the policy issue or problem, and provide configurable user portals for rapid access to needed resources and information**
- **Provide policy/decision maker with problem translation capability that translates their problem, issue and question into knowledge or intelligence resource collection requirements. Use data mining experts to rapidly decompose the search parameters.**
- **Provide profiling of historical need for similar search situations**

TIE LINE

Figure 3 Option Field Representation of second three Demensions

3.6 Grouping Dimensions into Clusters by Interdependencies

To discover the clusters associated with the new multi-source TPED dimensions, a look at the relationships among dimensions was performed. Interdependent dimensions would form multi-source TPED clusters. To do this, a trigger question was used:

"Does dimension A play a similar role in the new Multi-Source system as dimension B?"

Using comparative relationships associated with comparing each dimension, Table 9 was created to group dimensions and identifies clusters

Table 9 Comparative Assessment of each Dimension

Cluster #	Question	Answer
	Does Direct Tasking & Updates play a similar role in the new Multi-Source system as Information Awareness?	No
	Does Direct Tasking & Updates play a similar role in the new Multi-Source system as Collaboration?	No
C-1: Assured Satisfaction	*Does Direct Tasking & Updates play a similar role in the new Multi-Source system as Experts & Specialists?*	Yes
	Does Direct Tasking & Updates play a similar role in the new Multi-Source system as Tailored Products?	No
C-1: Assured Satisfaction	*Does Direct Tasking & Updates play a similar role in the new Multi-Source system as Problem Auto-translator?*	Yes
	Does Information Awareness play a similar role in the new Multi-Source system as Direct Tasking & Updates	No
C-2: Knowledge Awareness	*Does Information Awareness play a similar role in the new Multi-Source system as Collaboration?*	Yes
C-2: Knowledge Awareness	*Does Information Awareness play a similar role in the new Multi-Source system as Experts & Specialists?*	Yes
	Does Information Awareness play a similar role in the new Multi-Source system as Tailored Products?	No
	Does Information Awareness play a similar role in the new Multi-Source system as Problem Auto-translator?	No
	Does Collaboration play a similar role in the new Multi-Source system as Direct Tasking & Updates	No
C-2: Knowledge Awareness	*Does Collaboration play a similar role in the new Multi-Source system Information Awareness?*	Yes
C-2: Knowledge Awareness	*Does Collaboration play a similar role in the new Multi-Source system as Experts & Specialists?*	Yes
	Does Collaboration play a similar role in the new Multi-Source system as Tailored Products?	No
	Does Collaboration play a similar role in the new Multi-Source system as Problem Auto-translator?	No

Table 9 Compartive Assessment of each Demension (continued)

Cluster #	Question	Answer
C-1: Assured Satisfaction	*Does Experts & Specialists play a similar role in the new Multi-Source system as Direct Tasking & Updates*	Yes
C-2: Knowledge Awareness	*Does Experts & Specialists play a similar role in the new Multi-Source system Information Awareness?*	Yes
C-2: Knowledge Awareness	*Does Experts & Specialists play a similar role in the new Multi-Source system as Collaboration?*	Yes
C-3: Dynamic Interpretation	*Does Experts & Specialists play a similar role in the new Multi-Source system as Tailored Products?*	Yes
C-1: Assured Satisfaction	*Does Experts & Specialists play a similar role in the new Multi-Source system as Problem Auto-translator?*	Yes
	Does Tailored Products play a similar role in the new Multi-Source system as Direct Tasking & Updates	No
	Does Tailored Products play a similar role in the new Multi-Source system Information Awareness?	No
	Does Tailored Products play a similar role in the new Multi-Source system as Collaboration?	No
C-3: Dynamic Interpretation	Does Tailored Products play a similar role in the new Multi-Source system as Experts & Specialists?	Yes
	Does Tailored Products play a similar role in the new Multi-Source system as Problem Auto-translator?	No
C-1: Assured Satisfaction	*Does Problem Auto-translator play a similar role in the new Multi-Source system as Direct Tasking & Updates*	Yes
	Does Problem Auto-translator play a similar role in the new Multi-Source system Information Awareness?	No
	Does Problem Auto-translator play a similar role in the new Multi-Source system as Collaboration?	No
C-1: Assured Satisfaction	*Does Problem Auto-translator play a similar role in the new Multi-Source system as Experts & Specialists?*	Yes
	Does Problem Auto-translator play a similar role in the new Multi-Source system as Tailored Products?	No

Using the results from Table 9 above, I created Figure 4 to display the completed design in an Option Field Representation that shows clusters, dimensions, and options. Since Experts & Specialists dimension belonged to each cluster, I concluded that it was an independent dimension that formed a cluster. Further, since Tailor Products dimension only had an association with Experts & Specialists dimension, I concluded that it was also a independent cluster that formed a cluster.

Assured Satisfaction

Direct Tasking/Dissemination

- Direct tasking of intelligence resources
- Direct tasking of data mining and knowledge discovery resources
- Provide updated information and problem centric data on-demand
- Use trusted computer agents or similar technologies to find delta data updates pertinent to the policy/decision makers question, or issue

Problem Auto-translation

- Product creation based on ontology of problem or issue construct
- Provide Policy/decision-maker a multi-media shared area, direct interaction with data mining specialist and intelligence analysts supporting the policy issue or problem , and provide configurable user portals for rapid access to need resources and information
- Provide policy/decision–maker with problem translation capability that translates their problem, issue and question into knowledge or intelligence resource collection requirements Use data mining experts to rapidly decompose the search parameters.
- Provide profiling of historical need for similar search situations

Experts & Specialists

Experts & Specialists

- Provide facilitators to aid policy/decision makers in effectively using collaboration tools and resources
- Setup knowledge discovery through Enterprise Data Mining as an intelligence discipline
- Provide collaboration tool specialists to operate the tools and display the results based on policy/decision maker's request for information
- Use trusted computer agents or similar technologies to find delta data updates pertinent to the policy/decision maker's question, or issue

TIE LINE

Information Awareness

- Provide multi-source collection awareness
- Create common collection planning awareness that allows policy/decision makers to be automatically added for distribution to a currently planned resource collection in minutes
- Provide awareness of any source data collection from data mining specialist and traditional intelligence collection

Collaboration

- Provide collaboration tools between policy/ decision maker and resource planners
- Provide Policy/decision maker a multi-media shared area, direct interaction with data mining specialist and intelligence analyst supporting the policy issue or problem, and provide configurable user portals for rapid access to need resources and information
- Provide virtual reality meeting places where data mining specialist, intelligence analyst and policy/decision makers and staff can meet and work in near real time
- Create Policy Collaboration Centers with the communication, networking, processing and tools specialist support for problem and issue centric policy/decision making operations
- Provide problem centric locations configurable on policy/decision maker's portals

Knowledge Awareness

Tailored Products

- Provide for near real time reprocessing of tailored knowledge composite outputs based on selection of alternative need selection by the policy/decision maker mining
- Provide for near real time reprocessing of tailored multi-source intelligence composite outputs based on selection of alternative need selection by the policy/decision maker mining
- Provide for automatic or semi-automatic combining of tailored knowledge and multi-source intelligence composites outputs for the policy/decision maker
- Provide for auto extraction of feature data, object detection, elevation, location, candidate identification list, foundation data or other to be determined data, based on profiling of historical need for similar situations
- Provide formatting of data to support automatic co-registration data multi-source data, auto construction of composite feature 3D model or other NIMA defined Information Views
- Provide creation of event driven tailored product to automatically create problem centric views of provided data based on policy/decision maker alternative selections Shared desktops
- Automatically or semi-automatically generate knowledge and intelligence based composite products

Dynamic Interpretation

Figure 4 Multi-Source TPED Option Field Representation

3.7 2010 TPED Delta Chart

The final step in the generic design ISM process is to create a Delta Chart representation to present the sequential flow of clusters that make up the multi-source TPED. This is one of the most challenging tasks due to the complexity and flexibility of the multi-source TPED design. At this point, a designer would have to create a data chart for each possible process alternative. I will list some of the situations and create a delta chart for one. From the policy/decision-maker's point of view, the following are some of the situations that may occur, for which a separate delta chart needs to be developed:

1. A national or international crisis occurred in which the policy/decision maker needs detailed information concerning an incident. Examples are: national disaster, invasion or aggression into a country that supplies petroleum to the United States, terrorist attack of United States interest abroad, terrorist attack inside the United States, etc
2. Tracking of smuggling, narcotics trafficking, illegal or racketeering activities
3. Following activity associated with locations where suspicious building or production activities periodically occur

Figure 5 shows a delta chart for the first situation:

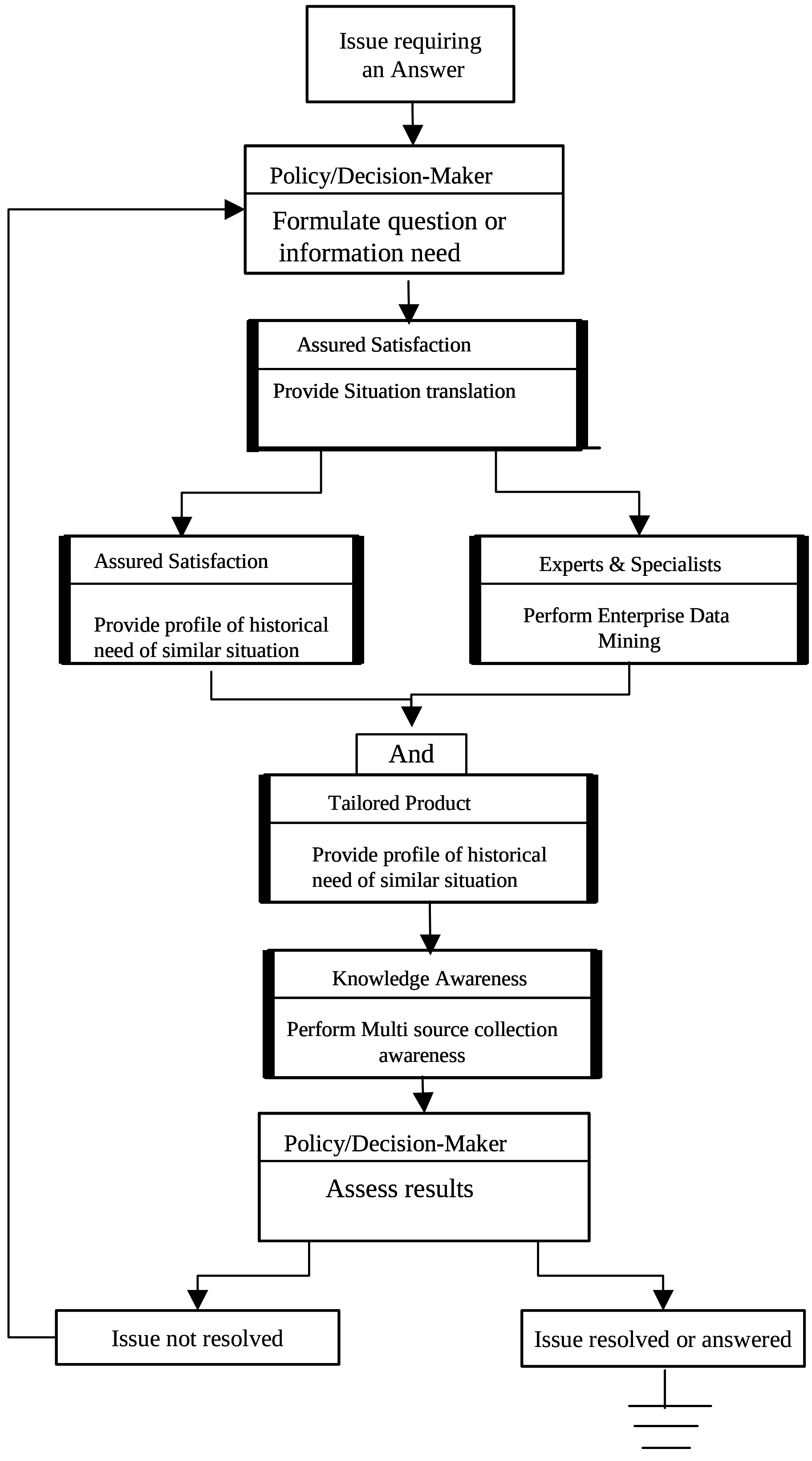


Figure 5 Multi-Source TPED Delta Chart Representing Sequential Flow

CHAPTER IV

A NEW MULTI-SOURCE TPED DESIGN

4.1 Defining a New Multi-Source TPED Domain

One of the challenges to NIMA is how to manage the significant increase in collection capability that results from current and future NTM, and commercial imagers or platforms, ensuring timely exploitation in the face of higher collection volumes, and finally, timely dissemination over communications paths. Since the new multi-source TPED is more than just imagery or multi-source space reconnaissance data, the biggest challenge to NIMA is how to rapidly integrate information collection and discovery of all types of critical knowledge to influence an outcome to a national or international issue. Thus, a new TPED domain emerges that assumes there is no resource scarcity and finding information or knowledge from massive volumes of data is the norm. This view redefines TPED as follows:

1. “Tasking” to be information discovery tasking in addition to resource tasking
2. “Processing” to be information correlation and formulation to create a multi-collector view
3. “Exploitation and Dissemination” to be the function of people with tools “seeing” patterns in information that is normally associated with information analysis and sharing this information through collaborative communications

The new multi-source TPED is an information business and emulates commercial information providers. Approved modernization architectures encourage intelligence information holdings to be “Web” accessible through Secret and Top Secret versions of Intelink, as well as Virtual Private Networks (VPN), and that applications be similarly web enabled and/or web-served. Here, the new TPED “processes” follow an e-business model to serve its intelligence consumers.

4.2 Decomposing the New Multi-Source TPED

The following section decomposes the multi-source TPED into its functional and component parts. Further, This section introduces the under laying system concept of operation that supports the multi-source TPED.

4.2.1 Decomposing Tasking Domain

The Tasking domain can be decomposed into User Tasking, Information Discovery, and Resource Tasking. Figure X shows the decomposition of tasking into its components parts. The components within User Tasking are Requests for Information, Request Translation, and Collection Database

Update. In the new TPED, Request for Information is intelligence problem centric that can be tied to one or more locations on the ground. Request Translation is the component that translates user problem centric requests into information collection needs and makes them available for discovery technology agents or Resource Tasking.

Information Discovery Tasking actively searches for related information associated with a given request for information using web enabled computer agents. The components within Information Discovery Tasking are Problem Keyword Constructor, Query Filtering and Distillation, and Information to Location Linker. In the multi-source TPED, Problem Keyword Constructor provides ontological creation of keyword search constructs that produce a high probability of query hits relevant to the problem of interest. The Query, Filtering, and Distillation component queries the World Wide Web, and all electronically accessible information repositories using the provided keyword searches and collect all hits into an index file. The index file filters false query hits out of the index file by zero size checking, bad link checking, and general context analysis, then geographically places valid hits based on person, place or thing location relevance to user centric problem. A recent study sponsored by the ADCI/Collection indicated that Geospatial Information System (GIS) tools that link diverse information to physical locations via layers would greatly improve understanding of intelligence problems.

The components within Resource Tasking are Opportunity Planning, Collection Resource Brokering, and Collector Commanding. The Opportunity Planning component assesses all available resources and interrogates all outstanding collection plans, and then creates a list of opportunities that could potentially satisfy that specific User Tasking. The Collection Resource Brokering component correlates and arbitrates resources through the use of web enabled computer agents that broker collection request and identify current collections that satisfy the request for information. Collector Commanding component issues a command list to any collector platform that is under the direct control of Resource Tasking.

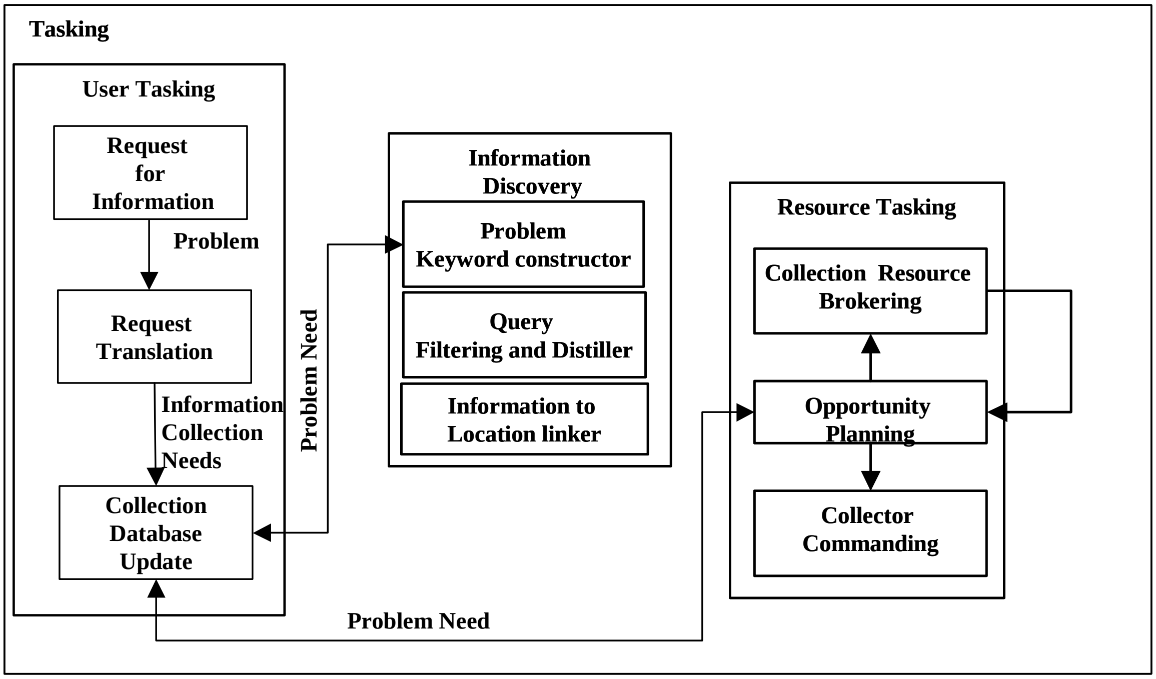


Figure 6 Functions and Components of the Tasking Domain

4.2.2 Decomposing the Processing Domain

The Processing Domain supports the results from both Resource and Information Discovery Tasking. The Processing Domain that supports Information Discovery Tasking decomposes into Pre-Processing, Data Formation/Correlation, Post-Processing. The Pre-Processing component receives indexed filtered data from Information Discovery Tasking and pulls each file in the index list. It performs detail content filtering and excerpt extraction. The Data Formation component receives the extracted information, performs content correlation, and then registers the information content to the location that corresponds to the location that relates to the information. Post-Processing components create and correlate displayable context overlay views containing the cue cluster of indexed information and a figure of merit associated with each entry. Figure 7 shows the processing components supporting information discovery tasking.

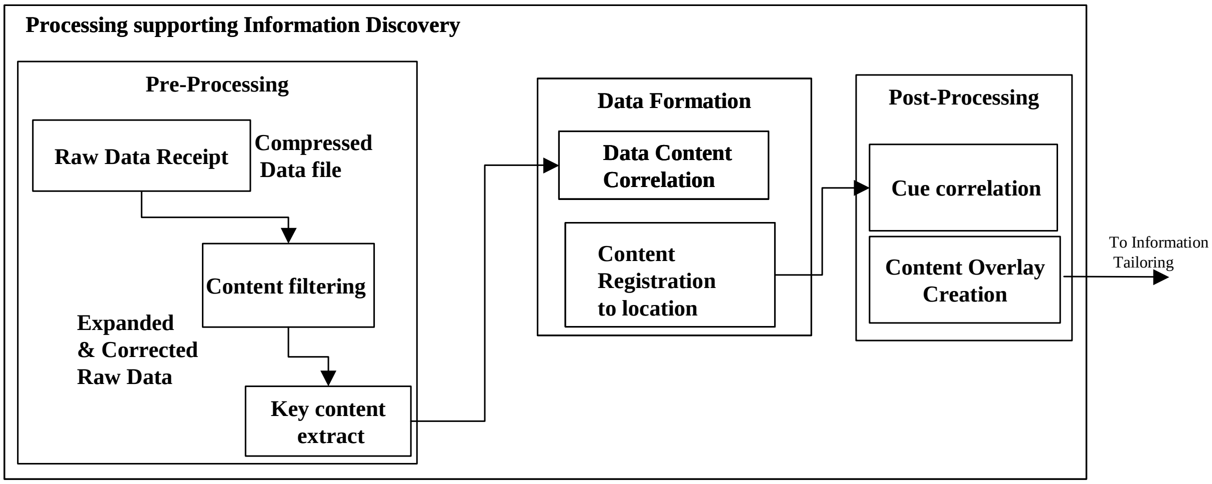


Figure 7 Components in Processing Domain for Information Discovery Tasking

The Processing domain supporting Resource Tasking also decomposes into Pre-Processing, Data Formation, and Post-Processing,. However, each component performs different processes or steps. The Pre-Processing component receives and decompresses data, if required, stores a copy of the unprocessed data file, and makes the data available for data formation. The Data Formation component forms the raw data into an usable data products used for basic analysis. Post-Processing components provide error and anomaly correction and data enhancement to the formed data product. Figure 8 shows the processing components supporting resource tasking.

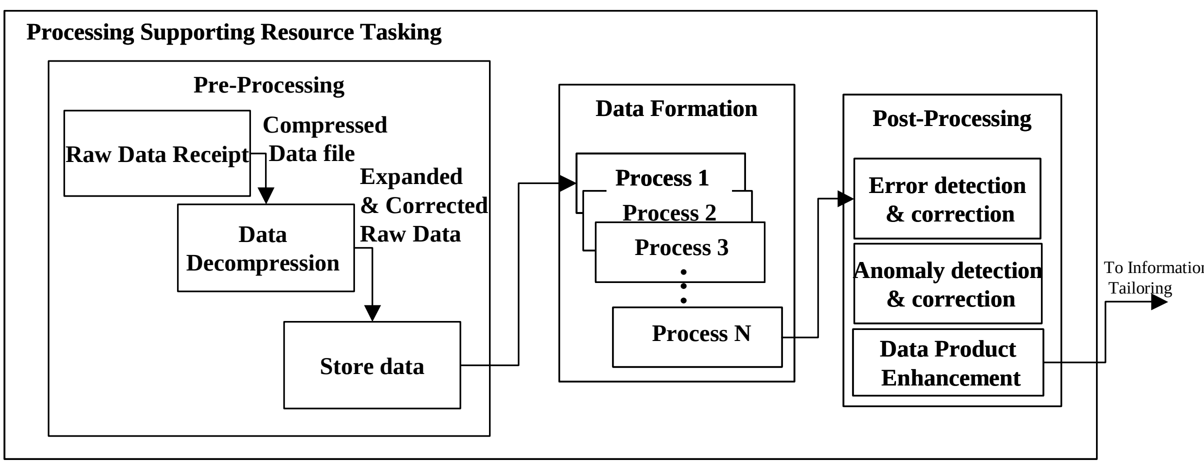


Figure 8 Components in Processing Domain for Resource Tasking

The Information Tailoring function is common to both processing supporting information discovery tasking and resource tasking. It combines the outputs from both information discovery and resource processing into a fused or composite information view. Further, it performs selected exploitation domain processes that were previously performed manually by an analyst but can be routinized and automated into the “upstream” Information Tailoring part of the Processing Domain. Information Tailoring provides the link in the chain that transforms “data” into “information” accessible to human analysts. Examples of Information Tailoring products are:

- Composite overlays of information and image views on vector-map foundation data
- Vector-map data (which are generally compact for the area covered) with imagery extracts of key visual features
- Automatic change detection
- Moving target indicators

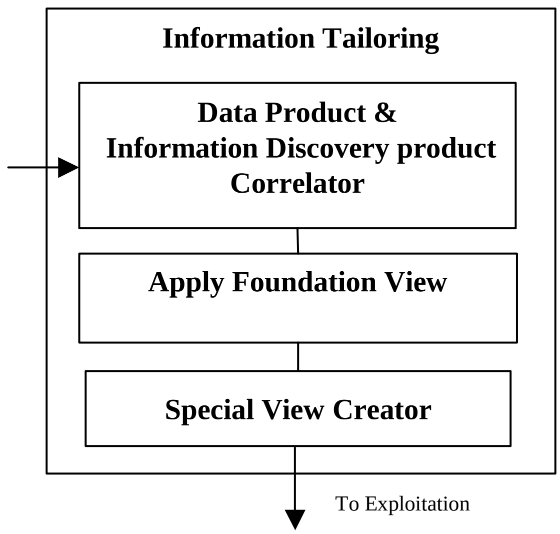


Figure 9 Components in Information Tailoring Domain

4.2.3 Decomposing The Exploitation and Dissemination Domain

The Exploitation and Dissemination domain can be decomposed into Exploitation Control, Data Analysis, Knowledge Sharing, and Knowledge Posting. Exploitation Control provides analyst and exploitation task assignment, work-package creation and transfer, workflow management. Data Analysis provides data display and viewing, information assessment, satisfaction notification, and retasking. Knowledge Sharing provides interactive collaborative data sharing between the data analyst and the user representative that initiated the request for information. Knowledge Posting provides for a problem set web portal for posting of information reviewed and approved by the user representative that satisfied the user request. Figure 10 shows the components supporting Exploitation and Dissemination.

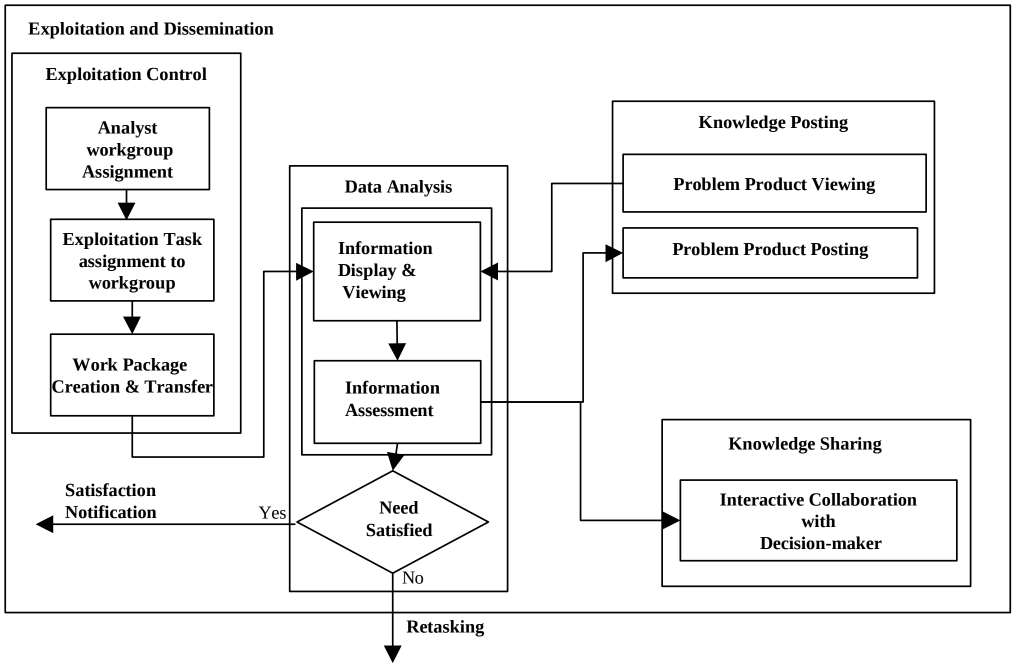


Figure 10 Components in the Exploitation Domain

Within the Exploitation and Dissemination Domain, Data Analysis has been manual and deserves further discussion. Evaluating each step in the seven-step exploitation process provided on page 3 an assessment was made to determine items to automate, to streamline, or to shorten information assessment and reporting timeliness, the limiting human factor considerations that need to be considered are:

- The analyst ability to rapidly read and comprehend intelligence requirements
- The analyst ability to type and spell
- Interpretation Task Complexity
- Target Familiarity
- Transition Time Between Tasks
- Task Backlog (Workload)
- Mental Fatigue
- Work Stress

Other limiting factors that needed to be addressed were:

- Transfer time of source image
- Transfer time of reference maps and images
- Display of all data to analyst
- Analyst wait time between available tasks

Four of the seven processes were evaluated for potential automation and the modifications to the exploitation process are contained in Table 10. The automation contributions were:

- Automatically Structure Analyst Tasks
- Simplify Tasks
- Provide Multi-Sensory Inputs
- Provide Automated Support Tools.

Table 10 Comparison Table of Current Exploitation Process to Futrue

No.	Original Steps	Transformations and Changes
1	Product Display to Analyst	Moved from step 1 to step 3
2	Analyst Orientation	Performed automatically moved From step 2 to step 4
3	Target Assessment	Performed automatically as part of Automatic Output Product Creation
4	Output Product Creation	Performed automatically as part of Automatic Output Product Creation
5	Security Downgrade of User Product	Moved from step 5 to step 2 and was automated
6	Review User Products	Moved from step 6 to step 5 and added a new step 6 for Analyst Modification request.
7	Transmit User Product	Unchanged

From my analysis, future analyst processes that would support rapid knowledge discovery and reporting results to users could use the same steps as today's image analyst processes. However, the order of the steps is different and many of the steps are automated to improve analysis. Figure 11 shows the new process with the automated steps highlighted.

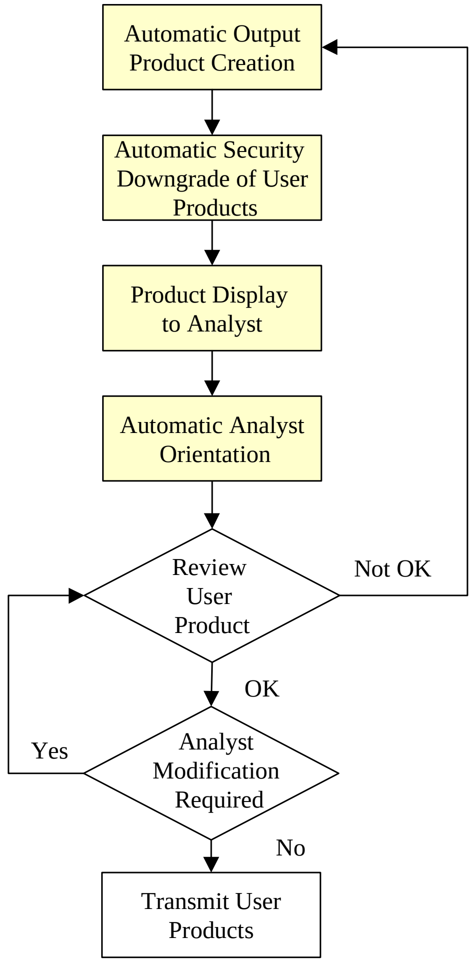


Figure 11 Multi-Sources Data Analyses Process

The following is an explanation of each analyst task in the new multi-source TPED step.

1. Output Product Creation - This automated step allows the processing segment to create a report item containing target specification information, image information and a remark statement containing answers to user request(s) in a USIGS compliant intelligence report format. Further, if a SID product is requested by the user, then the processing segment will automatically create a recommended SID product. Both the recommended intelligence report item and the SID product will be provided to the analyst product display (step 3).

2. Security Downgrade of User Products - This step allows for the segment processing to automatically downgrade report items and imagery based on a collateral data parameters file and security classification flags in the imagery support data. To allow validation of downgrade by the analyst, the header data file and all classification flags will be able to be reviewed by the analyst during step 5.

3. Product Display - Upon completion of the current analyst's task, the next task is automatically displayed. All reference data (maps etc.), analyst product(s), and recommended user product(s) are displayed. Further, all imagery and graphics are oriented to the analyst image product.

4. Analyst Orientation - This step automatically displays the problem centric knowledge discovery understanding to the analyst. This orientation can be as simple as a text display or as complete as a text display, image product with cues, and audio statement of problem centric knowledge discovery need(s).

5. Review User Product - The analyst performs this step. The analyst verifies that the user product answers the user need and that the product is properly downgraded to collateral. If the downgrading authority approves automatic downgrading and release or if all data is considered collateral, then security downgrade verification by the analyst can be deleted. If the output product does not meet the user need, the analyst has the capability to modify the output for release or request processing to generate a different tailored product.

6. Analyst Modification Required - This step provides the analyst with the capability to make changes to the user report product and/or the SID product.

7. Transmit User Product - This step allows the analyst to release all reviewed products to the user.

Using time-motion analysis and estimation of automated processing (assumed hardware performance increase based on trend forecast for the year 2008 timeframe presented in chapter 6) each step was estimated and tabulated. It was determined that an image analyst could perform this task in 3 to 5 minutes. Further study needs to be performed in this area to determine error rate due to fatigue, analyst relief cycle, and other human factor issues.

Knowledge Dissemination either occurs concurrent with exploitation or upon completion of exploitation. Knowledge Sharing Dissemination occurs concurrent with exploitation while Knowledge Posting Dissemination occurs *ex post facto* to exploitation. Knowledge Sharing Dissemination occurs in real time using collaboration tools and methods like video teleconferences and interactive sharing of the analyst exploitation monitor desktop.

4.2.4 TPED Component Interface Protocols

Since the new TPED is an information business and emulates commercial information providers, the possible protocol for the components that could guide new component development for inter and intra TPED communications and data sharing is JAVA, Enhanced JAVA Beans (EJB), Extensible Hypertext Markup Language (XML), and XML query language (XQL). In addition, World Geodetic

System (the standard by which points on the earth are measured in real space, National Imagery Transfer Format (NITF), JPEG and MPEG data standards are provided for imagery and video sharing. Finally, as other emerging data, communications, and protocol standards are realized, the architecture design must be adaptable to new standard insertion.

Chapter V

NEW TPED TECHNOLOGY PERFORMANCE AND SIZING ANALYSES

5.1 System Description

In order to facilitate visualization of the resource allocation process, consider the simplified Multi-Source TPED system functional block diagram presented in Figure 12.

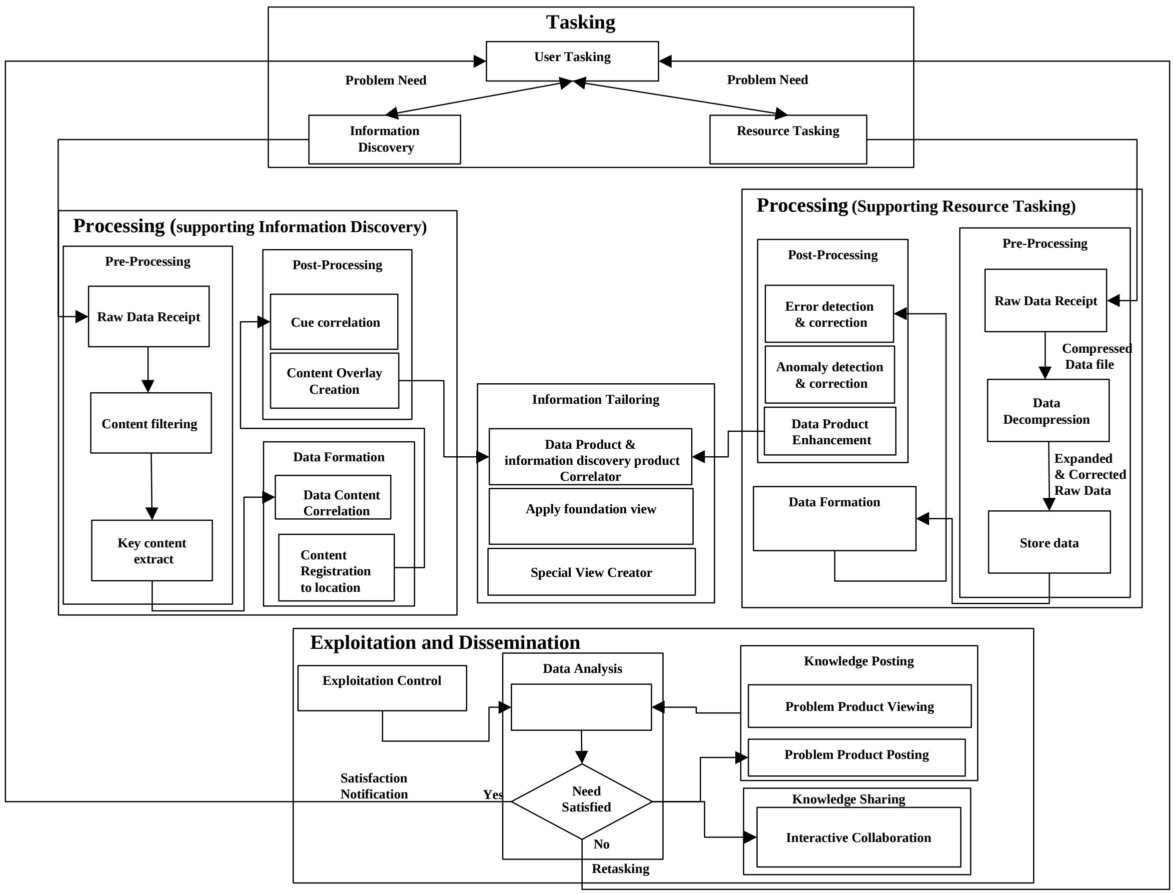


Figure 12 New Multi-Source TPED System Overview

The objective is to quantify the processing, storage and bandwidth required to successfully implement each of the major functions identified in Figure 12. The resource allocation for each of the multi-source TPED sub-domains is detailed in a separate subsection below.

Naturally, the resources required by a single multi-source TPED system depend on the operational scenario against which the sizing is made. For the purpose of sizing, the concept scenario shown in the delta chart at the end of Chapter III will be used. To support this crisis scenario, one problem centric user request shall generate 1000 data collects per minute. From the 1000 collects, 100 data collects shall be assumed for resource tasking and 900 collects shall be assumed for information discovery tasking. Further, this scenario shall assume satisfaction of a retask request within 10 minutes of initiation. The retask request assumes that a single interim data item is provided from resource tasking activities for each 9 interim data items provided from Information Discovery Tasking.

5.2 Tasking

Tasking includes User Tasking, Information Discovery, and Resource Tasking. For User Tasking, sizing is comparable to the current NTM Requirement Management System. The complexity of automation with improved results is assumed to add ten times the processing, ten times the memory, and three times the disk storage requirements of similar current day applications

Information Discovery Tasking is comparable to a number of currently available knowledge discovery tools like Silent Runner, Starlight, and Textwise except for the need for automation of these tools. The complexity of automation with improved results is assumed to add ten times the processing and five times the memory requirements of similar current day applications. Further, the disk storage requirement is assumed to be equal to the user tasking storage requirement.

Resource Tasking is comparable to the Command & Control, Mission Management, and Resource Allocation activities. While the data item throughput rate drives processing, the number of collection requests that must be successfully prioritized, deconflicted, scheduled and executed drives resource management. The number of data items captured per user request can vary dramatically due to user need requirements. However this paper assumes that 100 data items per minute corresponds to 100 collection requests per minute.

5.2.1 CPUs

Traditional user tasking approaches are currently capable of manually entering or updating several hundred user requests per day. However, many requests are created to satisfy a single-user problem. The translation of the user problem requests into information discovery and resource tasking collection requirements are performed manually and takes 12 to 24 hours to perform. These traditional user requests are created over a time period spanning several days prior to their actual execution, the resulting peak computational burden is very small (less than .5 Gigaflop). However, to automate this process is expected to require ten times the existing processing rate or five Gigaflops. Further, the translation time from user request until problem translation is complete should not exceed five minutes.

Therefore, user request translation must execute on a time line reduced from approximately 24 hours to 5 minutes. The computational burden for user request translation should increase compared to an assumed system loading by a factor of 288 to 1 (60 minutes*24hr/5 minutes). The peak computational burden for the worst-case scenario can thus be bounded at (288 x 5Gigaflops =) 1440 Gigaflops. Using a de-rating factor of 4 to1 due to algorithm parallelization inefficiencies yields a computational burden of (1440 x 4 =) 5760 Gigaflops.

There is no current information discovery tasking within the current TPED. Therefore, this paper assumes that the computational burden requirements are two times resource tasking. Today's resource tasking peak computational burden is estimated not to exceed one Gigaflop therefore, information discovery tasking is estimated not to exceed two Gigaflops. The worst case scenario requires insertion of information discovery retask events into the existing information mining schedule that meets the 10-minute satisfaction timeline from receipt of task request to data item receipt by data analysis (see exploitation). Successfully accomplishing this unique feature requires human intervention via interactive tasking tools. Moreover, it is assumed that three of the ten minutes allotted for information discovery retasking can be employed as exploitation time and collection time must also be budgeted within this constraint. Assuming that human interaction for adjudication requires around one minute, the time available for automated generation of the updated keyword schema and query, filter and distill activities is no more than two minutes. Therefore, information discovery tasking must execute on a timeline reduced from approximately six hours to two minutes. Since the complexity of the information discovery tasking function is not anticipated to significantly decrease, the computational burden should increase compared to assumed system loading by an approximate factor of 180 to 1. The peak computational burden for the worst-case scenario can thus be bounded at (180 x 2Gigaflops =) 360 Gigaflops. Using a de-rating factor of 4 to1, due to algorithm parallelizaton inefficiencies yields a computational burden of (360 x 4 =) 1440 Gigaflops.

Current resource tasking approaches are currently capable of scheduling several thousand imaging requests per day. The system works with a standing problem list and non-event driven tasks. These tasks can be validated, prioritized, scheduled, and collected employing traditional approaches to the resource management problem. Because these traditional event schedules are created via a moving window spanning several hours prior to their actual execution, the resulting peak computational burden is not excessive (less than one Gigaflop). The worst-case scenario requires insertion of a retasked event into the existing prioritized collection schedule that meets the ten minute satisfaction timeline from receipt of task request to data item receipt by data analysis (see exploitation). Successfully accomplishing this unique feature requires human intervention via interactive tasking tools. Moreover, at most only two of the ten minutes allotted for retasking can be employed for Resource Management, as exploitation time and collection time must also be budgeted within this constraint. Assuming that human interaction for adjudication requires around one minute, the time available for automated generation of the updated collection plan is no more than a minute. Therefore, Resource Tasking must execute on a time line reduced from approximately two hours to one minute. Since the complexity of the Resource Tasking function is not anticipated to significantly decrease, the computational burden should increase compared to current systems by an approximate factor of 120 to 1. The peak computational burden for the worst-case scenario can thus be bounded at (120 x 1 =) 120 Gigaflops. Using a de-rating factor of 4 to1, due to algorithm parallelization inefficiencies, yields a computational burden of (120 x 4 =) 480 Gigaflops.

The total tasking computation burden is expected to be the sum of each computational burden presented above which is (1440+360+120=) 1920 Gigaflops and the de-rated computation burden for tasking is (5760+1440+480=) 7680 Gigaflops.

5.2.2 Storage

Assuming at most 100 problem centric user requests or updates of approximately ten Mbytes are electronically received per day, these user requests translate into 1000 collection requirements, each collection requirement is also contained in a ten Megabyte file, and all files are archived for 60 days. User requests storage require (10 x 100 x 60)/1000 =) 60 Gigabytes of storage and the 1000 collection requirements require (1000 x 10 x 60)/1000 =) 600 Gigabytes. Applying a storage efficiency of 0.75 yields (660/0.75 =) 880 Gigabytes required storage. Projecting a commercially available disk capacity of 2450 Gigabytes in the 2008 time frame yields a requirement for (880/2450 =) 1 disk. This storage could be accommodated in the disk archive for processing. If a separate archive is established for this storage, the size will be driven by the choice of the type of disk system employed, e.g., RAID versus duplicate disks.

5.2.3 Bandwidth

Averaging the 100 user requests over a 12-hour period yields 8.33 requests per hour and thus the required bandwidth is approximately ((8.33 x 10 Megabytes x 8 bits per byte)/3600 seconds per hour=) 185 kilobits/ second or the speed of a current day DSL line.

5.3 Processing

The resource allocation for processing is performed for both information discovery tasking and resources tasking. Both types of processing consist of preprocessing, data formation and post processing operations.

For information discovery tasking, processing is driven by the requirement to support the overall maximum throughput of 900 data items per minute. Given the size of each data item in Megabytes and the number of operations required per data item to perform data item formation, one can compute the associated sustained processing rate. The estimates of data item size and computations required were not immediately available. Thus, assumptions regarding data item size and associated operations per file had to be made. The architecture must support all the candidate information discovery data item types. Therefore, these assumptions had to reflect the largest computational burden. Based on what was known of file ontology trends among the candidate data items, the data item size was set at 4 Megabytes and the computational burden set at 400 ops/byte for the purposes of resource allocation. It was felt that these parameters provided a reasonable upper bound for the most stressful computational case emerging from ongoing data mining technology as they pertain to processing in support of information discovery tasking.

For Resource Tasking, processing is driven by the requirement to support the overall maximum throughput of 100 data items per minute. Given the size of each data item in total pixels and the number of operations required per pixel to execute data item formation, one can compute the associated sustained processing rate. Firm estimates of data item size and computations required were not immediately available. Thus, assumptions regarding data item size and associated operations per pixel had to be made. The architecture must support all the candidate sensor types. Therefore, these assumptions had to reflect the sensor design mandating the largest computational burden. Based on what was known of design trends among the candidate sensors, the data item size was set at 11 K x 11 K pixels and the computational burden set at 4000 ops/pixel for the purposes of resource allocation. It was felt that these parameters provided a reasonable upper bound for the most stressful computational case emerging from ongoing design changes as they pertain to processing in support of resource tasking.

For Information Tailoring, processing is driven by the requirement to support the overall maximum throughput of 10 data items per minute from information discovery processing and 100 items per minute from resource processing. The estimates of data item size and computations required were not immediately available. Thus, assumptions regarding data item size and associated operations per file had to be made. The estimated size of 20 Megabytes for each information discovery processing output and 121 Mega-pixels for each resource processing output. Further it is estimated that 200 ops/byte and 200 ops/pixel must be performed for the respective data items types. It was felt that these parameters provided a reasonable upper bound for the most stressful computational case emerging from ongoing data mining technology as they pertain to processing in support of information discovery tasking.

5.3.1 CPUs

For processing in support of information discovery tasking, the required sustained computational rate is given by:

Equation 1

$$\left(\frac{900\ data\ items}{1\min}\right)\left(\frac{1\min}{60\sec}\right)\left(\frac{4000000 bytes}{1\ data\ item}\right)\left(\frac{400\ ops}{1\ byte}\right)\approx 24\ Gflops$$

For processing in support of resource tasking, the required sustained computational rate is given by:

Equation 2

$$\left(\frac{100\ data\ items}{1\min}\right)\left(\frac{1\min}{60\sec}\right)\left(\frac{(11000)^2\ \ pixels}{1\ data\ item}\right)\left(\frac{4000\ ops}{1\ pixel}\right)\approx 807\ Gflops$$

For processing in support of information tailoring, the required sustained computational rate is given by:

Equation 3

$$\left(\frac{10\ data\ items}{1\min}\right)\left(\frac{1\min}{60\sec}\right)\left(\frac{20000000 Bytes}{1\ data\ item}\right)\left(\frac{200\ ops}{1\ Byte}\right)+$$
$$\left(\frac{100\ data\ items}{1\min}\right)\left(\frac{1\min}{60\sec}\right)\left(\frac{121000000\ pixels}{1\ data\ item}\right)\left(\frac{200\ ops}{1\ pixel}\right)\approx 41\ Gflops$$

Thus the total sustained computational rate is (24+807 +41 =) 872 Gigaflops

In order to convert the sustained rate into a peak rate, as typically quoted for commercial computers, one must multiply by an appropriate de-rating factor. This factor accounts for the difference traditionally encountered between the quoted peak rate and the rate at which an application will actually execute. The de-rating factor is an experienced-based adjustment that reflects the dependence of quoted benchmarks on algorithmic structure. Based on my 18-years of experience, I have chosen a conservative de-rating factor of 4-to-1 for data item formation applications. Hence, the estimated peak computational rate for the data item processing function is given by:

Equation 4

$$(872\, Gflop\ Sustained\ Rate)\left(\frac{4\, Peak\ Rate}{1\, Sustained\ Rate}\right) \approx 3488\, Gflop\ Peak\ Rate$$

5.3.2 Storage

For storage in support of information tasking, the anticipated active collection time is continuous over 24 hours for all key storage repositories and web sites. The per day collection time is 1440 minutes per day , the local archival requirement for 30 day storage of the raw capture data is:

Equation 5

$$\left(\frac{900\, input\ items}{1 \min}\right)\left(\frac{4000000\, bytes}{1\, input\_item}\right)\left(\frac{1440}{1 day}\right)(30\ days) \approx 108\, Gigabytes$$

For storage in support of Resource Tasking, the anticipated active collection time is continuous over 24 hours for all key areas of interest. Per day, the imaging time is 1440 minutes per day. Assuming a duty cycle of 25% across all national and commercial collector systems and a 1-to-1 ratio between raw input pixels and output pixels, the local archival requirement for 30 days storage of the raw capture data is:

Equation 6

$$\left(\frac{100\, data\ items}{1 \min}\right)\left(\frac{(11000)^2\ \ pixels}{1\, data\_item}\right)\left(\frac{3\, bits\ compressed}{1\ pixel}\right)\left(\frac{1440}{1 day}\right)\left(\frac{1 \min collected}{4 \min available}\right)\left(\frac{1\, byte}{8\, bits}\right)(30\ days)$$
$$\approx 485{,}860\_Gigabytes$$

For storage in support of information output to information tailoring, the anticipated number of outputs are ninety 20 Megabyte files every minute collected over a continuous 24 hours (1440 minutes) for all key areas of interest and maintained in the local archive for 7 days:

Equation 7

$$\left(\frac{90\,Output\ items}{1\min}\right)\left(\frac{20000000\ bytes}{1\,data\_item}\right)\left(\frac{1440}{1day}\right)(7\ days) \approx 18{,}144\,Gigabytes$$

For storage in support of resource output to Information Tailoring, the anticipated number of outputs are ten 4 Gigabyte files every minute collected over a continuous 24 hours (1440 minutes) for all key areas of interest and maintained in the local archive for 7 days:

Equation 8

$$\left(\frac{10\,output\ items}{1\min}\right)\left(\frac{4\ Gigabytes}{1\,data\_item}\right)\left(\frac{1440}{1day}\right)(7\ days) \approx 403{,}200\,Gigabytes$$

For storage in support of Information Tailoring outputs to exploitation, the anticipated number of outputs are ten 8 Gigabyte files every minute collected over a continuous 24 hours (1440 minutes) for all key areas of interest and maintained in the local archive for 7 days:

Equation 9

$$\left(\frac{10\,output\ items}{1\min}\right)\left(\frac{8\ Gigabytes}{1\,data\_item}\right)\left(\frac{1440}{1day}\right)(1\,days) \approx 115{,}200\,Gigabytes$$

The total storage is the sum of all the stored items times an efficiency factor. The sum of all the raw stored data is (108+485,860+18,144+403,200+115,200 =) 1,022,512 Gigabytes. The efficiency factor for disk storage is 0.75. The actual disk capacity must be 25% larger than the volume to be stored. Hence, (1,022,512/0.75 =) 1,363,350 Gigabytes total disk capacity will be required for a 1-to-1 ratio between raw input pixels and output pixels.

5.3.3 Bandwidth

Processing Inputs

In support of information discovery, raw data inputs to process the maximum input bandwidth to processing is the bandwidth required to receive 900 four megabyte discovery input items per minute over a 24 hours period. The worst-case raw data input bandwidth to processing from information discovery is given by:

Equation 10

$$\left(\frac{900\_input\_items}{1\min}\right)\left(\frac{1\min}{60\sec}\right)\left(\frac{4\_Megabytes}{1\,item}\right)\left(\frac{8\_bits}{1\,byte}\right) \approx 480\ Mbits/\sec$$

In support of resource raw data inputs into processing, the maximum input bandwidth to processing is limited by the available downlink capacity of 3 Gigabits per second. Further, the data is sent compressed at 3 bits per pixel. The maximum output bandwidth from processing is given by:

Equation 11

$$\left(\frac{100\,images}{1\min}\right)\left(\frac{1\min}{60\sec}\right)\left(\frac{(11000)^2\ \ pixels}{1\,image}\right)\left(\frac{4\,bits}{1\ pixel}\right) \approx 807\ Mbits/\sec$$

Processing Outputs

In support of information discovery tasking processed data outputs to information tailoring, the maximum bandwidth from processing is the bandwidth required to send ninety 20 megabyte discovery items per minute to Information Tailoring over a 24 hours period. The worst case output bandwidth from processing is given by:

Equation 12

$$\left(\frac{90\_Output\_items}{1\min}\right)\left(\frac{1\min}{60\sec}\right)\left(\frac{20\_Megabytes}{1\,Output\ Item}\right)\left(\frac{8\,bits}{1\,byte}\right) \approx 240\ Mbits/\sec$$

In support of resource tasking processed data outputs to Information Tailoring, the maximum bandwidth from processing is the bandwidth required to send ten 4 Gigabyte output items per minute to Information Tailoring over a 24 hours period. The worst case output bandwidth from processing is given by:

Equation 13

$$\left(\frac{10\,output\ items}{1\min}\right)\left(\frac{1\min}{60\sec}\right)\left(\frac{4\,Gigabytes}{1\,output\_item}\right)\left(\frac{8\,bites}{1\,byte}\right) \approx 5333.4\ Mbits/\sec$$

Information Tailoring

In support of information tailoring output products to exploitation, the maximum bandwidth from information tailoring is the bandwidth required to send 10 eight Gigabyte output products per minute to exploitation over a 24 hours period. The worse case output bandwidth from information tailoring is given by:

Equation 14

$$\left(\frac{10\,output\ items}{1\min}\right)\left(\frac{1\min}{60\sec}\right)\left(\frac{8\,Gigabytes}{1\,output\_item}\right)\left(\frac{8\,bites}{1\,byte}\right)\approx 10666.7\,Mbits/\sec$$

The total network bandwidth is the sum of all the transmitted items times an efficiency factor. The sum of all transmitted data is (480+807+240+5333.4+10666.7 =) 17527.1 Megabits or 17.5 Gigabits. . Using an efficiency factor for Ethernet is 0.50. The actual network load is assumed to .20% larger than the data being sent. Hence, ((17.5 x 1.2)/0.80 =) 26.3Gigabits/sec

5.4 Exploitation

There are three types of exploitation. First phase exploitation is mission focused, time critical and provides the first look at imagery, signals, and information discovery text for determination of initial indications and warnings concerning priority situations. Second phase exploitation is detailed exploitation and provides detailed study and measurement of objects in an image scene, detailed analysis and measure of signal emissions, or detailed content correlation of text information discovered during information discovery. Third phase exploitation is long-term study and learning of location centric objects (e.g. factories, military installations, etc.). This sizing analysis only considers first phase exploitation and assumes it is synchronized with the rapid retasking capability. Regardless of the number of data items associated with the user request being serviced, the time criticality of the final answer on retasking remains invariant. Thus, the worst case scenario for the Exploitation function involves supporting the Retasking loop while ingesting data items at the maximum rate of 1000 data items per minute.

As no human can reasonably be capable of exploiting individual data items in less than a second, it is expressly assumed here that Information Tailoring function creates composite views that incorporates 10 data items acquired from processing of resource tasking items and 90 data items acquired from processing of information discovery tasking items. The Exploitation function ingesting ten 40.36 Gigabytes (10*4Gigabyte +90*4Megabyte) tailored products every minute sustained with the aid of as-yet-to-be-developed exploitation tools. However, even with this assumption operative, at most 3 to 5 minutes can be allotted for this exploitation task as stated in paragraph 4.2.3 Decomposing The Exploitation and Dissemination Domain above, as resource management and collection time must also be budgeted within the 10 minute retasking constraint. If 5 minutes analyst time is assumed, the implication is that it will take approximately 50 analyst work positions with automated exploitation tools to keep up with the exploitation workload from tailored product ingest. Further, one can assume that the worst-case retasking situation is that ten requests are adjudicated every minute.

5.4.1 CPUs

Exploitation management is responsible for efficiently allocating and transferring exploitation tasks, associated tailored products and supporting data to the correct client computer/analyst combination based on the analyst expertise, availability, and backlog. In addition Exploitation managemwnt tracks and reports individual task orders through every stage of exploitation from receipt of tailored product and task to delivery of the final product to dissemination. Finally exploitation management is responsible for continuously providing status of all exploitation resources and ensuring successful execution of all accepted tasks.

For this paper, it is assumed that the computational intensive part of exploitation management are accomplished via a COTS, Web-based, enterprise management software package executing on a high-end compute server requiring approximately 250 Gigaflops per work station. The worst-case situation is:

Equation 15

$$\left(\frac{50\ work\ \ positions}{1\,Exploitation\ \ function}\right)(50\,Gigaflops) \approx 2500\,Gflops$$

Therefore the resulting de-rated peak computational requirement for Exploitation Management is (2500Gigaflops x 4 =) 10000 Gigaflops.

The computational intensive part of the exploitation analyst task involves the application of the smart tools to assist the human in area-limited target detection based on established models and prior conditions. As these algorithms currently do not exist in the advanced form anticipated here, one must extrapolate their computational requirements from what is known about existing, albeit more primitive, exploitation tools. Such contemporary tools involve feature extraction based on texture variations, edges and linear geometric features. Typically, these algorithms require approximately 3000 operations per input pixel. For this sizing effort the following assumptions apply:

- Future algorithms will be on the order of 6000 operations per pixel
- Algorithms are applied to an area limited to 5k X 5k pixels

- Five algorithms will be executing simultaneously

Based on these assumptions, the exploitation sustained processing requirement is:

Equation 16

$$\left(\frac{10\,tailored\_products}{1\min}\right)\left(\frac{1\min}{60\sec}\right)\left(\frac{(5000)^2\ \ pixels}{1\,local\ area}\right)(5\,local\ areas)\left(\frac{(6000)\,ops}{1\ pixel}\right)\approx 1250\,Gflops$$

Therefore the resulting de-rated peak computational requirement for Exploitation is (1250Gigaflops x 4 =) 5000 Gigaflops.

5.4.2 Storage

The archival storage associated with exploitation involves the "problem set folders" that contain reference data items, graphics and textual reports. If one assumes that 1,000 active problem sets is a representative number and that each problem set folder contains on the average, six 40.36Gigabyte tailored products required storage. Thus the total storage is:

Equation 17

$$(1{,}000\ problem\ sets)\left\{\left(\frac{6\,tailored\ \ products}{1\ problem\ set}\right)\left(\frac{40.36Gigabytes}{1\,tailored\ \ product}\right)\right\}\approx 242{,}160Gigabytes$$

Applying a storage efficiency of 0.75 yields (242.16/0.75 =) 322880 Gigabytes.

5.4.3 Bandwidth

Assuming the maximum tailored product throughput of 10 per minute and that approximately 1 problem set folder must be retrieved per tailored product in support of exploitation. Since 50 analyst are being used to prevent exploitation backlog, it is assumed that each work position receive at most two tasks within 10 minutes. The worst-case scenario is the wait time for the first task after the analyst logs onto the work position. Further, to allow the analyst takes the maximum exploitation time, it is assumed that the analyst wait time from log on to display of first task is 15 seconds for the task and all supporting data. The amount of input data from processing to a single workstation yields an input bandwidth of:

Equation 18

$$\left(\left(\frac{8Gbytes}{1\,tailored\ \ product}\right)+\left(\frac{6\,tailored\ \_\ product}{1\ problemfolder}\right)\left(\frac{8\ Gbytes}{1\,tailored\ \ product}\right)\right)\left(\frac{8\,bites}{1\,byte}\right)\left(\frac{1}{15\,sec}\right)\approx 29.9\,Gbytes\,/\,\text{sec}$$

.

Using an efficiency factor for Ethernet is 0.50. The actual network load is assumed to .20% larger than the data being sent. Hence, ((29.9x 1.2)/0.80 =) 45 Gigabits/sec

5.5 Knowledge Sharing and Posting Dissemination

Knowledge sharing product dissemination occurs through interactive collaboration between the analyst's softcopy desktop and the decision makers softcopy desktop. The decision-maker can see the analyst's displayed results as the analyst data interpretation occurs. The decision-maker clarifies refinements to needed answers while that analyst initiates alternate tailored product views or retasking to resolve decision-makers questions. The average interactive session occurs within the 5-minute exploitation timeline. Network load in this situation consist of displaying delta change updates at 75 updates per second. Each display update is assumed to be 25% of all displayable dots on the screen. It is assumed that each high-resolution screen contains 1600 dots by 1280 lines of displayable data. Further, it is assumed that decision-maker receives a copy of the annotated information view containing his answer at the end of each session. Further, The annotated information view is automatically posted to the problem set web portal as pending final review and is accepted by the problem set responsible individual. All data will be encrypted and electronically distributed through a firewall via a virtual hub on a Wide Area Network. A copy of all disseminated products are maintained on a local website for 7-day data repository to allow community access for data recovery by any problem set responsible individual. The 7 day storage repository is provided by processing (see above). The knowledge discovery portal is maintained by the dissemination function.

5.5.1 CPUs

The processing requirements of the dissemination function consist of the need to host a collaboration and video conference Server program, a COTS inventory control program, perform rapid queries of the order entry database and retrievals from the local 7-day tailored product repository maintained by the processing function. Consequently, 25 Gigaflops peak processing capability per analyst position.

Equation 19

$$\left(\frac{25 Gigafolps}{1\, work\ position}\right)(50 work\_positions) \approx 1250\, Gigaflops$$

Therefore the resulting de-rated peak computational requirement for Dissemination is (1250 Gigaflops x 4 =) 5000 Gigaflops

5.5.2 Storage

Assuming the worst-case portal storage will be sized for 10 tailored products per minute, 1440 minutes per day, for a 7-day duration and that every tailored product with supporting information is 8 Gigabytes. It must be noted at this point that any data over 7-days old is retained in a robotic archive by problem set. The retention policy and size of this repository is beyond the scope of this paper and will require further investigation. This produced the following computation:

Equation 20

$$\left(\frac{10\, tailored\ products}{1 \min}\right)\left(\frac{1440 \min}{1\, day}\right)\left(\frac{8 Gbytes}{1\, tailored\ product}\right)(7\ days) \approx 806400 Gigabytes$$

Given a corresponding RI retrieve rate of 1000 per day for a 7 day period with each order requiring 10 Megabyte of storage results in a product order entry catalog of 210 Gigabytes. Thus the local 7-day archive requirement is (806400 + 210 =) 806610 Gigabytes. Applying a storage efficiency of 0.75 gives a requirement of (806610/0.75 =) 1075480 Gigabytes. Storing this on a striped RAID disk system, projecting year 2008 technology requires 441 disk drives with a capacity of 1080 Terabytes.

5.5.3 Bandwidth

Allowing for the maximum data item throughput of 100 data items per minute and assuming that all data items must be output with a corresponding intelligence report/graphics package of 100 Megabytes, results in an input bandwidth of

Equation 21

$$\left(\frac{10\, tailored\ products}{1 \min}\right)\left(\frac{1 \min}{60 \sec}\right)\left(\frac{8\ Gbytes}{1\, tailored\ product}\right) \approx 6.7 Gbyte / \sec$$

5.6 Infrastructure

All the above sized domains require a web enterprise Infrastructure. A normal web enterprise consists of a web server, applications server, a set of common services and a series of applications.

5.6.1External Resources

This function represents the system requirement to actively engage in collaborative networking with source experts during the exploitation-retasking cycle and to provide and receive Cross-INT tipoffs. While the functional requirement is reflected in this element the physical interfaces will occur through the Dissemination element of the architecture. This interface is anticipated to support collaboration and/or tipoffs with 25 distinct geographic sites simultaneously via T1 links. As this function is an interface only, there are no associated computational or storage requirements. The resulting bandwidth is thus (25 x 0.1875 Megabytes/sec =) 4.7 Megabytes/sec.

5.6.2 Workflow Management And Control

This is the overall workflow manager for the ground system and is responsible for efficiently allocating all the computational, memory, and bandwidth resources in addition to tracking individual task orders through every stage of processing from receipt of the initial request to delivery of the final product. This functional element is responsible for continuously statusing all system resources and ensuring successful execution of all accepted tasks.

5.6.2.1 CPUs

For this paper, it is assumed that the computational intensive part of workflow management are accomplished via a COTS, Web-based, enterprise management software package executing on a high-end compute server requiring approximately 250 Gigaflops of peak processing capacity.

5.6.2.2 Storage

The execution status files, day files, health and status files, optimization tables, etc. are assumed to to require a maximum of 250 Gigabytes of local storage.

5.6.2.3 Bandwidth

The estimated total bandwidth required for communication of commands and statusing with all the system components is assumed to be 2 Gigabytes/sec.

C H A P T E R V I

TECHNOLOGY ASSESSMENT AND FORECAST

6.1 Processor Performance

Moore's law states that CPU performance of will double every 18 months. Over the past 20 years processor performance has followed this trend very closely. During the decade of the 1990's processor performance has actually exceeded the Moore's law rate and has achieved performance doubling every 15 months. Figure 13 shows the processor performance improvement for RISC processors for the decade of the 1990's relative to the performance in 1992. As the chip design and manufacturing processes matured, there has been a convergence of processor performance across manufacturers. This has led to consolidation in the chip manufacturing industry. Six manufacturers are represented in the chart and because there has been significant consolidation some manufacturers have left the market. Currently only Intel and IBM have primary fabrication plants for high performance chips in the USA.

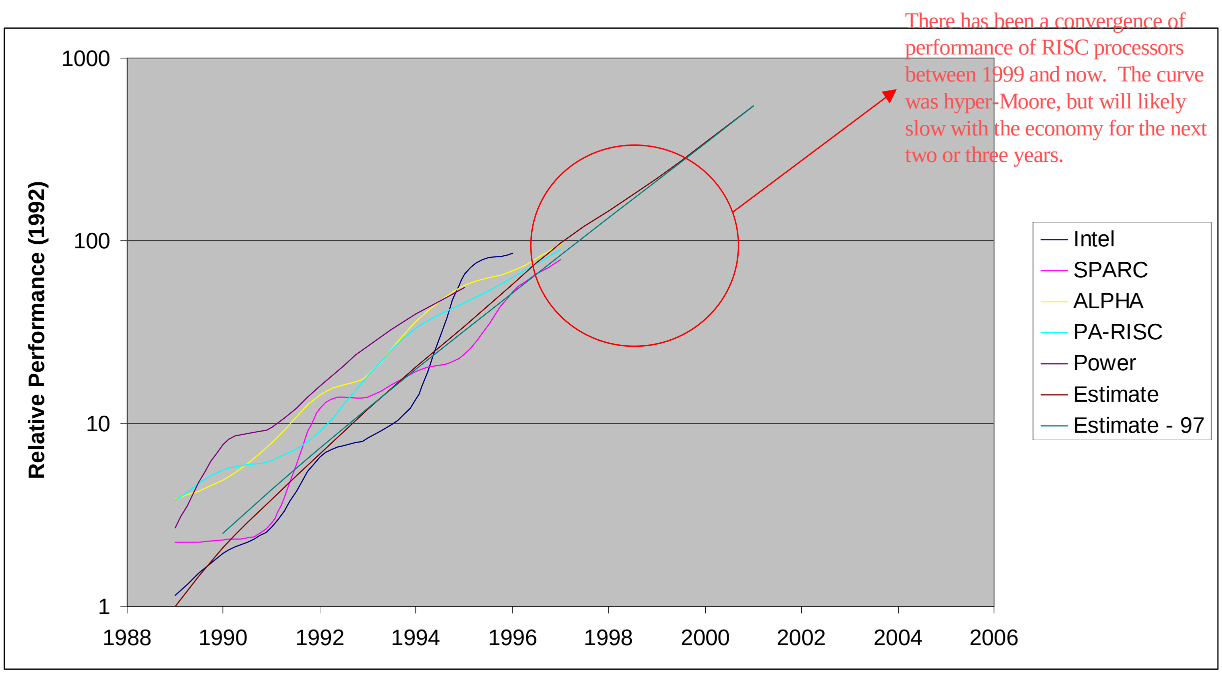


Figure 13 RISC processor performance.

The economy is slowing and there is a movement by the vendors toward the Intel processor. Because of high fabrication plant costs and a slowing economy it is expected that the processor improvement rate will slow to doubling every 18 to 24 months. The performance improvement rate could slow even more because vendors are consolidating on the Intel processor. With less competition the product development rate will slow.

In the year 2000, processor performance was 2.2 Giga Flops. We wish to build a system using processors purchased in the 2006 or later. In order to project a system design based on future processors, the performance was projected using an improvement factor of 1.5 per year. Figure 14 shows the processor performance through the year 2008 using this improvement factor.

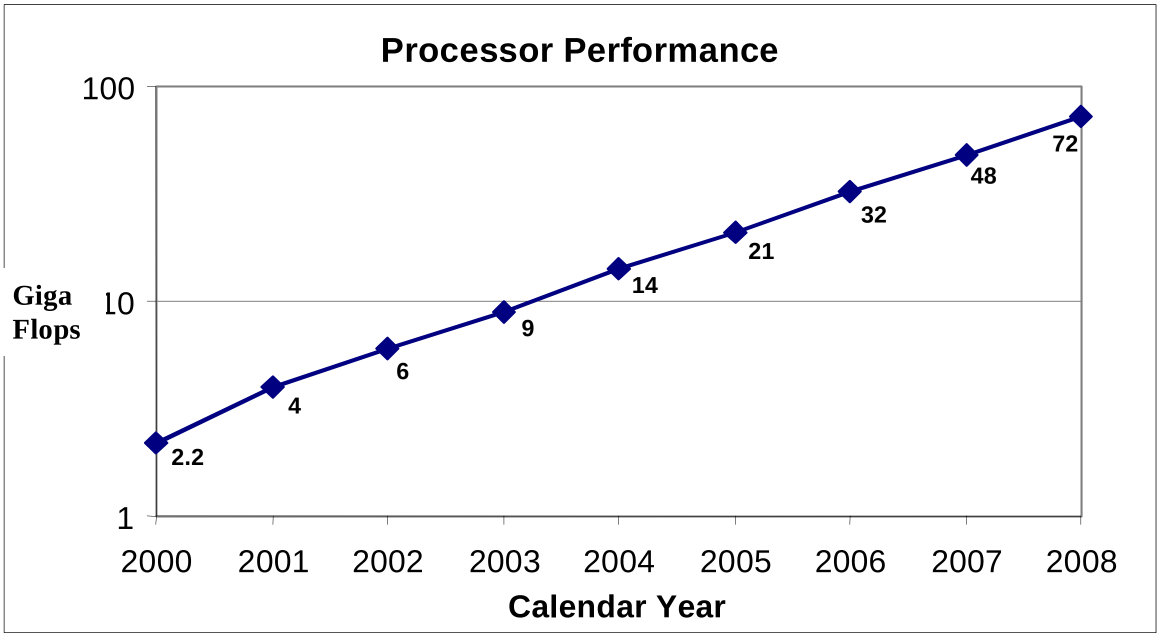


Figure 14 Projected processor performance.

6.2 Disk Capacity

The need for disk capacity is increasing much faster than the need for additional computing capacity. Currently, 80% of all new information generated each year is digital and digital information is increasing at the rate of 50% per year. The primary storage for this digital information is disk. Industry demand for disk storage is driving disk capacities higher at a rapid pace.

As more data is stored on disk, there is an increasing demand for reliable back-ups for the data on the disks. In the past, the primary back-up media has been tape. However, as databases increase in size to multiple hundreds of terabytes, tapes are not seen as viable for back-ups because tape back-ups require many hours to create and update. In order to decrease the volume of tape back-ups, the industry is rapidly moving to RAID disks. Storing data on a RAID disk system increases the required disk capacity for the same data volume. This has accelerated the demand for higher disk capacity.

Figure 15 shows a chart of disk capacities for the years 1997 to 2000 for three manufacturers. The curve which extends from 1992 to 2005 is a projection of disk capacity assuming a capacity increase factor of 1.5 per year starting in 1992. While Quantum and IBM have followed the projected curve, Seagate has had improvements will beyond the projection. It is unlikely that Seagate will continue to have capacity improvements at the same rate.

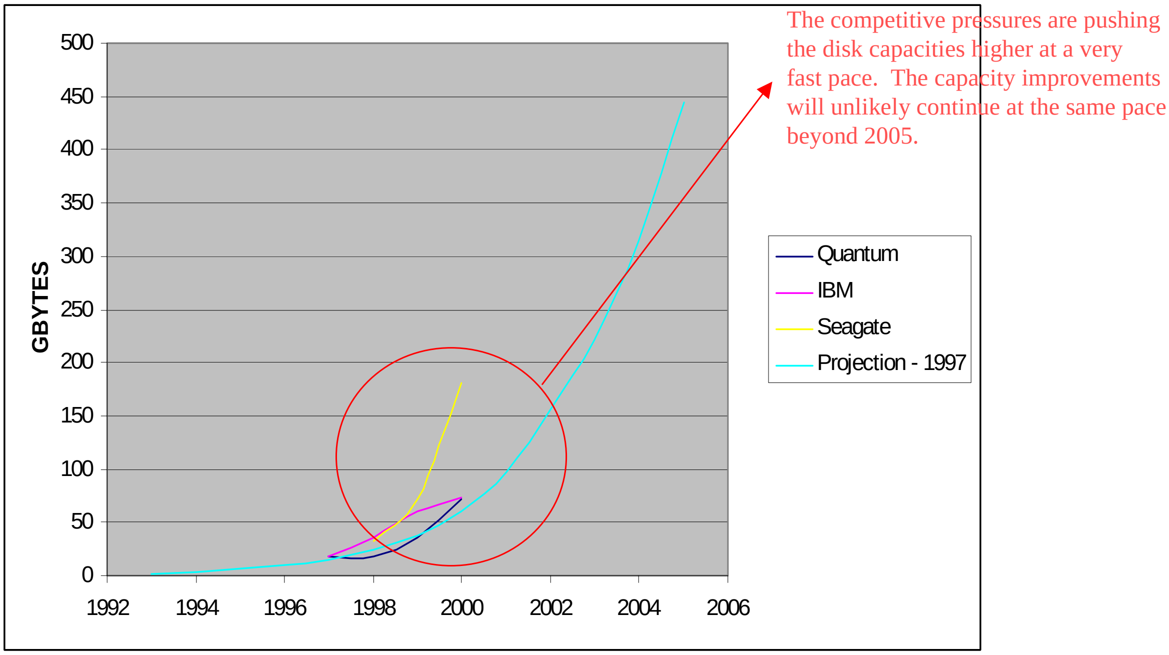


Figure 15 1997 to 2000 Disk capacities and the projection from 1992 to 2005.

In the year 2000, disk capacity was 72 Gigabytes. Using an improvement factor of 1.5 per year, a projection of disk capacity was made through the year 2008. Figure 16 shows this projection and this curve was used to project a system design based on future disk capacities.

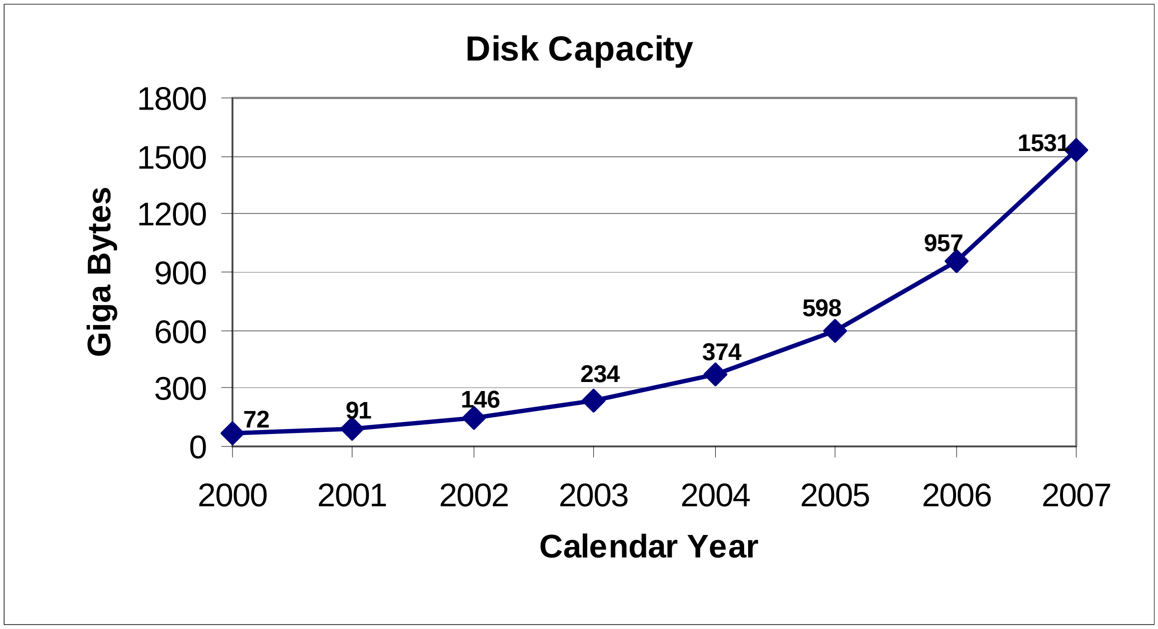


Figure 16 Projection of disk capacity through 2008

6.3 Ethernet Performance

Networking is an essential element of any processing system. Early in development of networking, Ethernet was introduced and quickly dominated the network market. Because of the fundamental nature of networking and communication, market share is important. Ethernet is well suited for networking and relatively easy to install. This has allowed Ethernet to dominate the network market against all competitors. Figure 17 shows the transmission rate of ethernet for a 25-year period. CISCO has produced a pre-standard 10 Gigabit ethernet and it should be standard by the year 2002.

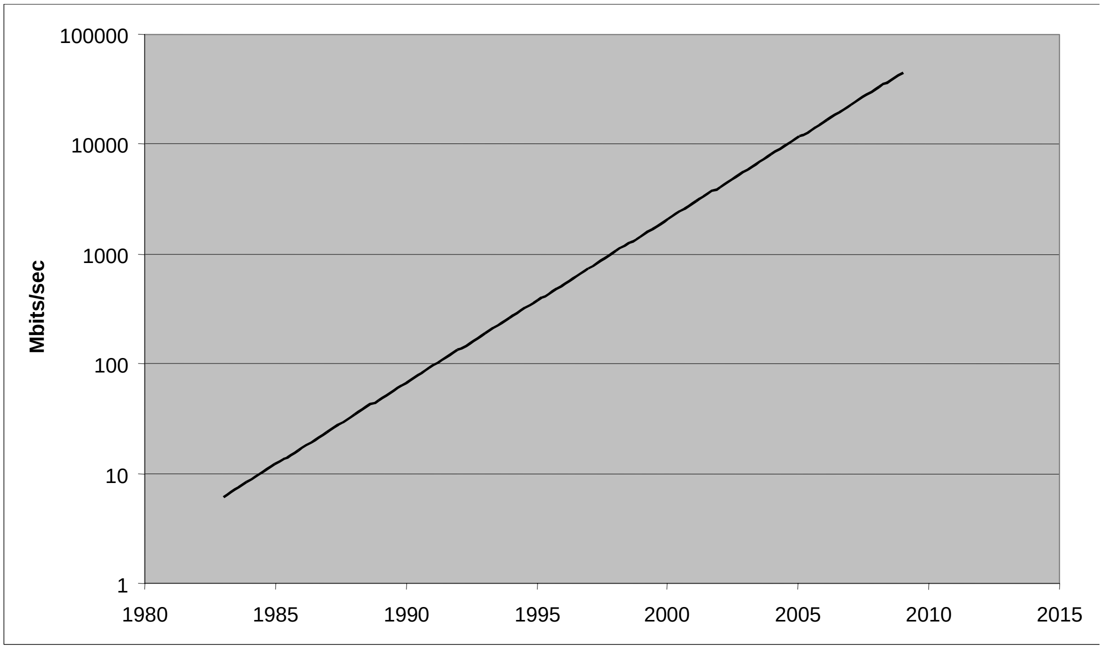
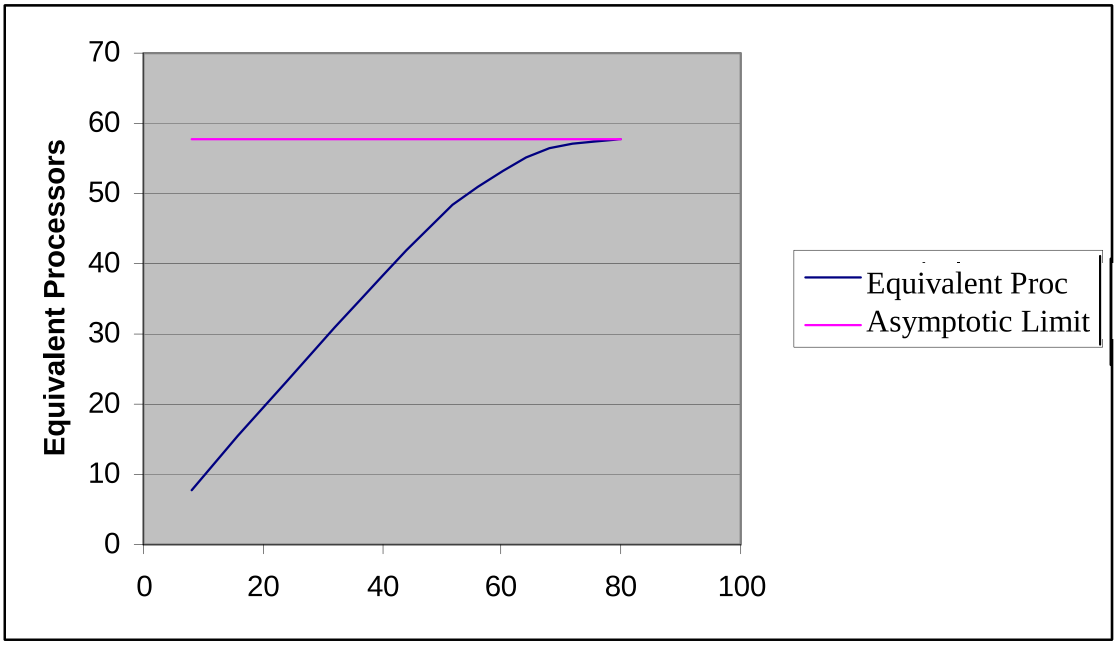


Figure 17 Transmission rate of ethernet from 1983 to 2008.

6.4 Computer System Performance

As the demand for processing power has increased, processing requirements have far outstripped the capability of single processors. As a result, there is an increased emphasis on system level performance. Computer systems with multiple CPUs are now commonplace. Systems have been built with hundreds and even thousands of processors. Currently, system designs are consolidating on CCNUMA (cache coherent, non-uniform memory access) for 32 or 64 processors. Overall system performance of multiple processor systems is dependent on both hardware design, operating system design, and on application software design. The efficiency is always less than the efficiency of a single processor multiplied by the number of processors in the system. Queuing theory and SpecRate performance are used to project how the overall system scales from a single processor to the top of the line system. The measure of scalable efficiency utilizes improvements in both the processors and the operating system. The efficiency of a system is measured in terms of equivalent processors. Figure 18 shows the mapping of actual number of processors to equivalent number of processors. As the number of processors increases the equivalent number of processors approaches an asymptotic limit of 58. For this reason, systems are typically limited to 64 processors.



* Based upon published SpecFP Rate numbers

Figure 18 Equivalent processors versus actual number of processors

After a new system is introduced, the scalable efficiency improves approximately 10% per year over the life of the system. The efficiency is measured by the maximum equivalent processor performance. The average scalable efficiency for most systems approaches 92% of the total number of processors and usually requires 2 to 3 years to reach this level of efficiency. When dramatic changes in product cause a drop in efficiency for existing systems, there is a significant effort to reconfigure the systems to regain acceptable performance. At this time, the asymptotic limit (for all systems) is always less than the maximum number of processors of the largest system. Stated another way, if the largest system uses 64 processors, the asymptotic limit is less than 64. Designing software for scalable systems is a difficult task. To optimize applications, a test environment is required.

6.5 Marketplace Trends

The High Performance Computing/scientific market is moving towards the small and mid-range systems. HP and SUN are de-emphasizing large systems and focusing on the mid-range systems. If the economy continues to slow, it is expected that more vendors will follow the lead of HP and SUN. Many vendors are migrating to the Intel processor for their mid-range systems in order to ensure market share.

The revenue for high performance computing systems has been growing for many years, see

Figure 19. High performance computers are a very small niche that is pursued for marketing and other reasons. The revenues for high performance computers has been almost constant. High performance computers used for Divisional and Departmental purposes have been responsible for most of the growth in high performance computing systems.

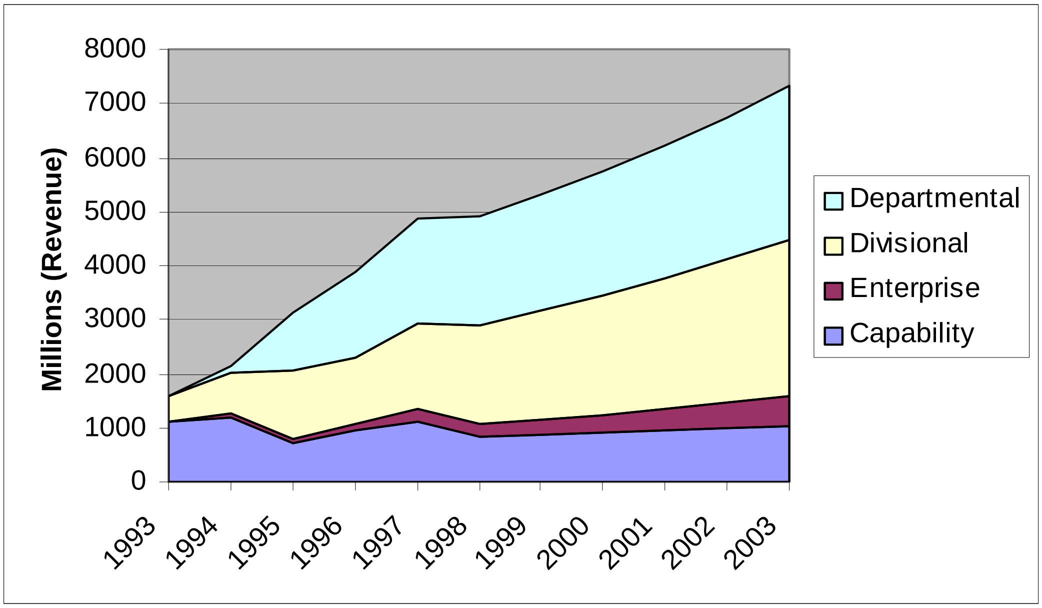


Figure 19 Revenues for HPC from 1993 to 2003.

6.6 Sales

Figure 20 shows the sales for five HPC manufacturers. HP, Compaq, IBM, and SUN seem to be profitable and competitive. SGI is rapidly losing market share. The main emphasis is on the mid-range systems by all of these vendors.

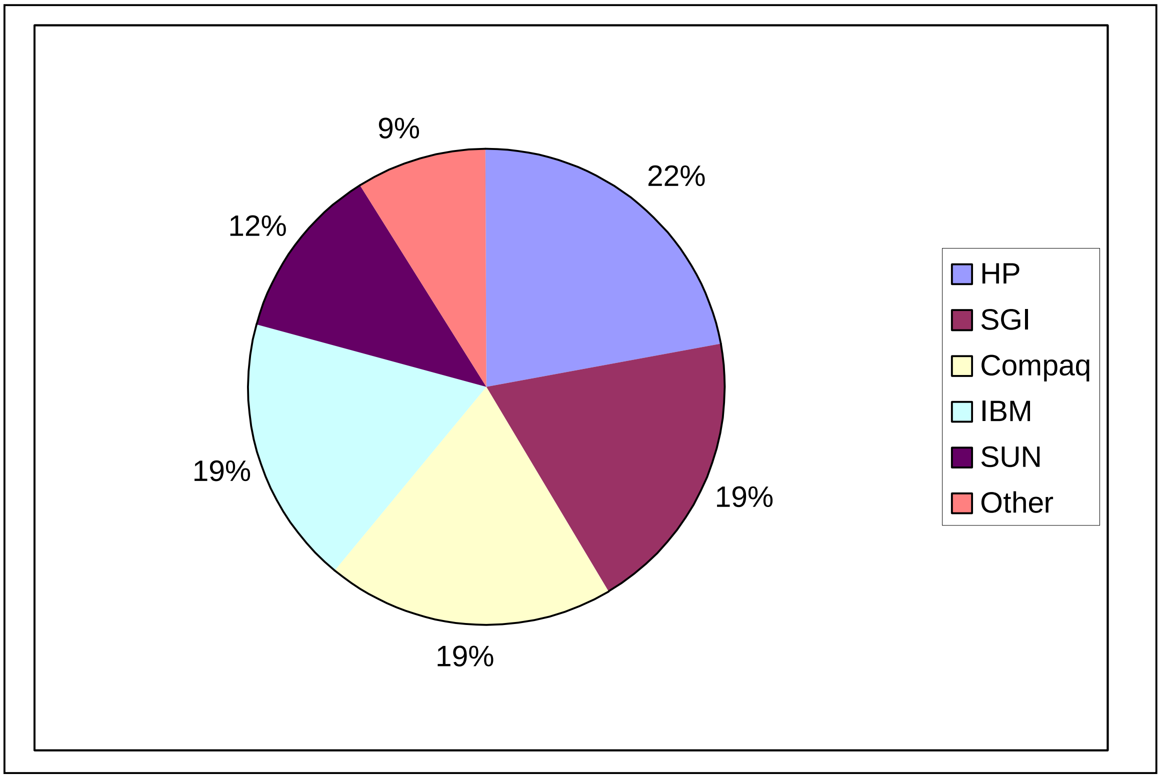


Figure 20 Sales of the largest High Performance Computing manufacturers.

CHAPTER VII

RESULTS AND CONCLUSIONS

7.1Comparison for Tasking

7.1.1 Tasking Resource Summary

Table 11 through Table 13 provides a summary of all the processor, archive and network needs for the tasking function as calculated in Chapter V. These numbers set the baseline performance requirements from which 2008 computers, disks and network projections are determined.

Table 11 Summary Of Tasking Compute Needs

Parameter	Units	Need
User Tasking		
User Problem Set Translations	Translations/minute	.2
Translations rate	Gigaflops	5
Computational improvement (10 to 1)	Multiplier	288
Computational Burden	Gigaflops	1440
De-rated Flops (4 to 1)	Gigaflops	5760
Information Discovery Tasking		
Current computational burden	Gigaflops	2
Keyword Generation Timeline	Minutes	2
Computational improvement (16 hrs to 2 min)	Multipliers	180
Computational Burden	Gigaflops	360
De-rated Flops (4 to 1)	Gigaflops	1440
Resource Tasking		
Current computational burden	Gigaflops	1
Collection update Timeline	Minutes	1
Computational improvement (2 hrs to 1 min)	Multipliers	120
Computational Burden	Gigaflops	120
De-rated Flops (4 to 1)	Gigaflops	480
Total Tasking Compute Needs		
Computational Burden	Gigaflops	1920
Compute Requirement	Gigaflops	7680

Table 12 Summary Of Tasking Storage Needs

Parameter	Units	Need
Number of User Requests or Updates/ day	Requests/day	100
File Size/ User Request	Gigabytes	.010
Translations/ user request	Collects/user request	1000
File Size/ Collection requirement	Gigabytes	.010
Number of days archived	Days	60

Table 13 Summary Of Tasking Network Needs

Parameter	Units	Need
Number of User Requests or Updates/ day	Requests/day	100
Number of User Requests or Updates/ Hour	Requests/day	8.33
File Size/ User Request	Gigabytes	.010
Bits	Bits/byte	.8

7.1.2 Tasking Function's System Projections

Based on technology trends provided in Chapter VI, it is reasonable to project commercially available 72 Gigaflops processors in the 2008-timeframe. One can project that a 64-processor machine will be the commercial standard. For system sizing, one can assume that the efficiency factor for 64 processors is 90%, one processor is used by the operating system, and two are used for the network. Therefore, the equivalent number of processors per system is ((64*0.9)-3 =) 54. Further, a prudent designer adds another machine to support future availability needs. To determine the number of computer systems, we take the number of required processors determined in Chapter V and then divide by the number of equivalent processors per system. Table 14 summarizes the 2008 computer needs for tasking.

Table 14 2008 Projection for Tasking

Parameter	Units	Need
Computer		
2008 Processor Performance	Gigaflops	72
2008 Processors/ Computer(see Chapter V)	72 Gigaflops	64
Equivalent 2008 Processors/ Computer	72 Gigaflops	54
Calculated number of processors	Integer	107
Num of Computers (|107/54| +1 redundant)		3
Storage		
Archive Size for User Requests	Gigabytes	60
Archive Size for Collection Requirements	Gigabytes	600
Total Tasking Archive size @.75 efficiency	Gigabytes	880
Network		
8 simultaneous users/hr for user requests	Kilobits/sec	.18

7.2 Resource Comparisons for Processing

7.2.1 Processing Resource Summary

Table 15 through Table 17 is a summary of all the processor, archive and network needs for the tasking function as calculated in Chapter V. These numbers set the baseline performance requirements from which 2008 computers, disks and network projections are determined.

Table 15 Summary Of Processing Compute Needs

Parameter	Units	Need
Processing Supporting Information Discovery Tasking		
Data Item Rate	Data Items/Minute	900
Information data Size	Megabytes	4
Computational Burden	Ops/byte	400
Computational Burden	Gigaflops	24
Processing Supporting Resource Tasking		
Data Item Rate	Data Items/Minute	100
Resource data Size (11K by 11K)	Millions of Pixels	121
Computational Burden	Ops/Pixel	4000
Computational Burden	Gigaflops	807
Processing Supporting Information Tailoring		
Data Item Rate from Information Discovery Proc	Data Items/Minute	10
Data Item Rate from Resource Proc	Data Items/Minute	100
Information data Size	Megabytes	20
Resource data Size (11K by 11K)	Millions of Pixels	121
Computational Burden	Ops/byte	200
Computational Burden	Ops/Pixel	200
Computational Burden	Gigaflops	41
Total Processing Compute Needs		
Computational Burden	Gigaflops	872
De-rated Flops (4 to 1)	Gigaflops	3488

Table 16 Summary of Processing Storage Needs

Parameter	Units	Need
Number of information collects/min	Info collects/ min	900
File Size/ Info Collect	Megabytes	4
Archive Size for Information Collects	Gigabytes	108
Number of Resource collects/min	Resource collects/ min	100
File Size/ Resource Collect	Millions of Pixels	121
Compression	Bits/Pixels	3
Archive Size for Resource Collects	Gigabytes	85860
Number of information outputs /min	Info output/min	90
File Size/ Info output	Megabytes	20
Archive Size for Information outputs	Gigabytes	18144
Number of resource output /min	Resource outputs/ min	10
File Size/ Resource Output	Gigabytes	4
Archive Size for Resource outputs	Gigabytes	403200
Number of tailored products/min	Resource outputs/min	10
File Size/ Tailored Product	Gigabytes	8
Archive Size for tailored product creates	Gigabytes	115,200
Number of days archived for input data	Days	30
Number of days archived for output data	Days	7
Num of days archived for tailored products	Days	1
Total Processing Archive data size	Gigabytes	1,022,512
Total Archive Size required @,75 efficiency	Gigabytes	1,363,350

Table 17 Summary of Processing Network Needs

Parameter	Units	Need
Information Discovery Raw Input	Megabits/sec	480
Resource Raw Inputs	Megabits/sec	807
Processed Information Output	Megabits/sec	240
Processed Resource Output	Megabits/sec	5333.4
Tailored Output	Megabits/sec	10666.7
Total amount of data over network @ .50 efficiency	Gigabits/sec	42.1

7.2.2 Processing Function's System Projections

Based on technology trends provided in Chapter VI, it is reasonable to project commercially available 72 Gigaflops processors in the 2008-timeframe. One can project that a 62-processor machine will be the commercial standard. For system sizing, one can assume that the efficiency factor for 64 processors is 90%, one processor is used by the Operating System, and two are used for the network. Therefore the equivalent number of processors per system is ((64*0.9)-3 =) 54. Further, a prudent designer adds another machine to support future availability needs. To determine the number of computer systems, we take the number of required processors determined in Chapter V and then divide by the number of equivalent processors per system. Table 18 summaries the 2008 computer needs for tasking.

The total number of Gigabytes of storage required for all processing was calculated in Chapter V as 1,363,350 Gigabytes. Projecting a commercially available disk capacity of 2450 Gigabytes in the

2008 time frame, yields a requirement for (1363350/2450 =) 557 disks. When stored in a striped RAID file system, 630 disk drives with a capacity of 1.364 Petabytes will be required (See Table 18).

Table 18 2008 Projection for Processing

Parameter	Units	Need
Computer		
2008 Processor Performance	Gigaflops	72
2008 Processors/ Computer(see Chapter V)	72 Gigaflops	64
Equivalent 2008 Processors/ Computer	72 Gigaflops	54
Calculated number of processors	Processors	49
Num of Computers (|49/54| +1 redundant)	Computers	2
Storage		
2008 Disk Size	Gigabytes	2450
Active Disks In RAID Stripe	Disks	8
ECC Disks In RAID	Disks	1
Total Archive Size	Gigabytes	1,363,350
Number of data Disks required	Disks	557
Number of ECC Disks	Disks	70
Number of RAID	RAID systems	70
Total RAID capacity	Petabytes	1.372
Network		
2008 Ethernet capability	Gigabit/sec	100
Ethernet de-rating factor	Percent	50
Total Network Load for processing	Gigabits/sec	26.3
Total percent of network utilization after de-rating	percent	49

7.2.3 Sensitivity testing through Modeling

Due to the computational, storage and network complexity of the processing function, the processing sizing calculations were modeled against 2008 forecast capabilities in order to better understand the range of sizes of different resources. A series of two-dimensional plots were created. Appendix A contains the Microsoft EXCEL model and visual basic spread sheet model applied to this problem. Three parameters are plotted as a function of input data size and operations per data item. They are:

- Archive Size
- Number of Compute Processors
- Concurrent Data Items in Process

Each of these plots show the technology forecast for 2008. The outputs from this model were compared against the preceding processing projections as a sensitivity test.

Archive Size

The plots in Figure 21 show the size in Terabytes of the disk storage archive system. The plot shows the volume archive for 2008. The actual number and size of the data items to be stored is the same for both plots. The striped disk system design used in this projection used 8 disks per stripe for both 2006 and 2008. The size of each disk increases from 957 Gigabytes to 2450 Gigabytes in the same period. While the size of the archive increases in terms of bytes, the number of physical disks in the system decreases.

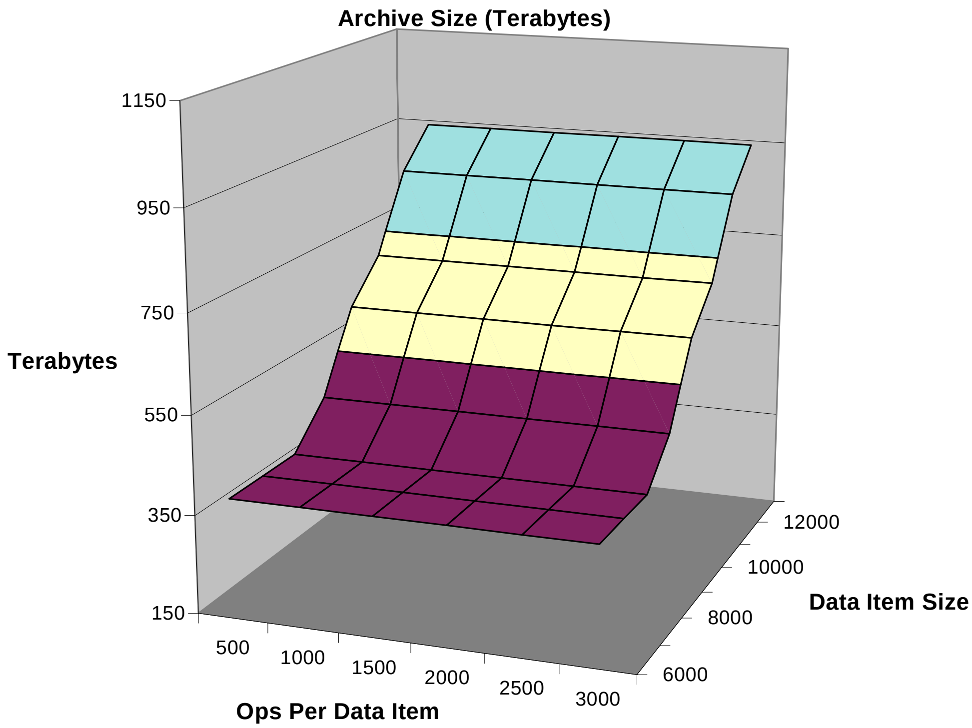


Figure 21 Disk Size Projections for the Year 2008

Number of CPUs

Figure 22 shows the number of FLOPS required to accomplish the required computations must be de-rated for the single processor efficiency and also for the system scalability efficiency. These factors are reflected in the plots.

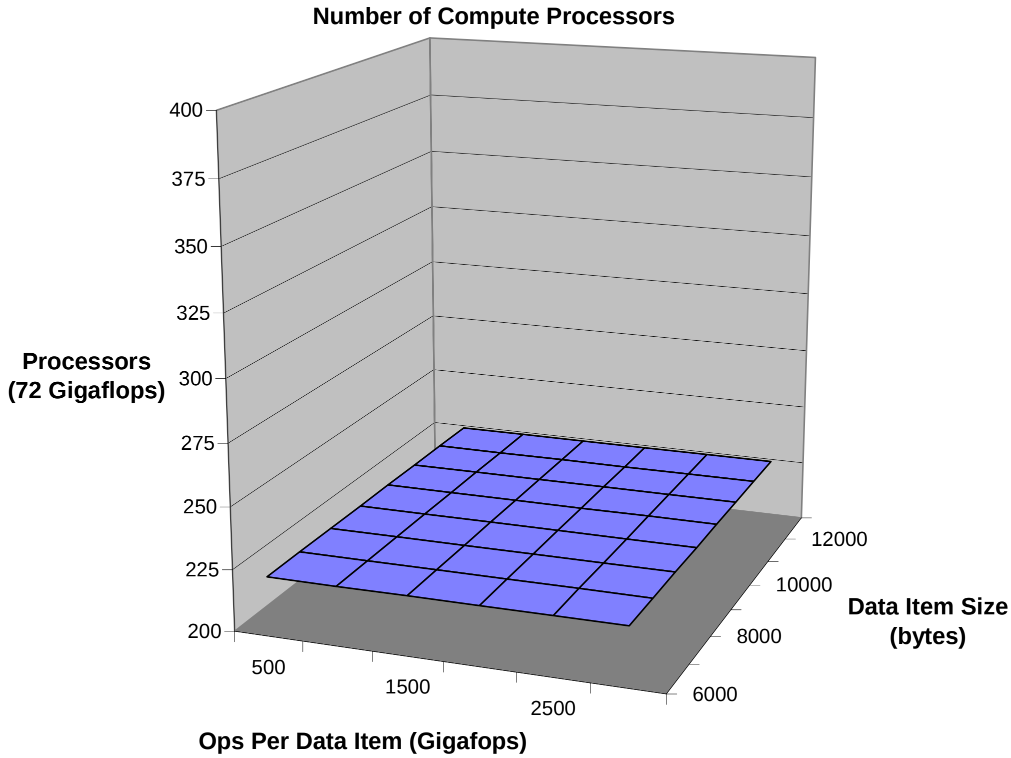


Figure 22 Projected Number of CPUs to support Processing in Year 2008

Number of Concurrent Data Items in Process

Because of throughput requirements, the time to process one data item, and the number of computed systems employed, a number of data items are being processed simultaneously. This number will vary with data item input rate, individual processor performance, and the number of compute systems in the architecture. Figure 23 shows the number of simultaneous data items processing on a single computer. The processor performance for the Year 2008 is much faster than for the Year 2006, number of data items in process is much smaller. Due to the small number of concurrently processing data items, the number of compute servers may need further investigation.

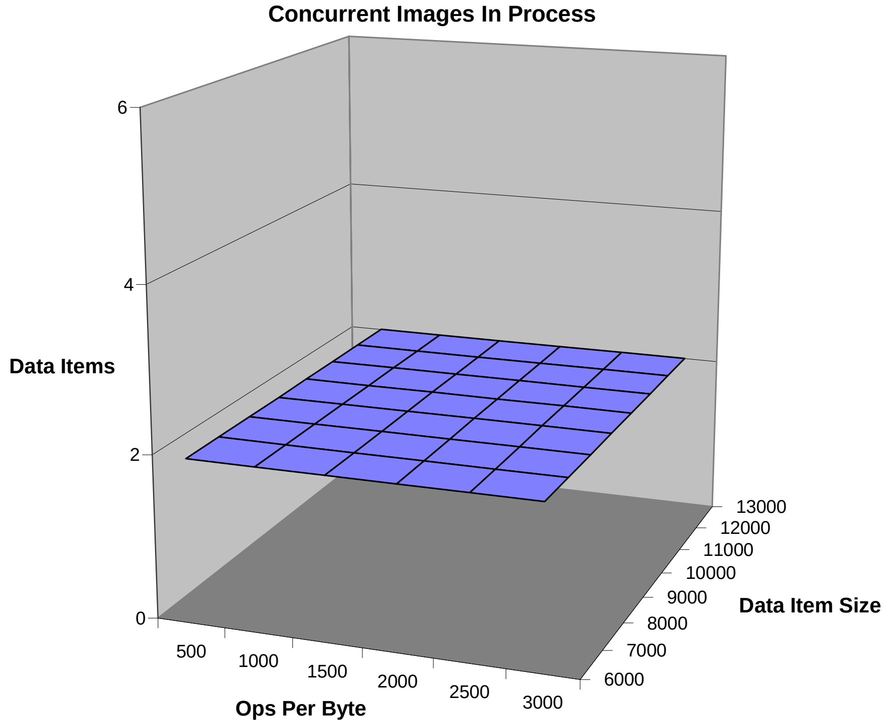


Figure 23 Number of Images processed concurrently in Year 2008

7.3 Resource Comparisons for Exploitation

7.3.1 Exploitation Resource Summary

Table 19 through Table 21 is a summary of all the processor, archive, and network needs for the exploitation function as calculated in Chapter V. These numbers set the baseline performance requirements from which 2008 computers, disks and network projections are determined.

Table 19 Summary Of Exploitation Compute Needs

Parameter	Units	Need
Exploitation Management		
Number of Analysts work positions	Positions	50
Management Computation Burden	Gigaflops/analyst position	250
Total Computational burden	Gigaflops	2500
Total de-rated computational burden		10000
Analyst Work Positions		
Number of Tailored Products	Tailored Products/minute	105
Size of tailored product to a work position	Mega Pixels	25
Simultaneous Software Executions	Applications	5
Number of Operations	Ops/pixel	6000
Computational Burden	Gigaflops	1250
De-rated Computational Burden	Gigaflops	5000
Total Exploitation Compute Needs		
Computational Burden	Gigaflops	11250
Total de-rated compute requirement	Gigaflops	15000

Table 20 Summary Of Exploitation Storage Needs

Parameter	Units	Need
Number of Active Problem Sets	Problems sets	1000
Number of products	Tailored products	6
Size of products and supporting data	Gigabytes	40.36
Archive Size	Gigabytes	242,160
Total Archive size required @.75 efficiency	Gigabytes	322880

Table 21 Summary Of Exploitation Network Needs

Parameter	Units	Need
Exploitation task inputs	Tasks/minute	10
Task size	Gigabytes/ sec	29.9
Transfer	Gigabits/sec	133.3

7.2.2 Exploitation Function's System Projections

Based on technology trends provided in Chapter VI, it is reasonable to project commercially available 72 Gigaflops processors in the 2008-timeframe. One can project that a 64 processor machine will be the commercial server standard and an eight-processor machine will be the commercial client standard. For system sizing, one can assume that the efficiency factor for multi-processors is 90%, one processor is used by the Operating System, and two are used for the network. Therefore the equivalent number of processors per server is ((64*0.9)-3=) 54 and number of processors client is ((8*0.9)-3 =) 4. To determine the number of computer systems, the number of required processors determined in Chapter V is divided by the number of equivalent processors per system. Table 22 summaries the 2008 computer needs for tasking.

With respect to archive sizing, Projecting a commercially available disk capacity of 2450 Gigabytes in the 2008 time frame yields a requirement for (322880/2450 =) 132 disk drives.

Table 22 2008 Projection for Exploitation

Parameter	Units	Need
Computer		
2008 Processor Performance	Gigaflops	72
2008 Processors/ Server Computer	72 Gigaflops	64
2008 Processors/ Client Computer	72 Gigaflops	8
Equivalent 2008 Processors/ Computer	72 Gigaflops	4
Calculated number of processors (|10000/72|)	Processors	138
Num of Server Computers (|138/54| +1 redundant)	Computers	4
Calculated number of client processors	Processors	3
Number of Client Computers (|3/4|)	Computers	1
Total Number of Client Computers	Client Computers	50
Storage		
2008 Disk Size	Gigabytes	2450
Active Disks In RAID Stripe	Disks	8
ECC Disks In RAID	Disks	1
Total Archive Size	Gigabytes	322880
Number of disk drives (322880/2450 =)	Disks	132
Number of ECC disks	Disks	17
Number of RAID	RAID Systems	17
Total Raid capacity	Gigabytes	333200
Network		
2008 Ethernet capability applied to Client	Gigabit/sec	10
Ethernet de-rating factor	Percent	5
Number of connections to each workstation	Connections	2
Network Load on one connection (|133/2|)	Gigabits	67
Transfer time per eight Gigabyte product	Seconds	13.4

7.4 Resource Comparisons for Dissemination

7.4.1 Dissemination Resource Summary

Table 19 through Table 21 is a summary of all the processor, archive and network needs for the dissemination function as calculated in Chapter V. These numbers set the baseline performance requirements from which 2008 computers, disks and network projections are determined.

Table 23 Summary Of Dissemination Compute Needs

Parameter	Units	Need
Load per Analyst work position	Gigaflops/ work position	25
Number of simultaneous positions	Analyst	50
Computational burden	Gigaflops	1250
De-rated Computational Burden	Gigaflops	5000

Table 24 Summary Of Dissemination Storage Needs

Parameter	Units	Need
Number of products	Products/ minute	100
Number of days retention on RAID	days	7
Product size	Gigabytes	8
Raw storage on portal RAID	Gigabytes	8064210
Total storage on portal RAID @.75 efficiency	Gigabytes	10752280

Table 25 Summary Of DisseminationNetwork Needs

Parameter	Units	Need
Desktop collaboration		
1600 dots by 1280 lines of displayable data size	Megabytes	2.048
Displaying delta change updates	Percent	.25
Refresh rate	Update/second	2
Total Collaboration Load	Megabits	8
Total amount over network@ .80 efficiency	Megabits/Second	10
Desktop Video-Teleconferencing		
250 dots by 250 lines of displayable data size	Megabits	0.5
Displaying delta change updates	Percent	100
Refresh rate	Update/Second	30
Load incoming	Megabits	120
Load outgoing	Megabits	120
Total Video-Teleconferencing Load	Megabits	240
Total amount over network@ .80 efficiency	Megabits/Second	300
Transfers to Portals		
Products	Products/5minute	1
Size products	Gigabytes	8
Product output rate	Megabits/Second	26
Total amount over network@ .80 efficiency	Megabits/Second	34
Total Dissemination		
Total amount over network	Megabits/Second	344

7.4.2 Dissemination Function's System Projections

Based on technology trends provided in Chapter VI, it is reasonable to project commercially available 72 Gigaflops processors in the 2008-timeframe. One can project that a 62-processor machine will be the commercial standard. For system sizing, one can assume that the efficiency factor for 64 processors is 90%, one processor is used by the operating system, and two are used for the network. Therefore the equivalent number of processors per system is ((64*0.9)-3 =) 54. Further, a prudent designer adds another machine to support future availability needs. In order to determine the number of computer systems, using the number of required processors determined in Chapter V and then divide by the number of equivalent processors per system. Table 22 summaries the 2008 computer needs for tasking.

Table 26 2008 Projection for Dissemination

Parameter	Units	Need
Computer		
2008 Processor Performance	Gigaflops	72
2008 Processors/ Server Computer	72 Gigaflops	64
2008 Processors/ Client Computer	72 Gigaflops	8
Equivalent 2008 Processors/ Computer	72 Gigaflops	4
Calculated number of client processors	Processors	.34
Number of Client Computers (|.34/4|)	Computers	1
Num of Exploitation in which dissemination reside	Client Computers	50
Storage		
2008 Disk Size	Gigabytes	2450
Active Disks In RAID Stripe	Disks	8
ECC Disks In RAID	Disks	1
Total Archive Size	Gigabytes	10,752,280
Number of disk drives (1072280/2450 =)	Disks	4289
Number of ECC disks	Disks	549
Number of RAID	RAID Systems	549
Total Raid capacity	Gigabytes	10,760,400
Network		
2008 Ethernet capability applied to Client	Gigabit/sec	10
Ethernet de-rating factor	Percent	.5
Number of connections to each workstation	Connections	2
Network Load on one connection (|.344/2|)	Gigabits	.172
Total collaboration update time over network	Seconds	.8

7.5 Projected TPED System

In summary, this paper described the current TPED system, its current domain and generic functions. Generic Design Science methods assured satisfaction, Experts and Specialists, Knowledge Awareness, and Tailored Products concepts were discovered to support a new multi-source TPED need for information superiority in time of crisis. The discovered concepts were used to develop a new design concept of operations that defines method and capabilities needed for the newly defined multi-source TPED systems using domain specific design and analyses methods. For each TPED domain, functions and components were derived and the system conceptual design was established. Using the multi-source TPED conceptual design, performance and sizing needs were defined. Finally, a commercial technology trend forecast was performed and a conceptual system architecture showing the number of computers, RAID storage devices, and network bandwidth needed to support a new multi-source TPED using projected 2008 capabilities. As shown by Figure 24, a new multi-source TPED system can be fielded that is a reasonable size using projected 2008 hardware technologies. However, further investigations in algorithm automation feasibility and development must be preformed to close the gap between today's capabilities and tomorrow's needs.

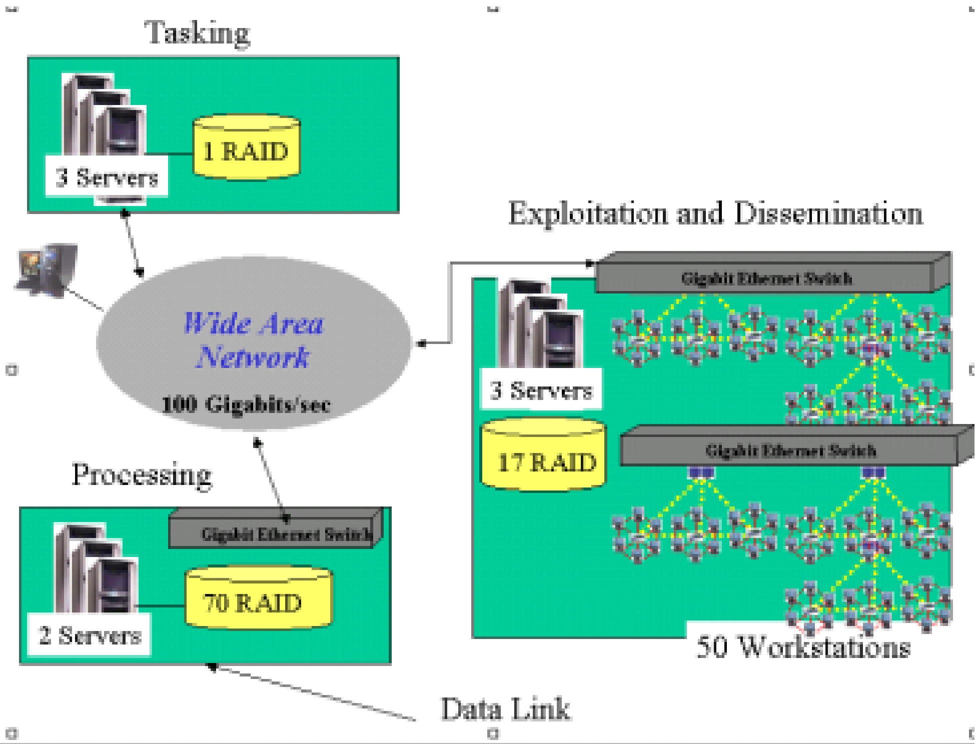


Figure 24 The New Multi-Source TPED project sizing with 2008 Hardware Technologies

REFERENCES

Raytheon 1997, *Systems Engineering Manual*, Reversion 2.1.

Plsek, P. E., 1997. *Creativity, Innovation, and Quality*. ASQ Quality Press

Osborn, A., 1953. Applied imagination. New York: Charles Scribner

Warfield, John N., A Science of Generic Design,. Second Edition, Iowa State University Press, 1994

Architecture Working Group, December 1997, Control, *Communications, Computers, Intelligence, Surveillance, and Reconnaissance* (*C4ISR) Architecture Framework, Version 2.0.*

National Imagery and Mapping Agency (NIMA), January 2000. *Dissemination vision 2010, Delivering the Information Edge Version 1.01*

National Imagery and Mapping Agency (NIMA), October 2000, *USIGS Information System CONOPS-2010, Version 2.0*

Independent Commission Final Report, *The Information Edge: Imagery Intelligence and Geospatial Information in an Evolving National Security Environment. 2000*

APPENDIX A

Spread Sheet Model Description

There are nine worksheets in the Microsoft EXCEL performance model. The first three worksheets are a group and each contains cells that are linked to cells in one or more of the other two. The changes in parameters are entered in the basic model worksheet. The Parameter Sizing worksheet contains embedded marcos which use the data from the first three worksheets to fill the tables. The four worksheets with surface plots reflect the data tables in the Parametric Sizing worksheet. In order to updates the tables in the Parametric Sizing worksheet, it is necessary to select Tools/Macros/Run macros.. The following is screen grabs of the first of worksheets and the visual basic program that performs the model computations.

BASIC MODEL

Hardware/Software Assumptions				
	Hardware Purchase Year	2008		
	Processor Improvements Per Year	1.5		
	Processor Efficiency	0.25		
	Disk Capacity Improvements Per Year	1.6		
	Active Disks Per Stripe	8		
	Max CPU Per System	64		
	Intrinsic System Scalability	0.9		
	Network Peak Channel Bandwidth	10	Gb/s	
	Network Channel Efficiency	0.6		
Data Assumptions				
	Comm Rate	5	Gb/s	
	Comm Rate Efficiency	0.8		
	Comm Bits/pixel or byte	3		
	Input:Output Pixel or Byte Ratio	1		
	Peak Output Data Item Per Second	1.5		
	Average Output Data Item Per Hour	6000		
	Average Output Data Item Height	13000		
	Average Output Data Item Width	13000		
	Sustained Operations/Pixel or Byte	3000		
	Sustained Gflops for Product Processir	760	Gflops	
Duty Information				
	Duty Cycle	0.5		
Data Item Product Timeline				
	Initial Processing			
	Collection - to - Capture on Disk	0.1014	sec	
	Capture - to - Processing	1.1014	sec	
	Image Processing	2.1304	sec	
	Processing - to - Disk	1.2704	sec	
	Total Time Available to Product Proces	4.6036	sec	
	Product Processing			
	Archive - to - Product Processing	1.2704	sec	
	Product Processing	1	sec	
	Product - to Archive	1	sec	
				Manual/forecast inputs =blue
	Total Time In Product Processing	3.2704	sec	Program Calculated +Red
	Total Time frm ProcTo Exploit	7.874	sec	

Data Items Assumptions

Capability				
PEAK				
	Output Images Per Second	1.5		
	Comm Efficiency	0.8		
	Input:Output Pixel or Byte Ratio	1		
	Comm Bits/pixel	3		
	Data Item Width	13000		Total Pixels Per Data Item
	Data Items Length	13000		
	Data ItemsPer Day	144000		
	Downlink Data Rate	5	Gbps	
	Image Size	0.507	Gbits	
	Image Downlink Time	0.1014	seconds	
	Ops Per Pixel	3000		
	Pixels/Sec	253500000		
	Operations/Sec	7.605E+11		
		760.5	Gigaflops/sec (Sustained)	
	Processor Efficiency	0.25		
	Peak	3042	Gigaflops/sec (Peak)	
AVERAGE				
	Output Images Per Hour	6000		
	Link Efficiency	0.8		
	Input:Output Pixel Ratio	1		
	Downlink Bits/pixel	3		
	Image Width	13000		Total Pixels or bytes Per Data Item
	Image Length	13000		
	Images Per Day	1548000		
	Downlink Data Rate	5	Gbps	
	Ops Per Pixel	3000		
	Duty Cycle	0.5		
	Total Downlink Per Day	98104.5	GB/day	
	Total Computed Per Day	98104.5	GB/day	
	Archive Days	7		
	Archive Efficiency	0.75		
	Total Required Storage	915996.67	GB	
		915.99667	Terabytes	

Computer Assumptions

	Price/Performance Factor										
Processors	1.5										
Memory	1.4										
Disk	1.6										
CAPACITY INFORMATION											
	Dec, 2000	July, 2001	July, 2002	July, 2003	July, 2004	July, 2005	July, 2006	July, 2007	July, 2008		
Disk Drive (Gbytes)	72	91	146	234	374	598	957	1531	2450		
Processor (Gflops)	2.2	4	6	9	14	21	32	48	72		
Archive Characteristics				Processing Characteristics					Product Processing		
	Active Disks In Stripe	8			Purchase Year	2008				Purchase Year	2008
	ECC Disks	1			Processor Peak	72				Processor Peak	72
	Purchase Year	2008			100% Efficiency Number o	42.25				100% Efficiency Number of Pr	42.25
	Disk Capacity	2450			Number Per System	64				Number Per System	64
	Total Number Of Disks	423			System Efficiency	0.9				System Efficiency	0.9
	Total Number of Stripes	47			System Effective Numbe	54	(1 for OS, 2 for network)			System Effective Number	54
					Number of Systems	1				Number of Systems	1
	Stripe Width For Capture	4			Additional For Failure	1				Additional For Failure	0.5
	Min. File Systems Per System	2			Total Number of Systems	2				Total Number of Systems	1.5
	Max. File Systems Per System	4									
	Minimum Number Systems	2	(Reliability)		Image Size	0.169	Gpixels				
					Flops Per Pixel	3000					
	Number Input Channels	1									
	Max Input Bandwidth	5	Gbps		Peak Flops Required	507	Gflops				
	Equivalent	625	MB/s		Effective System Peak Fl	3888	Gflops				
	Output Rate To Compute	625	MB/s		Processing Time	0.1304	sec			Processing Time	1
	Concurrent Outputs	2			Start-up Time	1	sec	(Not counted when doing multiple jobs since we will always			
	Number of File Transfers	3			Clean-up Time	1	sec	have one reading, one processing and one writing).			
	Number of File Systems So Low Prob of	5			Total Processing Time	2.1304					
	Total Number of Stripes	48									
	Total Disk Space	940.8	Terabytes								
Network Characteristics											
	Peak Network Bandwidth	10	Gbit/sec								
	Network Effeciency	0.5									
	Effective Throughput	5	Gbit/sec								
	Disk to Processing Image Transfer Time				Processing to Disk Image Transfer Time						
	Image Size	0.507	Gbits		Image Size	0.507	Gbits				
	Transfer Startup	1	sec		Transfer Startup	1	sec				
	Image Transfer Time	0.1014	sec		Image Transfer Time	0.2704	sec				
	Total Time	1.1014	sec		Total Time	1.2704	sec				

Parametric Sizing Tables

Automatically generated parametric sizing tables from assumption worksheets and visual basic program

Disk Sizing								
	6000	7000	8000	9000	10000	11000	12000	13000
500	392	392	392	470.4	627.2	705.6	862.4	940.8
1000	392	392	392	470.4	627.2	705.6	862.4	940.8
1500	392	392	392	470.4	627.2	705.6	862.4	940.8
2000	392	392	392	470.4	627.2	705.6	862.4	940.8
2500	392	392	392	470.4	627.2	705.6	862.4	940.8
3000	392	392	392	470.4	627.2	705.6	862.4	940.8

Processors (Including Product)								
	6000	7000	8000	9000	10000	11000	12000	13000
500	224	224	224	224	224	224	224	224
1000	224	224	224	224	224	224	224	224
1500	224	224	224	224	224	224	224	224
2000	224	224	224	224	224	224	224	224
2500	224	224	224	224	224	224	224	224
3000	224	224	224	224	224	224	224	224

SIP Timeline								
	6000	7000	8000	9000	10000	11000	12000	13000
500	7.16303	7.221901	7.28983	7.366817	7.45286	7.547961	7.652119	7.765334
1000	7.167659	7.228203	7.298061	7.377233	7.46572	7.563521	7.670637	7.787067
1500	7.172289	7.234504	7.306291	7.38765	7.47858	7.579082	7.689156	7.808801
2000	7.176919	7.240806	7.314522	7.398067	7.49144	7.594643	7.707674	7.830534
2500	7.181548	7.247107	7.322752	7.408483	7.5043	7.610203	7.726193	7.852268
3000	7.186178	7.253409	7.330983	7.4189	7.51716	7.625764	7.744711	7.874001

Number of Concurrent Images in Processing								
	6000	7000	8000	9000	10000	11000	12000	13000
500	2	2	2	2	2	2	2	2
1000	2	2	2	2	2	2	2	2
1500	2	2	2	2	2	2	2	2
2000	2	2	2	2	2	2	2	2
2500	2	2	2	2	2	2	2	2
3000	2	2	2	2	2	2	2	2

Visual Basic Program to Calculate model (1 of 2)

```
Option Explicit

Sub parmetricSizing()
Dim i, j, k, l, m, n, rowNum, colNum As Integer
Dim diskOffset As Integer
Dim processorOffset As Integer
Dim sipOffset As Integer
Dim concurrentOffset As Integer

Dim startOps As Integer
Dim endOps As Integer
Dim stepOps As Integer
Dim startImageSize As Integer
Dim endImageSize As Integer
Dim stepImageSize As Integer
Dim temp As Double

startOps = 500
endOps = 3000
stepOps = 500
startImageSize = 6000
endImageSize = 13000
stepImageSize = 1000

diskOffset = 2
processorOffset = 21
sipOffset = 41
concurrentOffset = 61
rowNum = 0
colNum = 1
'   Compute for peak of 3 data Items per second, 6000 data Items per hour
ThisWorkbook.Sheets("Basic Model").Cells(18, 3) = 1.5
ThisWorkbook.Sheets("Basic Model").Cells(19, 3) = 6000

For i = startOps To endOps Step stepOps
    rowNum = rowNum + 1
    colNum = 1
    ThisWorkbook.Sheets("Parametric Sizing 3x400").Cells(diskOffset + rowNum, 1) = i
```

Visual Basic Program to Calculate model (2 of 2)

```
    ThisWorkbook.Sheets("Parametric Sizing 3x400").Cells(sipOffset + rowNum, 1) = i
    ThisWorkbook.Sheets("Parametric Sizing 3x400").Cells(concurrentOffset + rowNum, 1) = i
    For j = startImageSize To endImageSize Step stepImageSize
        colNum = colNum + 1
        ThisWorkbook.Sheets("Parametric Sizing 3x400").Cells(diskOffset, colNum).Value = j
        ThisWorkbook.Sheets("Parametric Sizing 3x400").Cells(processorOffset, colNum).Value = j
        ThisWorkbook.Sheets("Parametric Sizing 3x400").Cells(sipOffset, colNum).Value = j
        ThisWorkbook.Sheets("Parametric Sizing 3x400").Cells(concurrentOffset, colNum).Value = j
                '    Put in the parameters into the model
        '
        ThisWorkbook.Sheets("Basic Model").Cells(21, 3) = j
        ThisWorkbook.Sheets("Basic Model").Cells(22, 3) = j
        ThisWorkbook.Sheets("Basic Model").Cells(24, 3) = i
        '
        '    Get the Disk Information
        '
        ThisWorkbook.Sheets("Parametric Sizing 3x400").Cells(diskOffset + rowNum, colNum) = _
            ThisWorkbook.Sheets("Computer Assumptions").Cells(35, 3)
        '
        '    Get the Processor Information
        '
        ThisWorkbook.Sheets("Parametric Sizing 3x400").Cells(processorOffset + rowNum, colNum) = _
            (ThisWorkbook.Sheets("Computer Assumptions").Cells(23, "G") + _
             ThisWorkbook.Sheets("Computer Assumptions").Cells(23, "L")) * _
             ThisWorkbook.Sheets("Basic Model").Cells(7, "C")
        '
        '    Get the SIP Timeline Information
        '
        ThisWorkbook.Sheets("Parametric Sizing 3x400").Cells(sipOffset + rowNum, colNum) = _
            ThisWorkbook.Sheets("Basic Model").Cells(52, 3)
        '
        '    Get the concurrent images Information
        '
        ThisWorkbook.Sheets("Parametric Sizing 3x400").Cells(concurrentOffset + rowNum, colNum) = _
            ThisWorkbook.Sheets("Computer Assumptions").Cells(23, "G")
    Next j
Next i
End Sub
```