5 October 2012

From: Rex A Buddenberg
2151 Trapani Circle
Monterey, Ca 93940

Subject: Comments on Notice of Inquiry
Docket No: 120928505-2505-01
RIN: 0660-XC002
Development of the Nationwide Interoperable Public Safety Broadband Network

I wish to comment on these areas in your Notice of Inquiry:

- Architecture
- Availability and survivability
- Quality of Service Control
- Anchor tenant, or public-private partnership
- Emergency services applications
  - General requirements
  - Specific applications.

Architecture.

"… comments on the FNN conceptual network design model presented at the FirstNet Board meeting. …. single, nationwide network architecture that evolves with technological advancements."

The term 'architecture' is a much-maligned one with many, rather slippery definitions. Since you use the term 'interoperability' in the NOI title, we should use 'architecture' to mean the modularization model that enables interoperability.

In internet communications systems, there are clear modular boundaries between the

- terrestrial wide area network,
- the radio wide area network and
- local area networks.

These boundaries are implemented today as routers. In Reference Model terminology, routers are layer 3 devices and provide agnosticity regarding network segments: a radio-WAN segment need know nothing about the terrestrial-WAN 'on the other side of the router'. This modularization model has resulted in 50 years of development of the internet as segments can be added and upgraded without cascading changes through the rest of the communications infrastructure.

Recommendation. FirstNet should structure its acquisitions and service agreements to foster the modularization model. For example, the contracting for both the terrestrial-WAN and radio-WAN should be from router-to-router rather than detailing myriad details inside each modular boundary. This means that the entire radio-WAN – both base stations and subscriber stations – should be under one acquisition authority.

The internet is also modularly constructed so that we have one infrastructure and many applications. Each application is identified by a port number in IP datagrams. As the set of applications evolves, we need this modularity boundary so that the evolution of applications is independent of the evolution of the infrastructure.

Availability and survivability.

"...ideas on how it can deploy a reliable, ubiquitous, redundant, and interoperable broadband network for public safety users."

The most important factor driving both the capitalization and operating budgets is availability[1]. And high availability is the most important difference between commercial internet and what emergency services needs.

The first step is to determine the required availability. Unquantified 'high availability' is not measurable and therefore not usable. Availability is defined as:

Availability = up time / total time

and that can be expressed more usefully as

Availability = (total time – down time) / total time.

What is the tolerable down time? The answer should be tied to street-level needs, not technology willingness.

Once you have a quantified requirement, you can express that to the service providers and enforce in contract specifications and service level agreements. Look for responses that include the three principles of high availability engineering:

1. Elimination of single points of failure[2].
2. Reliable crossover between primary and backup comms.
3. Prompt notification of failures, aka fault management.

The principles are applicable to any system. But if we apply them to communications systems, principle #1 generally boils down to alternate comms routes and backup power. Principle #2 is met by the stateless design of Internet Protocol – routers do this all day every day. Principle #3 is necessary because if the first two are faithfully executed, the user may never see an outage … but somebody needs to in order to direct maintenance (this principle will reappear in applications below).

1 Redundancy is one of several components of availability; the NOI confuses 'what' and 'how'.
2 Redundancy falls in here.

QoS control.

“… public safety’s requirements for priority, quality of service, and preemption ...”

The default behavior of internet technologies is to optimize bandwidth efficiency and to trade off other qos values such as latency and jitter to get the bandwidth efficiency. Since bandwith will always be a constraint in this environment, one should be chary about changing the defaults.

Further, since all 'broadband' technologies are packet switching ones (which are inherently non-blocking), there is nothing to pre-empt. So 'requirements for … preemption' is a statement meaningless to the infrastructure. It may have meaning for applications and human factors.

Let's back up a step and sort out the terminology and get the logic into proper order:

The most reliable congestion-control method in the internet for the past forty years has been overprovisioning. If there is no congestion in the internetwork, then no QoS scheme can possibly improve qos. In both the terrestrial-WAN and LAN components of a FirstNet infrastructure, this still remains the most reliable methodology. And overprovisioning is almost always a side effect of getting the availability engineering right. So QoS Control should never be discussed before availability – don't argue over a problem you may never have.

Overprovisioning is only a partial solution in the radio-WAN. The larger number of smaller-sized cell footprints allows for more aggregate capacity – good idea. But the capacity per segment is physics-limited by the amount of spectrum available. We will never get more that a few M bits/sec in a single radio-WAN segment (as compared to four orders of magnitude more capacity – 10s of G bits/sec – in both LANs and terrestrial-WANs). The lesson here is that radio is hard and it always will be; the more leverage we can get out of the terrestrial-WAN, the better.

Characteristics of radio-WAN protocols. LTE is a time division multiple access protocol. What that means is that every station (both base station and subscriber stations) in the network segment will get an opportunity to transmit something each frame. This confers a certain amount of deterministic behavior, which may, in fact, meet most requirements. If this is not adequate (I don't think we can tell for sure without some experience), stations need to 1) prioritize their own outgoing traffic so that the most urgent traffic goes first when transmit time arrives[3] and 2) request more bandwidth in subsequent frames.

The means for prioritizing its own traffic is found in Differential Services; applications need to set the Differential Services Code Point (the default setting gets you everyday internet behavior – routers work on a single queue, first-in-first-out). Enabling differential services gives urgent traffic 'head of line' privileges. The internet infrastructure largely does this today, but few applications are capable of setting the DSCP. While Differential Services works across the entire internet, it will be largely irrelevant outside of the radio-WAN.

There are several MAC-message means within LTE for requesting more capacity.

3 Treated below in the Applications section

Anchor tenant.

“...existing radio access network and core network infrastructure installed by commercial mobile operators”

I believe that the business model expressed by the NOI and slide deck makes sense.

There remains one question within and it's pertinent to the architecture discussion above. In the juvenile days of almost any protocol, we have found that multi-vendor interoperability within a network segment is difficult. Ethernet (IEEE 802.3), DOCSIS and WiFi (IEEE 802.11) are all good examples: for the first half dozen years that products were on the market, no one dared mix brands. Today, of course, these non-interoperabilities are past. I expect the same to happen in LTE. If you draw the modularity boundary at the routers, then the layer 2 interoperability problems are all in one place – the contractor's hands.

Applications.

"...framework for developing applications for public safety use. Commenters should: (1) provide suggestions for applications that would benefit public safety users; (2) address what interface requirements and other information innovators need in order to develop applications in an open environment; (3) address what specific security requirements public safety needs in its applications; (4) provide ideas as to what framework or organizational factors would allow for the development of the greatest number of quality applications; (5) provide specific suggestions for FirstNet's applications certification requirements; (6) discuss possible delivery methods (e.g., app store models) under the FNN conceptual architecture model presented at the FirstNet Board meeting or based on any alternative network design models that commenters propose; and (7) provide comment on any other issues that FirstNet should consider in facilitating the development of public safety applications."

and

"... voice services (cellular telephony and push-to-talk (PTT)) both within the FirstNet network as well as to/from other commercial networks..."[4]

General requirements. There are hundreds of applications that run over the internet today and an attempt to list all that public services would need or want would be an exercise in futility. But we can make progress into the topic this way:

There should be three categories of applications that FirstNet should recognize:

- Blue. Those applications that are supported by FirstNet and 'certified' both for interoperability and security.
- Green. Applications that are not supported by FirstNet but are acceptable. They do no harm.
- Red. Applications that should not be run over emergency services infrastructure. Applications with faulty security are examples that fall in this category.

Criteria for 'Blue' category. Whatever the specific application, there are a few must-haves that are not common in civilian applications that are important to emergency services:

- Interoperability. To write meaningful interoperability criteria, 'interoperability' needs to be parsed into communications interoperability, data interoperability, process interoperability, procedural and doctrinal interoperability. By limiting ourselves to applications that 'run over the internet' we achieve the first and most important communications interoperability requirement. Objective criteria for the remainder can easily be written and tested against – data interoperability is treated below.
- Security. There are no cases where authenticity of the data is not a requirement. Authenticity is a universal, ubiquitous requirement. There are many cases where confidentiality is also a requirement. These security requirements are always end-to-end which means that they should be independent of the infrastructure and implemented in applications. The issue of end-to-end is treated further below under the 'voice' heading. No applications should be supported by FirstNet unless they

4 Voice is an application and therefore should be classed here, not separately.

provide end-to-end security.

- Multicast[5]. Emergency services data needs to go more than one place. This is true most of the time and once you incorporate high availability requirements it almost becomes an 'always' case. This is a widespread characteristic regardless of the specific application. Further, true multicast is a bandwidth economy measure which we need in a capacity-constrained radio-WAN segment context. The internet infrastructure supports multicast IP and has for a decade, but few applications support it.
- Differential Services Code Point. Most routers (a notch above the cheapest home routers) know how to handle differential services – it is a configuration item in the routers' initialization (and routinely turned off by default). But routers can only give preferential QoS treatment to packets that have the DSCP set. Setting the DSCP is a task for end systems/applications. I know of no applications that either set or allow the user to set the DSCP. This is the one missing component of a preferential QoS structure in the commercial internet.
- Manageability. Recall from the availability discussion above that fault detection is the third principle of high availability engineering. This fault detection requirement applies to end systems and their applications just as much as anything else in the infrastructure. Configuration management is an important, and closely related, need – the use case of an officer down and ability to remotely turn on his microphone illustrates. Since many varied things from applications, to routers, to radios, to air conditioners need to be fault-managed, a vendor-specific solution will not suffice; multi-vendor interoperability is required. The means to this end is to incorporate a management agent that conforms to the Simple Network Management Protocol standards in each application and end system.

Delivery mechanisms. Open source operating systems distributions (including many applications) have been in operation for years, including secure (signed for authenticity) downloads. There is little difference in public safety requirements in this area. The 'app store' you mention in #6 is a commercialized variant. These download mechanisms work for both free software (e.g. open source) and commercial.

Common Alerting Protocol. CAP is not an application, but rather a data dictionary or data schema. It is designed to unify and replace various 'stovepipe' alerting systems (nuclear attack, heavy weather, etc) – several applications using CAP have been implemented. But CAP has much wider potential than its current uses – it could be used in the library of public warning messages for an Incident Commander and applications written for portable devices (e.g. smartphones) could display CAP messages in visual form. Another example would be to use the CAP schema as the data dictionary for a set of Blue Force Track (see below) interoperable implementations.

Unlike the other suggested applications here, CAP has adequate security features. Because CAP is XML-based (eXtensible Markup Language) it can use the XML-sign and XML-crypt primitives to provide authenticity and confidentiality respectively.

The next step of interoperability beyond communications interoperability is data interoperability and CAP provides the foundation for that.

5 In LMR, the analog is 'group' modes; I'm using internet terminology here.

Public Key Infrastructure. A PKI is necessary, and for true interoperability this must be a national level entity. Most of the security implementations in the applications noted below and in the plausible extensions to this list will use public key cryptography to attain both authenticity and confidentiality. Further, key pairs must be installed in management agents in order to securely manage the infrastructure so limiting the PKI to just people is not appropriate.

Specific applications. While the list of applications is open-ended, there are a few that are obviously needed. They are also applications where cross-jurisdiction interoperability is important, so the number of non-interoperable variants needs to be curbed.

Voice is undoubtedly at the top of the list of applications. And voice is a good example to exercise the criteria outlined above. A set of voice conversants needs authenticity – they need to know who they are speaking with and that the conversation is not spoofed with bogus speakers. Some voice conversations require confidentiality as well.

Some purportedly 'secure' implementations may indeed include signatures and encryption but do not provide them end-to-end. For example, Voice over IP involves the use of a SIP (Session Initiation Protocol) to set up voice conversations. If the implementation removes the security protection at the SIP server for any reason, then the server becomes a security vulnerability point (aka man-in-the-middle).

Legacy accommodation of LMR voice (with or without P25) is also likely to be required in many areas. There are two ways to accomplish:

- tunnel. It is feasible to simply use the internet infrastructure as an 'LMR tunnel'. This can be secured end-to-end, but at the price of loss of interoperability – an LMR conversant cannot reach someone outside the tunnel because the encryption is specific to LMR.
- Layer 7 gateway. Layer 7 gateways translate LMR voice into Voice over IP. This gets you the interoperability required – you can reach from an LMR handset any VOIP application anywhere on the internet. (This is how the telephone system accommodates POTS phone calls on its otherwise internet infrastructure. Only the local loop is non-internet). The security drawback is that in order for the translation to occur, the security measures specific to either VOIP or LMR (/P25) must be removed. Thus the Layer 7 gateway itself is a vulnerability point; the implementation must be a trusted one.

Most VOIP implementations today are point-to-point, not multicast. And few have management agents in them. In order to attain 'blue' status the implementations would need to be version-matured to gain these capabilities.

Blue force track. There are a variety of names for blue force tracking systems – asset tracking, position notification, etc. In the commercial maritime world, such are called AIS for Automatic Identification System[6]. But the essence of each is that a mobile unit reports its position to one or more locations (yes, multicast appears here too) where that information is displayed on a geographical plot (in the military often known as common tactical picture). FirstNet should support one or more (if >1, interoperable) blue force track implementations.

End-to-end security is again a requirement – both authenticity and confidentiality.

This is one application where your #6 query about delivery deserves special comment. There are several blue force track display implementations available in the military; one of the best supported and most mature is Global Command and Control System (GCCS) used by US Navy and US Coast Guard. In the interest of avoiding wheel-reinvention, suggest an adaptation of GCCS for emergency services use (the software itself is unclassified, only the

6 See http://www.marinetraffic.com/ais/ for a real world demonstration.

data is sensitive).

The DoD experience is that blue force track applications were so popular that at one time there were over sixty implementations. None, of course, were interoperable; this is a scenario FirstNet should not repeat. The first category of interoperability – communications interoperability – is dealt with by the architectural treatment above. The second category is a common data dictionary – we can live with multiple implementations, and they can be interoperable, if they share the same data dictionary (aka data schema).

Network management console. The need for network operations management has been discussed above, with the recommendation that all equipment and applications within the 'mission critical' perimeter have management agents for fault management reasons. To complete the picture, a management console where this data is collected, sorted, triaged, displayed, is needed.

Security, especially authenticity, of the management data is just as important as security of all other data in the system. Imagine the mayhem that someone could cause by dropping unauthorized 'set' messages on various equipments in the infrastructure. (This what version 3 of SNMP is designed to solve).

There are several implementations of management consoles including at least two that are open source (Nagios and OpenSNMP). Since there is a training curve for network operations personnel here, recommend a small number of consoles supported by FirstNet.

Mated with an SNMP console should be a trouble-ticket implementation (often referred to as 'issue tracking system'). The interoperability foundation here is the data dictionary; multiple possibilities exist, including several open source implementations.

Interoperability between an agency's management system and the service provider's management system is important. A service provider may not look kindly at a public service agency manager dropping 'get' messages on the provider's equipment, but that information may be very important to triaging and isolating problems so visibility into the provider's fault-management database, short of queries to equipment, may be an appropriate solution.

Instrumented ambulance. The instrumented ambulance application is a systems affair designed to telemeter patient data to a console at the hospital emergency room. Topology:

- the terrestrial-WAN infrastructure is identical to existing vision
- the radio-WAN infrastructure is identical except that the subscriber station does not terminate in a handset, but is connected to a router (similar to your cable TV internet connection).
- the router is installed in the ambulance.
- on the LAN side of the router, the medical diagnostic devices, a voice handset for EMT use, etc. are all attached. These devices only need a LAN interface (thereby controlling costs).

Requirements comments:

- data security is an obvious requirement – both authenticity and confidentiality. And the requirement is end-to-end.
- presumably the data would be of value to the dispatcher too, so multicast is an inherent requirement.

Instrumented fire engine. This application is similar to the instrumented ambulance above:

- the radio-WAN reaches to a router on the fire engine rather than an end system directly.
- On the LAN side of the router, at least two wireless LANs should be supported:
  - one is a closed LAN (similar to WiFi with the privacy measures enabled) to be used only by fire department personnel. This LAN can be optimized for building-penetration characteristics specific to firefighting needs.
  - The second wireless LAN should be conventional WiFi. It should certainly have an on/off switch (a configuration management item) and it should be configured as an 'anyone use' public hotspot. The reason for this is to provide instant public communications to, for example, allow distraught mothers to track down missing kids.
  - A third, private wired LAN (ethernet will do fine) may be useful to attach instrumentation such as a radionav receiver (see blue force track above) and fire pump controls within the fire engine itself.

In both these cases a local jurisdiction will not support enough vehicles to make an application economical. The economies of scale represented by FirstNet should help realize useful applications rather than one-off demonstrations.

Thank you for the opportunity to comment.