

THE DESIGN AND PERFORMANCE OF A PLUGGABLE PROTOCOLS FRAMEWORK FOR CORBA MIDDLEWARE

Fred Kuhns, Carlos O’Ryan, Douglas C. Schmidt, Ossama Othman, and Jeff
Parsons

Department of Computer Science, Washington University St. Louis, MO 63130, USA
{fredk,coryan,schmidt,othman,parsons}@cs.wustl.edu

Abstract

To be an effective platform for performance-sensitive real-time and embedded applications, off-the-shelf CORBA middleware must preserve communication layer quality of service (QoS) properties to applications end-to-end. However, the standard CORBA’s GIOP/IOP interoperability protocols are not well suited for applications that cannot tolerate the message footprint size, latency, and jitter associated with general-purpose messaging and transport protocols. It is essential, therefore, to develop standard pluggable protocols frameworks that allow custom messaging and transport protocols to be configured flexibly and used transparently by applications.

This paper provides three contributions to research on pluggable protocols frameworks for performance-sensitive communication middleware. First, we outline the key design challenges faced by pluggable protocols developers. Second, we describe how TAO, our high-performance, real-time CORBA-compliant ORB, addresses these challenges in its pluggable protocols framework. Third, we present the results of benchmarks that pinpoint the impact of TAO’s OO design on its end-to-end efficiency, predictability, and scalability.

Our results demonstrate how applying strategic optimizations to communication middleware can yield highly flexible/reusable designs and highly efficient/predictable implementations. In particular, the overall round-trip latency of a TAO two-way method invocation is ~ 125 μ secs using the standard CORBA inter-ORB protocol on a commercial, off-the-self Pentium II Xeon 400 MHz workstation running in loopback mode. The ORB middleware accounts for ~ 60 μ secs of the total round-trip latency. These results illustrate that (1) communication middleware performance is largely an implementation detail and (2) the next-generation of optimized, standards-based CORBA middleware can replace ad hoc and proprietary solutions.

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: [10.1007/978-0-387-35580-1_16](https://doi.org/10.1007/978-0-387-35580-1_16)

1. INTRODUCTION

Communication middleware resides between client and server applications in distributed systems. It simplifies application development by providing a uniform view of heterogeneous networks, protocols, and OS layers. At the heart of communication middleware are *Object Request Brokers* (ORBs), such as CORBA [6], which eliminate many of tedious, error-prone, and non-portable aspects of developing and maintaining distributed applications programmed using low-level mechanisms like sockets. However, the general lack of support in this off-the-shelf communication middleware for QoS specification and enforcement features, integration with high-speed networking technology, and performance, predictability, and scalability optimizations optimizations, has limited the rate at which applications have been developed to leverage advances in communication middleware.

To address the shortcomings of communication middleware we have developed *The ACE ORB* (TAO). TAO is open-source, standards-based, high-performance, real-time ORB endsystem communication middleware that supports applications with deterministic and statistical QoS requirements, as well as “best-effort” requirements. TAO is the first ORB to support end-to-end QoS guarantees over ATM/IP networks [3]. The source code for TAO is available at www.cs.wustl.edu/~schmidt/TAO.html.

We have used TAO to research many dimensions of high-performance and real-time ORB endsystems, including static and dynamic scheduling, request demultiplexing, event processing, ORB Core connection and concurrency architectures, IDL compiler stub/skeleton optimizations, systematic benchmarking of multiple ORBs, I/O subsystem integration, and patterns for ORB extensibility. This paper focuses on a previously unexamined dimension in the high-performance and real-time ORB endsystem design space: *the design and performance of a pluggable protocols framework* that supports high-speed protocols and networks, real-time embedded system interconnects, and standard TCP/IP protocols over the Internet.

The remainder of this paper is organized as follows: Section 2 outlines the CORBA protocol interoperability architecture; Section 3 motivates the need for a CORBA pluggable protocols framework and describes how TAO’s pluggable protocols framework is designed; Section 4 illustrates the performance characteristics of TAO’s pluggable protocols framework; Section 5 compares TAO with related work; and Section 6 presents concluding remarks.

2. OVERVIEW OF THE CORBA PROTOCOL INTER-OPERABILITY ARCHITECTURE

The CORBA specification [6] defines an architecture for ORB interoperability. Although a complete description of the model is beyond the scope of this paper, this section outlines the portions of the CORBA specification that are relevant to our present topic, *i.e.*, object addressing and inter-ORB protocols.

Object addressing synopsis: To identify objects, CORBA defines a generic format called the Interoperable Object Reference (IOR). An object reference identifies one instance of an object and associates one or more paths or routes by which that object can be accessed.

The same object may be located by different object references, *e.g.*, if a server is re-started on a new port or migrated to another host. Likewise, multiple server locations can be referenced by one IOR, *e.g.*, if a server has multiple network interfaces connecting it to distinct networks, there may be multiple network addresses.

References to server locations are called *profiles*. A profile provides an opaque, protocol-specific representation of an object location. Profiles can be used to annotate the server location with QoS information, such as the priority of the thread serving each endpoint or redundant addresses to enhance fault-tolerance.

Protocol model synopsis: CORBA Inter-ORB Protocols (IOP)s define interoperability between ORB endsystems. IOPs provide data representation formats and ORB messaging protocol specifications that can be mapped onto standard and/or customized transport protocols. Regardless of the choice of ORB messaging or transport protocol, however, the standard CORBA programming model is exposed to the application developers. Figure 1 shows the relationships between these various components and layers.

In the CORBA protocol interoperability architecture, the standard *General Inter-ORB Protocol* (GIOP) is defined by the CORBA specification [6]. In addition, CORBA defines a TCP/IP mapping of GIOP, which is called the *Internet Inter-ORB Protocol* (IIOP). ORBs must support IIOP to be “interoperability compliant.” Other mappings of GIOP onto different transport protocols are allowed by the specification, as are different inter-ORB protocols, which are known as *Environment Specific Inter-ORB Protocols* (ESIOP)s.

Regardless of whether GIOP or an ESIOP is used, a CORBA IOP must define a data representation, an ORB message format, an ORB transport protocol or transport protocol adapter, and an object addressing format. Below, we outline how GIOP defines each of these IOP elements.

GIOP synopsis: The GIOP specification consists of the following elements:

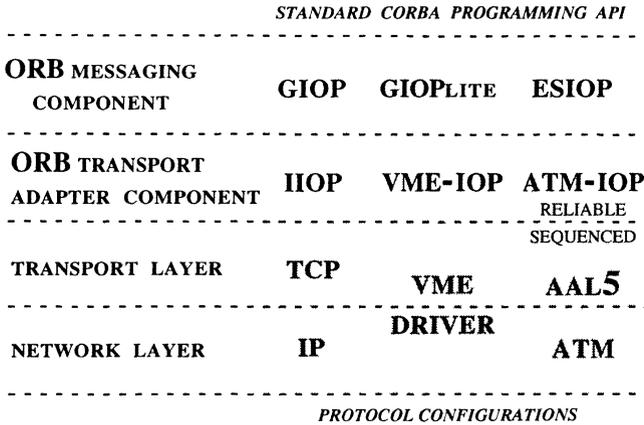


Figure 1 Relationship Between Inter-ORB Protocols and Transport-specific Mappings

- **A Common Data Representation (CDR) definition:** CDR is a transfer syntax that maps IDL types from their native host format into a low-level *bi-canonical* representation, which supports both little-endian and big-endian formats. CDR-encoded messages are used to transmit CORBA requests and server responses across a network. All IDL data types are marshaled using the CDR syntax into an *encapsulation*, which is an octet stream that holds marshaled data.

- **GIOP message formats:** The GIOP specification defines seven types of messages that send requests, receive replies, locate objects, and manage communication channels.

- **GIOP transport adapter:** The GIOP specification describes the features of an ORB transport protocol that can carry GIOP messages. Such protocols must be reliable and connection-oriented. In addition, GIOP defines a connection management protocol and a set of constraints for GIOP message ordering.

- **Object addressing:** An Interoperable Object Reference (IOR) is a sequence of opaque *profiles*, each representing a protocol-specific representation of an object's location. For example, an IIOP profile includes the IP address and port number where the server accepts connections, as well as the object key that identifies an object within a particular server. There may be multiple paths or routes to an object. Therefore, the same IOR can contain multiple IIOP profiles, along with profiles for other protocols, such as GIOP over ATM or non-GIOP protocols.

ESIOP synopsis: In addition to the standard GIOP and IIOP protocols, the CORBA specification allows ORB implementors to define Environment Specific Inter-ORB Protocols (ESIOP)s. ESIOPs can define unique data representation formats, ORB messaging protocols, ORB transport protocols or transport protocol adapters, and object addressing formats. These protocols can exploit the QoS features and guarantees provided in certain domains, such as telecommunications or avionics, to satisfy performance-sensitive applications that have stringent bandwidth, latency, and jitter requirements.

3. A PLUGGABLE PROTOCOLS FRAMEWORK FOR CORBA

The CORBA specification provides a standard for general-purpose communication middleware. Within the scope of this specification, however, ORB implementors are free to optimize internal data structures and algorithms. Moreover, ORBs may use specialized inter-ORB protocols and ORB services and still comply with the CORBA specification. However, an ORB *must* implement GIOP/IIOP, to be interoperability-compliant. This section identifies the limitations of current ORBs with respect to their protocol support, describes how TAO's pluggable protocols framework is designed to overcome these limitations, and then describes how TAO can be applied to develop middleware for high-performance multimedia applications.

Protocol Limitations of Conventional ORBs

CORBA's standard GIOP/IIOP protocols are well suited for traditional request/response applications with best-effort QoS requirements. They are not well suited, however, for high-performance, real-time, and/or embedded applications that cannot tolerate the message footprint size of GIOP or the latency, overhead, and jitter of the TCP/IP-based IIOP transport protocol. For instance, TCP functionality, such as adaptive retransmissions, deferred transmissions, and delayed acknowledgments, can cause excessive overhead and latency for real-time applications. Likewise, networking protocols, such as IPv4, lack the functionality of packet admission policies and rate control, which can lead to excessive congestion and missed deadlines in networks and endsystems.

Therefore, applications with more stringent QoS requirements need optimized protocol implementations, QoS-aware interfaces, custom presentations layers, specialized memory management (*e.g.*, shared memory between ORB and I/O subsystem), and alternative transport programming APIs (*e.g.*, sockets vs. TLI). Domains where highly optimized ORB messaging and transport protocols are particularly important include (1) multimedia applications running over high-speed networks, such as Gigabit Ethernet or ATM, and (2) real-time

applications running over embedded system interconnects, such as VME or CompactPCI.

Conventional CORBA implementations have the following limitations that make it hard for them to support performance-sensitive applications effectively:

- 1. Static protocol configurations:** Many ORBs support a limited number of statically configured protocols, typically only GIOP/IIOP over TCP/IP.
- 2. Lack of protocol control interfaces:** Many ORBs do not allow applications to configure key protocol policies and properties, such as peak virtual circuit bandwidth or cell pacing rate.
- 3. Single protocol support:** Many ORBs do not support simultaneous use of multiple inter-ORB messaging or transport protocols.
- 4. Lack of real-time protocol support:** Many ORBs have limited or no support for specifying and enforcing real-time protocol requirements across a backplane, network, or Internet end-to-end.

Overview of TAO's Pluggable Protocols Framework

To overcome the limitations described in Section 3, we identified logical communication component layers within TAO, factored out common features, defined general framework interfaces, and implemented components to support different concrete inter-ORB protocols. Higher-level services in the ORB, such as stubs, skeletons, and standard CORBA pseudo-objects, are decoupled from the implementation details of particular protocols, as shown in Figure 2. This decoupling is essential to resolve several limitations of conven-

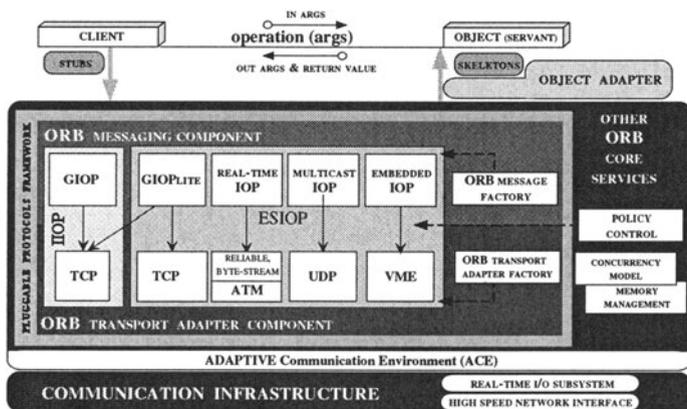


Figure 2 TAO's Pluggable Protocols Framework Architecture

tional ORBs outlined in Section 3.

In general, the higher-level components and services of TAO use a facade [2] interface to access the mechanisms provided by its pluggable protocols framework. Thus, applications can (re)configure custom protocols without requiring global changes to the ORB. Moreover, because applications typically access only the standard CORBA APIs, TAO's pluggable protocols framework can be entirely transparent to CORBA application developers.

Figure 2 also illustrates the key components in TAO's pluggable protocols framework: (1) the ORB messaging component, (2) the ORB transport adapter component, and (3) the ORB policy control component, which are outlined below.

ORB Messaging Component

This component is responsible for implementing ORB messaging protocols, such as the standard CORBA GIOP ORB messaging protocol, as well as custom ESIOPs. As described in Section 2, ORB messaging protocols should define a data representation, an ORB message format, an ORB transport protocol or transport adapter, and an object addressing format. Within this framework, ORB protocol developers are free to implement optimized Inter-ORB protocols and enhanced transport adapters, as long as the ORB interfaces are respected.

Each ORB messaging protocol implementation inherits from a common base class that defines a uniform interface. This interface can be extended to include new capabilities needed by special protocol-aware policies. For example, ORB end-to-end resource reservation or priority negotiation can be implemented in an ORB messaging component. TAO's pluggable protocols framework ensures consistent operational characteristics and enforces general IOP syntax and semantic constraints, such as error handling.

ORB Transport Adapter Component

This component maps a specific ORB messaging protocol, such as GIOP or DCE-CIOP, onto a specific instance of an underlying transport protocol, such as TCP or ATM. Figure 2 shows an example in which TAO's transport adapter maps the GIOP messaging protocol onto TCP (this standard mapping is called IIOp). In this case, the ORB transport adapter combined with TCP corresponds to the transport layer in the Internet reference model. However, if ORBs are communicating over an embedded interconnect, such as a VME bus, the bus driver and DMA controller provide the "transport layer" in the communication infrastructure.

TAO's ORB transport component accepts a byte stream from the ORB messaging component, provides any additional processing required, and passes the resulting data unit to the underlying communication infrastructure. Addi-

tional processing that can be implemented by protocol developers includes (1) concurrency strategies, (2) endsystem/network resource reservation protocols, (3) high-performance techniques, such as zero-copy I/O, shared memory pools, periodic I/O, and interface pooling, (4) enhancement of underlying communications protocols, *e.g.*, provision of a reliable byte stream protocol over ATM, and (5) tight coupling between the ORB and efficient user-space protocol implementations, such as Fast Messages [4].

ORB Policy Control Component

This component allows applications to control the QoS attributes of configured ORB transport protocols explicitly. It is not possible to determine *a priori* all attributes defined by all protocols. Therefore, TAO's pluggable protocols framework provides an extensible *policy control* component, which implements the QoS framework defined in the CORBA Messaging [5] and Real-time CORBA [7] specifications.

The CORBA QoS framework allows applications to specify various *policies* to control the QoS attributes in the ORB. The CORBA specification uses policies to define semantic properties of ORB features precisely without (1) over-constraining ORB implementations or (2) increasing interface complexity for common use cases. Example policies relevant for pluggable protocols include buffer pre-allocations, fragmentation, bandwidth reservation, and maximum transport queue sizes.

Policies in CORBA can be set at the ORB, thread, or object level. Thus, application developers can set global policies that take effect for any request issued in a particular ORB. Moreover, these global settings can be overridden on a per-thread basis, a per-object basis, or even before a particular request. In general, CORBA's Policy framework provides very fine-grained control over the ORB behavior, while providing simplicity for the common case.

A Pluggable Protocol Scenario

To illustrate how TAO's pluggable protocols framework can be applied in practice, we now describe a scenario where pluggable protocols can be used to support performance-sensitive and real-time CORBA applications. This scenario is based on our experience developing high-bandwidth, low-latency audio/video streaming applications and avionics mission computing systems. In previous work [3], we addressed the network interface and I/O system and how to achieve predictable, real-time performance. In the discussion below, we focus on ORB support for alternate protocols.

Low-latency, high-bandwidth multimedia streaming: Multimedia applications running over high-speed networks require optimizations to utilize

available link bandwidth, while still meeting application deadlines. For example, consider Figure 3, where network interfaces supporting 1.2 Mbps or

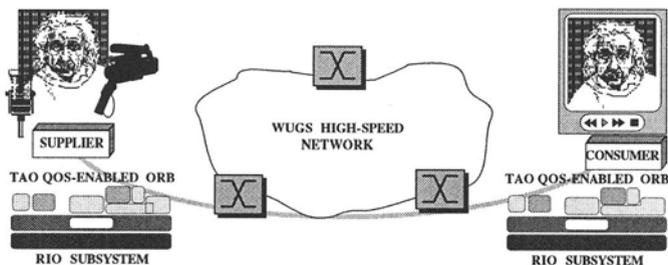


Figure 3 Example CORBA-based Audio/Video (A/V) Application

2.4 Mbps link speeds are used for a CORBA-based studio quality audio/video (A/V) application.

In this example, we can use TAO's pluggable protocols framework to replace GIOP/IOP with a custom ORB messaging and transport protocol that transmits A/V frames using TAO's real-time I/O (RIO) subsystem [3]. At the core of RIO is the high-speed ATM port interconnect controller (APIC) [1]. APIC is a high-performance ATM interface card that supports standard ATM host interface features, such as AAL5 (SAR). In addition, the APIC supports (1) shared memory pools between user and kernel space, (2) per-VC pacing, (3) two levels of priority queues, and (4) interrupt disabling on a per-VC bases.

We are leveraging the APIC features and the underlying ATM network to support end-to-end QoS guarantees for TAO middleware. In particular, pluggable ORB message and transport protocols can be created to provide QoS services to applications, while the ORB middleware encapsulates the actual resource allocation and QoS enforcement mechanisms. Leveraging the underlying APIC hardware requires the resolution of the following two design challenges:

1. Custom protocols: The first challenge is to create custom ORB messaging and transport protocols that can exploit high-speed network interface hardware. A careful examination of the system requirements along with the hardware and communication infrastructure is required in order to determine both the set of optimizations required and the best partitioning of the solution into ORB messaging, transport and policy components.

The A/V streaming application is primarily concerned with pushing data to clients (i.e., one-way method invocations) and with meeting a specific set of latency and jitter requirements. Considering this, a simple frame sequencing protocol can be used as the ORB's ESIOP. Moreover, because multimedia data has diminishing value over time, a reliable protocol, such as TCP, is not required. Thus, the overhead of GIOP is not required, nor are the underlying assumptions that require a transport protocol with the semantics of TCP.

The ORB messaging protocol can be mapped onto an ORB transport protocol using AAL5. The transport adapter is then responsible for exploiting any local optimizations to hardware or the endsystem. For example, conventional ORBs copy user parameters into internal buffers used for marshaling. These buffers may be allocated from global memory or possibly from a memory pool maintained by the ORB. In either case, at least one system call is required to obtain mutexes, allocate buffers, and copy the data. Thus, not only is an additional data copy incurred, but this scenario is rife with opportunities for priority inversion and jitter while waiting for ORB endsystem resources.

2. Optimized protocol implementations: The second challenge is to produce an optimized pluggable protocol that implements the design described above. For example, memory can be shared throughout the ORB endsystem, *i.e.*, between the application, ORB middleware, OS kernel, and network interface, by allocating memory from a common buffer pool [1]. This optimization eliminates memory copies between user- and kernel-space when data is sent or received. Moreover, the ORB endsystem can manage this memory, thereby relieving application developers from this responsibility. In addition, the ORB endsystem can manage the APIC interface driver, interrupt rates, and pacing parameters, as outlined in [3].

4. THE PERFORMANCE OF TAO'S PLUGGABLE PROTOCOLS FRAMEWORK

Despite the growing demand for off-the-shelf middleware in many application domains, a widespread belief persists that OO techniques are not suitable for real-time systems due to performance penalties attributed to the OO paradigm. In particular, the dynamic binding properties of OO programming languages and the indirection implied in OO designs seem antithetical to real-time systems, which require low latency and jitter. The results presented in this section are significant, therefore, because they illustrate empirically how the choice of implementation strategies and optimization principles makes it possible to implement very predictable, efficient, and scalable middleware *without* compromising non-functional requirements, such as portability, flexibility, reusability, and maintainability, offered by CORBA.

To quantify the benefits and costs of TAO's pluggable protocols framework, we conducted several benchmarks using two different ORB messaging protocols, GIOP and GIOPlite, and two different transport protocols, POSIX local IPC (also known as UNIX-domain sockets) and TCP/IP. These benchmarks are based on our experience developing communication middleware for avionics mission computing applications and multimedia applications.

POSIX local IPC is not a traditional high-performance networking envi-

ronment. However, it does provide the opportunity to obtain an accurate measure of ORB and pluggable protocols framework overhead. Based on these measurements, we have isolated the overhead associated with each component, which provides a baseline for future work in high-performance protocol development and experimentation.

Hardware/Software Benchmarking Platform

All benchmarks in this section were run on a Quad-CPU Intel Pentium II Xeon 400 MHz workstation, with one gigabyte of RAM. The operating system used for the benchmarking was Debian GNU/Linux “potato” (glibc 2.1) with Linux kernel version 2.2.10. GNU/Linux is an open-source operating system that supports true multi-tasking, multi-threading, and symmetric multiprocessing.

Our benchmarks were run using the standard GIOP ORB messaging protocol, as well as TAO’s GIOPLite messaging protocol, described in Section 3. For the TCP/IP tests, the GIOP and GIOPLite ORB messaging protocols were run using the standard CORBA IIOP transport adapter along with the Linux TCP/IP socket library and the loopback interface. For the local IPC tests, GIOP and GIOPLite were used along with the optimized local IPC transport adapter. This resulted in four different Inter-ORB Protocols: GIOP over TCP (IIOP), GIOPLite over TCP, GIOP over local IPC (UIOP) and GIOPLite over local IPC. No changes were required to our standard CORBA benchmarking tool, called `IDL_Cubit` [9], for either of the ORB messaging and transport protocol implementations.

Blackbox Benchmarks

Blackbox benchmarks measure the end-to-end performance of a system from an external application perspective. In our experiments, we used blackbox benchmarks to compute the average two-way response time incurred by clients sending various types of data using the four different Inter-ORB transport protocols.

Measurement technique: A single-threaded client is used in the `IDL_Cubit` benchmark to issue two-way IDL operations at the fastest possible rate. The server performs the operation, which cubes each parameter in the request. For two-way calls, the client thread waits for the response and checks that it is correct. Interprocess communication is performed over selected IOPs, as described above.

We measure throughput for operations using a variety of IDL data types, including `void`, `sequence`, and `struct` types. The `void` data type instructs the server not to perform any processing other than that necessary to

prepare and send the response, *i.e.*, it does not cube any input parameters. The sequence and struct data types exercise TAO's (de)marshaling engine. The struct contains an octet, a long, and a short, along with padding necessary to align those fields. We also measure throughput using long and short sequences of the long and octet types. The long sequences contain 4,096 bytes (1,024 four byte longs or 4,096 octets) and the short sequences are 4 bytes (one four byte long or four octets).

Blackbox results: The blackbox benchmark results are shown in Figure 4. All blackbox benchmarks were averaged over 100,000 two-way operation calls for each data type, as depicted by the bars in Figure 4.

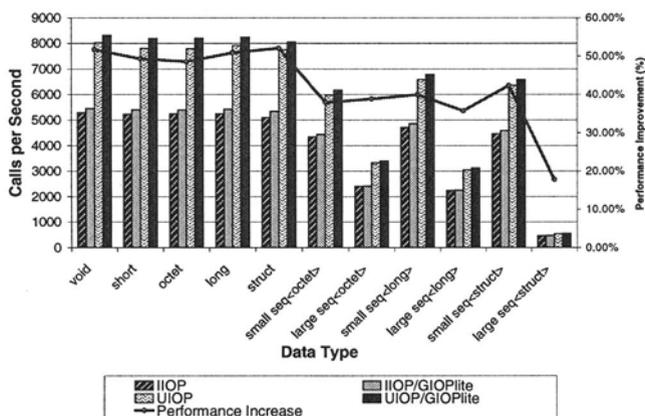


Figure 4 TAO's Pluggable Protocols Framework Performance Over Local IPC and TCP/IP

UIOP performance surpassed IIOP performance for all data types. The benchmarks show UIOP improves performance from 20% to 50% depending on the data type and size. For smaller data sizes and basic types, such as octet and long, the performance improvement is approximately 50%. However, for larger data payload sizes and more complex data types, the performance improvements are reduced. This is a direct result of the increasing cost of both the data copies associated with performing I/O and increasing complexity of marshaling structures other than the basic data types.

For certain data types, additional improvements are obtained by reducing the number of data copies required. Such a situation exists when marshaling and demarshaling data of type octet and long. For complicated data types, such as a large sequence of structs, ORB overhead is particularly prevalent. Large ORB overhead implies lower efficiency, which accounts for the smaller performance improvement gained by UIOP over IIOP for complex data types.

GIOPlite outperformed GIOP by a small margin. For IIOP, GIOPlite performance increases over GIOP ranged from 0.36% to 4.74%, with an average

performance increase of 2.74%. GIOPlite performance improvements were slightly better over UIOP due to the fact that UIOP is more efficient than IIOP. GIOPlite over UIOP provided improvements ranging from 0.37% to 5.29%, with an average of 3.26%.

Our blackbox results suggest that more substantial changes to the GIOP message protocol are required to achieve significant performance improvements. However, these results also illustrate that the GIOP message footprint has a relatively minor performance impact over high-speed networks and embedded interconnects. Naturally, the impact of the GIOP message footprint for lower-speed links, such as second-generation wireless systems or low-speed modems, is more significant.

Whitebox Benchmarks

Whitebox benchmarks measure the performance of specific components or layers in a system from an internal perspective. In our experiments, we used whitebox benchmarks to pinpoint the time spent in key components in TAO’s client and server ORBs. The ORB logical layers, or components, are shown in Figure 5 along with the timeprobe locations used for these benchmarks.

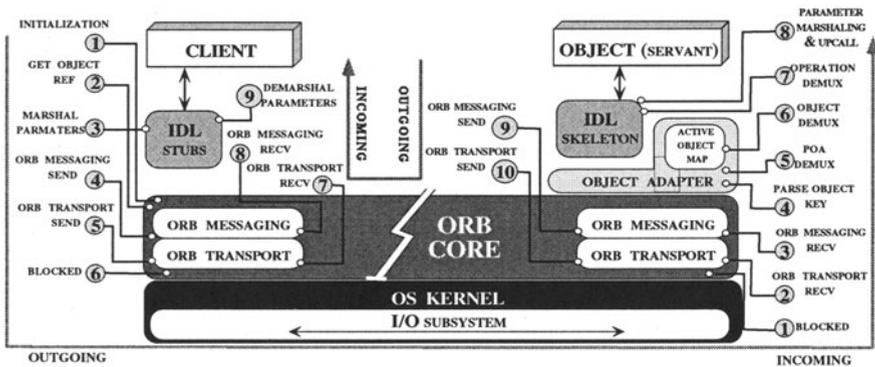


Figure 5 Timeprobe Locations for Whitebox Experiment

Measurement Techniques

One way to measure performance overhead of operations in complex communication middleware is to use a profiling tool like Quantify. Quantify instruments an application’s binary instructions and then analyzes performance bottlenecks by identifying sections of code that dominate execution time. Quantify is useful because it can measure the overhead of system calls and third-party libraries without requiring source code access.

Unfortunately, Quantify is not available for Linux kernel-based operating systems for which whitebox measurement of TAO’s performance is needed.

Moreover, Quantify modifies the binary code to collect timing information. Therefore, it is most useful for measuring *relative* overhead of different operations in a system, rather than measuring *absolute* run-time performance.

To avoid the limitations of Quantify, we therefore used a lightweight timeprobe mechanism provided by TAO to precisely pinpoint the amount of time spent in various ORB components and layers. The TAO timeprobe mechanism provides highly accurate, low-cost timestamps that record the time spent between regions of code in a software system. These timeprobes have minimal performance impact, *e.g.*, 1-2 μ sec overhead per timeprobe, and no binary code instrumentation is required.

Depending on the underlying platform, TAO's timeprobes are implemented either by high-resolution OS timers or by high-precision timing hardware. An example of the latter is the VMetro board, which is a VME bus monitor. VMetro writes unique TAO timeprobe values to an otherwise unused VME address. These values record the duration between timeprobe markers across multiple processors using a single clock. This enables TAO to collect synchronized timestamps and accurately measure communication delays end-to-end across distributed CPUs.

Below, we examine the client and server whitebox performance in detail.

Whitebox Results

Figure 5 shows the points in a two-way operation request path where timeprobes were inserted. Each labeled number in the figure corresponds to an entry in Table 1 and Table 2 below. The results presented in the tables and figures which follow were averaged over 1,000 samples.

Client performance: Table 1 depicts the time in microseconds (μ s) spent in each sequential activity that a TAO client performs to process an outgoing operation request and its reply.

Table 1 μ seconds Spent in Each Client Processing Step

Direction	Client Activities	Absolute Time (μ s)
Outgoing	1. Initialization	6.30
	2. Get object reference	15.6
	3. Parameter marshal	0.74 (param. dependent)
	4. ORB messaging send	7.78
	5. ORB transport send	1.02
	6. I/O	8.70 (op. dependent)
	7. ORB transport recv	50.7
	8. ORB messaging recv	9.25
	9. Parameter demarshal	op. dependent

Server performance: Table 2 depicts the time in microseconds (μ s) spent in each activity as a TAO server processes a request.

Table 2 μ seconds Spent in Each Server Processing Step

Direction	Server Activities	Absolute Time (μ s)
Incoming	1. I/O	7.0 (op. dependent)
	2. ORB transport rcv	24.8
	3. ORB messaging rcv	4.5
	4. Parsing object key	4.6
	5. POA demux	1.39
	6. Servant demux	4.6
	7. Operation demux	4.52
	8. User upcall	3.84 (op. dependent)
Outgoing	9. ORB messaging send	4.56
	10. ORB transport send	93.6

Depending on the type and number of operation parameters, the *ORB transport rcv* step typically requires the most ORB processing time. This time is dominated by the required data copies. By using a transport adaptor which implements a shared buffer strategy these costs can be reduced significantly.

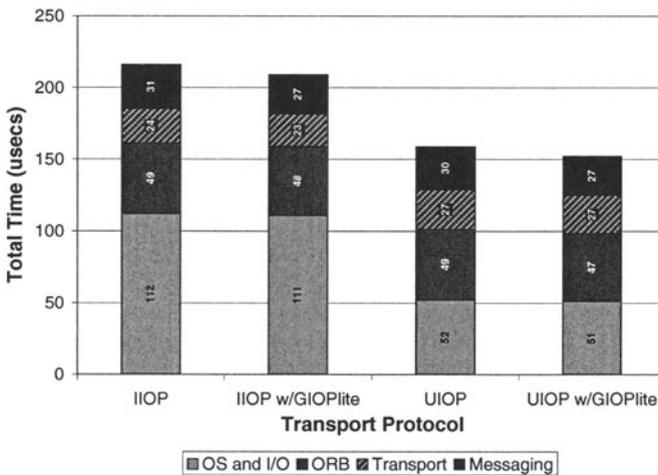


Figure 6 Comparison of ORB and Transport/OS Overhead Using Timeprobes

Component costs: Figure 6 compares the relative overhead attributable to the ORB messaging component, transport adaptor, ORB and OS for two-way

IDL_Cubit calls to the `cube_void` operation for each possible protocol combination. This figure shows that when using IIOp the I/O and OS overhead accounts for just over 50% of the total round trip latency.

It also shows that the difference in performance between IIOp and UIOP is primarily due to the larger OS and I/O overhead that TCP/IP has, as compared to local IPC.

The only overhead that depends on size is *(de)marshaling*, which depends on the type complexity, number, and size of operation parameters, and *data copying*, which depends on the size of the data. In our whitebox experiment, only the parameter size changes, *i.e.*, the sequences vary in length. Moreover, TAO's *(de)marshaling* optimizations incur minimal overhead when running between homogeneous ORB endsystems.

In Figure 7, the parameter size is varied and the above test is repeated. It shows that as the size of the operation parameters increases, I/O overhead

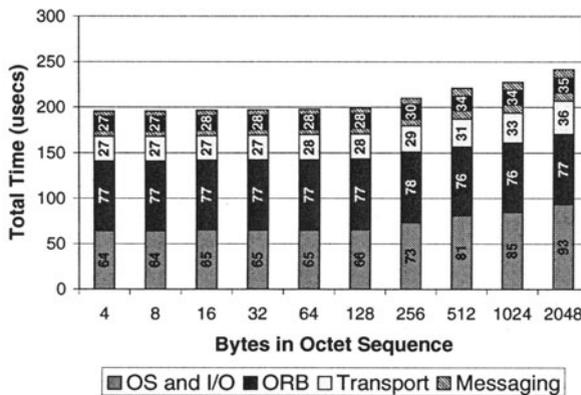


Figure 7 ORB and Transport/OS Overhead Versus Parameter Size

grows faster than the overall ORB overhead (including messaging and transport). This result illustrates that the overall ORB overhead is largely independent of the request size. In particular, demultiplexing a request, creating message headers, and invoking an operation upcall are not affected by the size of the request.

TAO employs standard buffer size and data copy tradeoff optimizations. This optimization is demonstrated in Figure 7 by the fact that there is a slight increase in the time spent both in the transport component and in the ORB itself when the sequence size is greater than 256 bytes. The data copy tradeoff optimization is fully configurable via run-time command line options, so it is possible to configure TAO to further improve performance above the 256 byte data copy threshold.

For the operations tested in the IDL_Cubit benchmark, the overhead of

the ORB is dominated by memory bandwidth limitations. Both the loopback driver and local IPC driver copy data within the same host. Therefore, memory bandwidth limitations should essentially be the same for both IIOP and UIOP. This result is illustrated in Figure 6 by the fact that the time spent in the ORB is generally constant for the four protocol combinations shown.

In general, the use of UIOP demonstrates the advantages of this framework and how optimized, domain-specific protocols can be deployed.

5. RELATED WORK

The design of TAO's pluggable protocols framework is influenced by prior research on the design and optimization of protocol frameworks for communication subsystems. This section outlines this research and compares it with our work.

The architecture of TAO's pluggable protocols framework is based on Open Communications Interface (OCI) [8]. The OCI framework provides a flexible, intuitive, and portable interface for pluggable protocols. The framework interfaces are defined in IDL, with a few special rules to map critical types, such as data buffers.

Defining pluggable protocol interfaces with IDL permits developers to familiarize themselves with a single programming model that can be used to implement protocols in different languages. In addition, the use of IDL makes it possible to write pluggable protocols that are portable among different ORB implementations and platforms.

Though the OCI pluggable protocols framework is useful for many applications and ORBs, TAO implements a highly optimized pluggable protocol framework that is tuned for high-performance and real-time application requirements. For example, TAO's pluggable protocols framework can be integrated with zero-copy high-speed network interfaces [1, 3].

However, TAO's pluggable protocols framework does not preclude the use of more general frameworks like the ORBacus OCI. In fact, we plan to implement OCI as a pluggable protocol into TAO, thereby allowing application developers to test and use OCI pluggable protocols. If applications have very stringent performance requirements, developers can use the internal TAO pluggable protocol framework to obtain the higher performance and greater predictability.

6. CONCLUDING REMARKS

To be an effective development platform for performance-sensitive applications, OO middleware must preserve communication layer QoS properties of applications end-to-end. It is essential, therefore, to define a pluggable

protocols framework that allows custom inter-ORB messaging and transport protocols to be configured flexibly and transparently by CORBA applications.

This paper identifies the protocol-related limitations of current ORBs and describes a CORBA-based pluggable protocols framework we developed and integrated with TAO to address these limitations. Benchmarking results are also presented which demonstrate that applying appropriate optimizations to communication middleware can yield highly efficient and predictable implementations, without sacrificing flexibility or reuse.

REFERENCES

- [1] Zubin D. Dittia, Guru M. Parulkar, and Jerome R. Cox, Jr. The APIC Approach to High Performance Network Interface Design: Protected DMA and Other Techniques. In *Proceedings of INFOCOM '97*, pages 179–187, Kobe, Japan, April 1997. IEEE.
- [2] Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, Reading, MA, 1995.
- [3] Fred Kuhns, Douglas C. Schmidt, and David L. Levine. The Design and Performance of a Real-time I/O Subsystem. In *Proceedings of the 5th IEEE Real-Time Technology and Applications Symposium*, pages 154–163, Vancouver, British Columbia, Canada, June 1999. IEEE.
- [4] M. Lauria, S. Pakin, and A. Chien. Efficient Layering for High Speed Communication: Fast Messages 2.x. In *Proceedings of the 7th High Performance Distributed Computing (HPDC7) conference*, Chicago, Illinois, July 1998.
- [5] Object Management Group. *CORBA Messaging Specification*, OMG Document orbos/98-05-05 edition, May 1998.
- [6] Object Management Group. *The Common Object Request Broker: Architecture and Specification*, 2.2 edition, February 1998.
- [7] Object Management Group. *Realtime CORBA Joint Revised Submission*, OMG Document orbos/99-02-12 edition, March 1999.
- [8] Inc. Object-Oriented Concepts. ORBacus User Manual - Version 3.1.2. www.ooc.com/ob, 1999.
- [9] Douglas C. Schmidt, Sumedh Mungee, Sergio Flores-Gaitan, and Aniruddha Gokhale. Software Architectures for Reducing Priority Inversion and Non-determinism in Real-time Object Request Brokers. *Journal of Real-time Systems*, To appear 1999.