Infrared Data Communications with IrDA

Charles D. Knutson, Ph.D.
Vice President, Research and Development
Chair, IrDA Test and Interoperability Committee

Glade Diviney
Manager, Test Tools and Services

Counterpoint Systems Foundry, Inc.
Corvallis, Oregon
knutson@countersys.com

Abstract

IrDA infrared communication is an inexpensive and widely adopted short range wireless technology that allows devices to "speak" easily to each other. Key protocol features make operation simple even for inexperienced users or devices with very little user interface. Digital cameras, phones, pagers, data collectors, set-top boxes, modems, kiosks, instruments, machinery, ID badges, watches, and computer peripherals are some of the natural users of this technology. This paper introduces IrDA infrared data communications and explores both mandatory and optional IrDA protocol layers and strategies.

1.0 Introduction

The Infrared Data Association was formed to enable universal point and shoot infrared connectivity between devices of all types. Today there are hundreds of devices that implement IrDA communication protocols and the dream of ubiquitous data transfer is becoming more of a reality. This paper briefly describes IrDA technology from low-level physical layers up to high-level optional protocols. It also describes IrDA Lite, a set of strategies for implementing minimal solutions for embedded systems.

2.0 The Infrared Data Association (IrDA)

The Infrared Data Association (IrDA) was formed in June, 1993. At IrDA's charter meeting, fifty companies came together to agree upon standard methods for communicating data via short range infrared transmission. Since that time, more than one hundred additional companies have joined IrDA, and hundreds of devices are currently available that implement IrDA communication protocols.

IrDA is administered by an Executive Director (John LaRoche) and an executive staff. The work of IrDA is conducted through three committees, whose chairs are elected each year. The Technical Committee is currently chaired by Dave Suvak of Counterpoint Systems Foundry, Inc. This committee refines and extends the hardware and software standards for IrDA. All new technical proposals come through this committee and its working groups. The Marketing Committee is currently chaired by Brian Ingham of IBM. This committee handles the marketing and promotional concerns of IrDA. The Test and Interop Committee is currently chaired by Charles Knutson of Counterpoint Systems Foundry, Inc. This committee deals with test specifications for hardware and software standards, as well as issues concerning the spread of interoperability and customer “out of the box” experience. The bulk of the work in IrDA is performed by special interest groups (SIGs) and working groups that carry specific charters.

3.0 The Promise of IrDA connectivity

One of the earliest motivations of the companies involved in IrDA was to eliminate wires and connectors with their accompanying limitations. Wires fray, wear, break, corrode, get tangled, and sometimes fail to reach far enough. The connectors on wires come lose, break, or otherwise become mangled and unusable. Wired connections clog desks with spaghetti-like entanglements, and are notoriously forgotten by the portable traveler who only later discovers that his mobile computer is useless without that one special wire that was left behind.

The core IrDA protocols were designed to replace wires with a “virtual wire” and some ability to accesses services over it. Originally, the focus of IrDA was to deliver this level of connectivity, and then let manufacturers worry about specific implementations above that. Since that time, it has become obvious that once the physical connector is gone (leaving in essence a “universal” connector), standardizing on higher level protocols can provide even greater levels of interoperability in the IrDA user space. Subsequently, a number of optional protocols have been approved, most of them derived from a specific vertical application model. Section 5.0 of this paper discusses the mandatory protocol layers of IrDA. Section 6.0 describes the optional higher-level protocols.

With these high-level protocols, a number of interesting and valuable use models are available to users. By standardizing on these protocols, application vendors can build systems that interoperate with systems of other vendors. This approach is quite similar to what we now see happening in the World Wide Web. The low-level TCP/IP protocol was not sufficient to provide interoperability. Without some higher-layer protocol, there were myriad ways of moving information, with a few garnering more common use than others (like FTP). However, when HTTP began to be used by web browsers with a standard file format (HTML), the elements were in place for a usage explosion. Suddenly anyone could build a web site, because they understood the file format. Anyone could access a web site using the common protocol. And anyone could build software to access anyone else’s web site. When users had the “Ah ha!” experience of what the web could do, independent entrepreneurs did the rest and the web exploded, to the benefit of users. That is the power of universal data access. In the short range, walk up, point and shoot space, IrDA offers that same promise, which is becoming more and more a reality.

4.0 High-Level Overview

The IrDA protocols are organized in a traditional layered or stacked architecture. These individual layers are described in the next two sections. Some of these layers are required for a device to carry the IrDA logo. These are treated in Section 5. In addition, there are optional layers that apply to specific use models. These are discussed in Section 6.

The current protocols provide connectivity at distances up to one meter, and at speeds up to 4 Mbps. IrDA is interested in extending both of these limitations, and is currently working on extending specifications in both cases.

In a typical scenario, a user might have a PDA that has phone and address lists. The user might walk up to another PDA user and beam selected items using IrDA's IrOBEX protocol. Two Palm III users might discover that they each have games that the other does not have. These can also be beamed to the other device on the spot. A laptop user might want to print a document, but lacks a parallel cable. With IrDA capability on both the laptop and the printer, the laptop selects the appropriate LPT port (the one redirected to the IrDA port) and prints. These and other use models extend to devices of all types, including digital cameras, LAN access devices, pagers, cell phones, laptops, PDAs, printers, scanners, medical devices, etc. Some of these will be discussed in the individual discussions in Section 6 where the protocol relates to a specific use model.

In the basic IrDA use model, there are two devices. One is the primary and the other is the secondary. The primary device is responsible for selecting a device within its visual space, establishing a connection, and maintaining the virtual wire or link. The secondary responds when spoken to. At the beginning of a typical IrDA operation, the primary initiates a process known as "discovery", in which it explores its visible space for devices. From those devices that respond the primary selects a device and attempts to connect to it. During connection establishment, the two devices negotiate to understand each other's capabilities. In this way a connection can be optimized despite the unpredictable differences between two disparate devices. Once they have negotiated, they will jump to their highest common transmission speed, and attempt to communicate in ways that optimize the throughput and reliability of their connection.

Having established a connection, the devices may now search the services of the other devices. If the other device supports a desired service, a connection can be made to the service. At this point, applications on either side of the connection can transfer data. Obviously there are considerably more details than have been presented here, and the IrDA specifications are the definitive source for that information.

5.0 Required Layers

In this section, we explore the required IrDA layers, starting from the bottom and working our way up. Each of these layers is described in painstaking detail in corresponding IrDA specifications. The references for these documents are provided in Section 9.

Figure 1 shows the basic organization of the IrDA stack.

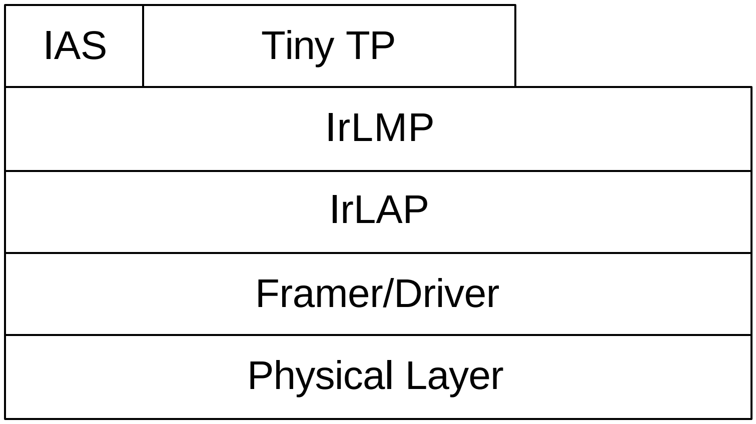

Figure 1. IrDA protocol stack layers.

5.1 Physical Layer

IrDA transceivers broadcast infrared pulses in a cone that extends from 15 degrees half angle to 30 degrees half angle off center. The IrDA physical specifications require that a minimum irradiance be maintained so that a signal is visible up to a meter away. Similarly, the specifications require that a maximum irradiance not be exceeded so that a receiver is not overwhelmed with brightness when a device comes in close. In practice, there are some devices on the market that do not reach one meter, while other devices may reach up to several meters. There are also devices that do not tolerate extreme closeness. The typical sweet spot for IrDA communications is from 5 cm to 60 cm away from a transceiver, in the center of the cone.

IrDA data communications operate in half-duplex mode. The reason is quite simple. While transmitting, a device's receiver is blinded by the light of its own transmitter. Because of this, full duplex communication is not feasible. The two devices that communicate simulate full duplex communication by quickly turning the link around. The primary device controls the timing of the link, but both sides are bound to certain hard constraints and are encouraged to turn the link around as fast as possible.

Transmission rates fall into three broad categories: SIR, MIR, and FIR. Serial Infrared (SIR) speeds cover those transmission speeds normally supported by an RS-232 port (9600 bps, 19.2 Kbps, 38.4 Kbps, 57.6 Kbps, 115.2 Kbps). Since the lowest common denominator for all devices is 9600 bps, all discovery and negotiation is performed at this baud rate. MIR (Medium Infrared) is not an official term, but is sometimes used to refer to speeds of .576 Mbps and 1.152 Mbps. Fast Infrared (FIR) is deemed an obsolete term by the IrDA physical specification, but is nonetheless in common usage to denote transmission at 4 Mbps. “FIR” is sometimes used to refer to all speeds above SIR. However, different encoding approaches are used by MIR and FIR, and different approaches are used to frame MIR and FIR packets. For that reason, these unofficial terms have sprung up to differentiate these two approaches. The future holds faster transmission speeds (currently referred to as Very Fast Infrared, or VFIR) which will support speeds up to 16 Mbps.

5.2 Framer/Driver

The framer and the driver are actually two separate functions, but have enough in common that they are typically grouped together (and commonly referred to simply as “framer”). The driver portion refers to the software that acts as a device driver for the system’s transceiver controller. This driver initializes the infrared hardware, changes transmission speeds, delivers data to the transceiver, and receives data from the transceiver.

The framer portion refers to the bundling of a data packet into a form that can be given to the hardware. This may include the calculation of cyclic redundancy check value, the addition of start and stop bits, and transparency for reserved bytes. Because the framing approach varies with the transmission speeds, it is most common for the framer and driver functions to be combined. In this way, all hardware dependencies in a system can be localized to one section of the IrDA stack.

5.3 IrLAP: Link Access Protocol

The IrDA Link Access Protocol (IrLAP) is responsible for performing device discovery and negotiation, and for preserving the physical connection, or “virtual wire”. It is at this level that the concept of primary and secondary devices is relevant. IrLAP is based on HDLC, adding features to facilitate the walk up, nature of IrDA connections.

IrLAP provides a reliable transmission medium on which to build additional communications. It facilitates error detection, retransmission of lost or damaged packets, and rudimentary flow control.

5.4 IrLMP: Link Management Protocol

The IrDA Link Management Protocol (IrLMP) allows one or more IrDA services to run over a single IrLAP connection. Applications using an IrDA stack can read/write directly to IrLMP, or can use other higher level protocols that, in turn, read/write to IrLMP. A typical service running on IrLMP might include the printing application on an IrDA-enabled laser printer. This application would register its service with IAS (described below), and then be able to service a print job via the IrLMP connection, should a user walk up and begin accessing the printer via the IrDA port.

5.5 IAS: Information Access Service

The Information Access Service (IAS) is the only required service available through IrLMP. It is the mechanism by which applications advertise and access services. Applications register their services when they load, and are given a specific selector, called an LSAP (Link Service Access Point) Selector, by which the service can be accessed by other devices. There are no pre-defined LSAP Selectors besides the IAS itself at selector 0. When a device connects to another, an application can query the IAS of the other device to determine what services it might have, and on which LSAP Selectors. Once the LSAP Selector is known, the application can connect and begin data transfer.

5.6 Tiny TP: Tiny Transport Protocol

Tiny TP is a transport protocol that provides two basic services: flow control and segmentation and reassambly (SAR). Tiny TP allows flow control per service channel, where the more rudimentary flow control provided by IrLAP controls the entire physical link. Flow control in Tiny TP is credit based, permitting an application to extend enough credit to the other side so that it won't be overwhelmed. The remote device uses these credits as packets are delivered. More credits are extended as space becomes available to receive. Segmentation and reassembly provides a mechanism for delivering large packets to the IrDA stack, allowing Tiny TP to break up packets (segmentation) on one side and, on the other side, put them back together (reassembly). This approach takes the burden off applications to be concerned with IrDA packet sizes. This feature is particularly relevant for IrLAN.

6.0 Optional Layers

The following layers are all high-level protocols, and are not strictly required by IrDA. However, some of them are essential within the context of certain use models, in the same way that HTTP is essential for web access, even if it's not part of a basic TCP/IP protocol stack.

6.1 IrOBEX: IrDA Object Exchange

IrDA Object Exchange (IrOBEX) can be viewed as essentially "HTTP for IrDA". IrOBEX was designed to resemble HTTP, and it leverages what it can from this internet protocol, adding capabilities that relate to the unique environment of IrDA. IrOBEX is best used in situations where objects of some kind need to be moved from one device to another. For example, two devices may exchange phone and address information, or calendar information in vCard and vCal formats. Or, a handheld scanner may capture a graphics image and beam it to a laptop to manipulate. Both of these are classic uses for IrOBEX. Because of its universal applicability for object movement, where applicable, IrOBEX is a required protocol for devices seeking interoperability certification.

6.2 IrCOMM

IrCOMM is designed to provide legacy support for applications that already run over COM ports. For example, assume we have a PDA with a cradle that plugs into the serial port of a computer. The desktop software for this PDA is designed to communicate using a serial cable connected to the PDA cradle. To allow synchronization directly between the computer and the PDA, the PDA could be enabled with IrCOMM. Then, by selecting a virtual COM port, the synchronization could take place over infrared without introducing any changes to the computer's software. This is an example of a legacy application for IrCOMM. While this works in legacy situations, IrCOMM is strongly discouraged as a platform for developing new use models, since it reduces IrDA's rich feature set to a virtual nine wires, requiring sophisticated applications to recreate many of the capabilities that are already present in the IrDA stack.

6.3 IrLPT

IrLPT is part of the IrCOMM specification, and is also referred to as IrCOMM 3-Wire Raw. It deserves special mention because it is the mechanism by which legacy printing is achieved between devices and IrDA-enabled printers. Support on desktops is achieved through a virtual LPT port that maps to the IrDA port. When an application or printer driver is configured with the virtual LPT port, infrared printing is enabled without changes to the printer driver or application. As with IrCOMM, IrLPT is intended for legacy support of existing applications. Because of its importance for legacy printing, where applicable, IrLPT is required for devices seeking IrReady interoperability certification.

6.4 IrTran-P

IrTran-P (IrDA Transfer Picture) represents a specific mechanism used by some manufacturers to transfer digital images between devices. IrTran-P is an IrDA application note, meaning that it represents a particular way of solving this problem, without carrying the mandate of IrDA as the only appropriate way to do it. IrTran-P is built on IrCOMM, and therefore requires the reconstruction of several key components to manage services and object exchange. Specifically, IrTran-P adds SCEP (Simple Command Execute Protocol) for service access and link management, and bFTP (Binary File Transfer Protocol) for digital image object exchange. In addition, IrTran-P defines its own digital image file format, UPF (Uni-Picture Format) so that IrTran-P devices can communicate effectively.

6.5 IrMC: IrDA Mobile Communications

IrMC (IrDA Mobile Communications) is a set of four protocols proposed by the IrDA's Mobile Communications Working Group. This group is fundamentally concerned with IrDA communication between telecom devices, such as pagers and cell phones. However, many of the features of IrMC are applicable to other devices, such as PDAs. Because of that, the scope of IrMC has expanded to include devices of all types. IrMC incorporates the following protocols: IrOBEX, IrCOMM, RTCON, and Ultra. IrOBEX, described in Section 6.1, is used in IrMC to exchange vCards, vCalendars, and similar objects. IrCOMM, described in Section 6.2, is used in IrMC to permit cellular phones to be used as external modems, via a virtual COM port connection between a laptop (or other device) and a cell phone. RTCON is described in Section 6.6. Ultra is a very small, connectionless communication mechanism that severely constrained devices can use for device programming and small object exchange using a connectionless version of OBEX.

6.6 RTCON

Real Time Transfer Control Protocol (RTCON) is used to transmit real-time voice and control data over an infrared link. In a typical use model, a cell phone can be placed in a cradle in a car, and an infrared link can be established between the speaker phone mechanism in the car and the cell phone. This permits dialing and talking without holding a cell phone while driving (which is illegal in some parts of the world—most notably Europe and Japan). The call is made with the cell phone, but the voice data in the phone is transferred via infrared to the in-dash speaker phone. This permits users to use their normal cell phones without having to have a special cell number specifically for their car (with accompanying duplicated fees).

6.7 JetSend

JetSend is a technology created and licensed by Hewlett-Packard for delivery of digital imaging information via a variety of transport mechanisms. The first two transports on which JetSend was implemented were TCP/IP and IrDA. JetSend permits devices to negotiate to their greatest common image handling capacity, eliminating the need for hardware-specific printer drivers. In a static, wired office environment, the accessible printers don't change that often. But for mobile devices such as laptops and PDA's, finding and installing the appropriate printer driver can be a significant problem. With JetSend, any IrDA enabled device can approach an IrDA-enabled JetSend printer, and render the best possible image.

7.0 IrDA Lite

Strictly speaking, IrDA Lite is not a protocol, but it is significant enough in the world of IrDA to deserve brief mention. The majority of devices incorporating IrDA are embedded devices. Most of these devices provide dramatically less memory than laptops or desktops. IrDA Lite renders the minimal implementation of IrDA that still interoperates with "full-featured" IrDA stacks. It does so by sacrificing speed, throughput, and non-essential features.

The effort would be similar to removing parts from a car, but requiring that it still seat one passenger and be sufficiently capable of traveling some distance under certain conditions. For the use models where memory savings are worth the loss in throughput, the fit is great.

Typical IrDA Lite strategies include limiting the packet size to 64 bytes, limiting window size to one, limiting transmission speed to 9600, and using a simplified state chart. By employing these and other IrDA Lite strategies, it is possible to achieve a two to five fold reduction in RAM and ROM requirements. For some small devices, throughput and features are not as important as memory, and the decision to use IrDA Lite is easy. For other devices, the tradeoffs are not so straightforward. In these cases, one need not employ all IrDA Lite strategies. A designer can employ those strategies that make the most sense, garnering the memory savings desired, without completely sacrificing throughput or feature set.

8.0 Summary

The number of IrDA devices available has been growing rapidly for the last five years and its growth continues to accelerate in the portable device market. The vision of ubiquitous point and shoot connectivity is becoming more commonplace. The continued spread of IrDA technology will be a key enabling factor in broadening acceptance of portable devices of all types and sizes throughout the world.

9.0 List of IrDA Specifications

IrDA Serial Infrared Physical Layer Link Specification, v1.2, November 10, 1997

IrDA Serial Infrared Link Access Protocol (IrLAP), v1.1, Jun 16, 1996

IrDA Link Management Protocol, v1.1, January 23, 1996

IrDA Tiny TP: A Flow Control Mechanism for Use with IrLMP, v1.1, October 20, 1996

IrDA IrCOMM: Serial and Parallel Port Emulation over IR, v1.0, November 7, 1995

IrDA Object Exchange Protocol (IrOBEX), January 22, 1997

Minimal IrDA Protocol Implementation (IrDA Lite), v1.0, November 7, 1996

IrDA Serial Infrared Physical Layer Measurement Guidelines, v1.0, January 16, 1998

IrDA Interoperability Test Plan and Process, v1.2, September 28, 1998