Protocol for self-alignment of laser-through-air ethernet links [unfinished, $Revision: 1.10 $]

Matthew Astley

5th September 2003

Cambridge, UK
matthew@fruitcake.demon.co.uk
http://www.fruitcake.demon.co.uk/

Abstract

Home-brewed hardware based on laser diodes is becoming more popular for point-to-point data links, but alignment problems have discouraged[1] common use. This article discusses some principles for controlling an ethernet-based transmitter with a steerable beam, then weighs the need for hardware simplicity against the possible theoretical efficiency. A mechanism for accurate hardware movement is assumed.

1 Introduction

Current recipes[Ronja, UPNLT] for optical data links require regular manual alignment. This is usually assisted by screw threads for fine adjustment, binoculars or perhaps a telephone link to the operator at the other end. If the transmitter mounting point moves, for example because of slight building subsidence, the beam will move off target and the procedure must be repeated.

1.1 Caveat on hardware assumptions

The author has no experience of building or aligning laser transceivers. Hopefully the ideas in this article will be useful despite this.

[1] It is considered normal to need to realign laser transmitters every month, even for buildings with solid foundations. This significantly reduces the usefulness of the link.
http://home.vicnet.net.au/~upnet/#physprobs

2 Steering the beam

It is assumed that the beam can be steered precisely and accurately, and without complicated hardware. This is not necessarily true, but without this assumption the paper requires prototype hardware - software is cheaper than a prototype. It is hoped that the presence of ideas or software will encourage someone to try the hardware.

2.1 Mechanism

Two perpendicular loudspeaker drivers or "voice coils" might make adequate drivers for a small transmitter.

> Figure: laser on a rod, pivoted at the back. Two driver coils at the front. Dimensions with names that I can refer to later.

It is not obvious that such apparatus will match the requirements below.

2.2 Steering requirements

Without certain properties, attempting to align the beam is likely to be futile in any case. In constructing a manually aligned transmitter, one may pick something which conforms to these requirements without realising why they are important. The ideas presented here will require them formally.

Positioning of the beam must be repeatable, at least over any "short" time period compared to movement of the mount. Hysteresis of positioning, or interference from a power supply will make control that much harder. The mount point will move slightly over a period, adding an unknown offset to the position. It is this offset which we wish to cancel to keep the beam on target. Since the offset changes relatively slowly, positioning accuracy can be maintained over "short" periods.

Beam positioning must also be precise. Specifically it must be steerable to within a "small" fraction of the active receiver element size or the spot size, whichever is larger. Using larger diameter[2] beams will make this easier, but requires good quality optics. Keeping the range of movement down will also aid precision, since the amplitude of the maximum permissible error becomes smaller relative to the full scale deflection.

2.3 Useful features

The system in 2.1 operates by scaling of triangles[3], so it will have a linear response. As a result the scanning speed on the target will be proportional to the scanning speed at the transmitter.

Scanning performance can be improved with fast beam control, so an area of the target can be scanned quickly. Fast scanning will require optimal damping, so the

[2] Larger diameter beams have other advantages where eye safety and disruption by birds are concerned, so it is becoming more common to expand the beam.

[3] Even for systems with an angular beam deflection, the angular movement needed to deflect the beam is small enough that $\sin\theta \simeq \theta$ (small angle approximation), because the laser target is generally at least 50 m away.

beam settles quickly after a change in movement. This follows from the requirement to maintain the positional accuracy.

3 Guiding principles

It is important to note that in steering the outgoing beam, the transmitter will have no immediate feedback on the received signal quality. If the beam fails to hit the receiver, no further information can be obtained[4] by either station. Even when the beam does hit the receiver, the transmitter will not know until some time later.

The situation is symmetrical since the link is full duplex when properly configured.

3.1 Data link speed allows for useful momentary contact

The data rate on a laser beam is sufficiently high that only a momentary alignment is necessary for a small transmission. "Momentary" for a 10Mbit/sec link may be as little as 13 μs for 64 bits of data preceded by 64 bits of clock synchronisation bits, or 26 μs if you allow two periods to cover the possibility of missing the start of a sequence.

The current transmit beam position can be encoded in a 64 bit data burst. If the receiver can return this data to the transmitter by some route, a link can be established.

There is also space in the data burst for a "last known good position" reply to the other station. This can provide a return path for coordinates, provided they are known at the time the beam hits the receiver.

Once one station returns a good set of coordinates to the other, by some means, a unidirectional link can be established. This can then be used to provide feedback for the other direction. For example,

1. Two stations, A and B, have just been set up[5] and do not know each other's position.
2. Station A happens to hit station B while transmitting coordinates (x_A, y_A).
3. Station A continues scanning, since it does not know it found the correct alignment.
4. Station B is also scanning, and when it happens to hit station A it will return the good coordinates (x_A, y_A) along with its current coordinates (x_B, y_B) .
5. Station A now knows a good position for itself, which it can then use to return (x_B, y_B) to station B.
6. Station B receives (x_B, y_B) and alignment is now complete, subject to small optimisations.

[4]The given advice[UPNLT] for manual alignment of links to around 500 m is to watch the laser spot with binoculars from the transmit station.

[5]This is not the stated purpose of the article, but it is a useful property that the system can bootstrap itself from incorrect alignment, even without an external (e. g. modem) connection.

Initial set-up will require an approximate alignment, but the larger the sweep area of beam driver the easier this will be.

The time taken to perform these steps will depend greatly on the target sweep speed of the beam driver and the distance between scan passes (assuming a raster-scan or spiral approach).

The great benefit of the short term repeatability is that even a delay of minutes before step 4 completes will not damage the protocol. The process can be short-circuited to some extent by returning (x_A, y_A) via a modem link, but this will not accelerate the establishment of a duplex link.

3.1.1 A carrier is easier to detect but less useful

The simple presence of a sine or square wave can be detected quickly and easily, and the arrival time could be used to deduce the position. With a contact time as low as 1 μs for 10 cycles of a 10 MHz signal, a 100 m^2 grid of 10 mm squares can be scanned in just one second.

Clock skew between the computers at the end points will be a major problem with this approach. Software such as NTPD is very good at synchronising the local clock to a remote source, but the errors may still be several hundred milliseconds - enough to ruin a calculation of position.

No data can be transferred[6] until the time of contact (i. e. position) has been returned to the transmitter. This is not such a great problem since a wired connection to the Internet is likely to be required for clock synchronisation anyway.

Finally, the raster scan rates of 1 kHz needed to cover the target so quickly will not be possible if the laser and optics are to move wholesale. A steerable mirror approach should run to such frequencies but inserting a standard glass mirror in the beam will cause losses of around 10%.

In an automated system, the speed gains would be lost in working round the other problems. Transmitting checksummed data seems to be a much more useful approach.

For manual alignment, a photographic flash unit triggered by detection of the carrier signal could prove invaluable for quick setup.

3.1.2 Methods for encoding position

The position encoding is not very important, since it is written and later read by the transmitter. In the absence of more advanced processing, the receiver merely relays the data back to the transmitter.

The simplest encoding is the Cartesian coordinate system of positions on the receiver, as seen from the transmitter. Precise time stamps, serial numbers or randomly generated cookies would work, but the transmitter must then store them for later recall.

3.2 Knowing the landing point on the receiver is not vital

If the receiver knew which part of its active area had detected a signal, it could guide the transmitter very precisely. When the beam wavered from a central position, the receiver could direct it back and alignment need never be lost.

[6] It should be possible to send approximately one bit per scan, but reliability will not be great.

Unfortunately the commonly used PIN diode detector does not allow for this. It may be possible to have an array of detectors with an amplifier each and some mixing after that, but even after the expense and complication there will be inactive gaps between receiver elements.

As shown in 3.3, however, an indication of signal strength can [TODO: something]

3.3 Knowing the received signal strength is useful

There will be a small range of acceptable misalignment, in which a signal can still be decoded. With expanded beams and optics there should be a gradual decline in signal strength as the alignment is lost, compared to a sudden cutoff when a small spot no longer lands on the active area of a naked PIN diode.

Feedback of signal strength from the receiver to the control system would be valuable in the case of expanded beam apparatus. Measuring the amplitude of the received signal .. calc $(\frac{\partial SNR(x,y)}{\partial x}, \frac{\partial SNR(x,y)}{\partial y}) = \nabla SNR(x,y)$?

TODO: I was suffering fuzzy-brain-itis when I wrote this, I think. There are notes in the LYX file.

3.4 The Ethernet protocol can be interrupted without data loss

The ethernet protocol[IEEE, EG3567] is tolerant of interruptions. If necessary, an alignment transmission can be made by shutting off the ethernet transmission, even in mid-packet, and sending some data. The higher levels of the OSI model will simply see a dropped packet and retransmit as appropriate.

TODO: this part might need explaining some other way.

4 Hardware for communicating alignment data

The design of hardware capable of sending, receiving and checking the integrity of short bursts of data is beyond the scope of this article. It seems safe to say it would have much in common with existing ethernet technology: a phase-locked loop, perhaps a preamble sequence with a packet start marker and a checksum at the end. There may be some benefit in running such a system at a different bit rate, e. g. 7Mbit/sec, so the ethernet and alignment receivers cannot lock onto each other's signals.

It is assumed that the system can send a continuous stream of bits containing packets about 64 bits long. There is no need for an inter-frame gap because there can be no collisions.

A system with a 64 bit preamble and 64 bits of data could send 78,000 packets/second. Only one complete packet is needed at the receiver, hence the figure of 26 µs in section 3.1.

4.1 Coping with existing hardware

Having illustrated the possibility of rapid back-to-back packets, we now consider what would be lost if the existing ethernet card is used to send and receive the alignment data.

4.1.1 Reduced sample rate

The minimum size of an ethernet[7] frame is 64 bytes, plus a 64 bit preamble and 9.6 μs inter-frame gap. This will carry the minimum size of data payload, 46 bytes. Although this is considerably more than we need, there is no advantage to sending less data because the frame size cannot be made any smaller.

The maximum transmission rate is 14,800 small frames per second[8], so the loss in performance will be at least a factor of $5\frac{1}{4}$. Performance is not critical however.

4.1.2 Increased data lag

Dedicated hardware could send a latched copy of the data used for steering the beam, for a latency of around one packet. This would be ideal, because the received coordinates would be as close to the correct value as possible.

Packets which have queued in the IP stack of an operating system are likely to be up to a millisecond late, so the alignment coordinates the receiver sees may well be wrong if the beam is moving quickly.

Stale data can by sending future coordinates with taking a second scan much more slowly over [TODO]

4.1.3 Benefits

The main advantages are in cost and simplicity. Mass produced ethernet cards are very cheap, for the functionality they provide - and the system will use one anyway. With custom transceiver hardware comes the need for another communications channel to the controlling computer, or even a dedicated microcontroller.

Inside the smallest ethernet frame, there is space for a UDP packet containing a of 18 bytes of data. This is plenty for alignment purposes, so there is no data-size cost to UDP. The use of standard UDP sockets will avoid the need for reading and writing custom ethernet frames. Furthermore, it should be much easier to route data back to the transmitter via a wired connection if the packets are of a standard routable form.

46	Ethernet data
−20	IP header
−8	UDP header
18	UDP payload

It would be very convenient to interleave alignment data with the bulk data on the beam, instead of interrupting data flow at inappropriate moments. Interleaving data would be hard to implement in custom hardware.

5 Calculations on acquisition time

I have done the calculations. I should write them here, but mostly once you ask the questions you find the answers are easy and suggest that the project will work.

[7] I take the normal meaning of Ethernet, i. e. 10Mbps. The 100Mbps variety is Fast Ethernet.

[8] Older computer hardware may not be capable of sending back-to-back packets, but a good quality ethernet card may help with this.

5.1 How long does it take to send the data?

5.2 How far can the beam move out of perfect alignment before the link is lost?

5.3 How big is the scan area and how long will it take to cover it?

6 Conclusion

TODO

7 Ideas and further work

The author intends to do 7.1 and 7.2.

7.1 Write a control daemon

The first version will need to

- listen on a UDP port
- accurately time stamp incoming packets
- log an analogue signal from (perhaps) a sound card, when the packet comes in
- be capable of transmitting a closely spaced flood of UDP traffic containing current positioning data. Outgoing traffic queues will not help here, and it may be necessary to do something more cunning than pile up outgoing packets of stale data.

Scanning the target with the path of a Hilbert curve instead of a more conventional raster scan may server to reduce requirements for freshness of data as it leaves the ethernet card, perhaps at the expense of a requirement for better damping.

It may turn out that the system time in the kernel for dealing with this flood of packets is too high for an older machine. This may make it worth writing some kernel level filter which can hijack the appropriate packets as they come in, but this is just an optimisation. In fact it could be bolted on to an existing daemon...

7.2 Build a steerable laser pointer

Mount a standard laser pointer at the intersection point of two small perpendicular loudspeakers and attempt to draw some pictures on a wall with it. The loudspeakers can be driven from a standard sound card. Creating some Lissajous figures should not be a difficult task.

7.3 Steerable mirror

On a smaller scale, a pair of piezo sounders could be used to drive a small mirror.

- The power consumption will be smaller.
- The system should be more responsive because of the reduced mass of the driven element and perhaps more precise by the nature of piezo crystals.
- Construction is likely to be more complex.
- Extra optical loss in the mirror.

7.4 Steering with worm drives

A pair of motors with high ratio worm drives would be capable of moving larger apparatus.

- Precision should be very good.
- Damping will not be a problem, but only because the movement will be quite slow.
- Knowing the absolute position of the beam at the other end then requires a counter on each drive motor.
- Static power consumption is zero.

7.5 Managing without absolute steering

If the requirement for repeatability is relaxed, construction of a high precision drive may be easier (7.4).

If absolute positioning is not available, timely feedback from the receiver becomes more important. Initial alignment may require a modem to report the first "hit", otherwise it will be harder to restore the correct position.

When the beams are only slightly misaligned, it will be possible to correct this with small relative adjustments, as in 3.3.

7.6 Multiple links

7.6.1 Multiple receivers

Allows for routing round birds (block beam for ~ 7 ms, so it's hardly worth it).

Ethernet MAC address assists figuring out who's who, and for free.

7.6.2 Multiple transmitters

It will all go horribly wrong if two transmitters point at one receiver. The collision detection and subsequent retries won't work because the transmitters don't see the collision.

If the receiver can detect the collision by signal strength/frequency analysis, it could use its own beam to cause collision detection at the transmitter - would require half duplex operation, which is otherwise not required.

A long thin tube with the detector at the bottom may be the answer here. Also keeps the sun out most of the time.

8 Acknowledgements

I would like to thanks the owners of the listed References for inspiring my interest in laser data links and assisting with technical details.

It is very likely that my ideas are based on something I have read somewhere, but I am not aware of any similar published work in this field. My "literature review" consisted of a few Internet searches, so I apologise if I have missed any previous work.

References

[Ronja] Twibright Labs, "Ronja", `http://atrey.karlin.mff.cuni.cz/~clock/twibright/ronja/`

[UPNLT] Realtime Control, "UPN Laser Transceiver", `http://www.alphalink.com.au/~derekw/upntcvr.htm`

[IEEE] TODO: 802.2 standard link

[EG3567] Godred Fairhurst, "3rd level EG 3567 course", `http://www.erg.abdn.ac.uk/users/gorry/course/`, ERG, University of Aberdeen, April 2002

[1] K3PGP. TODO: I can't get any sense out of `http://www.k3pgp.org/` at the moment, and I don't remember what it was I found there exactly.

[2] David A. Johnson, P.E., "Imagineering on-line magazine", `http://www.imagineeringezine.com/`, April 2002

[3] Matthew Skala, "On free networking", `http://ansuz.sooke.bc.ca/ofn.html`, May 1999

[4] Samuel M. Goldwasser, "Sam's Laser FAQ", `http://www.repairfaq.org/sam/lasersam.htm`, 1994 - 2002

Other links

- AFRL, “Development of Laser Data Link for Airborne Operations”, `http://www.dodccrp.org/Proceedings/DOCS/wcd00000/wcd00099.htm`
- Canon, “DT-50 Technology”, `http://www.canon.com/bctv/canobeam/canb_dt-50/tech.html`
- Optical Access, “Terescope”, `http://www.opticalaccess.com/products-ts.shtml`
- ...that company a couple of years ago that wanted to put multi-headed lasers on many buildings and have a network. Haven’t heard anything about them in a while.