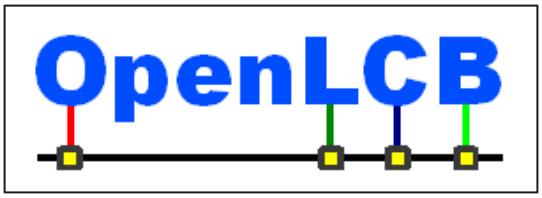


OpenLCB Working Note	
Traction Protocol	
May 2, 2013	Preliminary

1 Introduction

This working note covers the Traction Protocol, the way that OpenLCB handles moving objects such as locomotives, engines, and other rolling stock.

A Working Note is an intermediate step in the documentation process. It gathers together the content from various informal development documents, discussions, etc into a single place. One or more Working Notes form the basic for the next step, which is one or more Standard/TechNote pairs.

This protocol is for the long-term operation of really cool trains, not just "better control over DCC". Still, DCC is an important test case, both with new control hardware (OpenLCB native throttles connected to an OpenLCB-native command station that's generating the DCC signal) and when connected to legacy throttles and/or command stations.

1.1 Terminology

"DCC" refers to NMRA DCC; "Legacy" refers to all pre-existing protocols including DCC, TMCC, Marklin, DCS, etc.

"Trains": For our purposes, Train is anything which can be independently controlled. In addition to a model of a prototype train from locomotive to caboose, it might be just single caboose, a set of lit & controlled passenger cars, a diesel MU lash up, or basically anything that can take an OpenLCB "decoder" or a DCC decoder with a legacy attachment.

"Train Node": A Train is associated with a single, specific node. The Node ID is the fully-unique identifier for that Train.

"Throttles": For the purposes of discussion, we draw a distinction between three kinds of throttles that a user might encounter:

- "Legacy Throttles" refers to throttles designed for use with extant DCC systems, e.g. a Digitrax DT402 or Lenz LH100.
- "Full-Featured Throttles" refers to full-featured native OpenLCB throttles with multi-line color screens and effectively unlimited processing power, e.g. a software throttle implemented on an iPad.
- "Simple Throttles" refers to throttles which are native OpenLCB nodes like Full-Featured Throttles, but which have more limited capabilities, e.g. no text display, a limited array of physical buttons, and constrained processing resources.

"Proxies": In the long term, we expect that OpenLCB protocols will go all the way to the train. This has great advantages, because you're always in complete communication with the train, and

don't have to worry about only being able to configure the train when it's on a service track, storing information somewhere else so that it can be retrieved while the train is moving, etc. But until radio or other technologies mature to the point that this is possible, "proxy nodes" can be used as stand-ins for that capability. A throttle might communicate with a node that's serving as a proxy for the train, handling the communications, keeping track of status & configuration, etc. Out the back end of that proxy node is some other kind of communications, perhaps direct DCC or a connection to a legacy system that in turn makes DCC signals, or some other technology entirely. Due to the nature of those back side communications methods, the proxy may not be able to do everything that OpenLCB can, or only do some of it at certain times. The OpenLCB traction protocols need to take this reality into account. Anther use case for proxies, discussed in detail below, is to provide the mechanism for consisting.

“Command Stations”: Existing DCC and other control systems use “command stations” to create a track signal for controlling the trains. Usually the command station is controlled from the user side by some other network, to which throttles and other interface devices are connected. OpenLCB, it it's native form, has no such concept. Devices, like throttles, that want to talk to a train do so directly. Only when working with legacy systems does the concept of a command station enter, and usually through the form of a proxy node that is acting for the Train.

“Consisting”: The running of multiple items together, e.g. three coupled engines, each with their own NodeID or DCC address, as a single locomotive. DCC systems provide this now in various ways and with various names.

“Configuration”, “Functions”: Traditional DCC decoders provide “functions” for controlling accessories such as lights and sounds during operation, and provide a separate mechanism for doing long-term configuration via Configuration Variables (Cvs). OpenLCB makes the same distinction, providing access to “functions” via the traction protocol, while leaving “configuration” to the configuration protocol(s). The line between these is admittedly vague, and different node developers may implement some capability one way or the other. The general intent is that things that are changed in normal operation are considered functions, while things that are set once and forgotten are configuration.

1.2 Served Use Cases

1.2.1 Train Operation

Bill hasn't run his passenger train recently on his OpenLCB-equipped layout. He picks up a throttle, hits a few keys, sees his passenger train listed, selects it and starts to run it. Some configuration needs tweaking (e.g. volume too low), so he enters a configuration dialog on the throttle, finds the right item by reading through them, changes the value to be a few larger, and stores that back into the train as it's running on the main track. That makes it work immediately.

1.2.2 Large Modular Layout

Arnold has put his OpenLCB-equipped train on a large modular layout, where it is one of 500 pieces of equipment. He picks up a throttle, presses a few keys, sees his train, selects it and starts to operate it.

1.2.3 Train on New Layout

Jim takes his OpenLCB-equipped train to Bill's OpenLCB-equipped layout and puts it on the track. He picks up a throttle, hits a few keys, sees his train, selects it and starts to run it. On this layout, some configuration needs tweaking (e.g. volume too low), so he enters a configuration dialog on the throttle,

finds the right item by reading through them, changes the value to be a few larger, and stores that back into the train. That makes it work. When he gets back home that value is still present so he changes it back using the same procedure.

1.2.4 Legacy Train on New Layout

Jim takes his DCC-equipped train to Bill's OpenLCB- and DCC-equipped layout and puts it on the track. He picks up a throttle, hits a few keys, sees his train, selects it and starts to run it.

As an alternative, Jim takes his DCC-equipped locomotive to the layout, puts it on the track, enters the DCC address into a throttle, and starts to run it.

1.3 Unserved Use Cases

1.3.1 Multiple Independent Command Stations

Large modular layouts use multiple command stations to increase the effective bandwidth of the DCC bus. This is not an explicitly supported use case in the current work. Future work may make this possible as an extension.

1.3.2 Improved Legacy Addressing

DCC systems cannot run two locomotives with the same DCC address at the same time. This will still be true when running DCC-equipped locomotives via OpenLCB protocols to the command station.

1.3.3 Third-Party Communications

Node A is a throttle that is controlling train node B. Node C passively listens to the traffic and reacts to throttle commands and train status by taking various actions, such as providing appropriate sounds or preventing the speed from getting too high.

2 Specified Sections

This is the usual section organization for a Technical Note, to accumulate the Standard and Technical Note content in its eventual order.

2.1 Introduction

Note that this section of the Standard is informative, not normative.

2.2 Intended Use

Note that this section of the Standard is informative, not normative.

2.3 Reference and Context

NMRA S9.2 and NMRA RP9.2.1 define the formats for DCC addresses

Node implementing the Train Protocol must implement:

- Message Transfer Protocol
- Event Transfer Protocol
- Memory Configuration Protocol (optional?) and Datagram Protocol it depends on

- CDI
- SNII and/or ACDI?

Float-16 is the half-precision numeric format defined by IEEE 754-2008. It's available in the GNU tool chain via the -mfp16-format=ieee flag and __fp16 type.

2.4 Message Formats

AA.AA refers to an NMRA short or long address in the format defined by the NMRA. (Say a few words about short addresses in two bytes, maybe give examples)

2.4.1 Defined Event IDs

IsTrain: 01.01.00.00.00.00.03.03

IsIdleProxy: 01.01.00.00.00.00.03.04

IsInUseProxy: 01.01.00.00.00.00.03.05

IsProxiedDccAddress: 06.01.00.CC.AA.AA.03.03

EmergencyStopAll: 01.01.00.00.00.00.FF.FF

2.4.2 Traction Control Command Message

MTI: Priority 1, index 15, modifier 2, addressed => MTI 0x05EA, CAN frame [195EAsss] fd dd

The MTI modifiers are chosen to have the reply a higher priority than the request, ensuring replies to repeated instructions are always possible. The same priority and index are used for command and reply messages, changing only the modifier, to use less of the high-priority MTI space which is a scarce resource.

This message type and MTI is specific to traction control. The first byte of the content codes an “instruction”, which defines the rest of the format. The instruction codes were selected with the high nibble representing protocol (0x00 for OpenLCB full protocol; 0x80 for DCC legacy; others reserved)

Instruction	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
Set Speed/Direction	0x00	Speed and direction as signed float16						
Set Function	0x01	Address			Value			
Emergency Stop	0x02							
Query Speeds	0x10							
Query Function	0x11	Address						
Configure Proxy	0x80	Attach Node 0x01	Node ID					
		Attach DCC Address 0x81	DCC Address AA.AA		Speed steps (14, 28, 128)			
		Detach Node 0x02	Node ID					
		Detach DCC Address 0x82	DCC Address AA.AA					
Manage Proxy	0x82	Reserve 0x01						
		Release 0x02						
		Query 0x03						

Want to add “Query attached Node IDs” and “Query attached DCC addresses” as instructions.

The Set Function instruction uses a three-byte address for brevity; it's to be interpreted with a high byte of zero to make a four byte address in the function memory space (0xFA).

2.4.3 Traction Control Reply Message

MTI: Priority 1, index 15, modifier 0, addressed => MTI 0x05E8, CAN frame [195E8sss] fd dd

Higher priority to ensure can be sent immediately over Traction Control Command messages. Coding and structure similar.

Instruction	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7	Byte 8
Query Speeds Reply	0x10	Set Speed		Status	Commanded Speed		Actual Speed		
Query Function Reply	0x11	Address			Value				
Configure Proxy Reply	0x80	Attach Node Reply 0x01	Node ID						Reply Code
		Attach DCC Address Reply 0x81	DCC Address AA.AA		Speed steps (14, 28, 128)	Reply Code			
		Detach Node Reply 0x02	Node ID						Reply Code
		Detach DCC Address Reply 0x82	DCC Address AA.AA		Reply Code				
Configure Proxy Reply	0x82	Reserve Reply 0x01	Result: 0 == OK Non-zero == Failed						
		Query Reply 0x03	Max Number of Nodes		Number of Nodes Now	Max Number DCC addresses	Number DCC addresses now		

140 The Query Function Reply is in the format of the Set Function message, with a different MTI and the query bit set.

The Query Speed/Direction reply is almost in the Set Speed/Direction format, with the addition of the two additional speeds. If a node cannot provide any of those three speeds, it should use float16 NaN (not a number) 0xFFFF. “Set Speed” is the most recent speed received in a Set Speed/Direction instruction. “Commanded Speed” is the speed that the traction control is currently attempting to move, taking into account momentum and any other control modifiers. “Actual Speed” is the current measured speed of the locomotive. There is no accuracy guarantee for Actual Speed.

On CAN, the Query Speed/Direction reply does not fit in a single frame, so it's sent as two frames with start and end marked in the 1st data nibble (high part of destination address). The status byte was included so that the actual speed value would not be split across boundaries (though that's not necessarily guaranteed for other wire protocols that come along later) The status byte is reserved. Send 0x00, don't check on receipt. For example, from node with alias 123 to node with alias 456, all speeds equal to 0x4420 would be sent as the two frames:

195E8123 14 56 10 44 20 00 44 20

195E8123 24 56 44 20

which together are the single message

MTI=05E8 10 44 20 00 44 20 44 20

2.4.3.1 Proxy-Specific

The 0x80 and 0x82 instructions are only for proxy nodes. The MSBit is tentatively reserved to indicate proxy-specific instructions.

Attach reply code: Zero or not present means OK

Detach reply code: Zero or not present means OK

Do we want to return the number of attached nodes/DCC addresses?

Error codes need to be defined.

- Allocation Failed, temporary (try again) handles race conditions *(but we need to look carefully at that race condition in practice: With the current structure, does it end up adding two DCC addresses to the single proxy, instead of reserving it? Probably need to separate out reservation-for-configuration. In general, two different things like “assign” and “reserve” should be handled by two separate things)*
- Allocation Failed, permanent (out of proxies, etc)

Want to add responses to “Query attached Node IDs” and “Query attached DCC addresses”. These have unbounded length if the number of members in a consist is not bounded by the protocol. Could be a single large message, but everybody in the communication path still needs a maximum message size to ensure buffering. So a multiple message (or multiple datagram) response with an “last one” bit is the way to go. Using datagrams is more protocol work, but leverages existing protocol support and moves the response to a lower priority.

2.5 States

Full OpenLCB nodes:

- Set Speed – The speed set by a throttle, the content of the most recent “set speed” instruction
- Commanded Speed – the current speed that is intended. This may lag behind the set speed intentionally due to momentum settings, etc.
- Current Speed – a physical state, the speed at which the object is currently moving

Emergency stop is not a state.

Proxy Node States:

- Idle – not reserved and has neither nodes nor legacy addresses attached
- InUse – reserved and/or has node IDs and/or legacy addresses attached

The IsInUseProxy and IsIdleProxy Event IDs are used to indicate transitions in the state of a proxy node. Proxy nodes are created in Idle state.

2.6 Memory Spaces

2.6.1 Configuration Information

The configuration memory space holds the configuration of the train, such as how functions will work, how speed in scale meters/second will control motor operation, etc.

For a proxy node, the configuration memory space contains information on both the proxy operation, and also the proxied train(s) and DCC decoder(s). The configuration memory space is mapped to the CV configuration space for DCC decoders.

2.6.2 CDI

2.6.3 Function Information 0xF9

Functions, such as lights and sounds, can be operated by the Traction Control Set Function instruction, and their current value can be retrieved via the Traction Control Query Function instruction. The values are also available for reading and writing in the Function Information memory space. This allows multi-byte reads and writes using Configuration Memory Protocol access. That's a more efficient way of setting and reading large numbers of functions, for backup, initial setup, etc.

The NMRA 9.2.1 Recommended Practice describes DCC as having three separate sets of "functions". The most common one is the traditional F0-F28. In addition, the "Binary State Control Instruction long form" accesses 32767 addresses (confusingly called "states" in the NMRA doc) and the "Binary State Control Instruction short form" accessed 127 addresses (the NMRA document implies that these are overlapping address spaces, but at least one manufacturer has not implemented them that way). Finally, NMRA DCC WG Topic 9910241 never made it to the Recommended Practices due to internal Working Group politics, but has been widely implemented by decoder and command station manufacturers. It provides an "Analog Output Instruction" or "Analog Function Group" with 8 bits of address space and 8 bits of value for each address, for a total of 256 functions and a value in the range [0..255] for each function.

The OpenLCB 0xF9 space is allocated to cover all these by using the third byte of the address as a selector.

Type	Low Address	High Address	Values
F0-F28	0x0 00 00	0x0 00 1C	A non-zero value indicates "ON", a zero value is "OFF".
Binary State Controls (full space)	0x1 00 00	0x1 7F FF	"
Binary State Controls (short space, if separate)	0x2 00 00	0x2 00 7F	"
Analog Outputs	0x3 00 00	0x3 00 FF	8 bit unsigned quantity

The default Function Definition Information, when nothing is known about a particular decoder's capabilities, will just describe these as above. If the node serving the DCC decoder information knows more about the particular decoder's capabilities and/or preferred labeling, it can provide more information via the Function Definition Information.

There's nothing in the standard that says that the first three types of information need to be stored in 29+32768+128 bytes. Packing them into bits is certainly acceptable, given that the underlying DCC protocol can only send one bit for each. An individual node may not implement all of them, either.

2.6.4 Function Definition Information FDI 0xFA

“Function Definition Information”, similar in intent to Configuration Definition Information (CDI) is stored in XML format in address space 0xFA to provide user-oriented context. That includes:

- Memory layout of the function values, allowing for multiple data types from binary (one and off for lights) through integer values (for e.g. sound intensities) and strings (sign displays?).
- Function naming, so that a throttle can display useful names to the user such as "Bell", "Coupler Clank" and "Master Volume". This includes internationalization of those labels.

What else needs to be conveyed? "Make this prominent on the throttle"? "Have this there, just a little less prominent"? "Seriously, nobody cares about this option, bury it"?

At present, there are no default values that e.g. associate "Bell" with a particular location or function. These are thought to be too brittle, and there are just too many possibilities to be useful (see the unscientific and incomplete Survey of existing function names).

2.7 Interactions

2.7.1 Emergency Stop

Receipt of the Emergency Stop instruction stops the locomotive as fast as possible. This sets the set speed to zero (preserving existing direction) and the commanded speed to zero (preserving existing direction) regardless of any momentum, BEMF or other operations with the train node.

Emergency stop is not specific state. The next Set Speed/Direction instruction will act immediately to change the set speed, and start the commanded speed and actual speed moving toward that set speed.

2.7.2 Function Operation

Function values are stored in the 0xF9 memory space. They are written using the memory configuration protocol or using the Set Function instruction.

2.7.3 Train Configuration

Trains are OpenLCB nodes just like any other. As such, the Memory Configuration protocol can be used to configure them, the Configuration Description Information system can be used to make that process user friendly, etc. There's nothing traction-specific in these techniques, which are available any time that the train node is connected to the OpenLCB.

The configuration information in a train can include the user documentation that's sometimes referred to as "roster information". This might include owner name, prototype railroad and road number, information about the particular model's construction, etc. As yet, OpenLCB has no standards nor conventions on this information.

One approach to standards in this area would be extend ACDI/SNIP. Those currently have a block for manufacturer identification, and a block for user identification. Those are both versioned. We could take (one of) several approaches:

- Extend them with a third block, only present when the node implements the train protocol (as seen in PIP). To allow later introduction of more types, this block would have some versioning/type information, but that's straightforward.
- Create a version 2 of the user block, which holds additional data. (This would be using 2 as a format identifier, rather than a version number; that might make versioning complicated)

Suitable content for a (first version of) this might be (from [http://jmri.org/JavaDoc/doc/jmri/jmrit/roster/RosterEntry.html JMRI roster], see also similar concepts in [rocrail.net RocRail]):

- Road Name
- Road Number

(Manufacturer, model, owner description, comments, etc are already present in ACDI/SNI)

For DCC locomotives, more terms might be desired:

- DCC Address
- Decoder Type (Manufacturer, model)

2.7.4 Train Identification

OpenLCB Train nodes use:

- The Event Transport protocol to locate Train nodes
- PIP for enquiry about the support
- SNII and/or Memory configuration, CDI & ACDI for identification of a specific train node.

Trains are OpenLCB nodes just like any other. As such, they can take part in protocols such as Node Verification and Simple Node Information which allows other nodes to learn about them.

Train Acquisition Protocol is necessary because the train operator doesn't want to pick up a throttle and enter "06.011.00.02.1F.2D" (a node ID), or even "110 Long" (a DCC address), but rather just pick the desired locomotive from a list of those available. (A throttle should still allow the operator to directly enter the address, when that's what the operator wants to do.)

The train acquisition process simplifies locating desired train nodes so that small hand-held throttle nodes can efficiently take part. It does this using several approaches, which can be used as needed by throttles:

- Events are used to announce the existence and status of Train nodes
- Train nodes will generally implement the Simple Node Information protocol so that throttles can get basic, user-readable identification from them
- A search protocol is (being) defined to make it possible to locate individual Train nodes without having to read information from all of them

So that other nodes can find them, Train nodes must produce the well-known reserved event 01.01.00.00.00.00.03.03. This means they must produce that event in a Producer/Consumer Event Report message when they power up, and reply to requests for producers of that event. An IdentifyProducer request will therefore find all the train nodes on the OpenLCB, and further protocols can be used to get additional information on the individual Train nodes it locates.

SNIP will be used to carry both manufacturer-provided and user-provided information about the particular train node. In particular, the user (Node) Name and (Node) Description fields are to be used to hold train identification information that can be retrieved and presented by throttles for selection.

A search protocol maay eventually be defined to allow efficient location of specific train nodes on large OpenLCB installations. It provides a general or field-specific search over the SNIP information, returning the Node IDs of matching Train Nodes. One proposal, which is not based on memory configuration & SNIP, is available here in pdf format and here in OpenOffice Writer.

2.7.4.1 Proxy-Specific Details

For locating all proxies and just idle or in use proxies, two well-known Event IDs are used:

Proxy must produce the reserved event

- EventID 01.01.00.00.00.00.03.04 indicates the node is a proxy in Idle state
- Event ID 01.01.00.00.00.00.03.05 indicates the node is a proxy in InUse state

The node shall produce both events, which means that it will respond to IdentifyProducer and IdentityEvents requests by indicating that it produces these event IDs, marking the one the corresponds to the current state as valid, and the one that doesn't correspond to the current state as invalid. When the state changes, the node will transmit a Producer/ConsumerEventReport message with the event ID corresponding to the new state.

3 Background Information

3.1 Speed Control

For OpenLCB, the speed and direction to set is encoded as a half-precision floating point number (aka 'float16'), with positive numbers indicating forward direction, negative indicating reverse, and (signed) zero indicating full stop. The value specifies a speed in scale meters per second (scale-m/s).

Note that even the zero value is signed. This is needed because locomotives still have a direction, even when they are fully stopped. It's used to control the configuration of lights and sounds that the locomotive exhibits when stopped. “Negative zero” is well-defined in the IEEE float-16 standard, but not all libraries implement it well, and it's easy for code to convert the negative-zero value to the more common positive-zero by default. Setting the sign after all computations are done is one way to handle this.

Rationale: The use of a 16-bit floating point permits relatively precise speed commands, especially at lower speeds; such fine granularity ensures not just fine-grained control over the locomotive, but helps avoid aliasing issue that arise during the conversion to lower resolution system-specific speed commands (i.e. DCC's 14 or 28-step commands). Using 32-bit floats uses more bandwidth, more program and data memory and CPU cycles on small nodes, but is somewhat easier for large nodes. The conversion between float-16 and float-32 is very simple, though, and any node with native 32-bit support can handle float-16 easily. The converse is not true.

The use of meters per second is somewhat arbitrary, and reflects standard velocity units used throughout the metric-speaking world. By standardizing on specific units, we avoid any future unit conversion issues. By standardizing on metric units, we simplify future attempts to simulate and control train physics.

The use of *scale* meters per second has two distinct advantages. First, it permits us to transmit speed commands in a scale-independent way. Second, and because of this, it reduces the number of parameters that must be estimated when controlling a locomotive that has not yet been speed-calibrated (which, for new users using existing digital control systems, will be all of their models). For example, on a DCC system, if I issue a command to proceed at 30mph, the command station must convert the value in the speed command from 30mph to an integer in the range [0-26] (for 28-speed-step control). The command station need only estimate what a reasonable top speed for a locomotive might be: Let us say, 100mph. Thus, the command station could reasonably estimate that 30mph translates to speed step 8.

The alternative possibilities considered were absolute speed using real units (as opposed to scale units), and relative speed units. The difficulty with relative speed units (i.e., percentage of full throttle), is that they are ambiguous, and preclude the possibility of performing physical simulations in the cab

controller, at least without completely abandoning the particular interpretations assigned to sped values. The difficulty with using real (as opposed to scale) units is that it requires the estimation of an additional parameter for uncalibrated locomotives, specifically the train's scale. If I issue a command to a DCC locomotive to proceed at 0.1 (real)m/s, the command station must not only understand what a reasonable top speed for a train is, but how to scale the speed appropriately, as 0.1 m/s might be quite fast for Z scale, but quite slow for G. As there is really no reasonable scale to use as a default, users must configure their digital command station to set the scale for either the entire layout, or on a per-model basis—an additional configuration step that is easily avoided by the mechanism for scale units described above.

3.2 Function Control

"Functions" like "horn", "headlight", etc are key user features when operating modern decoders. But they're also configuration-like, in that they effect the operation of the device.

How to handle them for a native OpenLCB piece of rolling stock, and for legacy ones via e.g. DCC?

We could take the purist approach and say "configuration is configuration, it's all the same". But that ignores that many people are going to want "Bell" to appear (automatically) on their throttle, but not so many are going to want "Kp back-emf correction factor" to appear there.

People just think about operating and configuring their locomotives as separate things. (Though e.g. "Master Volume" can cross the line)

This doesn't mean that we can't use the same protocol for all other of them. A mixture of memory configuration and CDI should do just fine. It just means that we need to find a way to include clueing information for the throttles on e.g. what to present.

Function values are stored in the 0xF9 memory space. "Function Definition Information", similar in intent to Configuration Definition Information (CDI) is stored in XML format in address space 0xFA to provide user-oriented context. That includes:

- Memory layout of the function values, allowing for multiple data types from binary (one and off for lights) through integer values (for e.g. sound intensities) and strings (sign displays?).
- Function naming, so that a throttle can display useful names to the user such as "Bell", "Coupler Clank" and "Master Volume". This includes internationalization of those labels.

What else needs to be conveyed? "Make this prominent on the throttle"? "Have this there, just a little less prominent"? "Seriously, nobody cares about this option, bury it"?

At present, there are no default values that e.g. associate "Bell" with a particular location or function. These are thought to be too brittle, and there are just too many possibilities to be useful (see the unscientific and incomplete Survey of existing function names).

3.2.1 Outputs vs Functions

Tools like DecoderPro and its decoder-definition files make a distinction between "functions", which are the control commands sent via e.g. DCC, and "outputs", which are the things that a decoder can do: Control an electrical output, make a sound, etc. This distinction is useful because one of the configuration options in (some) DCC decoders is a mapping between the functions and the outputs,

useful in a world where throttles generally have only about a dozen buttons, but decoders have many output options.

OpenLCB makes a clean separation between functions, which are the control operations, and all configuration & physical information, which lives in the memory configuration and CDI. If there's to be a mapping, it's defined through the CDI.

3.3 Configuration

Trains are OpenLCB nodes just like any other. As such, the Memory Configuration protocol can be used to configure them, the Configuration Description Information system can be used to make that process user friendly, etc. There's nothing traction-specific in these techniques, which are available any time that the train node is connected to the OpenLCB.

The configuration information in a train can include the user documentation that's sometimes referred to as "roster information". This might include owner name, prototype railroad and road number, information about the particular model's construction, etc. As yet, OpenLCB has no standards nor conventions on this information.

One approach to standards in this area would be extend ACDI/SNIP. Those currently have a block for manufacturer identification, and a block for user identification. Those are both versioned. We could take (one of) several approaches:

Extend them with a third block, only present when the node implements the train protocol (as seen in PIP). To allow later introduction of more types, this block would have some versioning/type information, but that's straightforward.

Create a version 2 of the user block, which holds additional data. (This would be using 2 as a format identifier, rather than a version number; that might make versioning complicated)

Suitable content for a (first version of) this might be (from [http://jmri.org/JavaDoc/doc/jmri/jmrit/roster/RosterEntry.html JMRI roster], see also similar concepts in [rocrail.net RocRail]):

- Road Name
- Road Number
- (Manufacturer, model, owner description, comments, etc are already present in ACDI/SNI)
- For DCC locomotives, more terms might be desired:
- DCC Address
- Decoder Type (Manufacturer, model)

3.4 How to find human-readable loco information

e.g. SNII for getting the name of a real loco, DCC loco

3.5 Consisting

Consisting is a process familiar to model railroaders, who use it for many purposes:

- Run multiple diesel engines together as a model MU
- Run helper locomotives together due to a shortage of (real) engineers
- Connect multiple pieces of rolling stock (passenger cars, cabooses) together to make it easier to control sound and lighting effects
- Connect sound decoders with motion-control decoders so they work well together

OpenLCB handles consisting and similar “one acts for many” situations by having a node act as the front-man for the consisted group. This node may be a real physical object, but it's more likely to be a software construct somewhere within the hardware that's used to connect the main OpenLCB network to the actual trains.

Once the consisting proxy node has been initialized, it will handle the individual parts of the Traction Protocol series:

- Speed Control - this is perhaps the simplest, just passing the speed values through to the individual members of the consist. Because OpenLCB speeds are in scale meters/second, there's by definition no need to rescale them when forwarding them to disparate equipment
- Function Control - Although the setup process may be able to map or reassign functions in the consist-member nodes, in the end this protocol is just a pass-through of the function instructions and memory access protocol operations.
- Train Configuration - except for very limited configuration of the consist itself, the consist does not take part in any configuration operations. Those are done by talking directly to the nodes that control the individual pieces of the consist. (CDI for the consist node is going to require careful definition)
- Train Acquisition Protocol - the consist is a separately locatable thing, as throttles will want to be able to find it. As such, it will take part in producing the well-known events, providing human readable (and writable) information via SNII/ACDI, etc, and being the target of search operations.

Because there's no protocol difference between full-OpenLCB-node Trains and legacy equipment that uses OpenLCB proxy nodes, consisting works the same for all of those. It can even mix-and-match full nodes with various types of legacy trains, if that's electrically possible. Proxies can handle multiple node IDs (and/or DCC addresses), which allows implementation of consisting. Speed/direction just passes through; speed matching is automatic due to the use of scale meters/second for units.

3.6 Proxies

Proxies are used to have a single proxy node run multiple OpenLCB trains together and/or to run legacy trains as OpenLCB-aware trains.

3.6.1 Lifecycle & Location

Each proxy has a lifecycle:

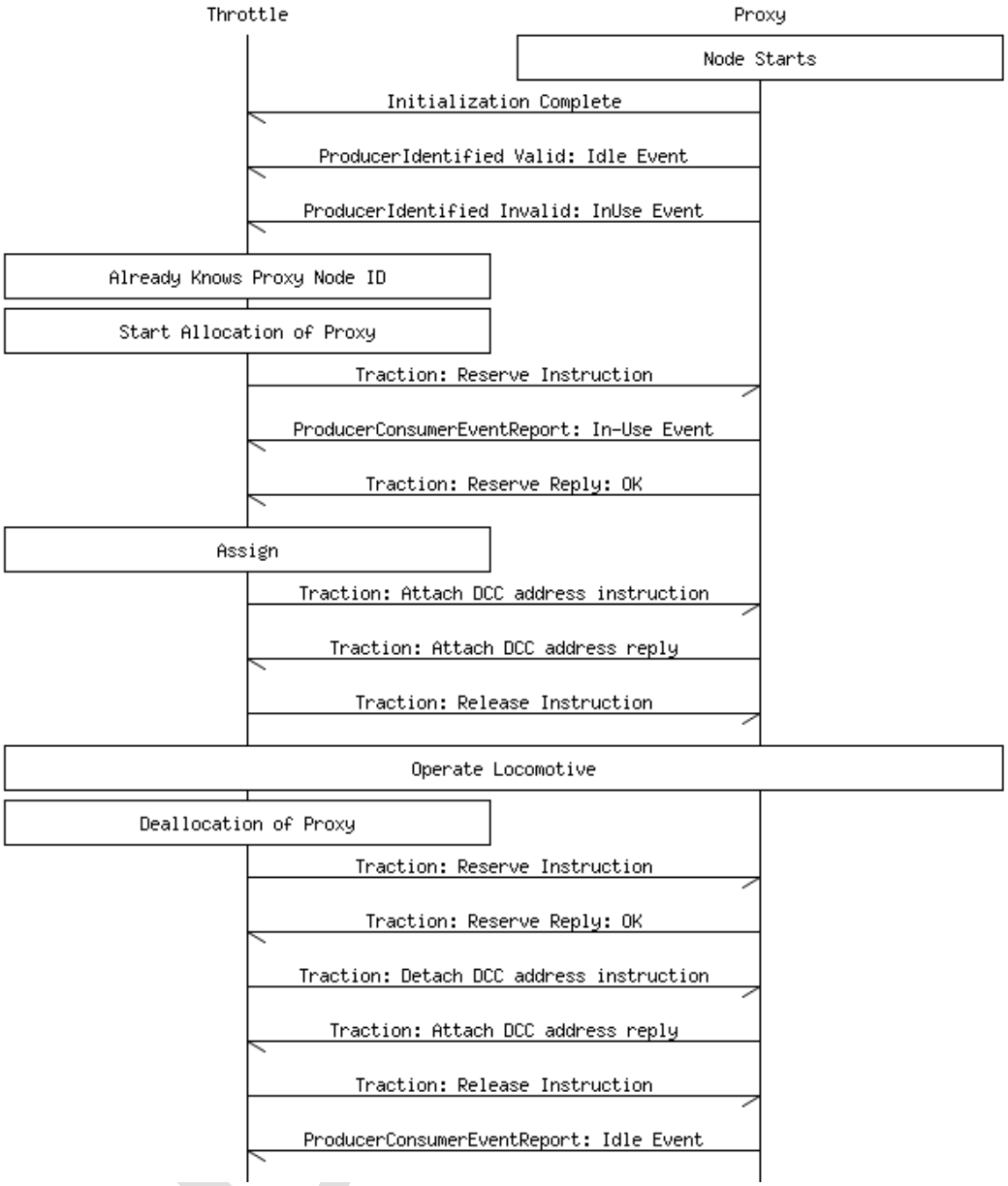
1. It is created. This might be when a physical node (the proxy or a physical node hosting a number of proxies within it) comes up, or it might be produced as needed.
2. A node reserves it. This ensures that only one node does the configuration of the proxy.
3. The reserving node attaches one or more DCC addresses or Node IDs to the proxy. These are the items that the proxy will operate.
4. The reserving node releases the proxy node.
5. The proxy is used to operate the equipment at that attached address(es). (Although this is optional, it's the whole point of having the proxy)
6. A node reserves it, to ensure it's the only node changing the configuration.
7. The reserving node deallocate the addresses from it.
8. The reserving node releases the proxy node. When/if the last address and Node ID have been deallocated, the proxy is now idle.
9. (Items 2 through 8 are a loop, and can happen multiple times)
10. Finally, the proxy is destroyed. (This is optional; a proxy may stay around until the layout is powered off)

So long as the proxy can do its job, there's nothing that requires the proxy to be resident in the command station hardware. It can be a separate board, in a computer somewhere that talks to pre-existing legacy equipment, or something else. One way for OpenLCB vendors to support DCC is to put the proxies and a command station in a single unit, but that's not required. Another approach would be an OpenLCB-compatible board that provides proxies, which then talk out through a LocoNet, XpressNet, TMCC or other connection to a legacy command station of a particular type.

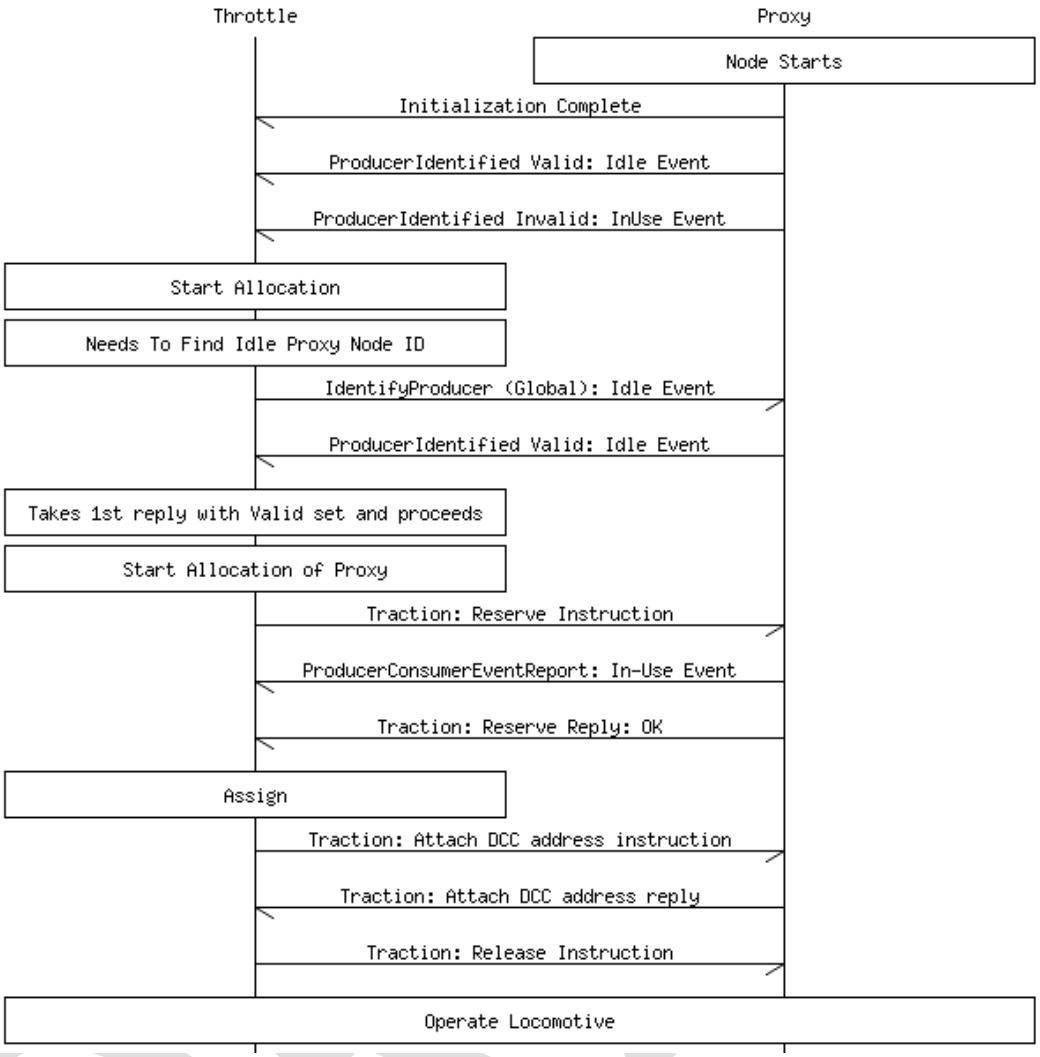
The simplest way to provide N proxies is to have N nodes appear when the layout turns on. At any given instant, only one idle proxy is needed, so another approach would be to create them as needed. In this case, the newly created ones will get link access (e.g. go through the process of getting an alias on CAN), then announce their existence with an InitializationComplete message followed by ProducerIdentified and ConsumerIdentified messages as needed. Those can be used by Throttle nodes to track the NodeID of available proxies without having to query for one. Note that this process needs to be very fast, so that multiple throttles needing to allocate proxies at the same time can reliably do it.

In the OLCB Train world, simply requesting which nodes are Train Nodes (asking for producers of isTrain event ID) is sufficient to begin communicating with a node through the Traction Protocol. In the Proxy Train world there needs to be an additional step to create a connection between the Train Proxy and the underlying DCC locomotive. This connection can not be static or you would need as many Proxy Train Nodes as there are possible DCC addresses (~10k). This connection step implies that a Train Proxy Node may or may not have a valid connection with the underlying protocol during its existence. Due to this two more states need to be added to the "IsTrain" state of a Train Node (isIdleProxy, isInUseProxy). Two are used because it is necessary to always have a defined response if

the Proxy Node is allocated or not. It would not be acceptable to have the Event be sent if it is associated with the underlying protocol and not sent if the node was not. How long does the caller wait before it decides the Event will not be sent?



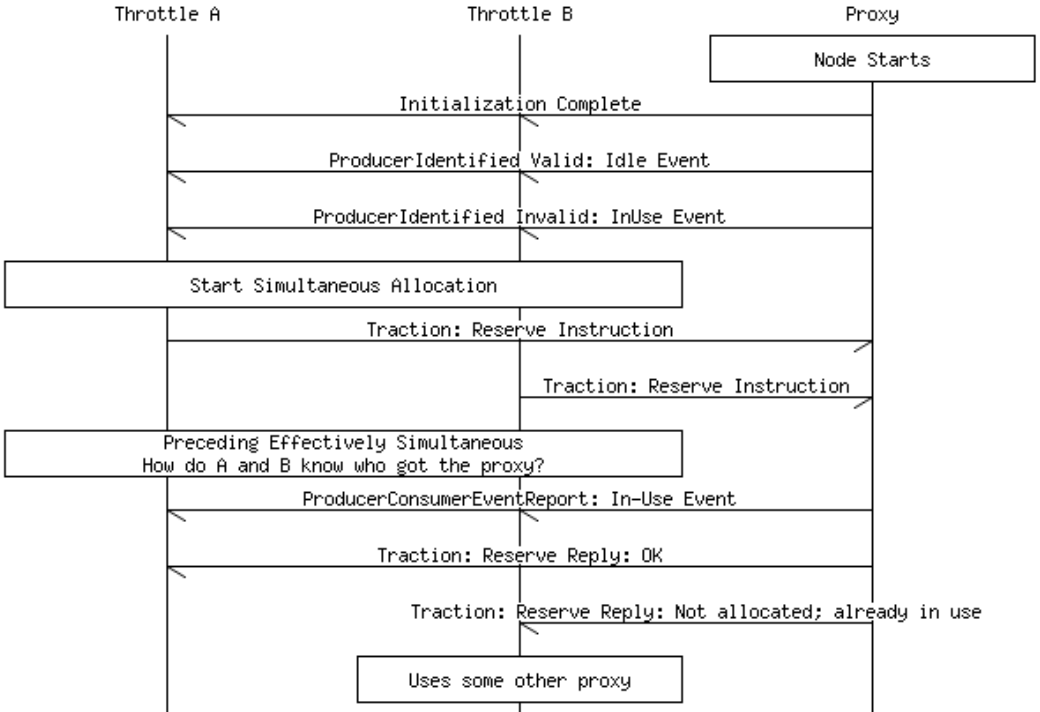
495 **Allocate, use and release a proxy**



Finding a proxy to allocate

3.6.2 Collisions During Allocation

You have to check the reply code and try again. That might mean inquiring for a new proxy idle proxy.



Handling a collision during proxy allocation

3.6.3 Legacy Protocol Proxies

The upper nibble of the Configure Proxy attach/detach instruction is intended to identify the specific protocol.

Query: Is 16 technologies sufficient? Here is a partial list of technologies we may have to support:

1. OpenLCB
2. DC
    1. With or without extensions for high-frequency constant lighting
    2. With or without extensions for three-rail control (eliminates need for reversing loops)
3. DCC
4. Märklin Digital (old and new)
5. Märklin DELTA (is this a subset of Märklin Digital, or something different?)
6. Märklin Digital (2004 w/ESU)
7. Selectrix
8. MTH DCS
9. Lionel LCS/TMCC
10. Gobs of legacy systems like Hornby Zero 1

There are two safety values for expansion beyond 16: It's possible to use almost all of the byte for this, not just the top nibble. Only the bottom bit is actually carrying any information. And some of these might be the same; the instruction is really only moving an address, which for DCC can be thought of as a 16-bit number. If command stations only deal with one protocol (but see the Digitrax command station line), then they don't really need to know about different protocols here, they just need to know the format of the address.

3.7 Legacy Train Operations

3.7.1 Configuration

For DCC trains, there are two parts:

- CV access (with the added complexity of indexed and double-indexed Cvs)
- Programmer (main, programming track) control.

In the special case of legacy equipment, for example DCC locomotives which are configured via CV values, the situation is a little more complex. The train may need to be on a special section of "programmer" track, or only certain values can be changed in certain ways, etc. These legacy devices may have to use a more restricted form of the configuration protocols. For more information, see the page on DCC CV Programming. Non-DCC roster information could be separately stored in the proxy node serving a particular locomotive.

OpenLCB has a protocol for handling configuration of nodes. It should work fine for native OpenLCB rolling stock. Can/should it be adapted to handle the way existing DCC locomotives are configured through Cvs?

For CVs, one idea that we'd have a memory space that maps straight to the decoder CV space.

That would allow decoder-specific CDI (if that was available) to customize what's presented. Like DecoderPro, the user wouldn't have to deal with CV 111, but rather with "Motor multisnarb angle offset" or whatever the user-fiendish manufacturer decides to provide as options.

There are two remaining issues. First, is handling "indexed" CVs. (Ones where you write 12 to CV 51, then 8 to CV52, and then CV54 is the value you want to read/write). QSI is the only extensive user of these, but they're starting to appear in other places too.

There are a couple ways to handle it. First, we could just map it as is: The configuring node would have to explicitly do those reads and writes. But that's a mess, well outside the OpenLCB model, and I've spent way to much time debugging weird failures with that.

A better approach, I think, is to use the large address space. E.g. CV 59 is found at 0x00 00 00 3B, while the one I mentioned above is found at 0x01 00 0C 08. (The 0x01 tells how to decode the address space, e.g. which CV the 2nd byte is written to, which the 3rd byte is written to. We'd have to extend CDI to carry that info, but it's within reason.

The other issue is packed CVs, e.g. mapping parts of one or more CVs to a single "variable". This could be just a single bit in one CV, or something more complicated split across two (like long addresses). I'm not sure how to map those as a general case. For bits, I'd suggest another mapping trick, where some other part of the space is actually bit mapped:

0xFF 00 53 3C

is the middle bits (the 3C mask) of CV 53, and the 0xFF is just an arbitrary key for this. That doesn't work with noncontiguous splits across generic CVs, though, as that takes a lot to configure.

(And don't get me started on CV1/CV29 sequencing; I think we just ignore that entirely at the CDI level & build it into the gateway to DCC)

In the end, these legacy pains can't really be avoided. We have to croft the OpenLCB protocols to do a reasonable job with them, but I don't think we have to go nearly as far into the weird special cases as e.g. DecoderPro does.

3.7.2 Legacy Function Control

Many trains offer independent control of various special effects (FX, sometimes called "functions") such as lighting or sound. The Set Function instruction permits a controller to control these effects directly. The first argument is the address of the FX to control as a 24-bit unsigned integer. This protocol does not define a semantics for FX addresses; that is, there is no particular address that is singled out as representing headlights or the air horn. Instead, the addresses are deliberately abstract, permitting the user to decide how to map addresses to FX for each train.

Additionally, each FX can take a 16-bit value. Current technology only permits binary FX; thus 0x00 should be interpreted as "off" for a binary FX, and any non-zero value as "on". Analog (non-binary) FX should treat the 16-bit value as an unsigned integer.

Rationale: Most digital systems offer multiple FX, each of which is numbered. Although currently most DCC throttles only permit access to 12 or 29 functions, DCC technically permits as many as 32796 different binary FX (see RP-9.2.1: Function Group One Instruction, Function Group Two Instruction, and Feature Expansion Command Instruction especially the Binary State Control Subcommand). For this reason, it seems prudent to go ahead and use a 24-bit value.

Likewise, although current DCC decoders only permit binary FX, some (for example SoundTraxx Tsunami sound decoders) actually permit a kind of analog control of FX by combining multiple DCC Functions. Thus, it seems likewise prudent to permit a wide range of values, and not simply a binary on and off.

One problem that faces the decision to use a single command with a numerical FX addressing system is that any kind of standardized assignment of FX addresses to particular FX is impossible in practice. DCC, for example, makes no such prescription, although by convention function F0 controls direction-sensitive headlights. Beyond this lies only manufacturers' conventions. Thus, any kind of mapping is best handled on a per-train basis, by configuring the mapping between particular FX (e.g. headlights, air horn) to FX addresses for each train.

One way to mitigate this problem is to not make fixed FX address assignments, but to map them directly onto the addressing scheme used by the various control systems, that is, to leave each address in the OpenLCB FX address space uninterpreted. In this way, the default behavior of each address will map directly onto the default behavior of the decoder in the train, giving some degree of predictability to the system, and permitting a digital command station the make reasonable guesses about the possible address-to-FX mappings. Nevertheless, users will often need to be exposed to this implementation detail, which is deeply unfortunate, but necessary given the ultimate flexibility of current train control systems.

Thus, it is recommended that the FX address space be mapped directly to the particular control system's address space; thus DCC F0 becomes FX address 0x000000, F1 becomes 0x000001, etc. The DCC Binary State addresses should be mapped to 0x010000 to 0x017FFF (i.e., to the 15-bit range beginning with 0b1.0000.0000.0000.0000). Other systems should be handled analogously.

RP-9.2.1 defines the following:

- Function Group One: 5 (F0-F4)
- Function Group Two: 8 (F5-F12)
- Binary State Control: 32767 (15bits) (note: different address space!)
- Feature Expansion Command 11110: 8 (F13-F20)
- Feature Expansion Command 11111: 8 (F21-F28)

And in the future, perhaps even more. Possibly a lot more. There are just under 20 bits of address space available in the Feature Expansion Command Instruction for the potential manipulation of Binary State Controls.

3.7.3 Legacy Throttles and Throttle Busses

The most straight-forward way to allow use of existing legacy DCC throttles and systems is to connect the DCC output of their command station to an OpenLCB node. That node then extracts the speed,

function, configuration, etc information from the DCC stream and transmits it either (1) directly to an OpenLCB DCC command station or (2) as Traction Protocol Messages after conversion.

Method (1) works for attaching the input (throttle) and output (DCC train driving) parts of a legacy control system to a new OpenLCB core, which in turn allows OpenLCB control of the layout.

Method (2) works more generally, and will be more valuable once full OpenLCB trains become available.

For certain legacy system throttle busses, e.g. perhaps LocoNet or XPressNet, it may also be possible to attach the throttle bus to an OpenLCB node for translation. That node then acts as an intermediary between Legacy Throttles and the OpenLCB, and can operate in either of the modes described above.

Table of Contents

