Response to Feedback from NSF Site Visit
Addendum to the Purdue Multimedia Support Infrastructure (MSI) Proposal #9972883

A. Elmagarmid A. Ghafoor T. Korb K. Park E. Spafford

Department of Computer Sciences
and
School of Electrical and Computer Engineering

Purdue University
June 3, 1999

Summary

In this addendum we respond to the comments and written feedback provided by Dr. Mahaney and the referees after the NSF site visit of May 5, 1999. The main concerns expressed in the feedback were: (1) specification of deliverables and the use of selected MSI applications, (2) synergism among project teams, (3) system-building experience by some of the investigators, (4) Fisk University as an effective minority institution partner, and (5) a project time line with deliverables.

We address each of these points in this addendum, as well as clarify and provide additional information beyond what was presented in the proposal or during the site visit. Since, the site visit appears to have resolved most of the concerns of the initial referee reports and panel summary, we do not respond to those concerns in this report.

Overview, Deliverables, and MSI Applications

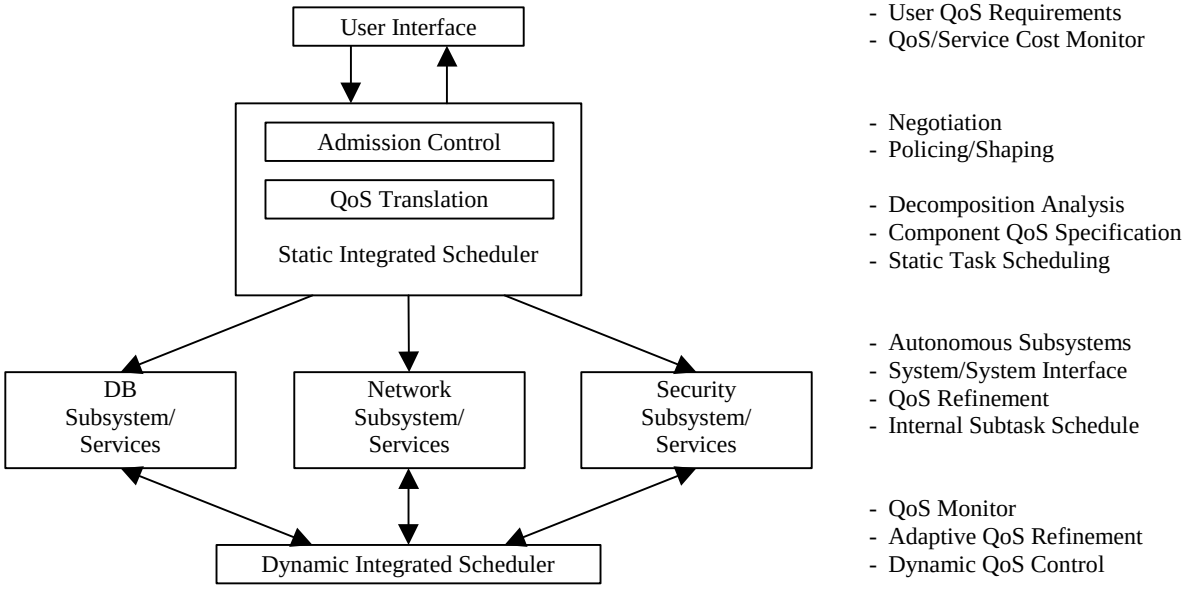
The principal activity of the Purdue Multimedia Support Infrastructure (MSI) project is to develop and evaluate a prototype system that will integrate key technology areas including database, networking, and security in order to support a comprehensive end-to-end Quality of Service (QoS) management framework for distributed multimedia document applications. The term QoS deals with the provisions for *user specified quality requirements*, including security, reliability, etc. We will evaluate the MSI architecture using the following three application domains. These domains have been selected due to their diverse demands on multimedia document access, storage, and delivery.

- Purdue On-Line: A distance learning activity using on-demand multimedia information distribution.
- Nuclear Engineering: A reactor safety project that requires high volume, high-resolution video transmission, in real-time.
- Veterinary Education: An educational program that needs large storage and indexing of high quality multimedia data including video, images, audio etc.

Since, MSI is planned to be a *software environment*, it will be able to support a wide variety of other application domains, such as electronic commerce.

Operational Overview of the MSI. At the highest level, MSI users select services such as accessing multimedia documents from specified applications, with desired QoS requirements. Users provide QoS requirements through a user/system INTERFACE, as shown in Figure 1. Based on such requirements, the system executes an admission control mechanism in order to determine the feasible levels of QoS provision for the user request. The admission control mechanisms can have varying degrees of sophistication ranging from simple best-effort services

to highly intelligent resource reservations schemes involving complex interplay among all the system components including network, multimedia database servers, operating systems, and security. The admitted user request is processed by such components in order to determine their QoS requirements. This process is called QoS translation. Through a set of pre-defined system/system interface calls, the system components interact with each other in order to determine a suitable system-wide operating point to satisfy the desired QoS levels. This static scheduling is carried out by the *Static Integrated Scheduler* (SIS). Static scheduling is



subsequently used by the system components in order to perform their own internal scheduling and allocation of resources to achieve the required QoS.

The *Dynamic Integrated Scheduler* (DIS), while monitoring the state of the system, uses the static schedule computed by SIS for adjusting the scheduling decisions, if necessary. .DIS can exploit the full power of *hybrid scheduling*, in which the same QoS may be achieved by different schedules that make non-uniform demands on database, networking, and security-related resources. Hybrid scheduling can provide adaptive, and cost-effective solutions , and can yield high utilization of system resources.

Deliverable Software Tools. As part of the MSI project, we expect to develop a number of software tools, specifications, and environments that will be portable to diverse platforms. A significant part of the MSI software will be based on expanding upon several on-going projects by the investigators in the areas of database, networking, and security, thus leveraging existing projects, wherever possible. In addition to extensive new development, substantial effort will be employed for the integration of projects from three technology areas. These deliverables are summarized below.

- *MSI Database, Networking, and Security Subsystems:* Prototype systems to be delivered will be an extensive integration of several sub-systems and modules. These prototype systems include distributed multimedia document management sub-systems; hierarchical storage management sub-system for multimedia data, network QoS scheduling modules; real-time CPU scheduling techniques; adaptive security systems using autonomous agents,

watermarking and compression. Our existing software base in MSI related areas is expected to accelerate the development of these deliverables.

- *Interface Development*. Two sets of interfaces will be developed to allow user/system, and system/system interface. The user/system interface will allow users to specify high-level services and their Quality of Presentation (QoP) characteristics. We plan to create a hierarchical classification of generalized QoS requirements that include QoP, network QoS, security, and reliability. The system/system interface will allow interaction between different MSI system components to achieve seamless integration among them.
- *QoS Translator and Integrated Scheduling Modules.* For the development of QoS translator, intelligent techniques will be designed to map QoP to QoS requirements for individual technology components. For resource scheduling, the architectures of both the static (SIS) and dynamic (DIS) scheduling modules will be developed. Their design will be largely platform independent making them portable and customizable to different environments. In addition to SIS and DIS modules, techniques for sophisticated hybrid scheduling involving both SIS and DIS will also be developed.
- *GUI-Based Application Interface*. We will use the three MSI applications described earlier to design and test a GUI-based user interface. This interface will allow the invocation of internal MSI sub-systems and will be highly extensible and customizable to individual users and allow web-based access.

The above list illustrates the core set of software tools and environments that we expect to develop as part of the MSI project. Additional items and details are given in the *Project Time Line and Deliverables* section at the end of this report.

MSI Applications and Deliverables. The three principal application domains, i.e. Purdue On-Line, Nuclear Engineering, and Veterinary Education, have been selected due to their diverse demands on multimedia document access, storage, and delivery. The main research contribution of the MSI project will be in the formulation, prototyping, and testing of an integrated framework that will ensure end-to-end QoS management for distributed multimedia applications. The main tangible output, above and beyond conceptual research and educational contributions, will be the development of software tools and environments described above.

Project Team Synergy

The technical challenges that underlie the proposed MSI project require strong interdisciplinary expertise in three key technological areas: database systems, networking, and security. Collectively, the researchers have strong track records in these areas, with extensive experience in designing and building experimental systems and managing several current projects relevant to MSI. Several of these projects involve collaboration between the investigators. The investigators co-author publications, jointly supervise Ph.D. students, serve on thesis committees, and serve on Purdue-wide technology committees. These wide-ranging professional relationships among the investigators will facilitate the MSI development through collaborations in the three key technology areas.

Previous and On-Going System Building Activities

Several investigators, including Delp, Spafford, and Park, have significant system development experience, as also apparent from comments by the reviewers. Description about such experience for the remaining investigators is described below.

Much of the PI's (Elmagarmid) work during the past several years has been in the area of developing systems for integration of heterogeneous multi-database systems. This work was supported by a five-year PYI award until 1994 and led to the creation of three multi-database

prototypes: InterBase, InterBase* and InterSect. Through technology transfer, InterBase and InterBase* have made significant impact on many commercial systems. For example, these systems have been deployed at Bell communications Research, GTE Government Systems, and at Bell Northern Research (now a part of Nortel), MCC, and InterSql. At BNR, InterBase has been used in daily operation for statistical analysis of databases distributed over many of their sites throughout the world. The InterSect project has resulted in a startup company, led to the development of a commercial product for transaction monitoring, with pending patents. Two books describing the development of InterBase and InterBase* were published in 1998.

In the area of multimedia database systems, the PI and the co-PIs have several ongoing experimental and theoretical research projects, some of which are listed below.

WVTDB—A Semantic Content-based Video Database System on the World Wide Web

Support: Intel, Bellcore, IBM, and Purdue
Duration: 1996-Present
Investigator: A. Elmagarmid

This project designed, developed, and is now performing application testing of a web-based video database system that demonstrates research on video data modeling, semantic content-based video queries, and a video database systems architecture. The system is based on multi-level video data abstractions and annotation layering, thus allowing dynamic and incremental video annotation and indexing, multi-user view sharing, and video data reuse. Users can query, retrieve, and browse video data based on semantic content descriptions and temporal constraints on the video segments. WVTDB employs a modular system architecture that supports distributed video query processing and sub-query caching. Several techniques, such as video wrappers and lazy delivery, were implemented specifically to address the network bandwidth limitations for this kind of web-based system. Most recently, we have been augmenting the system with mechanisms for access and authorization control, both discretionary and mandatory. Further details about this project may be found at `http://proximity.cs.purdue.edu:7000/vdb`.

Multilevel Representation and Query Processing in Multimedia Database Systems

Support: NSF
Duration: 9/1997-9/2000
Investigators: A. Ghafoor, R. L. Kashyap, S. Moni

This project is developing a video database system prototype with capabilities to handle heterogeneous media queries. The system caters to the computational and storage requirements while accommodating and exploiting the inevitable semantic and representational imprecisions. The design of the system is based on multilevel data models and search mechanisms. These methodologies allow the user to pose various types of queries, including low-level, such as finding objects in a multimedia database, and mid-level, based on spatio-temporal semantics, such as locating events associated with video data. A brief summary of this project is given at `http://shay.ecn.purdue.edu/~dmultlab`.

Multimedia Document Management System

Support: NSF, Siemens Corporate Research, Fuji Electric Company
Duration: 1996-present
Investigator: A. Ghafoor

Multimedia document management systems provide environments that allow the generation, representation, processing, storage, and dissemination of independent machine processable information composed of multiple media types such as text, video and audio. Part of this project

is to develop an object-oriented multi-layered architecture of a multimedia document management system. The architecture of the system has the following components:

- A multimedia user environment that provides interactive capabilities for authoring, querying and presenting multimedia documents. The environment allows users to specify document attributes including QoP parameters.
- A multimedia query processor that allows users to store, query and retrieve multimedia documents from the underlying multimedia database systems.
- A presentation manager responsible for scheduling the events and actions to play back and synchronize multimedia objects during document presentation.

Content-Based Modeling and Retrieval of Video Data

Support: NSF
Duration: 7/1999-7/2001
Investigators: A. Elmagarmid and A. Ghafoor

The objective of this research is to design and develop novel data models to provide the enhanced semantic expressiveness needed for content-based retrieval of video data. The key feature is to allow retrieval using keywords and video events involving object motion. This hybrid approach can substantially enhance the degree of precision and scalability of retrieval mechanisms for video data. This project will have a significant impact on the development of database tools for content-based indexing and retrieval of video data.

Women, Minorities, and Outreach Activities

The School of Science has a significant investment in attracting and retaining women and minorities as well as a substantial program in K-12 outreach. Two people are involved in the women and minorities programs; five full-time people are involved in outreach. The Department of Computer Sciences alone has one full-time outreach coordinator and other part-time faculty, staff, and student involvement in support of the women's program.

Women's Program

In 1997, the School of Science—committed to making careers in all areas of science accessible to female students—opened an office for the Women in Science Program (WISP). The goal of WISP is to increase the retention of female science students. WISP includes a Residence Hall Program, a Tutoring Center, Undergraduate and Graduate Mentoring Programs, and provides a limited number of scholarships to incoming freshman women.

Early statistics on the success of the programs are very promising. After four semesters, 93% of the 1997 beginners in WISP are still at Purdue and 61% of the beginners are still in the School of Science. Of the closely matched control group of women in science who didn't participate in WISP, 70% remain at Purdue and only 38% are still in science. The women in WISP also had higher grade point averages than the control group. The Women in Science Program almost doubled in size during the two years since it began and the percent of female computer science freshmen choosing to participate in the program went from 19.3% in 1997 to 34.3% in 1998. Although nationally, retention of women in science is a difficult problem, early evidence for success of the Purdue WISP program shows impressive promise.

Minorities

Fisk University and Purdue University participate in a research consortium (CIMEG) that is funded by EPRI/ARO. Prof. John Caulfield from Fisk and Prof. Ahmed Elmagarmid from Purdue are two of the co-PIs on this project. We are aware of the recent departure of Prof. Elaine

Eschen from Fisk and the problems they face with retention and recruitment of faculty. The attraction at Fisk for us is their track record of graduating excellent undergraduate students. Fisk has only 850 students with 70% females and 40% of those students in the Sciences. One in six of all African-American physicians are Fisk graduates and more Fisk graduates earn Ph.D. degrees, in proportion to enrollment, than any other school in the United States. The Purdue involvement with Fisk has two purposes: to recruit well-qualified students for graduate work at Purdue and to engage Fisk faculty in doing joint research. We anticipate the relationship with Purdue to indirectly help Fisk recruit new faculty. With the help of the minority offices at the School of Science and the office of the Vice President at Purdue, we are looking beyond Fisk. We have received notifications from Spelman College and Florida A&M University—two of the finest HBCUs in the nation—that they will participate in our program.

Our former alumnae, Dr. Andrea Lawrence, Chair of Computer Science and Associate Professor at Spelman, and Prof. Ahmed Elmagarmid, the PI of this proposal, have discussed Spelman student involvement in a Summer Institute program. We have received letters from the Honorable Frederick S. Humphries, President of Spelman, and the Honorable Audrey F. Manley, President of Florida A&M, committing their institutions to involvement in the Summer Institute at Purdue. The Summer Institute we outlined during the site visit will, in our opinion, go a long way towards improving the chances of success for the incoming minority students. It will focus on grounding the students in the areas of database systems, security, and networks and in the second year especially it will concentrate on preparing the returning students for our qualifying examinations. The follow up is a cornerstone of this summer institute program: Faculty mentors will follow up with the students after they return to their home institutions.

Outreach

The purpose of the School of Science K-12 Outreach Program is to increase achievement and interest in science and mathematics at all pre-college levels in Indiana. Coordinators from the six major departments in the School work with teachers and students in public schools all over the state and at Purdue, providing hands-on experiences in their respective science disciplines. The Outreach Program has made contact with 300 schools and 300,000 students over the past nine years.

Each year, the Outreach Program in the Department of Computer Sciences alone has visits with nearly 40 schools and conducts almost 100 workshops.

Project Time Line and Deliverables

In this section we provide a project time line in terms of activities to be undertaken for the five year grant period. We also provide detailed descriptions of deliverables for each year.

Year 1

During the first year, we plan to undertake activity in the following three areas:

- Deploy equipment related to key technologies for the project, including networking, database systems, and capture and presentation equipment. This equipment will be used in existing labs and as well as a new MSI Lab to be established in the CS building. Equipment management in these labs will be coordinated through dedicated, full-time staff.
- Design of consistent and extensible interfaces for the primary MSI subsystems. This design will include a web-based interface to allow application users to specify QoP requirements. In addition, we plan to design system modules for QoP-to-QoS translation for individual technology components, as well as integrated QoS (SIS and DIS) scheduling modules.

- Develop prototype systems in each of the three application areas. Several such prototypes are currently being developed by the co-PIs as part of their ongoing research. (Some were demonstrated at the site visit.) We plan to port and start integration of these prototypes using the proposed MSI interfaces.
- Develop a security services library for networking and database components.

In summary, the Year 1 goal is to develop the desired set of interfaces and prototype an initial system that will allow experimentation and testing of QoS-based research related to integration of the three application areas.

Year 2

The second year of the project will be to enhance and integrate the components of the initial system. This development will be achieved through the activities listed below.

- Implement QoP-to-QoS translation and integrated QoS scheduling systems and benchmark with generic QoP-sensitive multimedia document retrieval systems. For this purpose, we will develop multimedia document systems in the three application areas. Experimentation with such systems will be performed using user-specified QoP parameters supplied through a web-based interface (user/system interface). Such experimentation will be focused on evaluating interaction among the QoS layers.
- Enhance multimedia document database server by incorporating research results into developing high-performance hierarchical storage management (HSM) systems. This task will entail development of a RAID system coupled with one of the proposed tertiary storage systems and a clustered workstation environment integrated with the second tertiary storage system. The storage management, data migration, and disk scheduling will be tightly integrated with a QoP-to-QoS translation module as well as integrated QoS scheduling subsystem. We will develop compression techniques for multimedia document storage systems. Multimedia document database schema will be enhanced to incorporate detailed security features for multimedia document database systems allowing varying granularity of information access.
- Develop a distributed document management system. This activity will focus on developing a multimedia document meta-schema that will allow efficient, indexing and browsing of distributed documents. Such schema will be built upon a centralized directory system for linking distributed document components. Documents will be stored at six different servers, which will be connected using the interfaces provided by the GS/SBS network architecture. Integration of these servers over the network will start towards the end of Year 2.
- Implement enhanced GS/SBS network architecture. This activity will involve the deployment of a full-fledged GS/SBS architecture, including export of refined QoS-sensitive services based on end-to-end QoS amplification. Several resource allocation algorithms and security provision mechanisms will be implemented as part of this enhanced architecture. In particular, token-based resource allocation techniques will be implemented, which can further be used to include a pricing scheme within the user QoS requirement space. Development of a real-time operating system will be undertaken to incorporate features that will multithread security services.
- Incorporate an agent-based, reactive security system (e.g., AAFID and AdMon) into the MSI design for security, reliability, and availability. We will perform network measurement and collect audit data for intrusion detection.

Year 3

By the beginning of Year 3, the functionality of the major subsystems will be largely complete. The objective in this year is to finalize the integration of these components and to perform comprehensive testing and measurement. We will also pursue activities towards possible standardization of our research results in relevant forums (e.g., CORBA for distributed systems and databases, and the IETF for networking and security). Accordingly, the following activities are planned for this year.

- Complete the distributed multimedia document system. This activity will involve full-fledged integration of six distributed multimedia document servers with the GS/SBS network architecture. System features will include distributed query processing, data filtering by the servers' real-time operating systems, and multimedia data synchronization using intelligent network data caching techniques. Two of the document servers will consist of HSM systems developed during Year 2.
- Extend QoS translation and integrated QoS scheduling to include performance comparison of alternative scheduling algorithms. This activity will entail enhancement of QoP-to-QoS translation and integrated QoS scheduling modules to have a provision for implementation and testing of arbitrary scheduling techniques. In particular, we will conduct comparative evaluation of our techniques with other IETF DiffServ proposals around the year 2001. We plan to implement token-based pricing scheme for resource contention resolution and priority management. Network measurement data will be used to analyze and evaluate efficiency, stability, and "goodness" of resource sharing behavior.
- Develop database measurement tools to collect traces of multimedia document access behavior. The objective is to characterize document traffic profiles for traffic modeling and analyze the heavy-tailedness of multimedia document size distribution and its impact on traffic burstiness, in particular, self-similarity. For the development of such tools, we plan to use recently developed techniques for self-similar traffic control and load balancing
- Evaluate agent-based reactive security system (AAFID and AdMon). We will implement such a system for the proposed MSI and carry out extensive evaluation in terms of responsiveness, accuracy, and overhead under various forms of attack. We plan to extend such system to incorporate multiple levels of alarm, audit, and footprint features. Experimentation with such a system over the proposed MSI will allow us to analyze its impact on user-level, end-to-end QoS above and beyond basic security and availability. We will enhance integration of this system with real-time scheduling to tune and confine its performance impact to predictable levels.

Year 4

In this year, we plan to introduce a version of MSI into the public domain accessible via Internet2 and, in a more limited fashion, via the commodity Internet. The three principal application areas will be made available to users both inside and outside of Purdue. Such access to the MSI prototypes will allow users to experiment with techniques developed in the three key technology areas, including database systems, networking, and security. Access to network measurement data and selected audit data will allow the Internet community to study traffic and user behavior patterns in multimedia high-speed network environments.

A significant portion of Year 4 will be devoted to investigate deployability issues. In particular, we will address these issues in the context of interoperability between the external, non-MSI networks and systems developed at Purdue. Techniques such as QoS tunneling and end-to-end QoS amplification will be implemented to minimize the effect of "black box components" which are not under the control of MSI. This approach will also be integrated with active error control

performed at the information content level in variable video compression. A demonstration of some of the techniques was shown at the site visit. The efforts in QoS translation and integrated QoS scheduling are expected to lead to a rich set of research problems.

With respect to the public release of MSI, significant effort will be devoted to the dissemination of the lessons learned from our efforts. This information will be made available through publications, seminars, and an invited workshop. Concerted effort will be expended to introduce the architecture, tools, and algorithms to standards bodies for consideration.

Year 5

We expect to release the second public domain version of MSI to the research community and industrial affiliates in Year 5. This release will incorporate fixes and adjustments based on the users feedback and experience gained during the first release in Year 4. In essence, Year 5, in part, will be a continuation of Year 4, whereby learning from the comments and interaction with the larger research community will help in carrying out further refinements to MSI, and enabling technology transfer.

We expect to devote significant time and effort in planning follow-up activities, including the identification of new research problems, tackling new challenges, and formulating approaches to generate viable solutions. Although the bulk of our research activities as part of MSI will concentrate on system development, our efforts to obtain effective solutions for the integrated QoS management framework for MSI are also expected to yield the development of several interesting and valuable results in all the key technology areas of MSI.

The final, tangible output of the project will be a working, portable MSI system consisting of highly integrated prototype subsystems including: distributed multimedia document management sub-systems; hierarchical storage management sub-system for multimedia data, integrated QoS translation and hybrid scheduling modules; and adaptive security sub-systems.

In conjunction with research publications, seminars, educational outreach, and technology transfer to the research community at large, the MSI project is expected to yield significant contributions to pressing and challenging research problems that can only be satisfactorily addressed by an integrated framework for QoS management spanning the key technology areas of database systems, networking, and security.