Grid Computing
The Biometrics Grid: A Solution to Biometric Technologies
Anlong Ming and Huadong Ma • Beijing University of Posts and Telecommunications

By simplifying resource management and making data easy to access, the biometrics grid could benefit academic collaborations and enhance biometric systems’ efficiency in scientific research and commercial applications.

Biometric technologies use individuals’ unique, measurable biological and behavioral traits to automatically establish or verify their identities. These technologies are critical to domains such as person authorization in e-banking and e-commerce transactions or within the framework of access controls to security areas. These systems require not only advanced biometric technology interfaces but also the ability to deal with security and privacy issues. Integrating biometrics with access-control mechanisms and information security is another area of growing interest. The challenge to the research community is to develop integrated solutions to handle these problems in their entirety, from data acquisition and analysis to system design.

Grid systems gather distributed and heterogeneous resources. Grid computing’s ultimate goal is to create a global infrastructure for distributed system integration and resource sharing. The biometrics grid is a specialized system based on grid middlewares for biometric applications. Using grid functionalities, BMG can improve academic collaborations, provide biometric-data-sharing capabilities, and make biometric systems more efficient. BMG is feasible owing to the convergence of grid middlewares (such as the Globus Toolkit (http://www-unix.globus.org/toolkit/docs/4.0) and ChinaGrid Support Platform) and the development of biometric technologies.

How grids can help biometrics
In the age of digital impersonation, biometric techniques are increasingly providing a hedge against identity theft. Essentially, a biometric system is a pattern-recognition system. Single or multimodal biometric systems traditionally have three applications:

- **physical access** control to keep unauthorized people from obtaining access to specific places or rooms,
- **logical access** control for protecting networks and computers, and
- **time and attendance** control.

Challenges in biometric technologies
Other than performance issues such as failure to acquire (that is, failure to capture and extract biometric data) and failure to enroll (a system’s failure to form a proper enrollment template for an end user), biometric technologies face three principle challenges.

*Cooperation.* Biometrics spans various technologies such as fingerprint and face recognition, mathematics and statistics, system design and integration, and, last but not least, security and privacy. So, scientists and practitioners from the diverse fields of computing, sensor technologies, law enforcement, and the social sciences should exchange ideas and research results.
Data sharing. At present, most commercially available biometric technologies are either intraorganizational or operated only by people in the application domain—for example, Cognitec’s FaceVACS-SDK. The duplicated work in building test databases is not only wasteful but also causes difficulty in providing uniform performance standards.

Efficiency. Some biometric systems have various architectures for different scenarios. Independent biometric systems are limited to personal-identity authentication, however, principally because they aren’t considered sufficiently accurate and user acceptable in all scenarios. Also, a variety of heterogenous resources aren’t available in an independent biometric system. Usually, a biometric system is designed for a particular scenario or for use in a local place, but is used infrequently. We have all kinds of biometric systems and fast means to connect them—why not use them efficiently?

BMG’s advantages

The Web, CORBA, clusters, distributed computing, and peer-to-peer and virtualization technologies are examples of existing ways to undertake wide-area distributed computing. Grid computing (http://www-03.ibm.com/grid/about_grid/what_is.shtml) evolved from these methods and from client-server models that can provide fault-tolerant storage for data backup and recovery. (See the “Existing Grids and Applications” sidebar for examples of current Grid implementations.) Biometric technologies need a solution to enable the virtualization of numerous geographically distributed biometric systems and biometric data resources to create a single system image, granting users and applications seamless access to vast capabilities. Just as an Internet user views a unified instance of content via the Web, BMG can present to users a single, large virtual biometric system.

Scientific collaboration in biometrics. Grid computing is key to scientific collaboration. In a typical scientific collaboration, data is stored in many different locations. Grid computing lets scientists store and manage their files on local storage facilities and share them with remote participants. Grid computing aims to let users collaborate easily—that is, to locate each other, use asynchronous and synchronous messages, and share documents, progress, and applications.

Biometric resource sharing and on-demand biometric computing. Grid computing creates a virtual organization, enabling sharing of applications and data and linking of information stored at multiple locations. Furthermore, it enables on-demand access to computing, storage, and other devices. On-demand computing is essential to more fully leverage existing biometric systems by intelligently allocating finite resources to the appropriate applications on a grid.

Integration of biometric systems. Grid computing isn’t just about bringing systems together. It’s about integrating them onto a single large base, which presents a virtual view of the whole system. Grid computing can incorporate many biometric systems onto a single grid of interconnected ones, which is accessible through an entrance such as a Web portal. This lets users access biometric systems remotely.

Existing Grids and Applications

Several science and commercial applications use Grid technology. EGEE (Enabling Grids for E-Science) is a follow-up project of the European DataGrid project. Funded by the European Commission, it brings together experts from over 27 European countries. EGEE aims to give academia and industry researchers access to major computing resources, independent of their geographic locations. The Large Hadron Collider Computing Grid (LCG; http://lcg.web.cern.ch/LCG) aims to build and maintain a data storage and analysis infrastructure for the entire high-energy physics community that will use the LHC. TeraGrid (http://www.teragrid.org) aims to make it as easy and transparent as possible for researchers to move jobs and data freely among machines within and outside TeraGrid sites. Established by five Nordic academic institutes, the NorduGrid R&D collaboration aims to develop, maintain, and support the Advanced Resource Connector free grid middleware.
EGEE provides a production-quality grid infrastructure with more than 150 sites spanning more than 30 countries to a myriad of applications, from various scientific domains including earth sciences, high-energy physics, bioinformatics, and astrophysics. Although the EGEE project doesn’t impose any fabric-management policy, participating computing centers have expressed interest in adopting the Quattor framework (http://quattor.web.cern.ch/quattor). Also, TeraGrid is increasingly seen as a system that not only provides computing resources but also hosts services for communities. These emerging functions show grid computing’s role in supporting scientific collaboration. NorduGrid collaborative activity is based on the success of the Nordic Testbed for Wide Area Computing and Data Handling project (funded by the Nordic Council of Ministers via the Nordunet2 program). Both NorduGrid and the LCG’s support of physics experiments inspired the idea of academic testbeds for biometric algorithms in BMG.

References

A virtual collaborative environment
BMG creates a virtual collaborative environment, which provides advantages that could promote cooperation among scientists and practitioners in diverse fields. Although BMG is still in the conceptual stage, we have created a simple prototype system.

System design objective
BMG has several goals:

- Create a virtual collaborative environment linking distributed biometric systems, users, models, and data to simplify resource management.
- Offer a set of security mechanisms for access control and cross-domain storage of biometrics data.
- Provide an academic testbed for research on biometrics algorithms. The testbed will make researchers focus most or all of their attention on algorithm design by testing the modules they design on uniform databases.
- Support biometric applications with different QoS (quality of service) demands, including applications with large-scale databases or multimodal biometrics applications.

A BMG framework
Figure 1 shows a BMG framework. BMG has four layers:

- The portal. This controls the user interface and interaction and provides an aggregation service for users.
- BMG-specific services. The testbed is an important service for scientific research that lets researchers focus on algorithm design. Moreover, biometric systems are integrated by wrapping biometrics applications into Web services and are deployed into a Web services container.
- The platform. This provides remote, authenticated access to shared BMG resources. All the components are based on the Globus Toolkit.
- Resources. These include computational and data resources.
Portal services
The BMG portal aims to provide an aggregation service for a content provider's BMG applications. Generally, a portal lets you access multiple Web pages, automatically creates controls to link among them, and displays subsets of them on a single Web page. We assume that all data and information presented to users originates from a Web service.

Job scheduling
When several user jobs apply for concurrent execution, BMG must handle problems due to system failures or delays due to infected hardware, software vulnerability, or distrusted security policies. To enforce job-scheduling reliability under risky conditions, we use risk-resilient strategies. To determine which strategy to use, we first get a BMG job's security demand, and a resource provider at the BMG site provides the trust level. (The trust level quantifies how much a user can trust a site to successfully execute a given job.) When the security demand is lower than the trust level \( (SD < TL) \), the job can be successfully carried out. Second, BMG issues a security demand to all available resource sites and assesses the trust level of a job's resource site. The system could delay or drop a job if the security demand is higher than the trust level. Finally, the scheduler selects a strategy:

- **Failure** mode. The scheduler does nothing except return a failure report.
- **Preemptive** mode. The scheduler preempts a job when a site is inaccessible, then tries to carry it out at another site.
- **Replication** mode. The scheduler duplicates the job at multiple sites (all \( SD > TL \)). When a replicated job is successfully carried out, all the others will abort.
- **Delay-tolerant** mode. The scheduler reschedules the job to other sites after a period of time.

BMG's security mechanism
Security (including naming and authentication; secure communication; trust, policy, and authorization; and enforcement of access control) is critical in grid computing. BMG security will consider two security issues: enforcement of access control and cross-domain storage in secure communication.
Access control

The Open Grid Services Architecture-Data Access and Integration grid middleware provides role-based access control. However, OGSA-DAI by itself isn’t suitable for BMG because users and resources are dynamic in BMG.

In BMG, the Community Authorization Service,9 provided by the Globus Toolkit, complements OGSA-DAI’s access control mechanism. The CAS aims to control access to computational resources and file-based storage resources by recording user groups and their permissions for resources.

BMG users can obtain CAS credentials based on their user credentials. A certificate authority that all BMG entities trust issues the associated public and private keys. The CAS credentials contain users’ authorization information in terms of their virtual organization roles. We intend to extend OGSA-DAI to pass the CAS credential, to extend the server side to parse the credential to obtain the VO role in which the user is to be authorized, and to map the VO role to a local database role.

Cross-domain storage

When a user submits a job over BMG, he or she doesn’t know on which nodes it will be executed if the data reside on different sites. Furthermore, where will BMG store the resulting data when the job is done?

In a grid environment, distributing storage is more sensible than centralizing it.10 Cross-domain storage can be accomplished using third-party transfer, which moves data from a remote site to the disk cache or requesting computer by separating the data-control and data-movement functions in the archive. Special tools are needed to allow the data to be found on BMG, as well as to be copied or replicated. These tools should hide most of the complexity. Moreover, biometrics applications also need management because they might store additional metadata with their files.

GridFTP implements these tools by dividing the data to be transferred into portions and transferring each portion with a separate TCP connection. The GridFTP implementation in the Globus Toolkit consists of a family of tools: the GridFTP server, the client program, the client library, and the control libraries administer the data and control connections of an FTP session. The GridFTP driver manipulates and transports user data.

An academic testbed for biometric algorithms

Researchers only need to design some modules of feature extracting, feature matching, and information fusion according to the testbed interface and then submit them to BMG for comparison and evaluation.

Single-biometric test

A biometric system has a general structure (see figure 2).11 First, a sensor acquires a sample (a biometric measure that the user presents), such as a fingerprint or a face or iris image. The data-collection subsystem captures it as an image or signal. BMG stores the complete sample data in the storage unit. It uses and stores only a mathematical representation of the information, which the signal-processing module extracts from the sample; the system will use this representation to construct or compare the biometric feature with enrollment templates. If the extracted feature is stored (enrolled) into BMG, a template for future identification or verification (matching) is added. BMG has a measure of the similarity between features derived from a presented sample and a stored template. The measure produces a typical index called the matching score. The system can make a decision according to whether this score exceeds a certain threshold. A transaction is a user’s attempt to validate a claim of identity or nonidentity by consecutively submitting one or more samples as the system’s decision policy allows.12 Finally, a transmission process transmits the collected data to the signal-processing section.
BMG deploys this general single-biometric system on the testbed. Of course, BMG should first define the interfaces, such as the feature-extracting, feature-matching, and prefiltering interfaces. All these together are defined as part of the testbed interface description. For example, a simple feature-extracting interface can be defined as a binary file (such as FeatureExt) with a parameter (such as a sample’s filename). When a user submits FeatureExt to the BMG Web portal, FeatureExt itself will be wrapped into a Web service specified by the Web Services Resource Framework. BMG can then provide this service as a part of the testbed using the GRAM server. (GRAM is Grid Resource Allocation Management, which provides a single interface for requesting and using remote system resources for executing jobs.) The Globus Resource Specification Language provides a common interchange language to describe resources. The Globus resource management architecture’s various components manipulate RSL strings to perform their management functions in cooperation with the other system components.

**Multimodal biometrics test**

Multimodal biometrics fusion is possible when combining multiple biometric systems (see figure 3). Fusion can occur at three levels:

- The feature-extraction level concatenates features extracted using multiple sensors.
- The confidence level combines matching scores reported by multiple matchers.$^{11,12}$
- The abstract level consolidates multiple systems’ accept or reject decisions.$^{13}$
In the context of biometrics, fusion can take three modes:

- Single-biometric multiple-classifier fusion combines multiple classifiers on a single biometric indicator.\textsuperscript{14}
- Single-biometric multiple-matcher fusion combines scores generated by multiple matching strategies (on the same representation).\textsuperscript{15}
- Multiple-biometric fusion uses multiple biometrics.\textsuperscript{16-19}

Normalizing the scores obtained from different domain experts is important.\textsuperscript{19} Normalization typically involves mapping the scores obtained from multiple domains into a common framework before combining them.

BMG couldn’t directly deploy this general multimodal biometric system on the testbed as it could with a single biometric because fusion is absolutely necessary in a multimodal biometric system. To solve this problem, we try to view a multimodal biometric system from an operational aspect. Biometric systems can perform in three operational modes:

- \textit{Parallel} mode combines the modalities simultaneously.
- \textit{Serial} mode combines the modalities sequentially.
- \textit{Hierarchical} mode combines the modalities in a hierarchical scheme.\textsuperscript{20}
Figure 4 illustrates these three modes. If we view each classifier as a single biometric, then we can deploy a multimodal biometric system on the academic testbed as we do for a single biometric. For example, a classifier interface can be defined as a binary file (for example, Classifier $i$) with a parameter. When a user submits Classifier $i$ to the BMG Web portal, Classifier $i$ itself will be wrapped into a Web service.

![Diagram](image)

**Figure 4. Operational modes of multimodal biometric systems: (a) parallel, (b) serial, and (c) hierarchical.**

**Providing services with QoS demand**

Some biometric technologies, such as fingerprint identification and speaker verification, are relatively mature and can be deployed first on BMG for commercial applications. Globus provides the tools for us to wrap applications into Web services. For example, figure 5 illustrates the process of wrapping a fingerprint application into a Web service:

1. When the client needs to invoke the fingerprint Web service, it calls the client stub to turn the "local invocation" into a proper SOAP request.
2. The SOAP request is sent over a network using the HTTP protocol. The server receives the request and takes it to the server stub.
3. Once the SOAP request has been deserialized, the server stub invokes the fingerprint service, which then carries out the requested operation.
4. The requested operation's result is handed to the server stub, which converts the result into a SOAP response.
5. The SOAP response is sent over a network using HTTP. The client stub receives the SOAP response and turns it into content the client application can understand.
6. The application client then receives and uses the result of the Web service invocation.

![Diagram](image)

**Figure 5. A fingerprint service invocation.**
However, as a typical programming challenge, we must consider the application’s necessary latency and bandwidth to determine the most appropriate communication mechanism between modules. Given computation nodes’ bandwidth and capacity constraints, we can use grid computing to distribute computing applications while meeting QoS goals. The service deployment component deploys biometric Web services on BMG. Consider a set of $M$ applications $A = \{a_1, \ldots, a_M\}$ and a set of $K$ computation nodes $C = \{c_1, \ldots, c_K\}$. The bandwidth of the network connection for each computation node $c_j$ in $C$ is $B_j$ bits per second (bps). We can deploy more than one application to any computation node, but a given application runs entirely on a single computation node.

Each application has associated QoS goals. Such goals for application $a_i$ include its maximum average execution time $r_i^{\text{max}}$ and minimum throughput $x_i^{\text{min}}$. The workload intensity associated with $a_i$ is equal to $\lambda_i$ requests per second. The resource demand of $a_i$ is characterized by the tuple $(p_i, d_i, n_i)$, where $p_i$ is the application’s processing time on one of the computation nodes in $C$, taken as a reference computation node; $d_i$ is the application’s I/O demand measured in seconds of disk service time; and $n_i$ is the networking bandwidth measured in bps. Let’s assume that all computation nodes have the same capacity.

However, if the networking bandwidths of the applications allocated to a given computation node exceed that node’s bandwidth, not all possible allocations are feasible:

$$B_j \leq f(a_i) \leq c_j, \quad j = 1, \ldots, K$$

where $\Omega(c_j)$ is the set of applications allocated to $c_j$, $f(a_i) = n_i$. An allocation isn’t feasible if it violates at least one of the application’s QoS goals. Let $r_i$ be the average response time of application $a_i$. Then, for any feasible allocation, this relationship must hold:

$$r_i \leq r_i^{\text{max}}, \quad i = 1, \ldots, M$$

The average response time $r_i$ of application $a_i$ depends on its resource demands, its workload intensity, and the resource demands $(p_i$ and $d_i)$ and workload intensities of all applications allocated to the same computation node $(a_j)$. BMG can use a multiclass open-queuing network model to compute the average response time for each application in a given allocation. BMG must create a queuing network model for each computation node, with the applications allocated to that node for a given allocation constituting the various model classes.

**A case study**

We carried out two biometrics recognition processes for voiceprint and face recognition on BMG.

**The environment**

We used one voiceprint recognition approach and four face recognition approaches: the line-based face recognition algorithm, the improved line-based face recognition algorithm, PCA (Principal Components Analysis), and PCA+LDA (Linear Discriminant Analysis). Our development environments were Linux Fedora Core 4, Globus Tookit 4.0, Apache Tomcat 5.0, and MySQL 5.0. Figure 6 shows the Web portal for face recognition; figure 7 illustrates the flowchart of a face-recognition job.
Figure 6. A BMG prototype system’s Web portal.

Figure 7. The flowchart of a face recognition job.
For voiceprint recognition, we collected 24 samples from 44 people. We put the first 20 samples from each person in the training set and used the remaining four as the test set. For face recognition, we used a face database we established to evaluate our algorithm’s performance. We took pictures of 35 people (six pictures per person) using a standard camera under different lighting conditions (weak, medium, and strong). We selected three views of each person for training and used the other three for testing.

**Two biometrics recognition processes**

We defined three simple interfaces, which are Linux binary files, to run two biometrics-recognition processes for voiceprints and faces. Training, a binary file for biometrics data training, can be invoked using the command line mode: Training samples. FeatureExt, a binary file for extracting feature vectors using training results, can be invoked using FeatureExt a samples. FeatureMat, a binary file for matching feature vectors between two samples, can be invoked using FeatureMat sample_1 sample_2.

We coded each recognition method and built 3 × 2 binary files. The GRAM server then submitted these files to the grid as RSL (XML files) for recognition.

In the voiceprint-recognition test, we tagged one person as *imposter* and the other 43 as *client*. Every person entered as himself or herself, and the imposter also tried to enter as the other 43 people 20 times each. This gave us 44 × 43 × 20 verification results. In our voiceprint-recognition job, the false accept rate was 0.092 percent, and the false reject rate was 2.27 percent.

In the face-recognition test, we tested each method on the same (but individually processed) face database. Figure 8a shows different algorithms’ performance. As figure 8b illustrates, we were able to shorten the average execution times of the improved line-based face-recognition algorithm by adding more grid computation nodes.
On the basis of the results, we made three conclusions. First, it’s feasible to deploy biometrics applications on BMG. Second, BMG’s algorithm testbed can provide uniform interfaces to different algorithms belonging to either the same or different types of biometrics. Finally, BMG can meet some QoS requirements (execution times) by, for example, increasing the number of computation nodes.

BMG is designed to save processing time and improve resource management. However, when QoS is defined more broadly than bandwidth and capacity, the Web portal doesn’t guarantee that the QoS demands will match the QoS goals. Furthermore, BMG should develop under government guidelines. With the development of grid computing, we can perfect the BMG concept. In future work, we will focus on BMG security and the implementation of BMG academic testbeds.
Acknowledgments

The work is supported by grants from the National Natural Science Foundation of China (no. 90612013), the National High Technology Research and Development Program of China (no. 2006AA01Z304), and the Specialized Research Fund for the Doctoral Program of Higher Education (no. 20050013010), as well as the Ministry of Education’s NCET (New Century Excellent Talents in University) program, China.

References

1. D. Dessimoz et al., *Multimodal Biometrics for Identity Documents*, v. 2.0, research report PFS 341-08.05, Univ. of Lausanne and Swiss Federal Inst. of Technology Lausanne, 2006; (http://www.europeanbiometrics.info/images/resources/90_264_file.pdf).

**Anlong Ming** is a PhD candidate at the Beijing Key Lab of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications. His research interests include multimedia systems, computer vision, and grid computing. He received his master’s degree in software engineering from the Beijing University of Posts and Telecommunications. Contact him at the Beijing Key Lab of Intelligent Telecommunications Software and Multimedia, Beijing Univ. of Posts and Telecommunications, Box 139,10 Xitucheng Rd., Haidian, Beijing 100876, China; anthonyming@gmail.com.

**Huadong Ma** is a professor at and the director of the Beijing Key Lab of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications. His research focuses on multimedia systems and networking, grid computing, and sensor networks. He received his PhD in computer science from the Chinese Academy of Sciences. He’s a member of the IEEE and ACM. Contact him at the Beijing Key Lab of Intelligent Telecommunications Software and Multimedia, Beijing Univ. of Posts and Telecommunications, Box 139,10 Xitucheng Rd., Haidian, Beijing 100876, China; mhd@bupt.edu.cn.

**Related Links**
- DS Online's Grid Computing Community

**Cite this article:**