Networking Working Group
Internet-Draft
Intended status: Informational
Expires: January 9, 2009

K. Pister, Ed.
Dust Networks
P. Thubert, Ed.
Cisco Systems
S. Dwars
Shell
T. Phinney
July 8, 2008

Industrial Routing Requirements in Low Power and Lossy Networks
draft-ietf-roll-indus-routing-reqs-01

Status of this Memo

Abstract

Wireless, low power field devices enable industrial users to significantly increase the amount of information collected and the number of control points that can be remotely managed. The deployment of these wireless devices will significantly improve the productivity and safety of the plants while increasing the efficiency of the plant workers. For wireless devices to have a significant advantage over wired devices in an industrial environment the wireless network needs to have three qualities: low power, high

reliability, and easy installation and maintenance. The aim of this document is to analyze the requirements for the routing protocol used for low power and lossy networks (L2N) in industrial environments.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Table of Contents

1. Terminology

Actuator: a field device that moves or controls plant equipment.

Closed Loop Control: A process whereby a device controller controls an actuator based on information sensed by one or more field devices.

Downstream: Data direction traveling from the plant application to the field device.

PCD: Process Control Domain. The 'legacy' wired plant Network.

OD: Office Domain. The office Network.

Field Device: physical devices placed in the plant's operating environment (both RF and environmental). Field devices include sensors and actuators as well as network routing devices and L2N access points in the plant.

HART: "Highway Addressable Remote Transducer", a group of specifications for industrial process and control devices administered by the HART Foundation (see [HART]). The latest version for the specifications is HART7 which includes the additions for WirelessHART.

ISA: "International Society of Automation". ISA is an ANSI accredited standards-making society. ISA100 is an ISA committee whose charter includes defining a family of standards for industrial automation. [ISA100.11a] is a working group within ISA100 that is working on a standard for monitoring and non-critical process control applications.

L2N Access Point: The L2N access point is an infrastructure device that connects the low power and lossy network system to a plant's backbone network.

Open Loop Control: A process whereby a plant operator manually manipulates an actuator over the network where the decision is influenced by information sensed by field devices.

Plant Application: The plant application is a computer process running in the plant that communicates with field devices to perform tasks that may include control, monitoring and data gathering.

Upstream: Data direction traveling from the field device to the plant application.

RL2N: Routing in Low power and Lossy Networks.

2. Introduction

Wireless, low-power field devices enable industrial users to significantly increase the amount of information collected and the number of control points that can be remotely managed. The deployment of these wireless devices will significantly improve the productivity and safety of the plants while increasing the efficiency of the plant workers.

Wireless field devices enable expansion of networked points by appreciably reducing cost of installing a device. The cost reductions come from eliminating cabling costs and simplified planning. Cabling also carries an overhead cost associated with planning the installation, determining where the cable has to run, and interfacing with the various organizations required to coordinate its deployment. Doing away with the network and power cables reduces the planning and administrative overhead of installing a device.

For wireless devices to have a significant advantage over wired devices in an industrial environment, the wireless network needs to have three qualities: low power, high reliability, and easy installation and maintenance. The routing protocol used for low power and lossy networks (L2N) is important to fulfilling these goals.

Industrial automation is segmented into two distinct application spaces, known as "process" or "process control" and "discrete manufacturing" or "factory automation". In industrial process control, the product is typically a fluid (oil, gas, chemicals ...).

In factory automation or discrete manufacturing, the products are individual elements (screws, cars, dolls). While there is some overlap of products and systems between these two segments, they are surprisingly separate communities. The specifications targeting industrial process control tend to have more tolerance for network latency than what is needed for factory automation.

Irrespective of this different 'process' and 'discrete' plant nature both plant types will have similar needs for automating the collection of data that used to be collected manually, or was not collected before. Examples are wireless sensors that report the state of a fuse, report the state of a luminary, HVAC status, report vibration levels on pumps, report man-down, and so on.

Other novel application arenas that equally apply to both 'process' and 'discrete' involve mobile sensors that roam in and out of plants, such as active sensor tags on containers or vehicles.

Some if not all of these applications will need to be served by the

same low power and lossy wireless network technology. This may mean several disconnected, autonomous L2N networks connecting to multiple hosts, but sharing the same ether. Interconnecting such networks, if only to supervise channel and priority allocations, or to fully synchronize, or to share path capacity within a set of physical network components may be desired, or may not be desired for practical reasons, such as e.g. cyber security concerns in relation to plant safety and integrity.

All application spaces desire battery operated networks of hundreds of sensors and actuators communicating with L2N access points. In an oil refinery, the total number of devices might exceed one million, but the devices will be clustered into smaller networks that in most cases interconnect and report to an existing plant network infrastructure.

Existing wired sensor networks in this space typically use communication protocols with low data rates, from 1,200 baud (e.g. wired HART) to the one to two hundred Kbps range for most of the others. The existing protocols are often master/slave with command/ response.

2.1. Applications and Traffic Patterns

The industrial market classifies process applications into three broad categories and six classes.

- Safety
  - Class 0: Emergency action - Always a critical function
- Control
  - Class 1: Closed loop regulatory control - Often a critical function
  - Class 2: Closed loop supervisory control - Usually non-critical function
  - Class 3: Open loop control - Operator takes action and controls the actuator (human in the loop)
- Monitoring
  - Class 4: Alerting - Short-term operational effect (for example event-based maintenance)

  - Class 5: Logging and downloading / uploading - No immediate operational consequence (e.g., history collection, sequence-of-events, preventive maintenance)

Safety critical functions affect the basic safety integrity of the plant. These normally dormant functions kick in only when process control systems, or their operators, have failed. By design and by regular interval inspection, they have a well-understood probability of failure on demand in the range of typically once per 10-1000 years.

In-time deliveries of messages becomes more relevant as the class number decreases.

Note that for a control application, the jitter is just as important

as latency and has a potential of destabilizing control algorithms.

Industrial users are interested in deploying wireless networks for the monitoring classes 4 and 5, and in the non-critical portions of classes 3 through 2.

Classes 4 and 5 also include asset monitoring and tracking which include equipment monitoring and are essentially separate from process monitoring. An example of equipment monitoring is the recording of motor vibrations to detect bearing wear. However, similar sensors detecting excessive vibration levels could be used as safeguarding loops that immediately initiate a trip, and thus end up being class 0.

In the near future, most low power and lossy network systems will be for low frequency data collection. Packets containing samples will be generated continuously, and 90% of the market is covered by packet rates of between 1/s and 1/hour, with the average under 1/min. In industrial process, these sensors include temperature, pressure, fluid flow, tank level, and corrosion. Some sensors are bursty, such as vibration monitors that may generate and transmit tens of kilo-bytes (hundreds to thousands of packets) of time-series data at reporting rates of minutes to days.

Almost all of these sensors will have built-in microprocessors that may detect alarm conditions. Time-critical alarm packets are expected to be granted a lower latency than periodic sensor data streams.

Some devices will transmit a log file every day, again with typically tens of Kbytes of data. For these applications there is very little "downstream" traffic coming from the L2N access point and traveling to particular sensors. During diagnostics, however, a technician may

be investigating a fault from a control room and expect to have "low" latency (human tolerable) in a command/response mode.

Low-rate control, often with a "human in the loop" (also referred to as "open loop"), is implemented via communication to a control room because that's where the human in the loop will be. The sensor data makes its way through the L2N access point to the centralized controller where it is processed, the operator sees the information

and takes action, and the control information is then sent out to the actuator node in the network.

In the future, it is envisioned that some open loop processes will be automated (closed loop) and packets will flow over local loops and not involve the L2N access point. These closed loop controls for non-critical applications will be implemented on L2Ns. Non-critical closed loop applications have a latency requirement that can be as low as 100 ms but many control loops are tolerant of latencies above 1 s.

More likely though is that loops will be closed in the field entirely, which in most cases eliminates the need for having wireless links within the control loop. Most control loops have sensors and actuators within such proximity that a wire between them remains the most sensible option from an economic point of view. This 'control in the field' architecture is already common practice with wired field busses. An 'upstream' wireless link would only be used to influence the in-field controller settings, and to occasionally capture diagnostics. Even though the link back to a control room might be a wireless and L2N-ish, this architecture reduces the tight latency and availability requirements for the wireless links.

In fast control, tens of milliseconds of latency is typical. In many of these systems, if a packet does not arrive within the specified interval, the system enters an emergency shutdown state, often with substantial financial repercussions. For a one-second control loop in a system with a mean-time between shutdowns target of 30 years, the latency requirement implies nine 9s of reliability. Given such exposure, given the intrinsic vulnerability of wireless link availability, and given the emergence of control in the field architectures, most users tend to not aim for fast closed loop control with wireless links within that fast loop.

2.2. Network Topology of Industrial Applications

Although network topology is difficult to generalize, the majority of existing applications can be met by networks of 10 to 200 field devices and maximum number of hops from two to twenty. It is assumed that the field devices themselves will provide routing capability for

the network, and additional repeaters/routers will not be required in

most cases.

For most industrial applications, a manager, gateway or backbone router acts as a sink for the wireless sensor network. The vast majority of the traffic is real time publish/subscribe sensor data from the field devices over a L2N towards one or more sinks. Increasingly over time, these sinks will be a part of a backbone but today they are often fragmented and isolated.

The wireless sensor network is a Low Power and Lossy Network of field devices for which two logical roles are defined, the field routers and the non routing devices. It is acceptable and even probable that the repartition of the roles across the field devices change over time to balance the cost of the forwarding operation amongst the nodes.

The backbone is a high-speed infrastructure network that may interconnect multiple WSNs through backbone routers. Infrastructure devices can be connected to the backbone. A gateway / manager that interconnects the backbone to the plant network of the corporate network can be viewed as collapsing the backbone and the infrastructure devices into a single device that operates all the required logical roles. The backbone is likely to become an important function of the industrial network.

Typically, such backbones interconnect to the 'legacy' wired plant infrastructure, the plant network, also known as the 'Process Control Domain', the PCD. These plant automation networks are domain wise segregated from the office network or office domain (OD), which in itself is typically segregated from the Internet.

Sinks for L2N sensor data reside on both the plant network PCD, the business network OD, and on the Internet. Applications close to existing plant automation, such as wired process control and monitoring systems running on fieldbusses, that require high availability and low latencies, and that are managed by 'Control and Automation' departments typically reside on the PCD. Other applications such as automated corrosion monitoring, cathodic protection voltage verification, or machine condition (vibration) monitoring where one sample per week is considered over sampling, would more likely deliver their sensor readings in the office domain. Such applications are 'owned' by e.g. maintenance departments.

Yet other applications will be best served with direct Internet connectivity. Examples include: third-party-maintained luminaries; vendor-managed inventory systems, where a supplier of chemicals needs access to tank level readings at his customer's site; temporary

'Babysitting sensors' deployed for just a few days, perhaps during startup, troubleshooting, or ad-hoc measurement campaigns for R&D purposes. In these cases, the sensor data naturally flows to the Internet, and other domains such as office and plant should be circumvented. This will allow quick deployment without impacting plant safety integrity.

This multiple domain multiple applications connectivity creates a significant challenge. Many different applications will all share the same medium, the ether, within the fence, preferably sharing the same frequency bands, and preferably sharing the same protocols, preferably synchronized to optimize co-existence challenges, yet logically segregated to avoid creation of intolerable short cuts between existing wired domains.

Given this challenge, L2N networks are best to be treated as all sitting on yet another segregated domain, segregated from all other wired domains where conventional security is organized by perimeter. Moving away from the traditional perimeter security mindset means moving towards stronger end-device identity authentication, so that L2N access points can split the various wireless data streams and interconnect back to the appropriate domain pending identity and trust established by the gateways in the authenticity of message originators.

Similar considerations are to be given to how multiple applications may or may not be allowed to share routing devices and their potentially redundant bandwidth within the network. Challenges here are to balance available capacity, required latencies, expected priorities, and last but not least available (battery) energy within the routing devices.

2.2.1. The Physical Topology

There is no specific physical topology for an industrial process control network. One extreme example is a multi-square-kilometer refinery where isolated tanks, some of them with power but most with no backbone connectivity, compose a farm that spans over of the surface of the plant. A few hundred field devices are deployed to ensure the global coverage using a wireless self-forming self-healing mesh network that might be 5 to 10 hops across. Local feedback loops and mobile workers tend to be only one or two hops. The backbone is in the refinery proper, many hops away. Even there, powered infrastructure is also typically several hops away. So hopping to/ from the powered infrastructure will in general be more costly than the direct route.

In the opposite extreme case, the backbone network spans all the

nodes and most nodes are in direct sight of one or more backbone router. Most communication between field devices and infrastructure devices as well as field device to field device occurs across the backbone. From afar, this model resembles the WIFI ESS (Extended Service Set). But from a layer 3 perspective, the issues are the default (backbone) router selection and the routing inside the backbone whereas the radio hop towards the field device is in fact a simple local delivery.

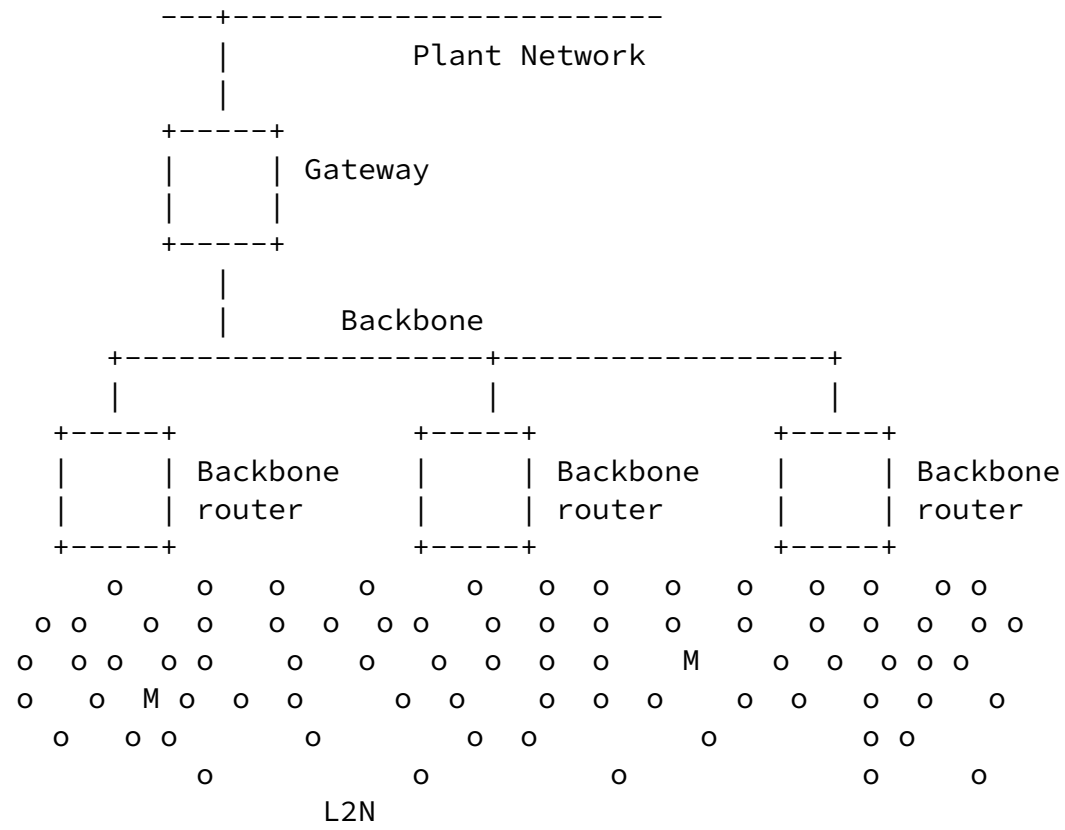

Figure 1: The Physical Topology

2.2.2. Logical Topologies

Most of the traffic over the LLN is publish/subscribe of sensor data from the field device towards the backbone router or gateway that acts as the sink for the WSN. The destination of the sensor data is

an Infrastructure device that sits on the backbone and is reachable via one or more backbone router.

For security, reliability, availability or serviceability reasons, it is often required that the logical topologies are not physically congruent over the radio network, that is they form logical partitions of the LLN. For instance, a routing topology that is set up for control should be isolated from a topology that reports the temperature and the status of the events, if that second topology has

lesser constraints for the security policy. This isolation might be implemented as Virtual LANs and Virtual Routing Tables in shared nodes the backbone, but correspond effectively to physical nodes in the wireless network.

Since publishing the data is the raison d'etre for most of the sensors, it makes sense to build proactively a set of default routes between the sensors and one or more backbone router and maintain those routes at all times. Also, because of the lossy nature of the network, the routing in place should attempt to propose multiple forwarding solutions, building forwarding topologies in the form of Directed Acyclic Graphs oriented towards the sinks.

In contrast with the general requirement of maintaining default routes towards the sinks, the need for field device to field device connectivity is very specific and rare, though the traffic associated might be of foremost importance. Field device to field device routes are often the most critical, optimized and well-maintained routes. A class 0 control loop requires guaranteed delivery and extremely tight response times. Both the respect of criteria in the route computation and the quality of the maintenance of the route are critical for the field devices operation. Typically, a control loop will be using a dedicated direct wire that has very different capabilities, cost and constraints than the wireless medium, with the need to use a wireless path as a back up route only in case of loss of the wired path.

Considering that though each field device to field device route computation has specific constraints in terms of latency and availability it can be expected that the shortest path possible will often be selected and that this path will be routed inside the LLN as opposed to via the backbone. It can also be noted that the lifetimes

of the routes might range from minutes for a mobile workers to tens of years for a command and control closed loop. Finally, time-varying user requirements for latency and bandwidth will change the constraints on the routes, which might either trigger a constrained route recomputation, a reprovisioning of the underlying L2 protocols, or both in that order. For instance, a wireless worker may initiate a bulk transfer to configure or diagnose a field device. A level sensor device may need to perform a calibration and send a bulk file to a plant.

For these reasons, the ROLL routing infrastructure MUST be able to compute and update constrained routes on demand (that is reactively), and it can be expected that this model will become more prevalent for field device to field device connectivity as well as for some field device to Infrastructure devices over time.

3. Service Requirements

The industrial applications fall into four large service categories [ISA100.11a]:

1. Periodic data (aka buffered). Data that is generated periodically and has a well understood data bandwidth requirement, both deterministic and predictable. Timely delivery of such data is often the core function of a wireless sensor network and permanent resources are assigned to ensure that the required bandwidth stays available. Buffered data usually exhibits a short time to live, and the newer reading obsoletes the previous. In some cases, alarms are low priority information that gets repeated over and over. The end-to-end latency of this data is not as important as the regularity with which the data is presented to the plant application.

2. Event data. This category includes alarms and aperiodic data reports with bursty data bandwidth requirements. In certain cases, alarms are critical and require a priority service from the network.

3. Client/Server. Many industrial applications are based on a client/server model and implement a command response protocol. The data bandwidth required is often bursty. The acceptable

round-trip latency for some legacy systems was based on the time to send tens of bytes over a 1200 baud link. Hundreds of milliseconds is typical. This type of request is statistically multiplexed over the L2N and cost-based fair-share best-effort service is usually expected.

4. Bulk transfer. Bulk transfers involve the transmission of blocks of data in multiple packets where temporary resources are assigned to meet a transaction time constraint. Transient resources are assigned for a limited period of time (related to file size and data rate) to meet the bulk transfers service requirements.

For industrial applications Service parameters include but might not be limited to:

- Data bandwidth - the bandwidth might be allocated permanently or for a period of time to a specific flow that usually exhibits well defined properties of burstiness and throughput. Some bandwidth will also be statistically shared between flows in a best effort fashion.

- Latency - the time taken for the data to transit the network from the source to the destination. This may be expressed in terms of a deadline for delivery. Most monitoring latencies will be in seconds to minutes.

- Transmission phase - process applications can be synchronized to wall clock time and require coordinated transmissions. A common coordination frequency is 4 Hz (250 ms).

- Service contract type - revocation priority. L2Ns have limited network resources that can vary with time. This means the system can become fully subscribed or even over subscribed. System policies determine how resources are allocated when resources are over subscribed. The choices are blocking and graceful degradation.

- Transmission priority - the means by which limited resources within field devices are allocated across multiple services. For

transmissions, a device has to select which packet in its queue will be sent at the next transmission opportunity. Packet priority is used as one criterion for selecting the next packet. For reception, a device has to decide how to store a received packet. The field devices are memory constrained and receive buffers may become full. Packet priority is used to select which packets are stored or discarded.

The routing protocol MUST also support different metric types for each link used to compute the path according to some objective function (e.g. minimize latency).

Industrial application data flows between field devices are not necessarily symmetric. In particular, asymmetrical cost and unidirectional routes are common for published data and alerts, which represent the most part of the sensor traffic. The routing protocol MUST be able to set up unidirectional or asymmetrical cost routes that are composed of one or more non congruent paths.

3.1. Configurable Application Requirement

Time-varying user requirements for latency and bandwidth will require changes in the provisioning of the underlying L2 protocols. A technician may initiate a query/response session or bulk transfer to diagnose or configure a field device. A level sensor device may need to perform a calibration and send a bulk file to a plant. The routing protocol MUST route on paths that are changed to appropriately provision the application requirements. The routing protocol MUST support the ability to recompute paths based on underlying link characteristics that may change dynamically.

3.2. Different Routes for Different Flows

Because different services categories have different service requirements, it is often desirable to have different routes for different data flows between the same two endpoints. For example, alarm or periodic data from A to Z may require path diversity with specific latency and reliability. A file transfer between A and Z may not need path diversity. The routing algorithm MUST be able to generate different routes for different flows.

4. Reliability Requirements

There are a variety of different ways to look at reliability in an industrial low power lossy network:

1) Availability of source to sink connectivity when the application needs it, expressed in #fail / #success

2) Availability of source to sink connectivity when the application might need it, expressed in #potential fail / available bandwidth,

3) Probability of failure on demand,

4) Ability, expressed in #failures divided by #successes to get data delivered from source to sink within a capped time,

5) How well a network (serving many applications) achieves end-to-end delivery of packets within a bounded latency

The common theme running through all reliability requirements from a user perspective is that it be end-to-end, usually with a time bound.

The impact of not receiving sensor data due to sporadic network outages can be devastating if this happens unnoticed. However, if sinks that expect periodic sensor data or alarm status updates, fail to get them, then automatically these systems can take appropriate actions that prevent dangerous situations. Depending on the wireless application, appropriate action ranges from initiating a shut down within 100 ms, to using a last known good value for as much as N successive samples, to sending out an operator into the plant to collect monthly data in the conventional way, i.e. some portable sensor, paper and a clipboard.

Another critical aspect for the routing is the capability to ensure maximum disruption time and route maintainance. The maximum disruption time is the time it takes at most for a specific path to

be restored when broken. Route maintainance ensures that a path is monitored to be restored when broken within the maximum disruption time. Maintenance should also ensure that a path continues to provide the service for which it was established for instance in

terms of bandwidth, jitter and latency.

In industrial applications, reliability is usually defined with respect to end-to-end delivery of packets within a bounded latency. Reliability requirements vary over many orders of magnitude. Some non-critical monitoring applications may tolerate a availability of less than 90% with hours of latency. Most industrial standards, such as HART7, have set user reliability expectations at 99.9%. Regulatory requirements are a driver for some industrial applications. Regulatory monitoring requires high data integrity because lost data is assumed to be out of compliance and subject to fines. This can drive up either reliability, or thrustworthiness requirements.

Hop-by-hop path diversity is used to improve latency-bounded reliability. Additionally, bicasting or pluricasting may be used over multiple non congruent / non overlapping paths to increase the likelihood that at least one instance of a critical packet be delivered error free.

Because data from field devices are aggregated and funneled at the L2N access point before they are routed to plant applications, L2N access point redundancy is an important factor in overall availability. A route that connects a field device to a plant application may have multiple paths that go through more than one L2N access point. The routing protocol MUST support multiple L2N access points and load distribution among L2N access points. The routing protocol MUST support multiple L2N access points when L2N access point redundancy is required. Because L2Ns are lossy in nature, multiple paths in a L2N route MUST be supported. The availability of each path in a route can change over time. Hence, it is important to measure the availability on a per-path basis and select a path (or paths) according to the availability requirements.

5. Device-Aware Routing Requirements

Wireless L2N nodes in industrial environments are powered by a variety of sources. Battery operated devices with lifetime requirements of at least five years are the most common. Battery operated devices have a cap on their total energy, and typically can report an estimate of remaining energy, and typically do not have constraints on the short-term average power consumption. Energy scavenging devices are more complex. These systems contain both a

power scavenging device (such as solar, vibration, or temperature difference) and an energy storage device, such as a rechargeable battery or a capacitor. These systems, therefore, have limits on both long-term average power consumption (which cannot exceed the average scavenged power over the same interval) as well as the short-term limits imposed by the energy storage requirements. For solar-powered systems, the energy storage system is generally designed to provide days of power in the absence of sunlight. Many industrial sensors run off of a 4-20 mA current loop, and can scavenge on the order of milliwatts from that source. Vibration monitoring systems are a natural choice for vibration scavenging, which typically only provides tens or hundreds of microwatts. Due to industrial temperature ranges and desired lifetimes, the choices of energy storage devices can be limited, and the resulting stored energy is often comparable to the energy cost of sending or receiving a packet rather than the energy of operating the node for several days. And of course, some nodes will be line-powered.

Example 1: solar panel, lead-acid battery sized for two weeks of rain. In this system, the average power consumption over any two week period must be kept below a threshhold defined by the solar panel. The peak power over minutes or hours could be dramatically higher.

Example 2: 100uA vibration scavenger, 1mF tantalum capacitor. With very limited storage capability, even the short-term average power consumption of this system must be low. If the cost of sending or receiving a packet is 100uC, and a maximum tolerable capacitor voltage droop of 1V is allowed, then the long term average must be less than 1 packet sent or received per second, and no more than 5 packets may be forwarded in any given second.

Field devices have limited resources. Low-power, low-cost devices have limited memory for storing route information. Typical field devices will have a finite number of routes they can support for their embedded sensor/actuator application and for forwarding other devices packets in a mesh network slotted-link.

Users may strongly prefer that the same device have different lifetime requirements in different locations. A sensor monitoring a non-critical parameter in an easily accessed location may have a lifetime requirement that is shorter and tolerate more statistical variation than a mission-critical sensor in a hard-to-reach place that requires a plant shutdown in order to replace.

The routing algorithm MUST support node-constrained routing (e.g. taking into account the existing energy state as a node constraint).

Node constraints include power and memory, as well as constraints

placed on the device by the user, such as battery life.

6. Broadcast/Multicast

Some existing industrial plant applications do not use broadcast or multicast addressing to communicate to field devices. Unicast address support is sufficient for them.

In some other industrial process automation environments, multicast over IP is used to deliver to multiple nodes that may be functionally-similar or not. Example usages are:

1) Delivery of alerts to multiple similar servers in an automation control room. Alerts are multicast to a group address based on the part of the automation process where the alerts arose (e.g., the multicast address "all-nodes-interested-in-alerts-for-process-unit-X"). This is always a restricted-scope multicast, not a broadcast

2) Delivery of common packets to multiple routers over a backbone, where the packets results in each receiving router initiating multicast (sometimes as a full broadcast) within the LLN. This is byproduct of having potentially physically separated backbone routers that can inject messages into different portions of the same larger LLN.

3) Publication of measurement data to more than one subscriber. This feature is useful in some peer to peer control applications. For example, level position may be useful to a controller that operates the flow valve and also to the overfill alarm indicator. Both controller and alarm indicator would receive the same publication sent as a multicast by the level gauge.

It is quite possible that first-generation wireless automation field networks can be adequately useful without either of these capabilities, but in the near future, wireless field devices with communication controllers and protocol stacks will require control and configuration, such as firmware downloading, that may benefit from broadcast or multicast addressing.

The routing protocol SHOULD support broadcast or multicast addressing.

7. Route Establishment Time

During network formation, installers with no networking skill must be

able to determine if their devices are "in the network" with sufficient connectivity to perform their function. Installers will have sufficient skill to provision the devices with a sample rate or activity profile. The routing algorithm MUST find the appropriate route(s) and report success or failure within several minutes, and SHOULD report success or failure within tens of seconds.

Network connectivity in real deployments is always time varying, with time constants from seconds to months. So long as the underlying connectivity has not been compromised, this link churn should not substantially affect network operation. The routing algorithm MUST respond to normal link failure rates with routes that meet the Service requirements (especially latency) throughout the routing response. The routing algorithm SHOULD always be in the process of optimizing the system in response to changing link statistics. The routing algorithm MUST re-optimize the paths when field devices change due to insertion, removal or failure, and this re-optimization MUST not cause latencies greater than the specified constraints (typically seconds to minutes).

8. Mobility

Various economic factors have contributed to a reduction of trained workers in the plant. The industry as a whole appears to be trying to solve this problem with what is called the "wireless worker". Carrying a PDA or something similar, this worker will be able to accomplish more work in less time than the older, better-trained workers that he or she replaces. Whether the premise is valid, the use case is commonly presented: the worker will be wirelessly connected to the plant IT system to download documentation, instructions, etc., and will need to be able to connect "directly" to the sensors and control points in or near the equipment on which he

or she is working. It is possible that this "direct" connection could come via the normal L2Ns data collection network. This connection is likely to require higher bandwidth and lower latency than the normal data collection operation.

Undecided yet is if these PDAs will use the L2N network directly to talk to field sensors, or if they will rather use other wireless connectivity that proxys back into the field, or to anywhere else, the user interfaces typically used for plant historians, asset management systems, and the likes.

The routing protocol SHOULD support the wireless worker with fast network connection times of a few of seconds, and low command and response latencies to the plant behind the L2N access points, to applications, and to field devices. The routing protocol SHOULD also

support the bandwidth allocation for bulk transfers between the field device and the handheld device of the wireless worker. The routing protocol SHOULD support walking speeds for maintaining network connectivity as the handheld device changes position in the wireless network.

Some field devices will be mobile. These devices may be located on moving parts such as rotating components or they may be located on vehicles such as cranes or fork lifts. The routing protocol SHOULD support vehicular speeds of up to 35 kmph.

9. Manageability

The process and control industry is manpower constrained. The aging demographics of plant personnel are causing a looming manpower problem for industry across many markets. The goal for the industrial networks is to have the installation process not require any new skills for the plant personnel. The person would install the wireless sensor or wireless actuator the same way the wired sensor or wired actuator is installed, except the step to connect wire is eliminated.

Most users in fact demand even much further simplified provisioning methods, whereby automatically any new device will connect and report at the L2N access point. This requires availability of open and

untrusted side channels for new joiners, and it requires strong and automated authentication so that networks can automatically accept or reject new joiners. Ideally, for a user, adding new devices should be as easy as dragging and dropping an icon from a pool of authenticated new joiners into a pool for the wired domain that this new sensor should connect to. Under the hood, invisible to the user, auditable security mechanisms should take care of new device authentication, and secret join key distribution. These more sophisticated 'over the air' secure provisioning methods should eliminate the use of traditional configuration tools for setting up devices prior to being ready to securely join a L2N access point.

There will be many new applications where even without any human intervention at the plant, devices that have never been on site before, should be allowed, based on their credentials and crypto capabilities, to connect anyway. Examples are 3rd party road tankers, rail cargo containers with overfill protection sensors, or consumer cars that need to be refueled with hydrogen by robots at future petrol stations.

The routing protocol for L2Ns is expected to be easy to deploy and manage. Because the number of field devices in a network is large,

provisioning the devices manually would not make sense. Therefore, the routing protocol MUST support auto-provisioning of field devices. The protocol also MUST support the distribution of configuration from a centralized management controller if operator-initiated configuration change is allowed.

10. Security

Given that wireless sensor networks in industrial automation operate in systems that have substantial financial and human safety implications, security is of considerable concern. Levels of security violation that are tolerated as a "cost of doing business" in the banking industry are not acceptable when in some cases literally thousands of lives may be at risk.

Security is easily confused with guarantee for availability. When discussing wireless security, it's important to distinguish clearly between the risks of temporary losing connectivity, say due to a

thunderstorm, and the risks associated with knowledgeable adversaries attacking a wireless system. The conscious attacks need to be split between 1) attacks on the actual application served be the wireless devices and 2) attacks that exploit the presence of a wireless access point that MAY provide connectivity onto legacy wired plant networks, so attacks that have little to do with the wireless devices in the L2Ns. The second type of attack, access points that might be wireless backdoors that may allow an attacker outside the fence to access typically non-secured process control and/or office networks, are typically the ones that do create exposures where lives are at risk. This implies that the L2N access point on its own must possess functionality that guarantees domain segregation, and thus prohibits many types of traffic further upstream.

Current generation industrial wireless device manufactures are specifying security at the MAC layer and the transport layer. A shared key is used to authenticate messages at the MAC layer. At the transport layer, commands are encrypted with unique randomly-generated end-to-end Session keys. HART7 and ISA100.11a are examples of security systems for industrial wireless networks.

Although such symmetric key encryption and authentication mechanisms at MAC and transport layers may protect reasonably well during the lifecycle, the initial network boot (provisioning) step in many cases requires more sophisticated steps to securely land the initial secret keys in field devices. It is vital that also during these steps, the ease of deployment and the freedom of mixing and matching products from different suppliers doesn't complicate life for those that deploy and commission. Given average skill levels in the field, and

given serious resource constraints in the market, investing a little bit more in sensor node hardware and software so that new devices automatically can be deemed trustworthy, and thus automatically join the domains that they should join, with just one drag and drop action for those in charge of deploying, will yield in faster adoption and proliferation of the L2N technology.

Industrial plants may not maintain the same level of physical security for field devices that is associated with traditional network sites such as locked IT centers. In industrial plants it must be assumed that the field devices have marginal physical security and the security system needs to have limited trust in them.

The routing protocol SHOULD place limited trust in the field devices deployed in the plant network.

The routing protocol SHOULD compartmentalize the trust placed in field devices so that a compromised field device does not destroy the security of the whole network. The routing MUST be configured and managed using secure messages and protocols that prevent outsider attacks and limit insider attacks from field devices installed in insecure locations in the plant.

Wireless typically forces us to abandon classical 'by perimeter' thinking when trying to secure network domains. Wireless nodes in L2N networks should thus be regarded as little islands with trusted kernels, situated in an ocean of untrusted connectivity, an ocean that might be full of pirate ships. Consequently, confidence in node identity and ability to challenge authenticity of source node credentials gets more relevant. Cryptographic boundaries inside devices that clearly demark the border between trusted and untrusted areas need to be drawn. Protection against compromise of the cryptographic boundaries inside the hardware of devices is outside of the scope this document. Standards exist that address those vulnerabilities.

11. IANA Considerations

This document includes no request to IANA.

12. Acknowledgements

Many thanks to Rick Enns, Alexander Chernoguzov and Chol Su Kang for their contributions.

13. References

13.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

13.2. Informative References

[I-D.culler-rl2n-routing-reqs]
Vasseur, J. and D. Cullerot, "Routing Requirements for Low Power And Lossy Networks", draft-culler-rl2n-routing-reqs-01 (work in progress), July 2007.

13.3. External Informative References

[HART] www.hartcomm.org, "Highway Addressable Remote Transducer", a group of specifications for industrial process and control devices administered by the HART Foundation".

[ISA100.11a]
ISA, "ISA100, Wireless Systems for Automation", May 2008, < http://www.isa.org/Community/ SP100WirelessSystemsforAutomation>.

Authors' Addresses

Kris Pister (editor)
Dust Networks
30695 Huntwood Ave.
Hayward, 94544
USA

Email: kpister@dustnetworks.com

Pascal Thubert (editor)
Cisco Systems
Village d'Entreprises Green Side
400, Avenue de Roumanille
Batiment T3
Biot - Sophia Antipolis 06410
FRANCE

Phone: +33 497 23 26 34
Email: pthubert@cisco.com

Sicco Dwars
Shell Global Solutions International B.V.
Sir Winston Churchilllaan 299
Rijswijk 2288 DC
Netherlands

Phone: +31 70 447 2660
Email: sicco.dwars@shell.com

Tom Phinney
5012 W. Torrey Pines Circle
Glendale, AZ 85308-3221
USA

Phone: +1 602 938 3163
Email: tom.phinney@cox.net

Full Copyright Statement

Intellectual Property