SANDIA REPORT
SAND2003-8803
Unlimited Release
Printed December 2003

Wireless Sensor Systems for Sense/Decide/Act/Communicate

N. Berry, J. Davis, T. Ko, R. Kyker, R. Pate, R. Stinnett, J. Baker, A. Cushner, C. Van Dyke, B. Kyckelhahn, D. Stark

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94-AL85000

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865)576-8401
Facsimile: (865)576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800)553-6847
Facsimile: (703)605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2003-8803
Unlimited Release
Printed December 2003

Wireless Sensor Systems for Sense/Decide/Act/Communicate

Research team: Nina Berry, Jesse Davis, Teresa Ko, Ron Kyker, Ron Pate, Regan Stinnett

Student team: James Baker, Adam Cushner, Colin Van Dyke, Brian Kyckelhahn, Doug Stark

Sandia National Laboratories
MS9915, P.O. Box 969
Livermore, CA 94551-0969

Abstract

After 9/11, the United States (U.S.) was suddenly pushed into challenging situations they could no longer ignore as simple spectators. The War on Terrorism (WoT) was suddenly ignited and no one knows when this war will end. While the government is exploring many existing and potential technologies, the area of wireless sensor networks (WSN) has emerged as a foundation for establish future national security. Unlike other technologies, WSN could provide virtual presence capabilities needed for precision awareness and response in military, intelligence, and homeland security applications. The Advance Concept Group (ACG) vision of Sense/Decide/Act/Communicate (SDAC) sensor system is an instantiation of the WSN concept that takes a "systems of systems" view. Each sensing nodes will exhibit the ability to: Sense the environment around them, Decide as a collective what the situation of their environment is, Act in an intelligent and coordinated manner in response to this situational determination, and Communicate their actions amongst each other and to a human command. This LDRD report provides a review of the research and development done to bring the SDAC vision closer to reality.

This page intentionally left blank

Contents

Figures

Tables

1. Introduction

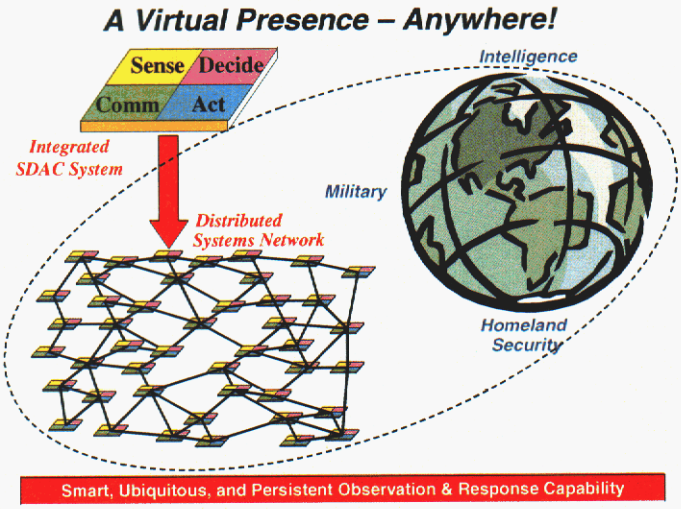
After 9/11, the observance and reactions of the United States (U.S.) to terrorist activities around the world suddenly changed. Like the fatal attack on Pearl Harbor in World War II, the U.S. was suddenly hurled into challenging situations they could no longer ignore as simple spectators. While the U.S. was no stranger to terrorism on U.S. interest abroad, the danger always seemed remote to the majority of the U.S. population. However as the event of 9/11 unfolded on televisions around the world everyone was left speechless and the response by the U.S. has been quick and continuous. The War on Terrorism (WoT) was suddenly ignited and no one knows when this war will end. While the government is exploring many existing and potential technologies, the area of wireless sensor networks (WSN) has emerged as a foundation for establish future national security. Unlike other technologies, WSN could provide virtual presence capabilities needed for precision awareness and response in military, intelligence, and homeland security applications. The Sandia National Laboratories Advance Concept Group (ACG) vision of Sense/Decide/Act/Communicate (SDAC) sensor system is an instantiation of the WSN concept that takes a "systems of systems" view to the creation, distribution, and functional usage of SDAC system in the WoT.

Each sensing nodes will exhibit the ability to: Sense the environment around them, Decide as a collective what the situation of their environment is, Act in an intelligent and coordinated manner in response to this situational determination, and Communicate their actions amongst each other and to a human command. Beyond the capabilities associated with distributed sensor networks in general, SDACs will incorporate distributed intelligence to not only collect data, but make sense out of it as well. As Gerold Yonas of the Sandia Advanced Concepts Group has pointed out, knowledge is not the same thing as data, and it is really the knowledge derived from the collected data that is of power and importance. With the incorporation of this knowledge construction into the system itself, an SDAC network will be aware of its surroundings, and will thus be able to adapt its behavior to dynamic environments in order to accomplish its missions. [1]

1.1 Sensor systems for the War on Terrorism (WoT)

While WSN provides the foundations of the SDAC vision, the conceptualization extends beyond a collection of wireless sensor nodes. The SDAC vision would create systems of sensors capable of detecting, locating, characterizing, and discriminating specific: actions, people, and other entities. The SDAC sensor systems would also be characterized by their ability to be: rapidly deployable, adaptive, autonomous, multi-modal, and globally integrated. The SDAC vision is innately a system level view. This view begins with the lowest components of the system – an individual sensor node and spirals outward encompassing all other sensor nodes creating a single SDAC sensor system (call it A). The vision extends beyond the single SDAC sensor system – A as it connects other SDAC sensor systems together creating an integrated distributed SDAC system, pictorially illustrated in Virtual Presence – Anywhere.

The SDAC "system of systems" vision SDAC LDRD could provide the WoT with collective intelligence, ability to locate terrorist targets, characterize, and report threat conditions and/or events, and to support or provide interdiction and protective response capabilities. The extended visions of the integrated SDAC sensor system are networked arrays of heterogeneous fixed and mobile sensor and potentially human responder (personnel with sensor devices embedded on their person).



The SDAC vision will bring together several pieces to collectively address expectations that are needed to combat the WoT. Among these are high-level concerns dealing with sensor system development, including: functional requirements, system integration, cost, reliability, security and authentication, and node size. Other concerns include deployment issues and human interfaces, which are not directly address in this LDRD. From the enabling side, issues are directed at actual technologies that address some of the high-level concerns and assist in producing overall SDAC systems, these include: micro sensors – physical, imaging, chem./bio, micro power or energy mining, signal processing, networking, collective intelligence, communications, situation awareness, command and control, data analysis, and interpretation. [2]

To address the problems associated with WSN in general requires a multi-disciplinary approach to achieve ideal design and development approaches. The SDAC LDRD brings together a team of software, hardware, and system engineers to produce an architectural tradeoff study and demo system that is geared toward making come of the conceptual views of SDAC into reality. The Embedded Reasoning Institute (ERI) at Sandia National Laboratories, California, has conducted the research and development of the SDAC LDRD demo system and architectural analysis. The ERI is a multi-disciplinary research initiative in the area of smart wireless sensing technology supported jointly by 8200 and 8900. The research and internship team collaborated jointly to provide the conceptual demonstration system and sensor tradeoff analysis.

1.2 LDRD objectives and approach

This LDRD addressed four inter-related objectives, which were applied to the requirements and concepts associated with the four mission areas (a) Military Operations in Urban Terrain, (b) Mobile Force Protection and Fixed Site Physical Security, (c) Intelligence Community Missions, and (d) Safe and Secure Borders. The objectives are listed below:

(1) Conceptually apply the SDAC platform to the four mission areas.

(2) Investigate current hardware and software architectures to establish a best fit for the SDAC requirements and the four mission areas. Propose a next generation SDAC sensor system.

(3) Determine metrics used to evaluate the appropriateness of a given sensor architecture for an application or mission space.

(4) Develop a conceptual demonstration of a SDAC wireless sensor node.

We approached these four objectives with an exploratory process that began with the creation of a set of basic requirements from the WoT mission space. In parallel we began an investigation of existing systems to map their capabilities and expose the differences between the systems. These system differences and WoT requirements were the initial starting points for research directed at establishing a methodology and tradeoff considerations for mapping sensor systems capabilities to application requirements. During this same time, the team flushed out conceptual demonstration in the MOUT domain.

1.3 Organization of LDRD report

This LDRD report has been developed as eight individually encapsulated chapters. Each chapter covers a specific related area that answers the four objectives stated in Section 1.2. Chapter 2 provides a brief overview of technologies related to general sensor systems and or nodes. This chapter also provides a sensor system capabilities table that illustrates known facts about current sensor systems. Chapter 3 provides an overview of SDAC systems in MOUT and border protection mission areas, with an overview of each area, a discussion of the challenges associated with each domain, and a table of potential applications where SDAC systems would improve the areas performance. The exploratory concepts behind matching applications to sensor system requirements are the incremental theme behind Chapter4. This chapter looks at sets of application requirements and attempts to match them with current and future sensor technology capabilities. Combining the results of the prior chapters is correlated in Chapter 5 as a proposed next generation SDAC architecture. This chapter is a summary of a prior SAND report, which details the proposed low power modular SDAC architecture in complete details. One critically area for wireless devices is the lack of good security for these devices. Chapter 5.5 discusses existing vulnerabilities for wireless devices with a specific emphasis on sensor networks. The chapter also includes an exploratory discussion of potential issues with the proposed SDAC architecture. The conceptual demonstration is detailed in Chapter 6, with an overview of the demo and the important concepts being shown as part of this demonstration. The final Chapter 7 concludes the report with results and future projects being explored as spin-offs of the original SDAC LDRD.

1.4 Chapter 1 References

[1] Pate, R., and Stinnett, R., SDAC 8-26-2003 presentation.

[2] Davis, J., Kyker, R., Berry, N., "A System Level Hardware Architecture for a Distributed Sensor Network Node", SAND2003-8209.

2 Wireless Sensor Network Background Material

This chapter covers two important issues of ad-hoc networking and existing systems evaluations as they relate to the review research covered for this LDRD.

2.1.1 A Protocol Guide for ad-hoc networks

The set of constraints that a mobile wireless device is typically under differs greatly from the desktop and server PC paradigm. A MANET (Mobile Ad Hoc Network) device first and foremost may be severely limited in power usage, needing to operate on a small fixed-energy source for a long period of time. Each device possesses a radio (half or possibly full duplex), processing capabilities, and either application specific hardware, such as sensors, or human interface hardware. Devices may be positionally static once placed, or they may be extremely mobile (though usually not self-propelled). MANETs are much more bandwidth constrained than fixed line networks; this makes the challenge of keeping the amount of network control overhead low in a mobile environment quite significant. Some fundamental characteristics are desired from all networks in varying degrees:

- High throughput – the ability to transmit large amounts of data per time
- Low latency – the ability to quickly transmit data
- Reliability – durability in hostile environments
- Security – resistance to human interception or disruption efforts
- Convenience – low complexity and easy implementation interoperability

The rest of the paper will be organized as follows. Section II will cover the PHY layer, discussing radio transmission methods. Section III will cover the DLC layer, explaining various access methods for sharing the wireless medium. Section IV will cover the NET layer, surveying a wide range of routing protocols. Various characteristics are more important for each layer, and thus the methods in each protocol layer are contrasted according to their appropriate differences and theories. Section V will conclude by addressing layer and protocol concerns and interactions.

2.1.1.0 The Physical Layer

Electromagnetic emission may carry or represent data in a variety of different schemes. The radio frequency bandwidth ranging from hundreds of megahertz to several gigahertz is most effective for short to medium range distances and is used in almost all cases. The hierarchy of PHY methods is an extension of the concept of modulation, which is shown in Figure 1.

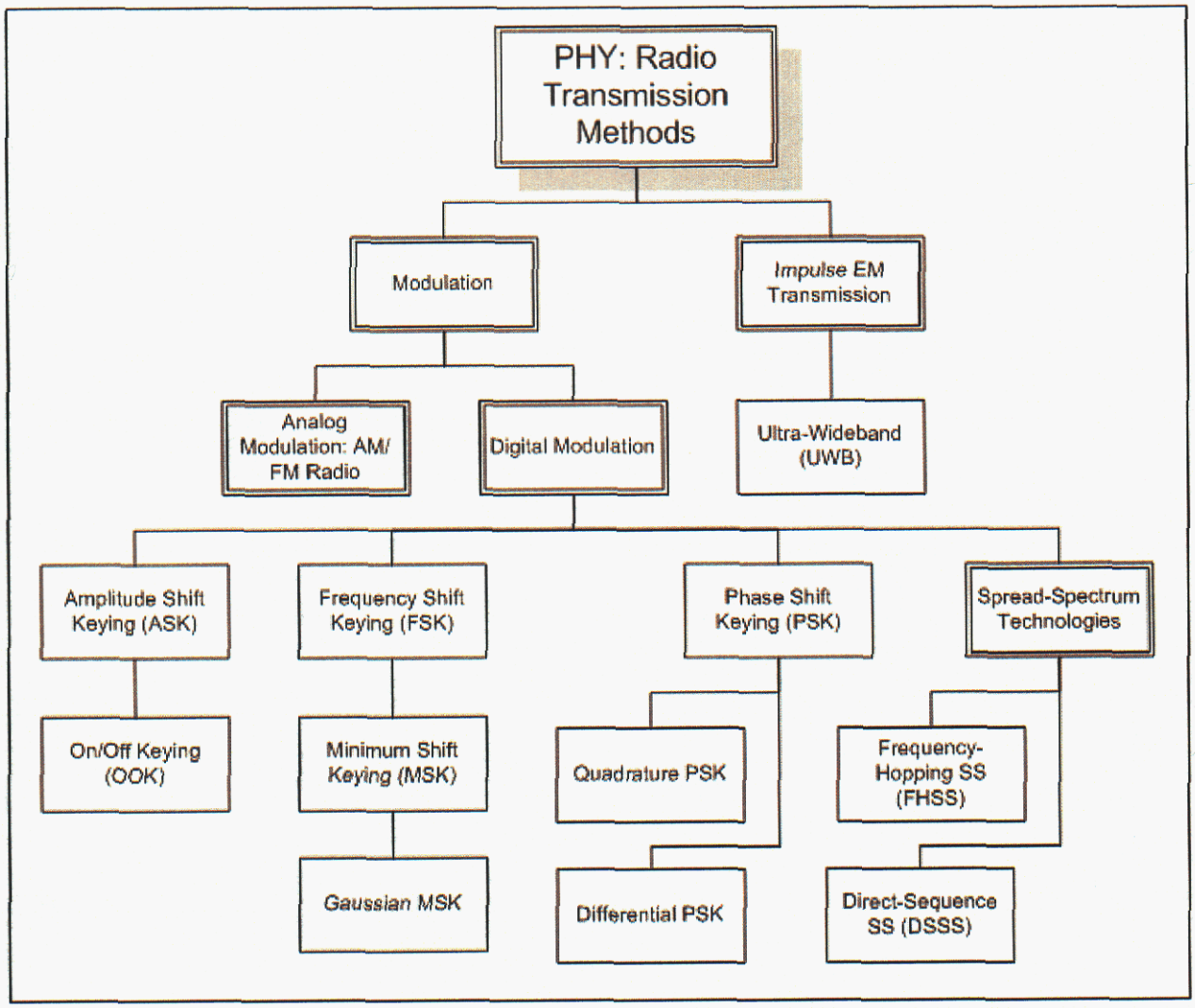


Figure 1: Hierarchical view of physical layer

Rated PHY characteristics are:

- Power – efficiency, transmit distance, power per bit
- Bandwidth – range of frequencies used
- Interference – susceptibility to signal degradation
- Throughput – efficiency of data encoding and data rate
- Security – detection, interception, and jamming characteristics
- Implementation – physical and conceptual complexity, cost

Characteristics common to almost all radio transmission, such as fading and multi-path effects, will not be covered in the following. PHY choice and design is also affected by receiver architecture and antenna choices.

Digital Modulation (category definition) [1]

Summary:	DM is radio transmission, which encodes digital information in a carrier wave via the modulation of some characteristic of that wave.
Power:	Transmission at a certain power level suffers losses according to basic propagation equations through free space and increasingly through obstacles such as walls or vegetation.
Bandwidth:	Concentrating power in a narrow bandwidth allows a signal to punch through noise easily but makes direct interference catastrophic. Spreading power over a large bandwidth makes interference more tolerable but may blend the signal more with noise. The amount of bandwidth used also affects the number of transmitters operating in exclusive frequency ranges that can coexist in the same physical space.
Interference:	Competing radio transmission or other radio frequency (RF) noise in the same frequency range as the transmitted signal may disturb or render impossible the reconstruction of the signal by the receiver. Unsynchronized and unassociated DM systems may not share bandwidth and will always interfere with one another.
Throughput:	Baseline throughput is dependant upon the rate of modulation. Throughput multipliers can be implemented by modulating a carrier in more than one way or by using multiple carriers.
Security:	Most DM methods are able to be detected and intercepted by a radio receiver tuned to the same frequency as the transmission. Transmission complexity may make a detectable transmission harder to intercept, and making detection difficult will in turn improve the security of a method that is easy to intercept once detected.
Implementation:	Radio construction and transmission implementation with DM is a baseline for simplicity and low cost.

Amplitude Shift Keying [2][3]

Summary:	ASK modulates, or varies, the amplitude of a carrier wave to transmit data. On/Off Keying is the simplest form, where amplitude is at full strength or no strength to represent binary data. Any number of fractional amplitude strength may be used in M-ary ASK. Powers of two work especially well to represent groups or strings of binary data.
Power:	Some fractional amount of power is conserved by ASK due to the usage of a complete absence of signal to represent a zero value. This effect is most prominent in OOK and less prominent as the number M of amplitudes increases.
Bandwidth:	Very narrow bandwidth consists of the carrier frequency plus sidebands at plus and minus the fundamental frequency of ½ the bit rate. Some frequency smearing due to bit transitions.
Interference:	RF interference is at its very worst here. Signal / noise ratio is at its best.
Throughput:	Modulation rate throughput is multiplied by the square root of the number of amplitude keys.
Security:	Very simple to detect and either intercept or jam.
Implementation:	Baseline complexity and cost.

Frequency Shift Keying [2][3]

Summary:	FSK transmits data through the modulation of the frequency of the carrier wave between discrete values. Binary FSK seems to be most common, though like ASK an arbitrarily complex system could be developed.
Power:	Baseline
Bandwidth:	Bandwidth at each distinct frequency is equal to ASK bandwidth.
Interference:	Somewhat less susceptible than ASK, as each frequency band can be considered as distinct OOK signal. FSK systems may interleave frequency ranges if the actual frequencies used are not in overlap.
Throughput:	Equivalent to modulation rate capability multiplied by the square root of the number of frequencies used.
Security:	Slightly better than ASK, but still very low.
Implementation:	Low cost.

Minimum Shift Keying [4]

Summary:	MSK is a special form of FSK. Continuous phase is kept between bit transitions and frequencies are set at the minimum spacing that allows two FSK signals to be orthogonally detected.
Power:	Slightly better due to increased spectral efficiency.
Bandwidth:	MSK occupies less bandwidth than FSK.
Interference:	Baseline.
Throughput:	Can gain a bit of a throughput advantage compared to FSK, but nothing to get excited about.
Security:	Low.
Implementation:	Uses a bit more hardware than FSK.

Gaussian Minimum Shift Keying [4]

Summary:	GMSK is nothing more than MSK with a pre-modulation filter added. The filter reduces the bandwidth used by the signal even more, at the cost of causing the individual pulses to smear together somewhat, creating inter-symbol interference (ISI). This method is used in GSM digital cellular systems.
Power:	Good efficiency due to the constancy of the spectral envelope.
Bandwidth:	Phase trajectories are smoothed, greatly reducing frequency side lobes and improving spectral efficiency.
Interference:	Baseline.
Throughput:	Identical to MSK.
Security:	ISI makes intercepting a detected signal slightly more complex, but security is still somewhat low.
Implementation:	Moderately complex in order to include the filter and to decode the smeared signal.

Phase Shift Keying [2][4]

Summary:	PSK is another DM system; this time the phase of the carrier wave is modulated to contain the data. Coherent PSK (with no instantaneous voltage shifts at bit transitions) provides better performance than non-coherence. Binary (BPSK), quadrature (QPSK), differential (DPSK), and other variants are used.
Power:	
Bandwidth:	Inefficient use of bandwidth, but efficiency goes up with the number of phases used.
Interference:	Quite robust to noise. Baseline jamming characteristics.
Throughput:	The square root of the number of phases used (bits per symbol) is the modulation rate throughput multiplier.
Security:	
Implementation:	Often used to modulate information in spread spectrum systems.

Direct Sequence Spread Spectrum [5][7][8]

Summary:	This method spreads its signal out over an entire frequency range at one time, earning the designation "spread spectrum". This is done by using a "chipping" bit sequence (an 11-bit code in 802.11) with a much higher frequency than the data rate to spread the bandwidth of the signal. DSSS fits with CDMA at the link layer quite well.
Power:	Data can be reconstructed at the receiver even if parts of the signal spectrum have become too weak to be detected through fading, etc.
Bandwidth:	Though bandwidth is very wide, DSSS is designed to coexist with other narrowband systems by keeping its signal strength low enough in any one band to stay near the range that would be considered noise by single frequency systems.
Interference:	Multiple DSSS signals in the same area will interfere if they are not otherwise differentiated (as in CDMA). Resistance to narrowband interference is good.
Throughput:	DSSS has the potential to operate at very high speeds, providing excellent throughput.
Security:	By appearing as noise to traditional radio signals, this method becomes very hard to detect and intercept.
Implementation:	Somewhat complex.

Frequency Hopping Spread Spectrum [5][7][8]

Summary:	A FHSS transmission is not spread spectrum in the same way as DSSS, but in it the signal is switched rapidly from frequency to frequency in order to decrease interference and to improve security. Both the transmitter and receiver must know the pseudo-random sequence of frequency hops to communicate.
Power:	Baseline. Perhaps interference resistance may translate into somewhat fewer packets being sent overall, slightly lowering power consumption.
Bandwidth:	Narrowband frequencies spread over a wide range.
Interference:	Narrowband interference in one or more frequency bands will hurt only a small fraction of FHSS transmissions. Likewise, multiple FHSS transmitters operating in the same area will, by virtue of the frequency sequences, largely avoid conflicting with one another for any length of time.
Throughput:	Less potential throughput than DSSS, but can be very good.
Security:	Quite good. Punches through wideband jamming better than DSSS, and is difficult for any outsider without the frequency schedule to intercept.
Implementation:	Usage of FHSS has become quite common, but is a bit more complex than basic keying schemes.

Ultra-wideband [6]

Summary:	UWB transmission does **not** use modulation for data transmission. Instead, it uses near-instantaneous non-sinusoidal impulses which carry data in their timing or presence. Still in development as of 2003, it purportedly "creates a new band of spectrum out of the noise floor."
Power:	Very low power. Short range communications only.
Bandwidth:	Ultra wide, usually in excess of a gigahertz. The 3-10 GHz range has been licensed to UWB operation.
Interference:	Designed to avoid multi-path effects in its intended applications. Coexistence of multiple UWB systems in one location remains an area of research, along with coexistence with GPS and other low-level radio systems.
Throughput:	Very high data rates available, though device timing requirements are also quite high.
Security:	UWB should be as undetectable as a transmission can possibly be. Signal strengths are well below the noise floor, and spectrum is only characterized by the shape of the antenna.
Implementation:	The sticky issue is that UWB, though conceived long ago, is still in the process of being developed and deployed in any commercial way. The designs in planning call for a complex design with extreme timing requirements. A device with lower throughput might possibly have a less complex implementation.

2.1.1.1 The Data Link Layer

Protocols exist at the data link layer to accomplish logical device communication across a medium that is in some way shared. This layer in the OSI model is typically subdivided in to the areas of Media Access Control and Logical Link Control. The easiest DLC situation is where every device has a point-to-point connection to every other device. Shared media networks must deal with multipoint connections. Ad-hoc networks must do without the presence of base stations or centralized communication controls. Finally, wireless networks must deal additionally with the fact that devices cannot access the entire medium and must make use of multi-hop communication. Figure 2 illustrates the DLC methods that vary from general to very specific, and are grouped most broadly by the nature of their assignment of media access.

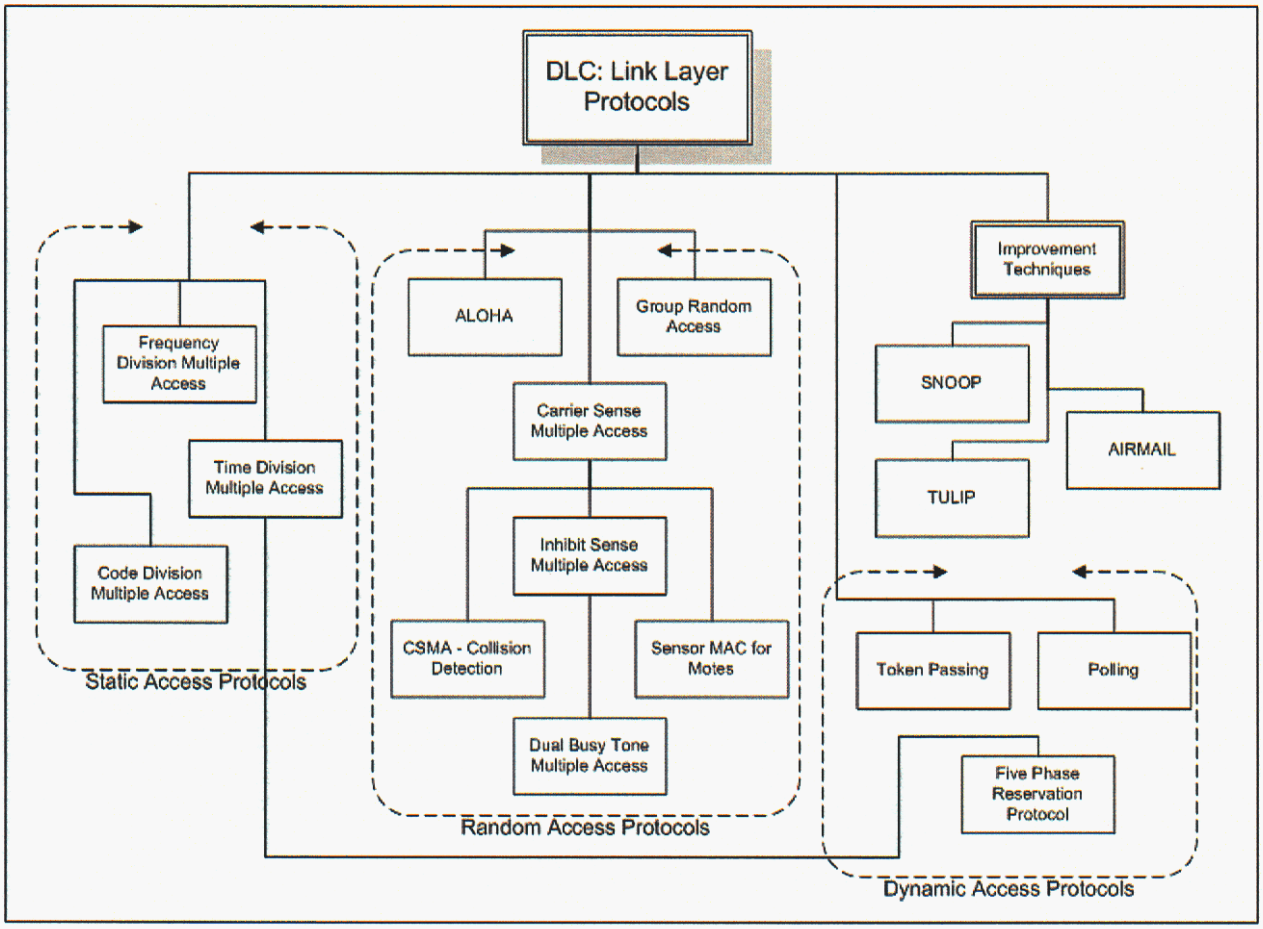


Figure 2: Data link layer

Rated DLC characteristics are:

- Throughput – efficiency of channel utilization for data transmission
- Fault Tolerance – interference, collision, fading, or other problems
- Overhead – medium usage, computation, and storage
- Latency – data transmission time, average and maximum
- Security – may be inherent or designed in

- Scalability – handling of devices, from two to infinity
- Interoperability – usability with PHY and NET protocols
- Complexity – ease of implementation logically and physically

Random-access multi-hop networks must deal with the 'hidden node' problem when two devices are too distant to communicate with each other but may collide in transmitting to a third device. Collisions may also occur in an 'exposed node' situation when device A, transmitting to device B, may interfere with nearby device C, which is receiving from device D.

Time Division Multiple Access [17][9]

Summary:	TDMA divides the medium into rounds made up of N discrete time units, each assigned to one device on the network.
Throughput:	Channel utilization under maximum load approaches 1, but as not all devices have something to say all of the time, this usually ends up being a very inefficient method.
Fault Tolerance:	No provisions.
Overhead:	Initial overhead in assigning times, very small maintenance overhead in static network. Dynamically adding devices to the network does not mesh well with the static nature of time division assignments.
Latency:	Small for networks with a few devices, the latency rises quickly with the number of devices, as each device has to wait until its prearranged time to transmit. Latency may be kept low in large networks only be drastically reducing throughput (i.e. the amount of time each device has to transmit).
Security:	No provisions.
Scalability:	Very bad.
Interoperability:	Good; this method may even be combined with other DLC methods.
Complexity:	Very simple.

Five Phase Reservation Protocol (TDMA base) [16]

Summary:	This method is a reservation system that turns TDMA into a dynamic protocol. The five phases mentioned are: reservation request, collision report, reservation confirmation, reservation acknowledgment, and the packing & elimination phase.
Other characteristics:	FPRP is designed to be completely distributed and scalable. Hopefully throughput goes up without unduly increasing latency. How exactly this affects all other attributes of TDMA is not explicitly stated.

Frequency Division Multiple Access [9]

Summary:	The channel is divided into N frequency bands, one for each device on the network. Conceptually, every device in the network could be transmitting at the same time with no interference. The problem, evidently, is to coordinate transmitting and receiving. Orthogonal Frequency Division Multiplexing uses frequency spacing to help cancel out interfering signals. Multiple antenna arrays can be constructed and used in Multiple Input, Multiple Output OFDM to achieve spatial multiplexing as well.
Throughput:	When all devices are transmitting, channel utilization approaches 100% x N, but this is very rarely the case. Actual throughput ability depends upon reception capability. As a standard radio can only listen to one frequency at a time, the essential problem of communication coordination is not lessened by FDMA.
Fault Tolerance:	No provisions.
Overhead:	Initial overhead in assigning times, very small maintenance overhead. Perhaps a frequency would need to be allocated to control information overhead?
Latency:	Very low transmit latency.
Security:	No provisions.
Scalability:	Dependent on frequency capabilities of hardware (spectral efficiency), but inherently upper-bound.
Interoperability:	Good; this method may even be combined with other DLC methods.
Complexity:	Seems to be quite tough and unpopular to implement due to hardware limitations.

Code Division Multiple Access [9][19]

Summary:	CDMA signals overlap in time and frequency. Their separation is achieved by the encoding of signals (via XOR) with a chipping signal (spreading out its frequency band) that is one of a set of orthogonal bit-codes. The signal is then mixed with all others in the channel. The receiving station can recover the signal transmitted from a specific device by repeating the XOR encoding process. The code XOR'ed to itself produces all 0's, leaving only signal, while all other orthogonal codes produce half 1's and half 0's, obscuring their signals in the noise background. CDMA forms the basis of 2G technologies, and a new wideband specification, W-CDMA, will be widely used in 3G wireless networks.
Throughput:	Very good, channel utilization greater than 1.
Fault Tolerance:	None inherently provided.
Overhead:	Processing of signals is more effort, but does not specifically take away any throughput bandwidth.
Latency:	Medium. Synchronization of codes required. (Combination with slotted ALOHA or other LDC methods can accomplish this.)
Security:	High. First, the signal is spread spectrum. Then the data must be decoded with the proper sequence.
Scalability:	Better than TDMA and FDMA, but still limited within its current implementations. There are a finite amount of orthogonal codes for any bit length chosen. Distributed coordination among MANET devices to spatially reuse codes could provide excellent scalability.
Interoperability:	Specific to PHY – DSSS.

Complexity:	High and somewhat proportional to scalability in non-multi-hop networks.

ALOHA [9]

Summary:	The simplest of all random medium access protocols, ALOHA by definition provides no channel control. Devices simply transmit whenever they have information to send. If acknowledgement from the receiver never comes, the transmitting device must assume that a collision occurred and resend. Slotted ALOHA restricts transmissions to defined time periods so that somewhat fewer collisions will occur.
Throughput:	Low. When all devices have an equal chance of transmitting, medium usage peaks at approximately 18% (pure) and 37% (slotted).
Fault Tolerance:	Retransmission will succeed eventually, but at the cost of throughput. A sufficiently busy network would have its throughput drop to zero.
Overhead:	Low. (ACKs) High if counting collision/retransmission.
Latency:	Low in low traffic, high in high traffic.
Security:	No provisions.
Scalability:	Extremely bad.
Interoperability:	Good.
Complexity:	Low.

Carrier Sense Multiple Access – Collision Avoidance [9][24]

Summary:	CSMA-CA is a random access method that's a bit more polite than ALOHA. A device must listen to the channel and detect that the channel is idle before attempting to transmit. If the channel is busy, the device will sit and wait until the channel is again free before retrying.
Throughput:	Much better than ALOHA. Collisions may still occur due to propagation delay and, in wireless networks, the hidden/exposed node problems.
Fault Tolerance:	Great at low error rates, no graceful degradation under link failure.
Overhead:	Little. (ACKs)
Latency:	Low, can degrade under heavy traffic.
Security:	No provisions.
Scalability:	Fairly good.
Interoperability:	Good.
Complexity:	Fairly low.

Carrier Sense Multiple Access – Collision Detection (used by IEEE 802.3 LAN) [9][24]

Summary:	In addition to the functionality and features of CSMA–CA, CSMA–CD provides for detection of a collision in progress by listening to the channel while in the process of transmitting. This reduces the effect of collisions by terminating a collided packet, but does not reduce their number and does not correct the hidden/exposed node problems.
Throughput:	A little bit better than CSMA–CA.
Fault Tolerance:	Great at low error rates, no graceful degradation under link failure.
Overhead:	Little. (ACKs)
Latency:	Low. Takes more traffic than CSMA–CA to degrade.
Security:	No provisions.
Scalability:	Fairly good.
Interoperability:	Good.
Complexity:	A bit more complex than CSMA–CA. Cannot be implemented with half-duplex radios.

Data (or Inhibit) Sense Multiple Access [14]

Summary:	This is a wireless version of CSMA designed to solve the "hidden node" problem. A base or receiving station will broadcast a busy signal during the times it can detect network traffic, helping devices to avoid otherwise unforeseeable collisions. This is best implemented in a centralized or cluster configuration where the inhibiting device is the only transmission target possible.
Throughput:	If inhibit signal can be constructed such that it does not interfere with reception by other nodes, then throughput should be equal to or greater than CSMA. Otherwise, the inhibit signal solves the hidden node problem only to exacerbate the exposed node problem.
Fault Tolerance:	Single point of failure at base station in centralized networks. Otherwise, standard ACK compensation.
Overhead:	Not significantly more than other CSMA methods.
Latency:	Somewhat low
Security:	Very bad. Provides built-in jamming method.
Scalability:	Between poor and good, depending upon implementation.
Interoperability:	Good.
Complexity:	Still fairly low.

Dual Busy Tone Multiple Access [15]

Summary:	DBTMA requires the use of two radio channels. If that can be achieved, this method provides an improvement beyond any other carrier sense or RTS/CTS method. Two busy signals are used, one for transmitter, one for receiver. DBTMA solves both the hidden and exposed node problems.
Throughput:	Good theoretical throughput due to lack of collisions.
Fault Tolerance:	Is not susceptible to the collision of RTS/CTS, but safeguards data packets above all else. Ack's remain as compensation beyond that.
Overhead:	Separate channel for overhead is both good and bad.
Latency:	Low
Security:	Susceptible to jamming by sine-wave busy signals.
Scalability:	Good, does not rely on a base station like the original ISMA concept.
Interoperability:	Specific physical requirements. Good otherwise.
Complexity:	Two radio channels mean hardware complexity.

Group Random Access [13]

Summary:	While other random access protocols use a random back-off feature to resolve collisions, GRA employs a binary-tree search method to enable smaller and smaller sets of devices until one can transmit without collision.
Throughput:	Though this technique is not widely used, I believe that throughput under a heavy load would be superior in this technique to any other random access protocol.
Fault Tolerance:	Single point of failure if one device controls search mechanisms. Will not completely fail at some density of network traffic.
Overhead:	High overhead; search packets and time take away directly from transmission bandwidth-time.
Latency:	Slightly higher than the average random back-off time, but transmit latency will never bog down at some channel load point.
Security:	No provisions.
Scalability:	Overhead and latency increases proportional to log(N) as N, the number of devices, is scaled up. Each device must have a fixed ID in the tree hierarchy, implying that mobile networks would not be easily handled.
Interoperability:	Fair vertical independence.
Complexity:	Moderate.

Token Passing (used in 802.4 & 802.5) [9][13]

Summary:	As a medium controller, Token Passing trades some latency for the ability to eliminate collision and contention altogether. A single logical token is passed around the network in some sort of sequential order, and a device may only transmit if-and-when it has the token.
Throughput:	Better channel utilization than random access methods at high load, and better than static channel division under low, asymmetrical, or bursty traffic.
Fault Tolerance:	Single point of failure at token.
Overhead:	Not much effort involved in token passing, but some channel overhead exists.
Latency:	Somewhat higher in low traffic situations. Fairness of latency becomes an issue in heavier traffic.
Security:	Possibly bad. Can the token be stolen by an intruder?
Scalability:	Infinite, at the cost of more latency.
Interoperability:	Good, will work with almost any physical method.
Complexity:	Rather simple.

Polling [11][21]

Summary:	Basic polling is very much like Token Passing, in that each device is asked in turn if it has anything to send. More sophisticated polling schemes can help in the areas of fairness and priority by conducting a 'reservation period' of polling before commencing data transfer authorization for that round.
Throughput:	Like Token Passing, there exists high channel utilization except for some organizational overhead.
Fault Tolerance:	Single failure point at polling device. Transmission faults can be quickly recovered from.
Overhead:	Possibly a bit more overhead even than token passing. Very controlled dynamic access.
Latency:	Traffic labeled important can be guaranteed a certain level of latency, but no low latency in general.
Security:	No provisions.
Scalability:	Fair, direct relationship to latency. One suggestion to improve scalability/latency is to split a network into two polling rings (active and inactive) and allow devices to move back and forth as needed.
Interoperability:	Good.
Complexity:	Moderately to highly complex as requirements dictate.

Asymmetric Reliable Mobile Access In Link-layer [10]

Summary:	Link protocol theory. It is actually defined not as a link layer method, but a set of actions to increase the performance of centralized wireless networks.
Throughput:	Forward Error Correction is called for at the bit, byte, and packet level to minimize wasted amounts of throughput.
Fault Tolerance:	Focus is on fault tolerance amidst noise and fading.
Overhead:	Computation at wireless devices is kept to a minimum, placing computational overhead at the base station
Latency:	
Security:	No provisions.
Scalability:	Hand-off of devices between base station cells/clusters is provided for.
Interoperability:	Questionable. It is assumed that AIRMAIL uses standard CSMA–CA as its core.
Complexity:	AIRMAIL is complexity added for the sake of performance.

Snoop [20]

Summary:	Link layer protocol designed to improve TCP over single-hop wireless links.
Throughput:	Corrects the tendency of TCP to assume that dropped packets are due to congestion, rather than loss. This improves wireless TCP hugely.
Fault Tolerance:	Excellent, performs caching of packets and handles all retransmission.
Overhead:	Inserts a large service in the link layer, reducing packet overhead at the cost of processing power.
Latency:	Dealing with losses at a low level may or may not be fast enough to avoid TCP timing out and trying to resend the packet at a higher level.
Security:	No provisions.
Scalability:	
Interoperability:	TCP specific, single use protocol built on top of CSMA–CA. Otherwise useless.
Complexity:	

Transport Unaware Link Improvement Protocol [12]

Summary:	Designed to improve TCP ala Snoop, but without requiring a specific version of TCP. Can be used in multi-hop networks without a base station present.
Throughput:	Is a bit quicker on the retransmissions, avoiding TCP timeouts.
Fault Tolerance:	Better than Snoop, which itself is much better than pure TCP.
Overhead:	
Latency:	
Security:	
Scalability:	
Interoperability:	Still TCP specific.
Complexity:	

Sensor-MAC for Motes [18]

Summary:	With a basis in CSMA–CD / 802.3 / 802.11, S-MAC is a recent design specifically for wireless sensor networks. Devices synchronize to their neighbors and then enter sleep state cycles. An RTS/CTS system is used to minimize the hidden node problem. Devices use the length indicator in packet headers to know how long to sleep for if a packet is not addressed to them.
Throughput:	Low normally, high when needed.
Fault Tolerance:	Good. Long messages are broken up into shorter messages (to make faults lets costly) but are then transmitted in burst to maximize throughput.
Overhead:	Control packets are passed around, but are kept short and few. Devices also perform regular time synchronization with their neighbors.
Latency:	High latency due to sleep states. Device level fairness is also reduced in favor of overall system performance
Security:	No provisions.
Scalability:	Very good scalability, completely distributed.
Interoperability:	Good.
Complexity:	Complex in concepts. No special hardware complexity.

2.1.1.2 The Network Layer

Where the Data Link layer handles communication on a device-to-device basis, the Network Layer handles the transfer of data streams from a point to any other point on a network. Routing through mobile ad-hoc networks is, like the lower layer services covered in this document, quite a challenge. Routing protocols can be classified as proactive/reactive as shown in Figure 3.

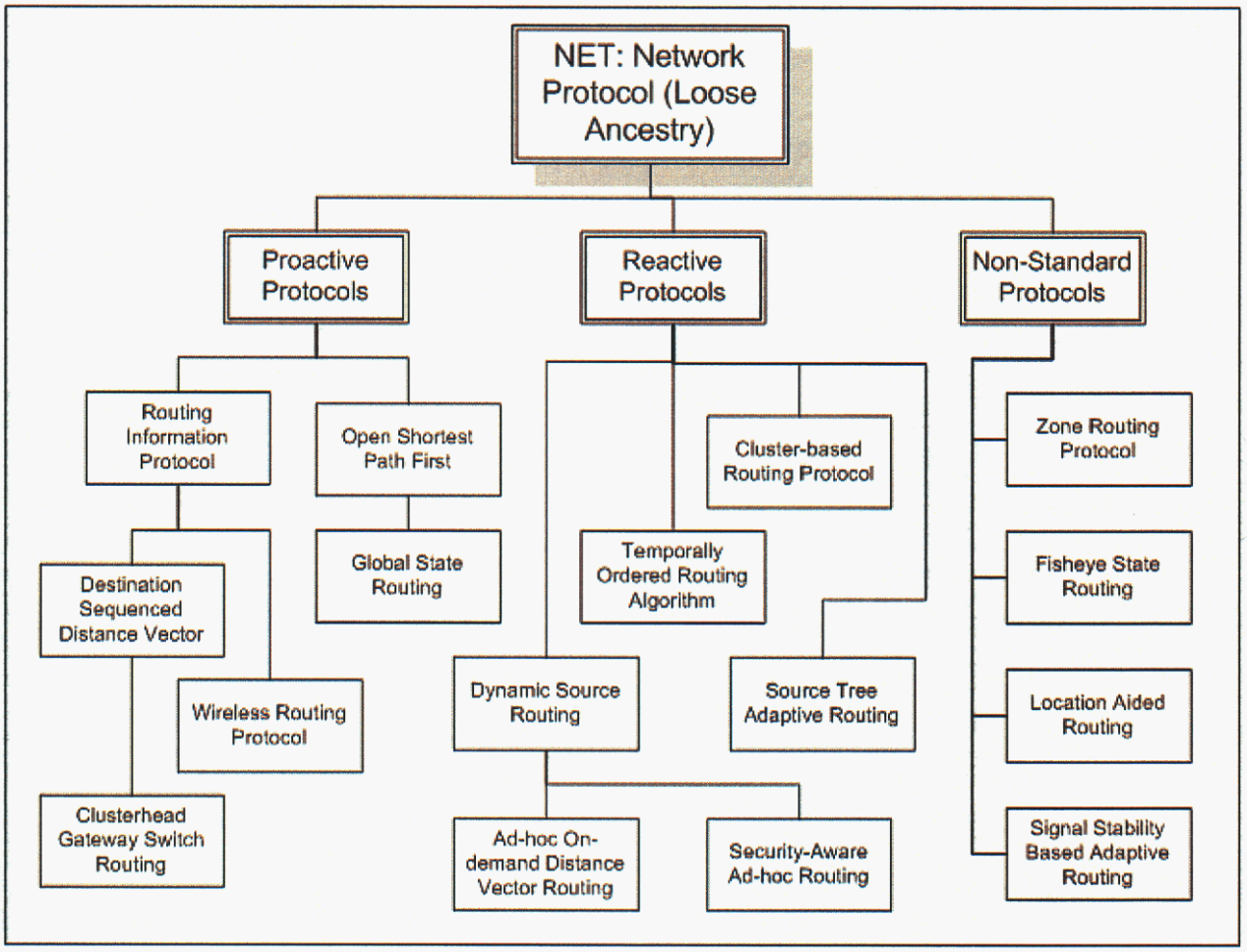


Figure 3: Routing protocols for network layer

A completely proactive protocol creates routes before they are needed, while reactive protocols create routes in response to route requests. Network routing protocols can also be categorized by the method by which a protocol constructs routes. Distance vector routing involves passing routes through the network for selection, while link-state routing passes neighbor-to-neighbor link status messages for each device to build a network topology and then create routes from it. Figure 4 provides a protocol comparison spread across these two factors.

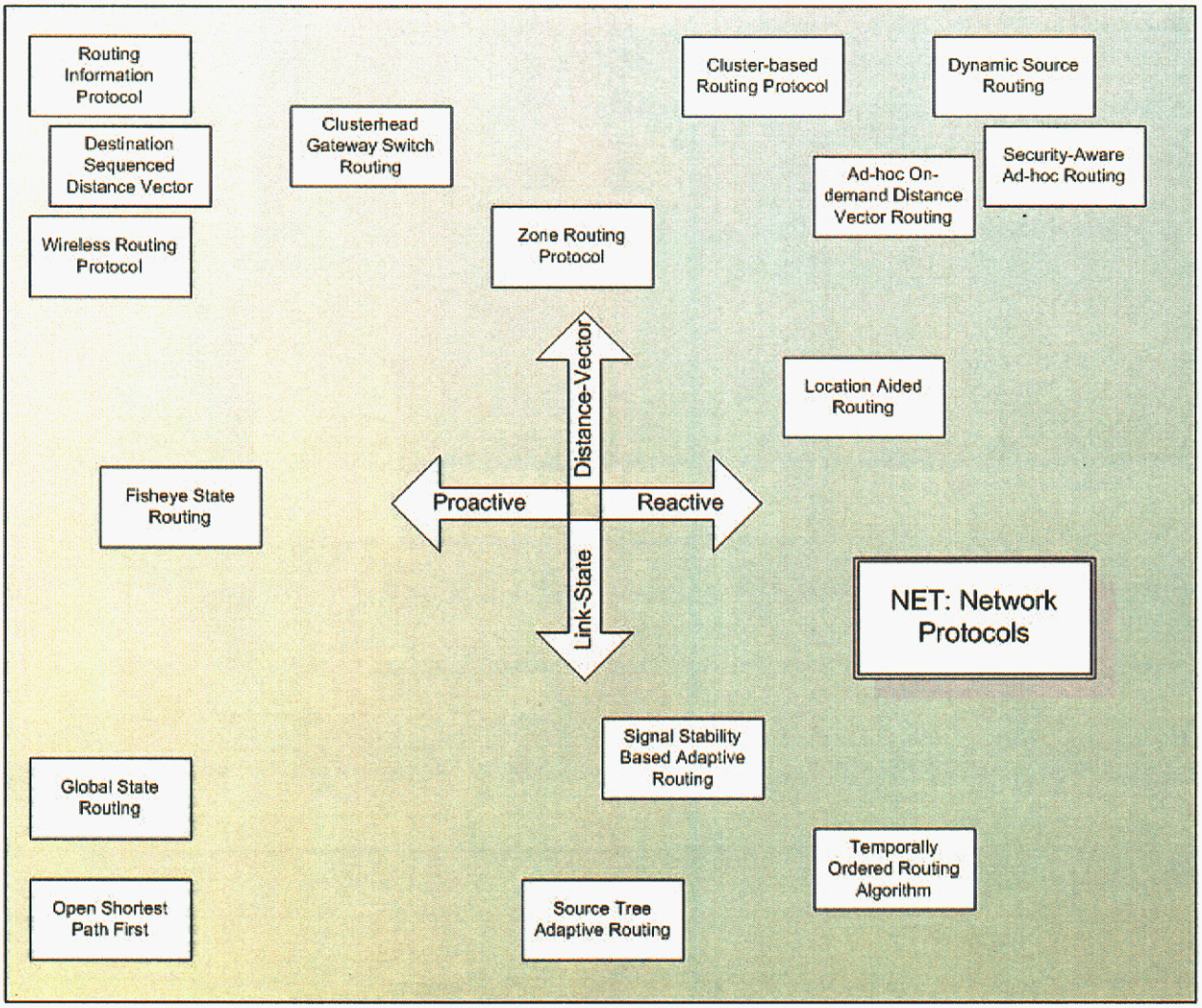


Figure 4: Network protocol comparisons for link-state, distance-vector, proactive, and reactive.

Rated DLC characteristics are:

- Throughput – data flow, successful routes
- Fairness / QoS – starvation, equality, service guarantees
- Overhead – computation, storage, and transmission bandwidth
- Latency – initial convergence ignored, all operational latency considered
- Security – anything to prevent intruders, detection, compromised insiders, etc
- Scalability – ability to operate and route through large networks
- Complexity – ease of implementation and development

The knowledge to make routing decisions must come in the form of network communication. Given the throughput limitations inherent in the hardware of current MANETs and the potentially large amount of information needed for effective routing, overhead and throughput have a critical relationship. The amount of mobility is a key factor to throughput, fairness, overhead and latency. All routing protocols which do in fact allow for mobility make

the assumption that mobile devices are moving at a speed slow enough that intelligent routing still improves the performance of the network (as the alternative is obviously uninteresting).

Routing Information Protocol [34]

Summary:	This is the first and most basic implementation of the Distance Vector routing algorithm developed in the 50's by Bellman-Ford. Each device keeps a table with an entry for every possible destination in the network. A table entry consists of a cost/distance/hops number and the ID of the first device on that optimal route. Broadcast updates are sent out periodically. If routing updates with shorter routes are received, the device updates its route table entries.
Throughput:	Ineffective routing information creates problem loops in the network. "Best" devices in key positions may also become a chokepoint and RIP would remain unaware (though this is common to almost all routing protocols).
Fairness / QoS:	No inherent unfairness other than physical proximity. No QoS provisions.
Overhead:	Significant useless overhead in a static network, and yet not able to route effectively through mobile networks.
Latency:	Optimal routes should be chosen, but bandwidth and traffic are not accounted for in cost decisions.
Security:	No provisions.
Scalability:	Absolutely do not use in networks with greater than 15-hop routes. Does not scale up well.
Complexity:	Very simple.

Destination Sequenced Distance Vector [30]

Summary:	This protocol improves on RIP in a few ways. It broadcasts route information immediately when there is a route change, and only when there is a change. It also adds a "sequence number" that increments with each broadcast that a device makes. More recent sequence numbers are given preference in routing decisions, and devices may also decide to delay transmission of a route update if it thinks a better route with the same sequence number may come along.
Throughput:	Loop-free paths are guaranteed at all times. Degrades quickly with increasing network mobility.
Fairness / QoS:	Baseline.
Overhead:	Route update traffic bursts may occur at times of connectivity changes.
Latency:	Traffic not accounted for, but latency should still be fairly low.
Security:	No provisions.
Scalability:	Perhaps slightly better than RIP, but still not good.
Complexity:	Somewhat simple.

Clusterhead Gateway Switch Routing [29]

Summary:	Designed for wireless; CGSR groups devices into clusters. Each cluster has one device that is selected as the Clusterhead, and may also have one or more devices that act as gateways to other clusters. All traffic flows through those devices. DSDV is used, with modifications, to handle routing.
Throughput:	Claims improvement over DSDV. Message delivery is improved, but ideal throughput is prohibited by excluding normal devices from traffic.
Fairness / QoS:	Major issue here. Clusterheads have top priority and all traffic flows through them.
Overhead:	Channel access overhead is reduced by ordering everything through clusterheads.
Latency:	Using priority token scheduling and gateway code scheduling, latency can be quite low in most cases.
Security:	No provisions.
Scalability:	Good scalability until throughput requirements bog down clusterheads.
Complexity:	Uses CDMA between clusters and Polling within clusters.

Wireless Routing Protocol [35]

Summary:	Designed as a wireless improvement on RIP; devices transmit second-to-last hop (utilizing a path-finding algorithm) as well as the distance to the destination. This brings faster route convergence through better path-finding.
Throughput:	
Fairness / QoS:	
Overhead:	Periodic update messages are required in addition to route change updates.
Latency:	
Security:	No provisions.
Scalability:	
Complexity:	

Global State Routing [33]

Summary:	Designed for wireless networks, desiring low overhead and mobility.
Throughput:	Routing accuracy not as good as an ideal link-state algorithm at low mobility, but better than RIP in high mobility conditions.
Fairness / QoS:	Can implement bandwidth function for QoS purposes.
Overhead:	Periodic route updates only (accounting for link-state being better in low mobility), keeping overhead at reasonable levels.
Latency:	There exists an optimum route update interval for each network's size & mobility conditions where the latency of routing accuracy is balanced by the latency of control packet overhead.
Security:	No provisions.
Scalability:	Fair.
Complexity:	A bit more complex than basic RIP, somewhat less than link-state.

Fisheye State Routing [32]

Summary:	The eye of a fish captures a high level of detail at its focal point, but detail decreases as the distance from the focal point increases. This principle is used in FSR to improve upon link-state routing. A topology map is kept at every device, but flooding is not used for propagation. Rather, devices communicate with their neighbors frequently using a DSDV-type sequenced update system. Long distance packets are sent with full link routes but may be corrected along the way by devices with a more precise picture of the remaining portion of the route.
Throughput:	Throughput is slightly reduced due to the initial route inaccuracies. After the number of fisheye scope levels becomes greater than two, throughput becomes insensitive to further scope gradation and overhead is reduced to its optimum under this system.
Fairness / QoS:	Can implement bandwidth function for QoS purposes.
Overhead:	Reduced from plain link-state, as the link-state changes propagate at defined intervals rather than instantaneously. Updates to neighbors are frequent; updates to distant devices are infrequent.
Latency:	Stored routes may be very inaccurate, but as the packet progresses, each device has a good idea of what is around it and corrects the packet's route path.
Security:	No provisions.
Scalability:	Good.
Complexity:	More complex than other proactive systems, but not orders of magnitude so.

Dynamic Source Routing [31]

Summary:	DSR was the first departure from proactive routing protocols. It operates completely on demand through the mechanisms of Route Discovery and Route Maintenance. Full source routes are passed in control messages and stored by devices.
Throughput:	Very good in all networks with up to moderate mobility. Can store multiple routes to a source.
Fairness / QoS:	Research to add adaptive QoS reservations and resource management is ongoing.
Overhead:	Zero in static networks, increases with mobility. Control packets contain full source routes.
Latency:	High any time that a route has not already been cached, requiring a new discovery sequence to complete before the packet can take that same trip.
Security:	No provisions.
Scalability:	Fair -- there may come a point where the length of routes starts to noticeably increase route discovery overhead.
Complexity:	Initial concept is quite simple. There are a number of optimizations and modifications available.

Ad-hoc On-demand Distance Vector Routing [28]

Summary:	AODV uses a method similar to DSR in creating and maintaining routes. It does not store source routes though, just the next hop for any destination ala the proactive Distance Vector protocols.
Throughput:	Cannot handle unidirectional links like DSR. Less overhead than DSR.
Fairness / QoS:	
Overhead:	Zero in static networks, increases with mobility. Less overhead than DSR due to less information transmitted in route control packets. Nodes may broadcast regular update packets when not being used for traffic.
Latency:	Comparable to DSR.
Security:	No provisions. It is expected that security is implemented in a higher layer.
Scalability:	Slightly better than DSR with respect to number of devices, but cannot handle unidirectional links and thus, weak points in a network, as well as DSR.
Complexity:	Moderate.

Cluster-based Routing Protocol [44]

Summary:	Devices are grouped into clusters for purposes of route discovery and information storage. Routing based on DSR.
Throughput:	Comparatively rather vague. Collision avoidance undefined.
Fairness / QoS:	
Overhead:	Targets route requests to cluster heads, rather than flooding them. This reduces some overhead.
Latency:	
Security:	
Scalability:	
Complexity:	Several considerations remain undefined in this specification.

Source Tree Adaptive Routing [36]

Summary:	Table based link state routing protocol where 'source trees' are kept. Route updates are disseminated only when absolutely necessary, and routes are allowed to deviate from optimum paths as long as permanent loops are not created.
Throughput:	Favorable comparison by author to DSR and ALP.
Fairness / QoS:	
Overhead:	Strives for less overhead than any table-based or on-demand protocol.
Latency:	
Security:	
Scalability:	
Complexity:	

Temporally Ordered Routing Algorithm [39]

Summary:	"decouples the generation of potentially far-reaching control message propagation from the rate of topological changes"
Throughput:	Good. Performance will degrade with mobility across time, as no global refresh on link-state information will ever happen.
Fairness / QoS:	Does not congest optimal routes with single route path choice.
Overhead:	Very low. Multiple routes are kept and any link change that does not compromise overall connectivity is ignored.
Latency:	
Security:	
Scalability:	Overhead does not scale with network size, while storage space at each device does (linearly).
Complexity:	

Zone Routing Protocol [41][42]

Summary:	ZRP combines proactive and reactive routing in an effort to reduce control overhead and still maintain low latency. Each device keeps and updates a small routing table for all neighbors within hop count N, defining its routing zone. When a message needs to be sent outside of that zone, a dynamic route discovery is performed by sending the route request only to devices on the edge of the routing zone.
Throughput:	Predicted to be very good.
Fairness / QoS:	
Overhead:	Selection of an adequate routing zone size produces less overhead than both table routing and source routing.
Latency:	Better than other reactive methods, not quite as good as proactive methods.
Security:	No provisions.
Scalability:	This is a strength of ZRP, as it is a flat protocol that scales very well.
Complexity:	Optimizations to route request method are necessary for optimal performance. Moderate complexity.

Security-Aware Ad-hoc Routing [38]

Summary:	Existing MANET routing schemes are, at their core, trusting and naïve in regards to security considerations. Devices must trust their neighbors to supply and carry data and route information. SAR was developed to surround a basic on-demand protocol (such as DSR or AODV) and provide the security properties of: Timeliness, Ordering, Authenticity, Authorization, Integrity, Confidentiality, and Non-Repudiation.
Throughput:	Directly affected by the number of security measures implemented. A secure route may not be an optimal hop count route.
Fairness / QoS:	Routes are chosen based on the provisions that a secure device can give. This "QoP" method would mesh well with any additional QoS requirements.
Overhead:	Large overhead placed more on device computation/power than on bandwidth.
Latency:	Latency increase is entirely computational, based upon the amount of

	encryption desired.
Security:	As much as desired. SAR does assume, however, that some method for secret key distribution is pre-existing.
Scalability:	
Complexity:	In a way, as high as possible.

Location Aided Routing [37]

Summary:	LAR is a dynamic routing protocol which uses as its basis knowledge about each device's location. This is intended to be implemented using GPS receivers on each device. Accumulated knowledge about a device's location is used to limit the broadcasting of new route requests to a smaller area and cut down on network control traffic.
Throughput:	Better than flooding. Unclear exactly how effective it is.
Fairness / QoS:	
Overhead:	Less than flooding.
Latency:	Location predictions may fail in high mobility situations, increasing latency by requiring multiple sequential route requests.
Security:	No provisions.
Scalability:	Better than flooding, but not too much of an improvement.
Complexity:	There is a compromise that exists between overhead (along with scalability) and latency as the size of the location predictions is scaled. This scheme requires increased power consumption if GPS modules are not previously implemented / in use.

Signal Stability Based Adaptive Routing [40]

Summary:	SSA makes two assumptions. First, that radio signal strength is proportional to link stability. Secondly, it assumes that a stable link will be more likely to remain in service. Using those criteria, SSA routes as much traffic as it can over stable links, trading hop count for a drop in route reconstruction cost. SSA (with FP and SRP components) is also known as SSR (components called DRP and SRP).
Throughput:	Dubious, even in the inventor's presentation of results.
Fairness / QoS:	
Overhead:	A subset of non-mobile nodes will quickly become an ad-hoc backbone, bearing the brunt of overhead and power dissipation.
Latency:	Latency is reduced in those networks that fit the mobility conditions targeted by SSA.
Security:	No provisions.
Scalability:	The bandwidth of more mobile nodes that is ignored by SSA may prove prohibitive when scaled up to some size.
Complexity:	Moderately complex. Forwarding Protocol and Dynamic Routing Protocol are each simple components.

Open Shortest Path First [43]

Summary:	OSPF is a Link-State protocol. In contrast to Distance Vector schemes, each device maintains a complete network topology according to the local topologies each device floods out to the network. Not designed for MANETs, this method is replacing RIP in much internet routing.
Throughput:	
Fairness / QoS:	
Overhead:	Significant storage at each device
Latency:	
Security:	
Scalability:	
Complexity:	

2.2 Current sensor system overview

Wireless sensor systems have been around for over a decade; while there are differences between these systems, there are several similarities. This section provides a cumulative collection of information on some select wireless sensor systems. The purpose of this review was to (1) determine existing set of sensor networks, (2) evaluate how those networks can be applied to SDAC, and (3) identify the problems that were encountered in developing the networks. During the creation of this review we found (3) to be the hardest factor to uncover, since most developmental teams do not generate 'lessons learned' documents. We also discovered that there are plenty of sensor systems, for which no technical information was available on the web. A list of the sensor systems and potential contact can be found in Appendix, Section 8.3.

To begin the evaluation process we generated a list of sensor system enabling capabilities and technologies. This list included issues associated with networking, hardware, software, communication, power, deployment, and other related items. Each of these higher-level technologies is further decomposed into more precise items to be evaluated. For the high-level area of software technology we considered (1) operating system or software architecture, (2) the extensibility of the architecture, (3) ability to process local and remote data, (4) power awareness of software, and (5) ability to support high-level applications for decision-making.

Network

1. Routing algorithms: type of algorithm, latency, robustness
2. Network architecture: homogeneous vs. heterogeneous, centralized or decentralized
3. Robustness: avoid single point of failure, rapidly reconfigurable

Hardware

1. Node architecture: single processor vs. multiprocessor, expandability, modularity
2. Reconfigurability: general architecture suitable to rapid prototyping, variety of applications
3. Upgradeability: easily able to introduce new technology
4. Sensors: implemented sensors and interfaces

Software

1. Architecture: RTOS based
2. Extensibility: easily expandable for new applications
3. Data processing: localized vs. distributed, collaboration with other nodes
4. Power aware: APIs built into code for power reduction capabilities
5. Intelligence: possibilities for higher-level decision-making application development

Communication

1. Wireless: speed, reliability, error correction,
2. Range: maximum and optimum node separation
3. MAC: always on vs. timeslots

Power

1. Lifetime: battery life
2. Power consumption: current draw

Deployment

1. Rapid deployment: by hand, remotely, automated
2. Configurability: adaptable to several deployment methods

Other

1. Size: physical size of nodes
2. Cost: cost per node or network
3. Application: purpose of network, broad vs. general applications

This complete list of high-level and supporting technologies was used to evaluate four existing sensor systems. The systems being evaluated in this report include: (a) Sensoria sGate, Crossbow Motes, Ember, and Sandia Hybrid Emergency Radiation Detector (HERD). The results of the evaluation for the Crossbow Motes and HERD are shown in Table 1 and the remaining two systems are in Section 8.1. Missing from the table was the deployment category, which seemed to be an overlooked point for the majority of the existing sensor systems.

Table 1: Existing sensor system evaluation

	Crossbow Motes	HERD
Network		
Routing algorithms	Broadcast – active messaging	Ad-hoc source routing
Network Architectures	Homogeneous. One node becomes a gateway by installing a piece hardware that allows the mote to connect to a PC	Semi-homogeneous. Requires one gateway node
Robustness	Tolerant to network changes. Does not require routing tables	Network is dynamically created and maintained. Routing tables are continuously updated as the network topology changes
Hardware		
Node architecture	Communications processor with 51-pin expansion bus suitable for a host processor	Two processors: wireless communication and application
Reconfigurability	Designed to act as a platform for rapid prototyping	New sensors can be added easily
Upgradeability	Standardized interfaces should allow easy upgrading	Moderately portable to new processors. Sensors can be upgraded easily
Sensors	Light, temperature, acceleration, magnetic, acoustic, vibration	Uses sensor specific interfaces. Currently uses GPS and radiation sensor
Software		
Architecture	Tiny OS RTOS	Communications processor: schedule and interrupt based. Application processor: RTOS based
Extensibility	Possibly limited by 4k RAM. Software is open source with considerable community support	Both processors are near their data and processing power limits, but there is room for additions
Data processing	undefined	Centralized on a PC
Power aware	RTOS includes power management features	Application processor idles at a low frequency when not in use. Radio and sensors are shut down when not in use
Intelligence	Processing power likely limits these activities	
Communication		
Wireless	900 MHz, 38.4kbs	76.8kps on 1.8s intervals, very reliable, single hop error checking, 900MHz
Range	150m	100 – 300m
MAC	SMAC	Nodes active during a 10% duty cycle of a 1.8s period. Timeslots govern communication within the period
Power		
Lifetime	1 year on AA batteries	1 week to 1 month
Power consumption	Processor: 8mA under load, 15uA in sleep. Radio: 27mA transmit, 10mA receive, 1uA sleep	~8mA idle, ~80mA full speed while GPS is active
Other		
Size	5.8cm x 3.2cm x .7 cm	6cm x 6cm x 4.5cm.
Cost	Wireless sensor network platform.	$400/node
Application		Rapidly deployable radiation detection system

2.3 Chapter 2 References

Physical

[1] Radio Fundamentals.
http://www.cs.berkeley.edu/~randy/Courses/CS294.S96/Fundamentals.pdf

[2] Amplitude, Frequency, and Phase Shift Keying.
http://www.cs.uidaho.edu/~krings/CS420/Notes.F02/02-420-06.pdf

[3] OOK, ASK, and FSK Modulation in the Presence of an Interfering Signal.
http://www.rfm.com/corp/appdata/ook.pdf

[4] MINIMUM SHIFT KEYING, DIFFERENTIAL PHASE SHIFT KEYING

http://www.winlab.rutgers.edu/~narayan/Course/Wless/Lectures02/lect11.pdf

[5] SPREAD SPECTRUM TOPICS

http://www.sss-mag.com/sstopics.html

[6] ULTRA WIDEBAND WORKING GROUP

http://www.uwb.org

[7] PHYSICAL LAYER

http://www.informit.com/isapi/product_id~{85189561-10E7-4D4D-861D-C7030DE2AB09}/content/articlex.asp

[8] SPREAD SPECTRUM: FREQUENCY HOPPING VS. DIRECT SEQUENCE

http://www.wireless-nets.com/articles/whitepaper_spread.htm

Data Link

[9] Chapter 5: The Data Link Layer
http://www.cs.umd.edu/~shankar/417-F01/Slides/chapter5a-aus/

[10] AIRMAIL: A Link-Layer Protocol for Wireless Networks
http://citeseer.nj.nec.com/ayanoglu95airmail.html

[11] An efficient polling MAC for wireless LANs
http://portal.acm.org/citation.cfm?id=504646&coll=portal&dl=ACM&ret=1

[12] TULIP: A Link-Level Protocol for Improving TCP over Wireless Links
http://citeseer.nj.nec.com/parsa99tulip.html

[13] Group Random Access Protocols
http://www-net.cs.umass.edu/cs653-1998/notes/ch7-1/sld025.htm

[14] Data Sense Multiple Access
http://www.soi.wide.ad.jp/class/20000002/slides/05/15.html

[15] Dual Busy Tone Multiple Access (DBTMA): A New Medium Access Control for Packet Radio
http://citeseer.nj.nec.com/deng98dual.html

[16] A Five-Phase Reservation Protocol for Mobile Ad Hoc Networks
http://www.ieee-infocom.org/1998/papers/03b_3.pdf

[17] IEC: Time Division Multiple Access
http://www.iec.org/online/tutorials/tdma/

[18] An Energy-Efficient MAC Protocol for Wireless Sensor Networks
http://www.isi.edu/scadds/projects/smac/

[19] CDMA Development Group
http://www.cdg.org/index.asp

[20] The Berkeley Snoop Protocol
http://nms.lcs.mit.edu/papers/hari-phd/

[21] Investigation of a Polling MAC Protocol with Capture Effect for Wireless LANs
http://www.csupomona.edu/~cs/ms/thesis/TJang.doc

[22] OFDM Tutorial
http://www.wave-report.com/tutorials/OFDM.htm

[23] OFDM Forum
http://www.ofdm-forum.com/

[24] Carrier Sense Multiple Access
http://www.erg.abdn.ac.uk/users/gorry/course/lan-pages/csma-cd.html

Routing

[25] Routing Protocols for Mobile Ad-Hoc Networks
http://citeseer.nj.nec.com/348952.html

[26] Routing Protocols for Ad Hoc Mobile Wireless Networks
http://www.cis.ohio-state.edu/~jain/cis788-99/adhoc_routing/

[27] Ad Hoc Wireless Networks
http://www.cpe.ku.ac.th/~anan/courses/204529/document/09-Adhoc_Wireless_Networks.pdf

[28] Ad hoc On-Demand Distance Vector (AODV) Routing
http://citeseer.nj.nec.com/382730.html

[29] Routing in Clustered Multihop, Mobile Wireless Networks with Fading Channel
http://www.ics.uci.edu/~atm/adhoc/paper-collection/gerla-routing-clustered-sicon97.pdf

[30] Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers
http://citeseer.nj.nec.com/perkins94highly.html

[31] DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks
http://citeseer.nj.nec.com/johnson01dsr.html

[32] Fisheye State Routing: A Routing Scheme for Ad Hoc Wireless Networks
http://citeseer.nj.nec.com/316997.html

[33] Global State Routing: A New Routing Scheme for Ad-hoc Wireless Networks

http://citeseer.nj.nec.com/60636.html

[34] Routing Information Protocol
http://www.faqs.org/rfcs/rfc1058.html

[35] An Efficient Routing Protocol for Wireless Networks
http://citeseer.nj.nec.com/murthy96efficient.html

[36] Source-Tree Routing in Wireless Networks
http://citeseer.nj.nec.com/garcia-luna-aceves99sourcetree.html

[37] Location-Aided Routing (LAR) in Mobile Ad Hoc Networks
http://citeseer.nj.nec.com/43769.html

[38] A Security-Aware Routing Protocol for Wireless Ad Hoc Networks
http://www-sal.cs.uiuc.edu/~rhk/pubs/SCI2002.pdf

[39] Temporally-Ordered Routing Algorithm
http://tonnant.itd.nrl.navy.mil/tora/tora.html

[40] Signal Stability based Adaptive Routing (SSA) for Ad-Hoc Mobile Networks
http://citeseer.nj.nec.com/dube97signal.html

[41] A New Routing Protocol for the Reconfigurable Wireless Networks
http://citeseer.nj.nec.com/haas97new.html

[42] The Performance of Query Control Schemes for the Zone Routing Protocol
http://citeseer.nj.nec.com/haas98performance.html

[43] Open Shortest Path First
http://www.ietf.org/rfc/rfc2328.txt

[44] Cluster Based Routing Protocol (CBRP) Functional Specification
http://www.ietf.org/proceedings/99mar/I-D/draft-ietf-manet-cbrp-spec-00.txt

[45] Routing and Multicasting Strategies in Wireless Mobile Ad hoc Networks
http://citeseer.nj.nec.com/lee00routing.html

Other
A Flexible and Reliable Radio Communication Stack on Motes
http://www.isi.edu/scadds/papers/commstack.pdf

3 War on Terrorism (WoT) Mission Areas

The LDRD investigated the following four-mission areas: fixed/mobile site protection, military operations in urban terrain (MOUT), intelligent community, and borders. The objective of this investigation was to identify application space(s) that represent significant overlap between the different mission areas. Figure 5 provides a brief overview of some high-level application space (inspection, detection, tracking, identifying, Surveillance, reconnaissance, perimeter, and security) that correlates significant concepts from all four-mission areas. Attached to each high-level application space are sensor related technologies like RF-ID tags and X-rays. Included in these lists are lower-application areas like sniper locater and chemical detection, which would use a variety of sensor technologies to achieve these applications.

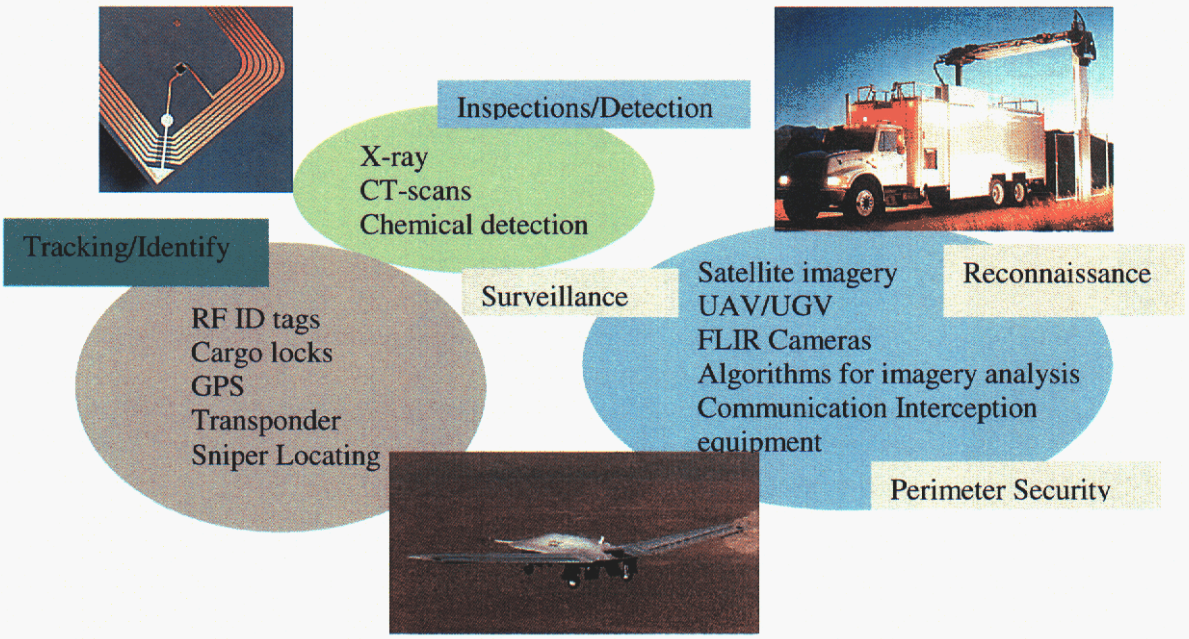


Figure 5: High-level application spaces for all four-mission areas

After these initial investigations the team revised and reduced the list back to concentrate on only two areas of the original mission spaces. The removal of the intelligence community was due to the inability to discover details about this specific domain. While the fixed/mobile site protection area was eliminated, aspects of this area were incorporated into the two remaining mission spaces for borders and MOUT. This chapter provides an overview of the motivation, challenges, potential usage of SDAC sensor networks, and detail scenario for the MOUT and border domains.

3.1 Military Operations in Urban Terrain

The military community has long recognized Military Operations in Urban Terrain (MOUT) as an area for which there has been insufficient preparation. Only recently has the community finally accepted that MOUT is not something that can be avoided, and furthermore that old warfare techniques, training, and technology are not applicable and need to be

completely revamped. Within the past ten years a great deal of theoretical research has been conducted to explore this terrain type, but no definitive solutions have yet been found to give the United States a solid upper hand in this arena. This overview discusses the motivation for MOUT research, the challenges of the MOUT environment, current MOUT military strategy, and the potential impact of technology in the MOUT domain.

3.1.1 Motivating MOUT research

There are several motivators for developing preparedness for MOUT. First, the world's urban population is growing disproportionately quickly in comparison with other environments. Besides the natural population increase, more and more people are moving away from their rural communities to the cultural, social, political, infrastructural, and economic "centers of gravity" of urban areas. Furthermore, urban zones are continually expanding and building up previously under-developed areas. United Nations estimates that by 2025, 60% of the world's population (5 billion people) will be in urban environments [1]. It is an assumption that areas with more people will have inherently more military conflict. Second, the well recognized MOUT failures in places such as Mogudishu, Grozny, and Jenin have called a great deal of attention to the lack of preparedness for such a complex landscape [12].

Third, MOUT is an equalizing terrain in the sense that the technological and warfare prowess of highly developed militaries do not readily transfer from other terrain types. Most warfare technology is geared towards long-range combat, but in MOUT 80% of all engagements take place in under 100 meters [20]. Also, most advanced military troops are simply not trained in MOUT environments, or the training that does exist is highly insufficient. Fourth, Military Operations Other Than War (MOOTW) will likely be conducted more often in urban terrain simply because that is where there are the highest concentrations of people [4]. Fifth, warfare simulation in other terrain types does not nearly capture the high degree of variability, complexity, and urbanization of the MOUT environment; so besides not having very much actual empirical data to work with, not even simulated MOUT results can be used for technique and technology development [20]. Finally, the increasing hazards of world terrorism will likely be concentrated in urban terrain because this will be the area where the most infrastructural and symbolic damage can occur [12].

3.1.2 Challenges and strategies of the MOUT domain

It should be apparent that there is a great need for MOUT development, but it is important to understand the unique challenges that make MOUT such a highly complex and interesting terrain. Aside from the extensive man-made constructions, perhaps the most unique feature of MOUT is the presence of non-combatant populations [14]. Not only does this add to the difficulty of identifying the enemy without injuring civilians or committing fratricide, but also it increases political, social, cultural, and economic tensions, which can govern military engagements. Urban terrain also has the very unique characteristic of changing in response to the military operations that are conducted within this domain [1]. An avenue through which troops moved yesterday might have been blocked overnight, and battlefields can quickly open, close, and shift from the razing of structures and creation of rubble fields. In this way, urban terrain can also create severe mobility restrictions beyond the already inherent problems of horizontal and vertical movement through built up areas.

Another feature of urban terrain is that it creates difficult Command, Control, Communication, Computer, Intelligence, Surveillance, Target Acquisition, and Reconnaissance (C4ISTAR) issues [4]. Besides the physical restraints on radio communications imposed by structural interference, multi-path, and fading effects, the lack of line of site surveillance of soldiers and battlefields by commanders grossly impedes their ability to lead coherent operations. Since fighting may occur from building to building, or even room to room, the fast pace, high casualty rate, high ammunition usage, and close combat situations further compound command issues. Finally, as Marine General Charles Krulak has pointed out, MOUT can be thought of in the context of a three-block war. In neighboring urban blocks, soldiers may be conducting humanitarian, peace keeping, or high intensity warfare operations. Not only do these create extreme psychological tension, but also necessitate highly dynamic troops. All of these issues and restrictions combine to create a very dense battle space with acute levels of physical, psychological, communication, and social interference not found in any other terrain.

To deal with this extreme terrain, military analysts have developed three main strategies [14]. The first MOUT strategy is simply not to engage in warfare in urban terrian. This is a serious strategy that has been proposed primarily because of the lack of military preparedness. This is quite obviously not a sufficient long-term solution, but the high casualty and destruction rates that currently accompany MOUT lead several analysts to believe it is a terrain that is simply unmanageable.

The second MOUT strategy is called attrition style warfare. This strategy is accompanied by a methodology to Isolate an enemy, Retain control of an area, Contain an enemy, Deny an enemy outside assistance, and then Reduce an enemy's material and human assets (IRCDR) [4]. This style has been employed in numerous operations throughout the world and has been seen to incur high numbers of casualties and leave complete infrastructural destruction in its wake. This style can be characterized as a "ring of fire" or "shock and awe" created to surround and then completely level enemy strong holds with large amounts of ammunition and firepower. It is agreed that attrition style warfare is asset intensive, both in money and troops, and for this reason is not an attractive MOUT solution. The immense infrastructural damage also creates a huge after battle cost to rebuild the demolished urban environment. Furthermore, attrition style warfare does not take advantage of urban terrain; it instead levels it in order to create a terrain for which the military is more readily prepared.

Recently, a new, and as yet untested, strategy has been developed called maneuver style warfare. This style has been developed to overcome the shortcomings of attrition style warfare. It seeks to appreciate the urban terrain and leverage its unique features in order to dominate an opposing force. It is characterized by a fast tempo, more precise and directed destruction and attack in order to reduce an enemy's mobility, funnel enemy troops into "killing zones", and reduce an enemy's assets through iterative attacks on its weakest links. Along with this new style of warfare also comes a new operational methodology to Understand the unique urban terrain of an engagement, Shape the battle space by moving assets into strategic locations, Engage enemy forces with integrated and synchronized attacks, Consolidate areas that have been gained, and ultimately Transition control back to local authorities (USECT) [7]. The maneuver style warfare

has a great promise to revolutionize and "clean up" MOUT, but many significant technological advances must be developed in order to make this strategy effective and realizable.

3.1.3 SDAC for MOUT

The particular areas that technology can be most effective have been fairly well defined from operational perspectives, but the list is large and growing. Table 2 provides a brief list and description for a set of nine different operational sub areas in MOUT where SDAC wireless sensor networks could be used. Aside from developments such as precision short-range and non-lethal weaponry [16], identification of friend, foe, and non-combatant populations [4], mobility in the vertical, horizontal, and subterranean domains [1], deception operations to control the behavior of opposing forces [16], and simulation research needed to develop more efficient MOUT techniques and procedures [20], sensor system technology specifically can make a significant and wide-spread impact in various different parts of the environment [8]. In case studies of recent MOUT failures, one of the most widely given reports was that the MOUT situations are extremely confusing and complex, and it was easy to lose track of what people were supposed to be doing when in response to various fast paced and close range changes in the operating environment [21].

Table 2: SDAC capabilities for MOUT scenarios

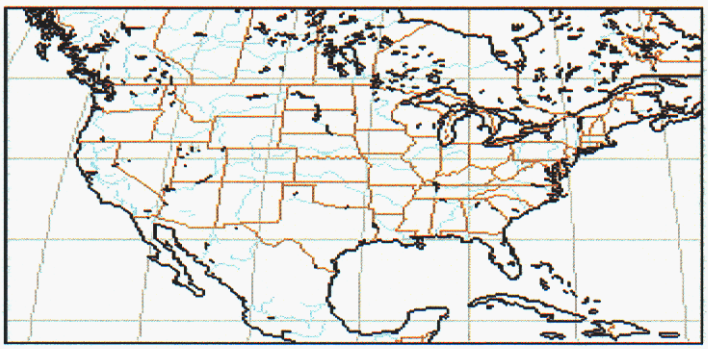
Area where SDAC apply	Description of potential SDAC capabilities
Identify Combatants from Non-combatants	Urban terrain has an interesting socio-economic mix and military operations should seek to avoid civilian damage. Beyond simply identifying the infrastructural layout of an area, information about the social and human factors in an area have been said to make an equal impact as to the success of a MOUT mission.
Mapping an area	Video data from a variety of nodes can be combined with satellite and human assistance to map the area. The video data from closely spaced nodes that know their location can be theoretically fused into a featured map.
Surveillance of an area	The same nodal network could be used to create a robust video and audio surveillance network to track enemy movement. This helps in the **retaining**, **containing,** and **reducing** aspects of operations.
Create a decentralized sensing, observation, and control network	The distribution of thousands of nodes in an area to act as sensors and routers to gain **situational awareness** of their surroundings by collaborative data processing and then relaying this information back to a central control. This will address the difficulty of communication in an urbanized area through redundancy and also be robust against destruction by enemy forces. (Maybe broadband is the way to go for robustness issues, security can

	be taken care of with encryption.)
Perimeter construction	Since **denying** and **isolating** are two other primary missions, a perimeter around the urban terrain should be established. In an urban area this could be more challenging than in non-urban terrain because of the ability of covert enemy intrusion. An electronic perimeter could be set up with SDACs to aid human forces in maintaining of this perimeter.
Deception operations	SDACs could be used to jam enemy signals, create audio chimeras, startle forces with loud sights and sounds, deploy fog or tear gas, etc.
Logistic support	Creating data maps of an area after a WMD release, large-scale fires, floods, or weather could be conducted by SDACs to give logistic support to controllers of the area.
Physiological Sensing Units	"Smart" clothing that can sense physiological, logistical, and positional aspects of soldiers and their gear and relay this information to commanders

Sensor networks, if distributed across this terrain, could identify movement of opposing troops, localize snipers, create a communications backbone for command and control, and provide soldiers with a heightened level of connectivity and situational awareness. This will allow soldiers to dynamically see and know their operational conditions and eliminate confusion. This type of capability would also allow for commanders to track and control their troops more efficiently, and be able to disseminate information or mission changes to their troops on the fly. This type of sensor network could also be extended and used to monitor physiological conditions of friendly troops while tracking and observing enemy and non-combatant populations to provide an even wider scope of battle space understanding. Though sensor networks could have a variety of different applications in the MOUT domain, the situational awareness enhancement and communications connectivity would probably be the two largest possible benefits.

3.2 Border protection

3.2.1 Motivation for Border Research



The United States (US) mainland has a mixture of both terrain and maritime borders, each providing unique challenges for secure and stable border control. With over 6000 miles of terrain borders surrounding its mainland and over 350 ports of entry, the US has a lot of land to protect. Adding to this situation is the recent integration of the Customs/INS, Border Patrol, and Coast Guards into a single agency, which makes the problem even more complex. To the South the United States shares

1604 miles of mostly terrain borders with Mexico, where illegal aliens and drugs have been a problem for decades. To the North, Canada borders the US with 4329 miles of both terrain and maritime (great lakes region) borders. Unlike the Southern border, Canada was always seen as the stable neighbor, even with steady increases in drug trafficking.

However with the initiation of the WoT, Canada's continuous hills and dense tree terrains have proven to be a growing problem for the United States border patrol. After the "December arrest of Ahmed Ressam as he attempted to enter the U.S. from Canada with hundreds of pounds of sophisticated bomb-making materials", the US/Canadian causal relationship became tenser [22]. Suddenly it was apparent to the United States that Canada and not Mexico had proven to be a better haven for terrorist attempting to enter the United States both legally and illegally.

Smith's immigration subcommittee heard terrorism experts from both the United States and Canada cite the Canadian public's historic lack of concern about terrorism and the growing realization refugees are taking advantage of Canada's lenient policies.

3.2.2 Challenge of the border domain

The security challenges of the United States borders is threatened by several variables from difficult terrain issues to the need to promote efficient open trade within the continent. The *political challenges* for the US border represents somewhat of a nightmare for the government, which must assert relationships with allies to broaden influence and catch problems before they reach the US shores. By combining political issues with the *challenge to detect illegal substances, weapons of mass destruction*, and *other weapons* the government has learned to broaden and exercise its influence with significant trade partners. The homeland security agency has devised screen methods (X-ray, radiation testing, etc.) to test containers at their origins before being placed on a ship headed to the US shores [23][24]. However, on Aug 23, 2003 ABC News exposed a potential flaw in the screening of containers entering the US borders. The news agency exposed U.S. screeners' failure to detect a 15-pound shipment of depleted uranium in a container sent from Jakarta, Indonesia [25].

The *challenging terrain and climate* issues that reside at the US Northern and Southern borders require different solutions for each area. The ill-defined Northern border represents a unique problem for the US, which has traditionally treated these borders as less of a threat over the Southern terrain [26]. The Northern border represents open dense trees and difficult hilly terrain that makes timely location of people by plane or on foot difficult. Complicating the problem are the altering weather conditions in the winter that makes the terrain almost impassable for the border patrols. The Southern border represents the US most watched terrain, due to the *challenge of detecting and halting the movement of illegal aliens* across these areas. Unlike the Northern borders the US-Mexico border represents flatter open terrain that is ideal for aerial surveillance and automated camera units. Traditionally the border patrol sets up sensor suites on the Southern border, which represents a camera and a set of buried trigger sensors (seismic and/or acoustic). The trigger sensors pick up vibrations and sound that moves the camera in the direction of the sensor. The *challenge of covert sensor placement is vital* to the success of the trigger sensor units. Especially since smuggles look for disturbed ground to locate the buried sensor units and attempt to disable these devices.

Some problems that plague the border situation are *limited personnel and resources* needed to cover the 6000+ miles of terrain. Other *challenges* include *sensor placement* and *sensor node power usages*, which require large batteries for the buried trigger sensor units and solar power for the camera units. The US border situation has become an ever evolving situation with varied degrees of complexity that range from political negotiations to stricter monitoring of movement between the US-Canadian and US-Mexican terrains.

3.2.3 SDAC for borders

The applications of SDAC sensor networks to the border domain are fairly extensive. Table 3 provides a brief list and description of six potential applications where SDAC wireless sensor networks could be used. The applications include cargo container monitoring [27][28], detection of weapons of mass destruction, and integrated sensor platforms that will reduce false alarms and improve patrol investigation process.

Table 3: SDAC capabilities for borders domain

Area where SDAC apply	Description of potential SDAC capabilities
Reduce false alarms and improve efficiency	Numerous video images sent back to the patrol area has no significant information and represent false alarms. Intelligent data fusion within the sensor system, which combines imaging and sensor data, could assist in the reduction of false alarm.
Automate border patrol triangle investigation process	The same intelligent data fusion capability stated above can be used to automate the investigation process used by the patrol. However as we add a larger collection of sensor nodes with multiple sensors on a given unit, we will be able to better assess the situation.
Integrate multiple sensors onto single platforms	Replace current single sensor units with multi-sensor units on a single platform. Use intelligence situated in the sensor network to provide improved information back to the human
Detection of weapons of mass destruction	Sensor networks suites containing chem.-bio sensor and/or radiation sensor can be placed on containers to monitor the leakage of agents. These sensor suites can also be used at portals to pick up hints of these agents as people pass through these areas.
Cargo container monitoring & tracking	Placement of secure tags on cargo containers can be used to track movement of the container in the US borders. It could also be used to detect tampering of the container prior

	to entry to the US shores.
Underwater surveillance	Establishing a web of wireless underwater sensor buoys that detect and identify potential divers or non-aqua life in the area.
Power Management	By establishing advanced architectures that conserve power through intelligent, scaled, and sufficient resources, lifetimes can be extended
Advanced Sensors	By applying sensors with orthogonal or advanced sensing properties, the false alarm rates can be reduced essentially increasing the signal to noise ratio and improving detection capability.

The distribution of sensor networks across the border terrain will improve the border patrols' ability to better assess the Northern and Southern borders with units that are uniquely design for these different areas. The sensor could potentially eliminate confusion over false positive sensor reading and classify the cause of movement for a given image. Tagging sensors can be used to monitor flow of cargo and misuse of tampering of containers while in route to the US shores. Collectively, the incorporation of sensor technology into the border domain will improve the overall abilities of the patrol to achieve their mission of safe and secure borders.

3.3 Chapter 3 References

[1] Military Operation On Urbanized Terrain (MOUT)
http://www.globalsecurity.org/military/library/policy/army/fm/90-10/toc.htm

[2] An Infantryman's Guide to Combat in Built-Up Areas
http://www.globalsecurity.org/military/library/policy/army/fm/90-10-1/default.htm

[3] Combined Arms Operations in Urban Terrain
http://www.globalsecurity.org/military/library/policy/army/fm/3-06-11/toc.htm

[4] Handbook for Joint Urban Operations
http://www.dtic.mil/doctrine/jel/other_pubs/juohdbk1.pdf

[5] The Urban Operations Journal
http://www.urbanoperations.com/

[6] A Concept for Future Military Operations on Urbanized Terrain
http://www.concepts.quantico.usmc.mil/mout/docs/moutfinal.PDF

[7] Report by the RTO Study Group into Urban Operations in the Year 2020
http://call.army.mil/products/spc_prod/nato/uo2020.htm

[8] Emerging Technologies and MOUT
http://www.sandia.gov/ACG/pubs-pres/MOUTpaper.pdf

[9] Task Force Ranger: A Case Study Examining the Application of Advanced Technologies in Modern Urban Warfare
http://www.at.y12.doe.gov/ranger.pdf

[10] Challenges of Urban Warfare as We Transform
http://www.ausa.org/www/armymag.nsf/(all)/6FE5198C73934EFD85256C1D00692549?OpenDocument

[11] Can We Fight In Cities?
http://www.urbanoperations.com/hills1.htm

[12] Soldiers in Cities: Military Operations on Urban Terrain
http://www.carlisle.army.mil/ssi/pubs/2001/cities/cities.pdf

[13] Urban Warfare and the Urban Warfighter of 2025
http://carlisle-www.army.mil/usawc/Parameters/99summer/hahn.htm

[14] Military Operations in Urban Terrain: A Survey of Journal Articles
http://www.urbanoperations.com/articlesurvey.pdf

[15] The Urban Awareness Concept
http://www.urbanoperations.com/urbanawareness.htm

[16] Technology and MOUT
http://www.rand.org/publications/MR/MR1007/MR1007.chap5.pdf

[17] Marquet, Louise and Ratches, James. Sensor Systems for the Digital Battlefield. In Digitization of the Battlefield II, Raja Suresh, Editor, Proceedings of SPIE-The International Society for Optical Engineering Vol, 3080, 1997

[18] Suresh, Raja editor. Battlespace Digitization and Network-Centric Warfare. Proceedings of SPIE-The International Society for Optical Engineering Vol, 4396, 2001.

[19] Ellefsen, Richard. Characteristics of Urban Terrain. Naval Surface Weapons Center. June 1979, San Jose State University.

[20] Peters, B.A.; Smith, J.S.; Medeiros, D.J; and Rohrer, M.W. Representation of Urban Operation in Military Models and Simulations. In Proceedings of the 2001 Winter Simulation Conference.

[21] Kumagai, Jean. Fighting in the Streets. IEEE Spectrum, February, 2001. pp. 68-71.

[22] Panel considers terrorist threat on U.S.-Canada border, CNN Web posted at: 2:43 a.m. EST (0743 GMT), *From Producer Terry Frieden* January 27,2000. http://www.cnn.com/2000/US/01/27/us.canada.border/

[23] U.S. Expands Container Security Initiative to South Africa. U.S. customs agents in Durban to screen cargo for terrorist weapons, U.S. says. http://hongkong.usconsulate.gov/csi/2003/120201.htm

[24] U.S. Borders Still Open to Nuclear Smuggling, NewsMax.com Wires, Oct 18 2002. http://www.newsmax.com/archives/articles/2002/10/17/160918.shtml

[25] Border Breach? Customs Fails to Detect Depleted Uranium — Again. http://abcnews.go.com/sections/wnt/Primetime/sept11_uranium030911.html

[26] A&E Border Patrol video

[27] CARGO SECURITY-A PARADIGM SHIFT, W. Gordon Fink Emerging Technology Markets http://www.senate.gov/~gov_affairs/120601fink.htm

[28] Sensors may aid in tracking cargo RAE Systems, container maker to run test of low-cost monitor, Andrew F. Hamm. http://www.bizjournals.com/sanjose/stories/2003/12/01/story2.html

4 Matching Scenarios with Sensor Systems Requirements

4.1 SDAC capabilities needed for WoT mission space

The four WoT mission domains provide a challenging collection of potential capabilities that must be applied across a vast number of potential sensor technologies. While we did not fully develop all four-mission domains, the initial sensor system requirements are based on high-level scenarios from each domain. Table 4 provides a large list of potential capabilities broken down by high-level sensor technology categories including mission, physical sensor, imaging sensors, environmental sensors, communication, tags, emplacement or mobility, power, control, data processing, networking, and algorithms. Each of these categories is decomposed into sample technologies, which are applied to five scenarios for the different mission domains. The binary yes/no answers can assist the sensor system developer into eliminating any initial misconceptions of capabilities verses application needs.

To examine these capabilities listed further we extended the concept to a set of current sensor systems to determine how the desired technology capabilities of existing systems compared to the WoT mission space. Table 5 contains the results of these comparisons for the Berkeley Motes, SteelRattler, Acousid III, and MIDS. While the existing systems appeared to match the physical sensor list, they did not fair well in the communication category. This lack of reliable and secure communication is also illustrated in the overall wireless field as a major issue that concerns may researchers in the wireless domain.

While this initial attempt at understanding the capabilities and requirements it only serves as a way to initially compare and reject obvious incapable systems. The results of these comparisons cannot determine which systems that passed these tables are better than others. What is needed are more in depth requirements and capabilities for a more realistic matching and metric system. This initial set of evaluations provided foundation for understanding potential parameters and issues that feed directly into the tradeoffs and metrics covered in Section 4.2 and Section 4.3.

Table 4: SDAC capabilities for mission space

SDAC Capabilities Needed For Missions					
Capabilities	MOUT snipers or friends	Seal Long Borders	Find SCUDS Before launch	Force/Facility Protection	ID Terrorists, WMD at Portals
Mission					
operational life	0.5 mo.	24 mo.	6 mo.	3 mo.	indef.
environment	outdoor	outdoor	outdoor	outdoor	in/out
real time or delayed	real time	real time	real time	real time	real time
covert, small size	yes	yes	yes	yes	no
persistent with wake up capab	yes	yes	yes	yes	no
Physical Sensors					
acoustic	yes	yes	no	yes	no
magnetic/EM	no	yes	yes	yes	yes
seismic/accelerometer	yes	yes	yes	yes	yes
meterological/weather	no	no	no	no	no
Imaging Sensors - important all apps					
optical imaging	yes	yes	yes	yes	yes
thermal imaging	yes	yes	yes	yes	yes
3-d optical radar	yes	yes	yes	yes	yes
penetrating mm radar	yes	no	no	no	yes
Environmental Sensors					
chemical/expl	yes	yes	no	yes	yes
biological	yes	yes	no	yes	yes
gamma/neutron	yes	yes	no	yes	yes

Communications					
local rf comm link	yes	yes	yes	no	no
long haul rf comm link	no	yes	yes	yes	yes
GPS	yes	yes	yes	yes	yes
LPI/LPS/authentication	yes	yes	yes	yes	yes
encription	yes	yes	yes	yes	yes
distributed array antenna	yes	yes	yes	yes	no
smart w/memory for delayed comm (covert)	yes	yes	yes	yes	yes
monitor and id rf comms across full spect	no	no	no	no	no
Tags					
passive or active rf	yes	yes	yes	yes	yes
chemical tags	yes	yes	yes	yes	yes
Emplacement/Mobility					
airdrop	yes	yes	yes	no	no
ground mobility	yes	no	?	yes	no
air mobility	yes	no	no	yes	no
Power					
low power comms	yes	yes	yes	yes	no
wake-up capability, maybe after 5 years	no	yes	yes	yes	no
low duty fact/rotating	yes	yes	yes	yes	no
energy mining/photovoltaics/RTGs needed	yes	yes	yes	yes	no
Data Proc,Networking, Algorithm, Cntrl					
for low false positives	yes	yes	yes	yes	yes
beam forming	no	no	no	no	no
rotating power off	yes	yes	yes	yes	no
reprogrammability, adaptability	no	yes	yes	no	yes
high level processing	yes	yes	yes	yes	no
biometric recognition	no	no	no	yes	yes

Table 5: Comparing mission domain capabilities with existing sensor systems

	Application Requirements					System Capabilities			
	MOUT	Force/Facility	Seal Long	ID Terrorists,	Secure Shipping				
Capabilities	snipers/friends	Protection	Borders	WMD at Portals	Contain./no WMD	Acousid III	MIDS	SteelRattler	Motes
Mission									
operational life	0.5 mo.	3 mo.	24 mo.	indef.	indef.	2 mo.	3 mo.	3 mo.	1 mo.
environment	outdoor	outdoor	outdoor	in/out	in/out	out	out	out	in/out
Physical Sensors									
acoustic (Basic)	yes	yes	yes	no	no	yes	no	yes	yes
acoustic (ATR/Tracking)	yes	yes	yes	no	no	no	no	yes	no
seismic/accelerometer	yes	yes	yes	yes	yes	yes	yes	yes	yes
magnetic	no	yes	yes	yes	yes	no	yes	yes	no
meterological/weather	no	no	no	no	no	no	yes	yes	no
chemical/expl	yes	yes	yes	yes	yes	no	no	no	no
biological	yes	yes	yes	yes	yes	no	no	no	no
gamma/neutron	yes	yes	yes	yes	yes	no	no	no	no
Imaging Sensors									
optical imaging	yes	yes	yes	yes	no	no	no	yes	no
thermal imaging	yes	yes	yes	yes	yes	no	no	yes	no
Communications									
local rf comm link	yes	no	yes	no	yes	no?	yes?	no	yes
long haul rf comm link	no	yes	yes	yes	yes	yes?	no?	yes	no
GPS	yes	yes	yes	yes	yes	no	no	yes	no
LPI/LPS/authentication	yes	yes	yes	yes	yes	no?	no?	no?	no
encryption	yes	yes	yes	yes	yes	no?	no?	no?	no
Emplacement/Mobility									
hand emplace	yes	yes	yes	yes	yes	yes?	yes?	no	yes
airdrop	yes	no	yes	no	no	no?	no?	yes	no
ground mobility	yes	yes	no	no	no	no?	no?	no	no
air mobility	yes	yes	no	no	no	no?	no?	no	no
Power									
wake-up capability	yes	yes	yes	no	no	no?	no?	no	yes
low duty cycle	yes	yes	yes	no	no	no?	no?	no	yes
energy mining	yes	yes	yes	no	yes	no?	no?	no	no

4.2 Fundamental Tradeoffs in Wireless Sensor Networks and Application Architectures

The multitude of WSN architectures creates problems as well as solutions. With many architectures suitable for each newly developed application, two major questions arise: which architecture is most appropriate for a particular mission?, and how are the existing architectures different from each other? The first issue essentially questions how to analyze the requirements of an intended application and match them with the capabilities of potential architectures. The second issue essentially questions what architectural tradeoffs can be made in WSNs. In order to address these issues, a fundamental parameterization of WSN design is suggested. The parameterization allows a quantitative analysis of application difficulty and architecture capability. It also provides a means to quantitatively determine the best-suited architectures for particular applications through a proposed matching metric. A demanding application parameterization that cannot be matched by existing architectural technologies may also elucidate necessary engineering developments. Furthermore, through several quantified relationships between the proposed fundamental WSN design parameters, tradeoffs in the WSN design space are explored. Ultimately, an understanding of WSN tradeoffs provides a common language which WSN architectural designers and WSN operational designers can use to develop robust and efficient applications.

Wireless sensor networks (WSNs) have been touted by many industry and military leaders to be a revolutionary technology that has potential to completely revamp the way we live our lives and conduct our business. One of the primary difficulties in WSN design is that the engineers do not understand the operational necessities or "killer app" scenarios, whereas the institution leaders do not understand the engineering capabilities and limitations. This leaves the industry leaders calling for technologically unrealistic systems, and the engineers building operationally unrealistic systems. In order to give a common ground on which the two sides can meet, this section discusses some of the fundamental tradeoffs in WSN design and operation.

There are several fundamental parameters that must be understood and addressed when designing WSNs, but power is generally considered the most important. Energy storage technology lags behind computational capability development, which results in the need for significant detail of thought to be given to power minimization and conservation in any wireless system [1]. The power consumption of a node is inversely proportional to node (and hence network) operational lifetime, thus power concerns are justified to properly address operational requirements. Power is controlled in many different aspects of a WSN from the individual node hardware and software to the networking protocols at several different layers of the networking stack. Many innovative approaches have been used in order to trade quality reductions in various WSN parameters for lifetime extensions. These parameters include performance, size, security, network bandwidth, network latency, network fairness, network reliability, network throughput, algorithmic decision accuracy, algorithmic and sensor data resolution, communications range, system flexibility, node cost, network cost (due to density considerations), and potentially many

more. The remainder of this section will discuss each of these parameters and how they affect each other and system operation as a whole.

4.2.1 Performance

The power/performance tradeoff is very well known, but difficult to specifically define. The main difficulty is that performance can take many different meanings. Better performance can refer to more accurate algorithms, faster computational speed, collection of higher resolution data sets, or a number of other metrics. This ambiguity does not allow for a formal discussion of the tradeoff unless a specific performance metric is chosen. Since it is not otherwise listed in the parameters to be discussed, faster computational speed is the measure of performance discussed here. Although the power consumption of a WSN node is generally due mostly to the radio, the power of high-speed processors used to perform complex data manipulation can be a significant power drain as well. The power of a processor is made up of two main components, one component is a power overhead, and the other is the power that can scale with processor frequency. The power overhead is a static power that results mainly from memory (though sections of the memory can be disabled in modern processors), I/O, and support circuitry. The scalable power results from capacitive loads on switching transistors. As the frequency of transistor switching increases (i.e. as clock speed increases), power consumption increases proportionally. A fundamental, simplified equation for a single transistor switching a capacitive load, C, at a 50% duty cycle frequency, f, and supply voltage, V, is:

$$P = fCV^2$$

Processor design can significantly impact the actual total processor power consumption, but the linear relationship between scalable power and processor frequency holds generally. A scale factor of about 1mW/MHz is a common specification for power aware processor designs. In other words, a 1GHz processor will consume about 10 times more power than a 100MHz processor when both are running at full speed. As processor speed gets lower, this linear relationship breaks down since the power overhead becomes a more significant proportion of the total power consumption.

As an example of the impact of power versus performance Table 6covers potential tradeoff issues.

Table 6: Power verses performance tradeoff table

Processor	Cygnal 8051	XScale PXA250
Speed	25MHz	400MHz
Power Consumption	20mW	400mW
Lifetime on 2 AA Batteries	750 Hours	37.5 Hours
	Just over 1 month	Under 2 days

As can be seen, the lifetime of a system can be affected immensely by an order of magnitude change in processor speed. This is an extremely important factor to consider when determining what type of performance is really necessary in a network.

Other: As performance increases, algorithmic decision accuracy, algorithmic resolution, system flexibility, and system cost also increase. Higher performance processors can handle a wider range of tasks and more complex, higher order algorithms, but they come at a higher cost. For comparison, a 1.6GHz Intel Pentium 4 is about $200, a 40MHz Rabbit Semiconductor RCM3400 is about $40, and a 4MHz TI MSP430 is about $1.

4.2.2 Size

As the power consumption of a node increases, its size must increase also if the node lifetime is to remain unaffected. This relationship derives from the fundamental chemical limits in energy storage technology. For conventional commercial batteries, the theoretical upper limit of energy density is 300Wh/kg. [2] Lithium ion batteries, popular high energy density choices, have about 200Wh/kg. From a volume standpoint, common energy densities for lithium ion batteries are about 1000Wh/L or 1Wh/cm^3. WSN platforms generally use on the order of 100mW to 1W of peak power, but power consumption greatly depends on the types of sensors needed. Additionally, peak power consumption is rarely used since the power consumption of high power components, such as radios or high-resolution imagers, is heavily duty cycled. As an example of what all these considerations imply, at 100mW average power (a moderate power, moderate performance system), a WSN node would require a *theoretical* minimum of 2.4cm^3 of battery volume weighing 8g for every day of operation. For a three-month mission, this node would require a *theoretical* minimum of 216cm^3 of battery volume weighing 720g, or a 6cm-sided cube battery. The following formula will give the theoretical minimum battery volume required where V is the volume in cm^3, P is the average power consumption of the node in Watts, H is the number of hours of operation required, and D is the battery type density in Wh/L:

$$V = 1000\frac{PH}{D}$$

Other: As size increases physical covertness and thus security decreases. At a cost of about 2.4¢/Wh or at least 2.4¢/cm^3, as battery size increases, node and network cost will also increase. Size increases also mean that larger antennas and lower transmission frequencies can be reasonably used which results in increasing communications range.

4.2.3 Security

The relationship between power and security is somewhat hard to formally determine since levels of security are not generally quantifiable. An additional complication is that security can take multiple forms since hardware, software, and network security each required for a robust WSN. Network communications security is the most WSN specific, whereas hardware and software security are issues general to all computing systems, and so network security is what will be discussed here. There are several different security issues that can be addressed for WSNs including authentication, encryption, anti-spoofing, and anti-jamming measures. Network authentication requires that each node prove its right to communicate on the network. This can

be done only when a node joins the network, or instead could occur each time the node needs to communicate with the rest of the network. Out of the four security issues listed, authentication takes by far the most power since it is entirely a wireless overhead. More frequent authentication increases both security and power consumption.

Encryption is another major issue that can significantly affect the energy per bit of transmissions, but only over short distances (<~20m). Encryption requires more computational time and therefore more power than sending unencrypted messages, and this processor usage can be a significant addition when using extremely low power, short range radios. For example, the energy per bit of communication over short distances (<~20m) may be on the order of 500mW, whereas the energy per bit of a 128-bit RSA encryption and decryption is on the order of 150uJ in a StrongARM processor. [3] (At 50m, transmission energy goes up 10 fold, and encryption energy becomes insignificant. ElGamal encryption requires almost 10 times the power of RSA. The energy per bit changes with packet length – these results hold for packets of about 100 bytes. Longer packet length would decrease the impact of the encryption since the communications energy per bit asymptotes.) [4] Additionally, the length of the key used for encryption scales the power consumption of encryption approximately linearly, so power can be reduced by reducing the key length, and hence reducing the security.

Anti-spoofing and anti-jamming measures are meant to ensure that an outside observer would be unable to mimic a nodes' operation by retransmitting packets, replacing a friendly node with one of his or her own, or attempting to create communications noise in order to prevent system operation. These security measures are essentially implemented in the authentication and encryption already in addition to robust messaging routines, adaptable routing protocols, and spread spectrum physical layer networking technologies. Messaging routines can affect power consumption if they increase chatter in the network. For example, simply requiring an acknowledge (ACK) from a receiver would almost double network power consumption for fixed length message structures or short variable length messages. As will be discussed below, routing protocol overhead can also have large power consumption impact when network traffic is low. Finally, physical layer changes to the networking stack can drastically affect power requirements. Traditional spread spectrum radio uses more power than single frequency radio, but is also more secure against eves dropping, jamming, or spoofing. Complex types of spread spectrum, such as Orthogonal Frequency Division Multiplexing (OFDM), increase security further, but again by exacting a power cost. A promising new technology that breaks this security at the price of power mold is Ultra-Wide Band (UWB). UWB uses extremely short radio pulses to spread the frequency of its transmission out over a wide frequency swath (>~200MHz). This makes the transmission harder to jam, harder to intercept or detect, and also far lower power. Short-range UWB radios have been demonstrated in the nano-watt range.

Other: Increases in security have little affect on other network parameters besides power. Communications latency will be slightly increased across the network, but only by the amount of time required to encrypt and decrypt the message (potentially at each hop). The computational speed of the nodes might be slightly lower per data point as well since security measures implemented in the code might slow it down also. There is a potential for network throughput to be reduced if security protocols are not designed efficiently, with a secure connection between

two nodes tying up network traffic for extended periods of time, but this affect should be able to be minimized.

4.2.4 Network Bandwidth

Bandwidth is defined as the width of the frequency range used by or possible with a particular communications link. For example, AM radio uses 10KHz bandwidths since each channel (station) uses a center frequency +-5KHz to encode audio signals (e.g. 1020KHz-1010KHz = 10KHz); FM radio uses 200KHz bandwidths (e.g. 94.5MHz-94.3MHz = 200KHz); Ultra-Wide Band (UWB) technology uses on the order of 200MHz bandwidth theoretically allowing it to encode 20,000 times as much data as AM radio and 1,000 times as much as FM radio. A frequency hopping radio uses a bandwidth equal to its highest used frequency minus its lowest used frequency, even though communication actually only takes place on single frequencies.

The bandwidth of an RF channel is related to the transmission power required through the signal to noise ratio, S/N:

$$B \propto \frac{P_{tx}}{\left(S/N\right) r^2}$$

where P_{tx} is the transmit power, B is the bandwidth, and r is the transmission distance. All communications systems have a minimum allowable S/N for successful transmissions, and thus for a given maximum range and transmission frequency, the power and bandwidth are linearly related. In other words, if a 1Mbit/s communications system consumes 1W, a similar 10Mbit/s system will consume 10W. The amount, complexity, and sample rate of data required from a WSN directly relates to bandwidth needed and thus it also has a direct linear impact on the power and lifetime of the system.

Except for cameras, most processors and simple sensors consume less than 200mW of power, so the radio, commonly operating at peak powers of at least 1W, is generally the largest power consumer. As a result, many techniques have been tried in order to reduce its power consumption by trading bandwidth. One method of bandwidth reduction is to perform data fusion and data analysis on the nodes themselves and only transmit information rich messages. Unless raw data needs to be correlated amongst many different nodes, this on-node processing can drastically reduce bandwidth, and hence power requirements.

Two other techniques of power reduction are based on an analysis of the type of wireless traffic common in WSNs. The first technique takes advantage of the fact that there are two main types of messages in WSNs: those containing sensor data and those used for topology management. The data messages generally require considerably more bandwidth than the control messages since control messages can be made extremely short and efficient. Using this observation, Feng [1] has suggested using two radios to separate the data and control channels. Unless data needs to be streamed out of the network constantly, data transmissions are usually much more infrequent than control transmissions. Using a low bandwidth, low power radio for the control communication allows a high bandwidth, high power data radio to remain off for the

majority of the node's lifetime. Feng's simulations show between a 50% and 80% reduction in radio power consumption using her bandwidth to message matching scheme.

The second technique is called Sparse Topology and Energy Management (STEM) developed at UCLA. STEM takes advantage of the observation that a node does generally not need continuous reception. Therefore, instead of continuously monitoring a communications channel with a high bandwidth, high power radio, this function can be accomplished by a low bandwidth, low power radio. If a transmission does need to be received, the low power radio will turn on the high power radio to receive it. The savings depend on the amount of time spent in the monitoring state versus the transmitting state and also on the length of transmissions, but for 500ms transmissions, simulation results show a factor of 2-power reduction.

Other: As bandwidth increases, network latency decreases correspondingly. Bandwidth directly relates to the amount of data that can be shipped per time period and faster transmissions mean that data can propagate faster thus decreasing network latency. As can be seen in the bandwidth formula above, if all other communication system factors remain equal, as bandwidth requirements go up, communications range decreases. For higher bandwidth necessities, node density might therefore have to be higher which increases total network cost. Finally, higher bandwidth systems are generally more expensive, so as bandwidth requirements are increased, individual node costs will also increase.

The bandwidth of a wireless link depends greatly on the type of wireless transmission method used. Although wireless communication is generally associated with radio frequency (RF) transmission, there are two other notable wireless communications schemes developed for WSNs: ultrasonic and ultraviolet (UV). Ultrasonic communication can be more secure and covert than RF because it is more directional and harder to effectively jam. It has far less bandwidth, however, is severely limited in transmission distances being effective to at most about 30 meters line-of-sight (LOS), and is blocked by most materials. UV communication is in early stages of development and works on the principle that air scatters UV rays. Transmitters emit a conical beam straight up, and with a receiver pointing straight up also, reception has already been demonstrated at up to 10 meters. The details of the channel characteristics are currently under research, but UV communications will likely have similar restrictions to ultrasound except the bandwidth will be considerably higher. Ultrasonic and UV communications are only for very local transmissions, but in short-range links, they use less power than traditional RF methods.

4.2.5 Network Latency

Network latency trades off with power primarily due to (Medium Access Control (MAC) and routing networking methods. The MAC layer is responsible for arbitrating access to the wireless channel used by the network. Three main MAC classes can be identified: static access, dynamic access, and random access. The routing layer is responsible for transferring data packets from node to node throughout the network. Two main routing classes can be identified: proactive and reactive. The power consumption associated with a protocol is associated with the communications overhead required to run the protocol and any additional resource support needed from the system. As will be discussed, at the MAC layer, the network latency generally varies inversely with communications overhead for high traffic conditions or multiple packet

messages and directly with communications overhead for low traffic conditions or single packet messages. At the routing layer, the network latency varies inversely with communications overhead. Since the wireless link is the most power intensive resource in WSN nodes, the amount of overhead of the protocols is a very significant factor and can make a large impact on node lifetime.

Static access MAC protocols rely on a fixed schedule of access in which each node has a communications time slot assigned to it. (In order to make these protocols scalable, nodes are generally organized into clusters. Clusters communicate with each other either through nodes on cluster boundaries, or between elected cluster heads.) Since a node may only transmit during its slot, the best case latency of single packet messages associated with static access protocols is greater than for other methods. Since the throughput of static access protocols is very high compared with other methods (33.3% of physical layer channel bandwidth for ideal Time Division Multiple Access (TDMA) [5], however, multiple packet messages or high traffic situations will yield a lower average latency than with other protocols. During periods of low network traffic, static access protocols have a power overhead greater than other protocols since they must continually manage the communications schedule as nodes come in and out of connectivity. Since the nodes are synchronized, however, they may turn off their transceivers on a periodic basis (duty cycling) to reduce the average power consumption thus compensating for this overhead increase, but also increasing the total latency of messages in the network. The benefits of static access protocols are that they are fair across all nodes, they allow a high per node throughput compared with non-static methods, and they maintain a high degree of network reliability. Static access protocols are best suited to networks in environments where there is expected to be high traffic distributed evenly across the network. An example of such a network would be a distributed data collection system.

Dynamic access protocols are more data-centric than static access protocols in that they allow a node access to the channel when the node requires access in order to send a message. There is no master schedule that is kept by the nodes, instead nodes must request the channel, be assigned a slot, and then transmit in that slot on a one time only basis. (Alternatively, a token passing scheme may be used in which nodes pass a token between each other signifying that the token holder has current access to the channel.) The power overhead of these methods are therefore low during the initial phases of network organization, but can become much higher than for static access protocols when network traffic is high. Additionally, either a channel arbitrator must be assigned or a distributed arbitration method developed in order to manage network requests and slot assignments. Network latency in a dynamic protocol can be unbounded if a data prioritization scheme is used, but in general, for low traffic, non-prioritized situations, network latency will be lower than for static access protocols. The latency is non-deterministic, however, as it will depend on the specific network traffic, and this indeterminism can be a major disqualifier if a real-time system is needed. Turning off the radios periodically as in the static access situation can compensate for the overhead associated with the access scheduling, but again this will increase total message latency. The benefits of dynamic access protocols are that they can give prioritized access to nodes with important data (reducing network fairness), they require no background periodic power overhead (e.g. schedule maintenance), and they maintain a moderate to high degree of network reliability. Dynamic access protocols are best suited to networks in environments where there is expected to be low, sporadic, and unevenly distributed

traffic, but where a high reliability must be maintained at the cost of some power overhead. An example of such a network would be an event detection system monitoring an area with highly important fine-grained information. Dynamic access protocols are also well suited to environments where there may be node mobility or frequent changes in node connectivity. An example of such a network would be a mobile robot mounted sensor system or a sensor system in a military operations in urban terrain (MOUT) environment.

Random access protocols require the least overhead of all the MAC schemes. Accordingly, for low traffic conditions, they also have the lowest single packet latency. Random access is so named because the access to the channel is done at random without the use of any global arbitration or schedule. The most basic random access method is pure ALOHA. Under pure ALOHA, nodes simply transmit packets whenever they have a packet to transmit. This can result in contention and collision, but there is absolutely no overhead involved in the scheme. Slotted ALOHA increases the probability of successful transmission (but also increases the protocol overhead) by allowing communications to occur only in globally synchronous time slots. In order to avoid using the channel at the same time as another node, a collision avoidance scheme such as Carrier Sense Multiple Access (CSMA) or the incorporation of Collision Avoidance as well (CSMA/CA) (802.11 is based on CSMA/CA) may be used. In CSMA, a node will listen to the channel before it transmits to see if the channel is available. If no other node is transmitting, it immediately transmits. If another node is transmitting, it performs a contention resolution algorithm which may involve methods such as listening until the transmission is complete, or waiting a random amount of time before trying again. CSMA/CA adds a Ready-to-Send/Clear-to-Send (RTS/CTS) transaction between nodes effectively muting all potentially contending nodes within the region until the message transaction is complete. (This is described in more detail in the Network Reliability, Quality of Service section.) The only overhead in these more advanced types of random access is during the carrier sensing or RTS/CTS phase of a transaction. The latency of random access protocols is again non-deterministic, but in low traffic networks, single packet latency will be the lowest of any of the schemes. Since the throughput of random access methods is far lower than scheduled access protocols (4.19% of physical layer bandwidth for slotted ALOHA, 7.7% of physical layer channel bandwidth for ideal CSMA, 5.5% of physical layer channel bandwidth for ideal CSMA/CA [5], the latency of multiple packet messages may be higher than for other methods, however.

Since random access protocols have no global arbitration, they are not as reliable as the static or dynamic access protocols unless additional overhead is expended. High node density to give sensing redundancy helps solve this problem, but decreases throughput and thus multiple-packet or high traffic latency even more. In general, random access protocols are not as robust as schedule-based schemes. Another issue is that since the nodes are completely unsynchronized, their radios must remain continually on since it is always unknown when any other node will make a transmission. This is compared with both the dynamic and static cases where the synchronization allows nodes to schedule periods in which radios are turned completely off to reduce average power consumption. As a result, in most network traffic situations, random access protocols will require more power than static or dynamic access protocols. Random access protocols are best suited to networks in environments where there is expected to be low, sporadic, and unevenly distributed single packet traffic, and where reliability can be guaranteed by sensing redundancy as opposed to communications robustness. Random access protocols are

also best suited to environments in which power and lifetime are not of utmost concern, but single packet latency must be as low as possible. An example of such a network would be an event detection system monitoring an area with simple coarse-grained, extremely time sensitive information.

A simulated performance comparison of different MAC schemes is given in [6]. As can be seen, the throughputs given here are much higher than the theoretical values given above. The reason for this is a difference in definition. In [6], throughput is defined as the ratio of successfully transmitted packets to the channel transmission rate (in packets per second) over the entire network. In the theoretical values given in [5], throughput is defined as the ratio of number of successful transmissions per frame per node to the length of the frame. (A frame is the period (measured in slots) of repetition of transmission slots.) The reason for this difference in definition is that in [6], simulations were conducted for a fixed number of nodes all using the same channel and all in range of each other, whereas in [5], theoretical maximum data transfer per node in a multi-hop environment where not every node is in range of each other was the purpose.

Some of the findings of [6] are given in the Figure 6.

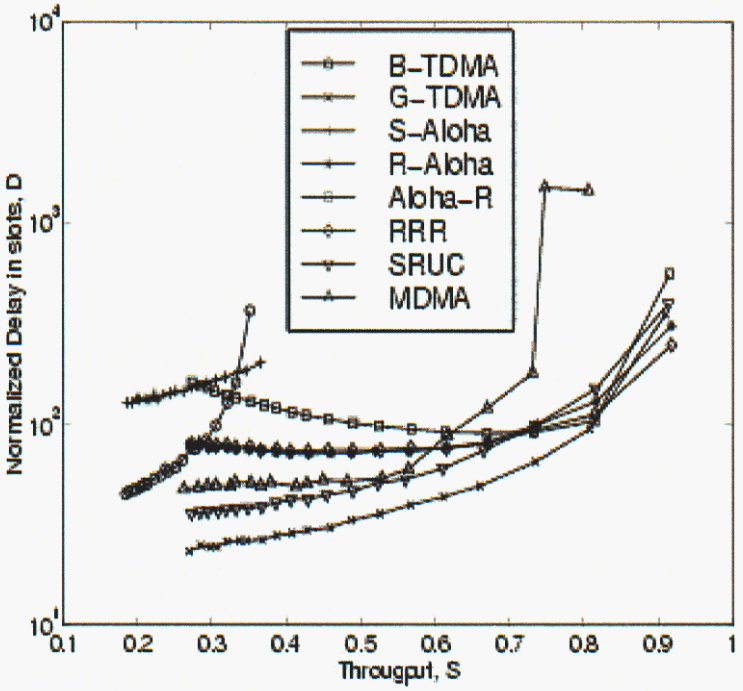


Figure 6: Throughput vs. Delay for several MAC protocols
(Message burst lengths of 5 packets where 1 slot allows 1 packet to be sent.)

Table 7: Characteristics of MAC protocols

MAC Protocol:	Characteristics:
B-TDMA (Static)	Basic-TDMA. TDMA as described above
G-TDMA (Static)	Generalized-TDMA. Width of slots adjusted to bandwidth requirements of individual nodes. Widths are pre-determined and static.
S-ALOHA (Random)	Slotted ALOHA. ALOHA in which nodes may only transmit during specific time slots rather than at any asynchronous time.
R-ALOHA (Hybrid of random and dynamic)	Reservation – ALOHA. Starts as S-ALOHA, but if a node has a successful transmission in a given slot, it continues to transmit only in this slot, i.e. moves towards B-TDMA.
ALOHA-R (Hybrid of random and dynamic)	ALOHA-Reservation. Uses S-ALOHA to reserve slots (during a preliminary slot) and then transmits its message only during successfully reserved slots.
RRR (Hybrid of random and dynamic)	Round-Robin Reservation. Slots are initially assigned using B-TDMA. Unused slots are given to other nodes that require them in a round-robin fashion.
SRUC (Adaptive)	Split-Channel Reservation Upon Collision. Starts as S-ALOHA, but if there is a collision, transitions to G-TDMA. When all messages have been sent, it transitions back to S-ALOHA.
MDMA (Adaptive)	Minimum Delay Multiple Access. Nodes transmit in slots with a certain probability, f, that is dynamically adjustable and also make a secondary reservation. If f=1, this becomes S-ALOHA, if f=0, this becomes a reservation based dynamic access.

The findings in [6] concluded that for low traffic (termed there low throughput) conditions, random access protocols gave lower latency (termed there delay) characteristics, whereas for moderate to high traffic conditions (termed there moderate to high throughput), schedule-based protocols (either static or dynamic) gave lower delay (termed there delay) characteristics. As can be seen in Figure 6, G-TDMA gave the best latency characteristics, but it relies on the assumption of a completely static network, traffic pattern, and network topology. As a result, ultimately recommended SRUC as a best general solution since it dynamically adjusts itself to traffic conditions, and can adapt also to topology changes. The findings in [8] did not analyze communications overhead of the various protocols, it only analyzed complexity of on-node computation required to manage the protocols.

Since power overhead is of utmost concern in WSNs, the overhead of various MAC protocols have been analyzed in several reports. In [7], results showed that R-TDMA (Reservation-TDMA, much like ALOHA-R) gave significantly lower overhead than S-ALOHA in all but low traffic conditions. In other words, a random access protocol requires less overhead than a dynamic access protocol for low traffic conditions only. The reason for this is mostly due to contention in the random access protocol when traffic increases. The reprint of **Figure 7** show the total power dissipation of the two MACs under varying loads [7]. (A load of .5 indicates that a node will attempt to transmit a packet in 50% of the slot periods.) Note that the number of nodes in the random access protocol (S-ALOHA) affects the total power dissipation because of possible contention issues and a need to retransmit when collisions occur. In the dynamic access protocol (R-TDMA), the number of nodes does not affect the total power dissipation because contentions are avoided due to the channel reservation process.

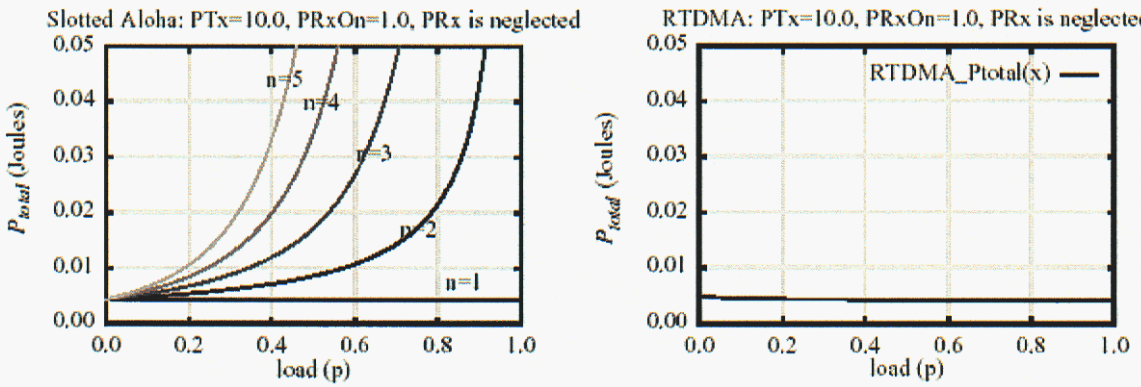


Figure 7: Total Power Per Packet Required by S-ALOHA and R-TDMA MAC Layers

Another paper that analyzed MAC power overhead [8] found similar results. This paper analyzed 802.11 (CSMA/CA, a random access protocol), PRMA (Packet Reservation Multiple Access, much like R-ALOHA, a dynamic access protocol), MDR-TDMA (Multi-services Dynamic Reservation TDMA, much like ALOHA-R, a hybrid protocol), DQRUMA (Distributed Queuing Request Update Multiple Access, much like RRR using ALOHA rather than B-TDMA for reservations, a hybrid protocol), and EC-MAC (Energy Conserving Medium Access Control, much like ALOHA-R using B-TDMA instead of ALOHA for reservations, a hybrid protocol). The results showed that for a network of only a few nodes or under low traffic, PRMA (a random access protocol) had the lowest power consumption, but for a higher node count or moderate to high traffic loads, the reservation based protocols (especially EC-MAC which is completely schedule driven and contentionless even in the reservation phase) gave lower power consumption. Interestingly, 802.11, the other random access protocol tested, had the highest power consumption of any protocol under low traffic loads, but had lower power than PRMA at high traffic loads. The reason for this higher power for 802.11 at low traffic loads is because it must perform collision avoidance before each transmission. At high traffic loads, 802.11 has a lower total power than PRMA because it requires less retransmission of messages than PRMA does. The reprinted graphs shown in Figure 8 and Figure 9 indicate the total power used for transmission and reception per node for different numbers of nodes in the network for the different MAC protocols tested. The research also showed that for multiple packet messages, the

effects of traffic load were amplified, but that the same general power consumption ordering and trends were maintained.

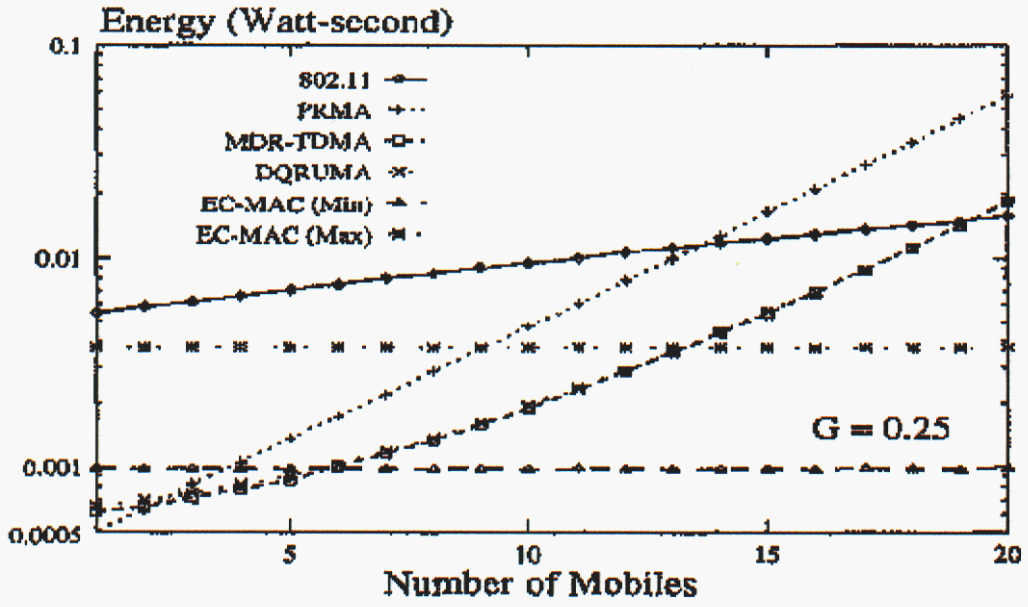


Figure 8: Required Transmission Power Per Node Per Packet (Single Packet Messages)

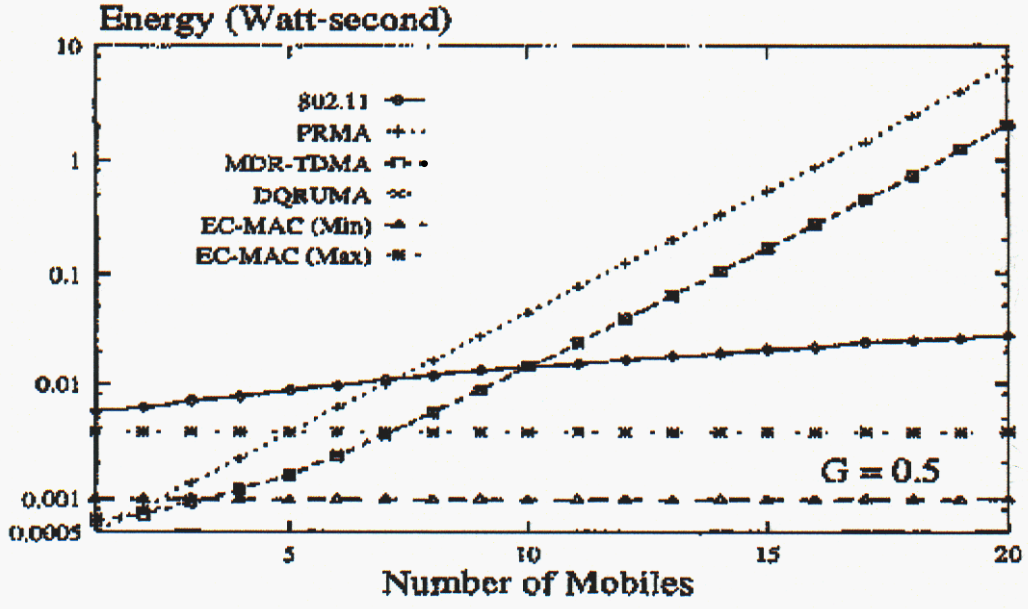


Figure 9: Required Reception Power Per Node Per Packet (Single Packet Messages)

At the network layer, routing protocols are also a means by which latency may be traded with power consumption. There are two general classes of protocols with relevance to this

tradeoff: proactive and reactive. Proactive protocols maintain network routes throughout the course of network lifetime even if the routes are not intended for immediate use. Reactive protocols establish network routes only when they are intended for immediate use, and routes are not maintained over network lifetime. Proactive protocols give lower latency than reactive protocols since routes are already established whenever a node needs to send information. The power overhead that reactive protocols suffer to establish a route before each transmission is generally far less than the power overhead incurred by the periodic maintenance of proactive protocols, but this depends on the network traffic, the time period of route maintenance, and the mobility of the nodes among other factors. In general, for low traffic networks or networks in which there is high node mobility, reactive protocols are preferable for the power savings they give if the slight decrease in latency (involving only the time it takes to establish a route) can be tolerated. When latency is of utmost concern, a proactive protocol should be chosen.

Several comparative studies have been performed of different routing protocols. One study that illustrates important differences between the four most commonly studied protocols is given in [9]. The protocols analyzed were Destination-Sequenced Distance Vector (DSDV), Ad-Hoc On-Demand Distance Vector (AODV), Dynamic Source Routing (DSR), and Temporally-Ordered Routing Algorithm (TORA). DSDV is the only one of these protocols that is proactive. AODV and DSDV are both based on the concept of a distance vector, which keeps track of the number of hops away an intended receiver node is. The path that is ultimately chosen will use the shortest possible path through the network by using these hop distances to direct the routing. AODV generates these distance vectors "on-demand", i.e. only when a route is required. AODV works by flooding the network with a route request and propagating a shortest path back to the sender once the receiver hears the request. Each node along the path that receives the request updates a hop counter in the request and thus the receiver can ultimately determine which route is the shortest path. This shortest path is then chosen and stored on the intermediate nodes so that no single node knows the entire route, just the next node in the chain. This route information is stored in a table on each node. If a neighboring node is lost, a node will erase all dependencies in its routing table on the lost node, but this change is kept local and not propagated. DSR works by flooding the network and tracing the shortest path to the requested receiver, much like AODV, but unlike AODV, the source node learns the entire path to the receiver and intermediate nodes do not store any state. The route is then sent along with the message to tell intermediate nodes to which subsequent nodes they need to route the message. TORA, unlike all of the other protocols, sacrifices the guarantee of a shortest possible path for a gain in time of path generation. It finds multiple paths to a receiver by creating a directed flow graph through the nodes. Intermediate nodes add an incremental cost to paths based on a metric (hop distance, power remaining, etc...), and advertise these costs to the source. The source picks a path with least cost. TORA is also locally proactive in the sense that if it loses a link to a neighbor, any routes depending on that neighbor will be erased. Additionally, an update will be sent to the network to erase dependency on nodes upstream from the lost node to the particular receivers it was routing to. This proactivity is unlike DSDV because it is localized to the lost node, whereas in DSDV, routes throughout the entire network are periodically updated.

The following two graphs shown in Figure 10 and Figure 11are taken from routing/routing protocol [9]. The x-axis of each of these graphs represents a degree of mobility based on what is known as the random waypoint mobility model. Without going into details of the model, a

"pause time" of 0 seconds indicates constant motion, a pause time of 900 seconds indicates no motion, and a pause time in between represents intermittent motion. These simulations were done for 20 nodes, and the results indicate the network wide totals over a full 900-second simulation.

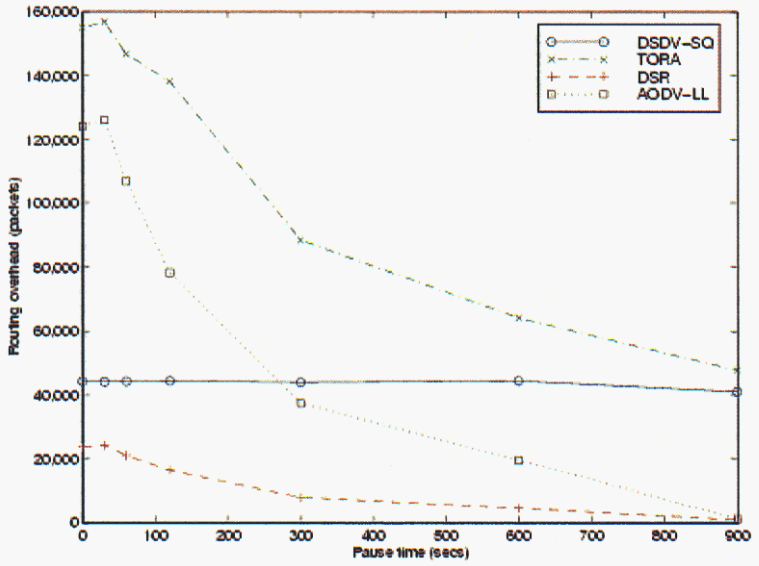


Figure 10: Routing overhead in packets for four routing protocols in a mobile network

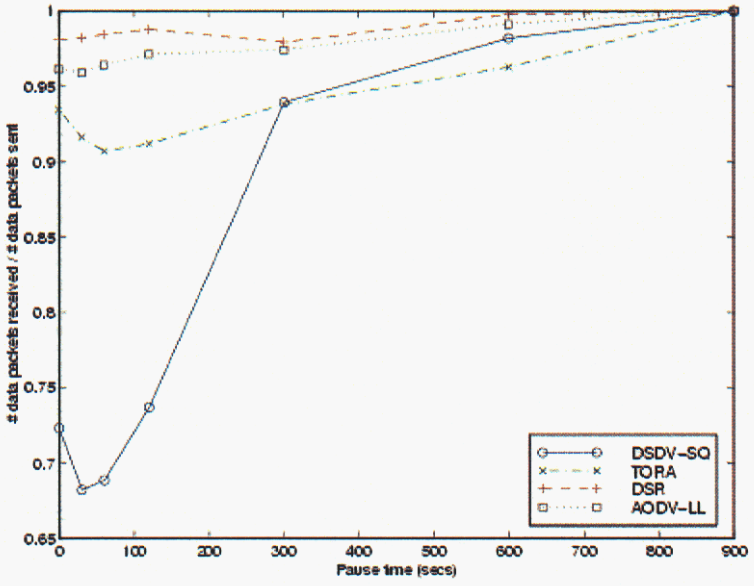


Figure 11: Successful packet reception rates for four routing protocols in a mobile network

The most notable observations of these simulations for simple, static sensor networks is that in the immobile case (i.e. pause time = 900 seconds), all of the routing protocols gave fully successful communication, and the fully reactive protocols, AODV and DSR, gave by far the least overhead. Of these two, other simulations not given here showed that DSR required the least overhead in total packets whereas AODV required the least overhead in total bytes. As will be discussed shortly, power consumption per transmitted bit (i.e. per transmitted byte) depends on packet length. For short packet lengths, the power consumption can change drastically, thus the determination of which protocol is truly the least power consuming requires further investigation.

One other simple power conservation method used widely in WSNs is radio duty cycling. Since radios consume large amounts of power relative to other system resources, even when they are in idle or receiving modes, they must be put into a low power sleep whenever possible in order to extend network lifetime. In order to ensure neighboring nodes have their radios on at the same time, a radio power management schedule must be disseminated into the network requiring a periodic power overhead. This overhead is vastly outweighed by the benefits of even a small duty cycling, and should be of minimal concern in the network power consumption over its lifetime. Radio duty cycling is most applicable to static and dynamic access protocols since network synchronization is already required by these two methods. If an overhead is going to be consumed to synchronize and duty cycle the radios throughout the network, there is little sense in using only a random access protocol. Radio duty cycling can give extreme power benefits, but it can also degrade network latency severely. Essentially, if radios are on only 10% of the time, the worst case latency can increase by a factor of 10, but the overall power consumption of the system may reduce by almost a factor of 10 (since the radio takes up most of the node's power budget). Duty cycling therefore gives a method of making power and latency approximately inversely proportional and can be used in conjunction with MAC and routing schemes to optimize a power versus latency tradeoff. Since the optimal tradeoff between duty cycle, MAC protocol, routing protocol, and power consumption depends on network traffic patterns, there is a suggestion throughout the research community to develop dynamically adjustable full communications system management algorithms, but thus far, no results have been published.

A final power versus latency tradeoff is affected through packet length. For networks in which data is complex enough so that multiple packets need to be sent for a single transmission, longer packet length is actually more power conservative as well as latency reducing. For a given message, fewer packets will be needed if a longer packet length is used. It is apparent that sending fewer packets will yield a lower latency since less time will be required to deconstruct and reconstruct messages. Although initially reducing the number of packets would not seem to decrease the overall power per bit necessary for communication, the key observation is that the power overhead associated with the header, and also the power overhead associated with the start-up time of the radio, will be needed less for fewer packets. This overhead can be considerable in certain routing schemes where the full route information is sent along with each message. When fewer packets are sent, the header overhead is needed less, and power is thus decreased. The power per bit of a transmission reaches an effective asymptote as packet length increases, however, so network fairness can still be maintained through packetizing without losing the power benefits. Figure 12 is taken from [4]. The particular parameters used to generate

this graph are unimportant; the general observation of the energy per bit dependency on packet size is the important point to be noted.

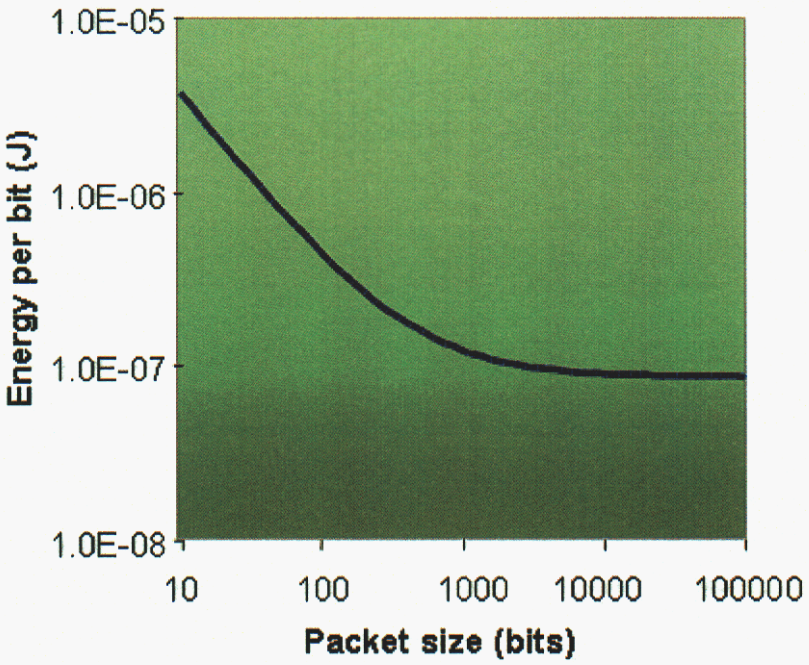


Figure 12: Example communication energy per bit versus packet size

Other: Changes in latency requirements can also affect other aspects of the WSN. Security and latency are directly proportional to each other. As security is enhanced through measures such as authentication, network latency will increase correspondingly due to the increased communications overhead. For similar reasons, as reliability is enhanced through measures such as CTS/RTS or message acknowledgements (ACK), overall latency is increased also. As eluded to in the above paragraphs, fairness and latency are also directly proportional to each other. As nodes are given more of an equal opportunity to communicate over a channel, more network management through message scheduling is necessary, and as the schedule becomes less flexible, in order to give equal opportunity to each node, the latency of data communications from an individual node increases. For the same reason that fairness and latency are proportional, throughput and latency are also proportional. As schedules become less flexible, the throughput from individual nodes increases because a node's slot cannot be overridden by another node. Finally, bandwidth and latency are inversely proportional. As bandwidth goes up, more data per unit time can be sent, and thus latency will go down.

4.2.6 Network Fairness

Network fairness can be an extremely important issue when multiple nodes each have mission-critical data to transmit. If a network is tracking enemy troop movement in an urban terrain environment, for example, many parts of the network may be collecting vital data simultaneously. Fairness trades with power in that it takes more inter-node collaboration to allow each node similar access to the wireless medium. Fairness is thus a direct product of the type of

MAC used: static access, dynamic access, or random access. As more power is used to communicate and synchronize schedules between nodes, fairness of access to the channel increases.

In a random access MAC, there is no global arbitration of the medium. If a particular node accesses the wireless channel, and maintains a need for it, that node can dominate control of the channel for as long as it chooses. Random access thus gives minimal fairness across the network. In order to combat this difficulty, many random access protocols implement a self-arbitration, so that if a node has controlled the channel for over a certain amount of time, it must give up the channel and wait some delay before attempting to access the channel again. This self-arbitration scheme is helpful in eliminating the node domination issue, but it still allows no guarantee that a node will ever have access to the wireless channel if it is in a high traffic cluster. As was described in the Network Latency section, the lack of inter-node communication and collaboration makes random access MACs the least power intensive.

Dynamic access MACs are fairer than random access MACs because nodes can schedule access to the channel. Although this scheduling scheme guarantees nodes access to the channel, the length of time a node must wait is non-deterministic since any number of nodes may be in the schedule at a time. In a dynamic access system in which certain nodes or certain data can be given prioritized access, the non-determinism problem is even worse. However, since a primary driver for network fairness is that high priority data be allowed access from any node fairly, a prioritized access scheme based on data is a decent solution to the fairness issue. During initial phases of network organization, dynamic access schemes take more power than random access schemes because they have to synchronize the nodes. As data traffic on the network increases, dynamic access schemes are the most power intensive of any MAC since overhead must continually be expended to schedule channel access. (The scheduling can be done either on a separate wireless channel or, as is most common, in a specified time slot on a periodic basis.) Thus for an increase in network fairness over the random access methods, a power increase will be incurred as well.

Static access MACs are the most fair of any of the MAC schemes. They give each node a static slot assignment, which recurs on a periodic basis. This slot is the only time during which a node may ever transmit, but as a result, it is guaranteed a deterministic wait time for access to the channel. The power consumption of static access protocols includes an initial synchronization overhead when the network first organizes, but in non-mobile, robustly connected networks, this is the only additional power required. Thus static access protocols may provide the highest fairness to power ratio of any of the methods. Although, the power consumption of static access protocols does not depend on network traffic, it does depend on node mobility and connectivity. If the connectivity of a network changes often, the nodes must continually reestablish an access schedule, which could greatly increase power consumption. From a fairness and power perspective, a static access protocol is thus preferable over a dynamic access protocol if connectivity is stable and mobility is negligible, but if these conditions do not hold, a dynamic access protocol should be chosen instead.

Other: Since network fairness directly relates to what type of MAC layer is used in the networking stack, only those factors that are also directly related to MAC tradeoff with fairness.

Latency is one major system parameter that is affected by the MAC layer as described in the previous section. As latency goes up when a more static schedule based MAC is used, fairness also increases, so latency and fairness are directly proportional. The MAC layer also directly affects the throughput of the system. As protocols become more static schedule based, throughput increases, thus fairness and throughput are also directly proportional. Finally, security and fairness are related through MACs as well. In a system where access to the medium is either dynamic or random, a malicious outsider could potentially jam the network traffic by continually requesting the medium. In a static access network, this type of jamming would be impossible, however. Thus as fairness increases with more statically scheduling of node access, security will increase as well.

4.2.7 Network Reliability, Quality of Service

Three primary methods of ensuring network reliability and quality of service involve communication precursors and acknowledgements, error checking, and error correcting communications. Each of these methods involves increasing the amount of data sent on the network, which will in turn increase the power consumption of each node. Thus as network reliability increases, so does power consumption of each node. Network reliability and quality of service will also depend on type of MAC used. Static access MACs will give the highest reliability since there is no worry of nodes attempting to transmit on top of each other. Dynamic access MACs will give the same reliability as static access MACs since nodes will also have schedules access. Random access MACs, even with collision avoidance methods, may suffer from a hidden terminal problem, or other types of network contention and so they are the least reliable of any of the methods. The power consumption of each of these access methods has been described in the Network Latency and Network Fairness sections above.

The first of the direct network reliability enhancers are communication precursors and acknowledgements. A communication precursor involves an initial transaction between sender and receiver in order to establish that a data stream is coming. The most basic of these precursors are the Ready-to-Send/Clear-to-Send (RTS/CTS) messages standard throughout many communication protocols. The sender transmits a RTS message across the network to the receiver. If the receiver receives the message and is not currently busy, it will respond with a CTS message. In a multi-hop network, the CTS messages can cause any node that hears it to mute itself until the completion of the sender/receiver transaction. This muting effectively clears the communications route between the sender and receiver so no contention will result. This increases reliability of communication further since it ensures a message will not be overridden by other messages during its transmission. If a sender does not hear back from the receiver, it can either try to establish communication again, or simply give up on the transmission. The RTS/CTS scheme is still prone to the hidden terminal problem, however, in which a node out of reception range of a CTS, but within transmission range of the receiver, transmits concurrently with the intended sender. This contention eclipses the intended transmission and will create a communications failure.

Once a message is received, an additional message can be sent by the receiver back to the sender to acknowledge or not-acknowledge (ACK/NACK) proper reception of the message. A NACK would be sent by the receiver if either a transmission was corrupted by contention or noise or if a transmission was not received within a certain specified timeout. If an ACK or

NACK is never received by the sender, the sender can either assume the transmission failed and retransmit or attempt to reinitiate contact via another round of RTS/CTS transaction. In sensor networks with high node mobility or choppy connectivity, the retransmitting of the RTS/CTS signals, although it might occasionally be unnecessary, is a better logistical choice for power minimization and to clear the communications channel again. Both the precursor messages and these acknowledgement messages can be very short in order to minimize their impact on power consumption. Any bit sent across the wireless channels may have to be multi-hop routed to its destination, however, so the power impact of these messages affects not only the sending and receiving nodes, but all other nodes in between.

Error checking is another method of ensuring network reliability and quality of service by helping to ensure that any message sent has not been corrupted during its transmission. The most straightforward implementation of error checking involves appending a checksum to each packet. A checksum is a bit or string of bits appended to a message that is generated by combining the bits that make up the message in a certain mathematical way. The simplest form of checksum is a single parity bit. With a series of 0's and 1's making up the message, a parity bit will be assigned a value such that the total number of 1's will always be odd (odd parity) or even (even parity). A single parity bit can detect single bit errors, but may fail for more complicated error patterns. In general, the more bits that are sent to check the message, the more likely any errors of any length will be caught. One of the most widely used error checking methods is known as the Cyclic-Redundancy-Check (CRC). This algorithm applies a generator polynomial with special properties to any message and results in a binary number that always has the same length. Most commonly used commercial CRC lengths are 16 bits which can detect any single point error in a message as well as up to a 16 or fewer consecutive bit error. The DoD uses a 32 bit CRC for additional error protection. Since the CRC is a fixed additional amount of data to be sent with each message, the per bit overhead it causes depends on packet length, but the total overhead it causes remains fixed. As packet length increases, the CRC overhead has less and less of an effect. A 16-bit CRC appended to a 128 bit message will increase the power required to send the message by approximately (128+16)/128-1 = 12.5%; appended to a 1024 bit message, the power increase will be only approximately (1024+16)/1024-1 = 1.6%.

Finally, error-correcting codes can be implemented to automatically detect and correct errors in sent messages. Whereas error checking only allows errors to be detected in which case a complete packet retransmission is necessary, error correction allows errors to be both detected and corrected without retransmission. The most famous and basic of error correcting codes is the Hamming code. Hamming codes can be developed to detected n-bit errors in messages of any length. As the number of error bits that can be detected goes up, or the length of each packet goes up, the number of additional bits that need to be sent for error correction goes up as well. As an example, the impact of a single-bit error correction Hamming code is analyzed. A fundamental property of single-bit error correcting codes for m check bits and a coded message length (original message plus check bits) of n is:

$$2^m \geq n+1$$

This specifies a minimum number of check bits for any coded message length. Manipulating this formula reveals that the longest original message length, r, for m check bits is:

$$r = 2^m - m - 1$$

Following this result, the worst-case factor increase in length, f, of a message when the check bits are added will be for:

$$r = 2^c - c$$

where c is some integer greater than 0. For this message length, m = c+1 bits will be needed for error checking causing an increase in length of:

$$f = \frac{2^c + 1}{2^c - c}$$

These factors will define the upper envelop of message length increase factors, but the general factor length increase of any message of length r will be:

$$f = \frac{r + \left\lceil \log_2\left(1 + r + \log_2\left(1 + r\right)\right)\right\rceil}{r}$$

Table 8 and Figure 13 summarize a few of these factor increases in length, and these length increases approximately correspond to factor increases in power.

Table 8: Effects of length and power for r and f

Original Message Length in Bits, r	Length Factor Increase, f
4	1.75
8	1.5
16	1.313
32	1.188
64	1.109
128	1.063
256	1.035
512	1.02
1024	1.011

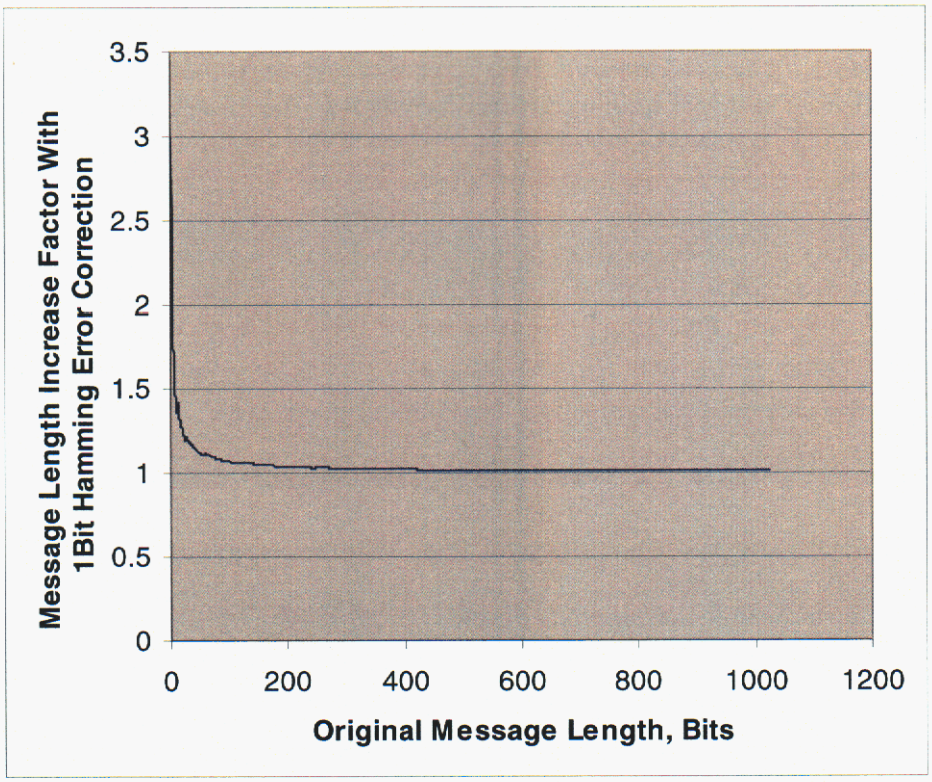


Figure 13: Message Length Increase Factor for 1-Bit Error Correcting Code

As can be seen in the table and accompanying figure, as packet length increases, the incremental increase on power consumption goes down drastically. (With the error correcting coding scheme indicated, the error correcting encoder must have access to the entire message. If a byte-by-byte scheme is needed, the power increase will be approximately 150% as indicated in the table above for 8 bit increase.) As was pointed out earlier in the Security section as well as the Network Latency section, this evidence provides yet another reason to increase packet lengths to reduce overall power consumption.

Other: Other than power, reliability also trades off with security and latency. The error checking and error correcting reliability enhancements make interception and data faking much more difficult. If a malicious outsider were to change any data in the message, the error check bits or error correcting bits would have to be updated as well. Since the intruder may not know the formula by which the check bits or correcting bits are calculated, these bits can essentially serve as a very simple form of data tamper detection. As reliability goes up with more complex error checking or correction, security would thus go up also. Reliability and latency are related through the communication precursors and acknowledgments. The additional time it takes to initiate and complete a message communication when precursors or acknowledgements are implemented directly increases the latency of data flow out of the network. As a result, as reliability increases, latency increases as well.

4.2.8 Network Throughput

Throughput is defined as the amount of data per node per time that can be extracted from a system. The maximum possible throughput would be equal to the bandwidth of the communications channel, but due to possible contentions and communications overhead, it is generally less than this maximum in any multiple node setting. A thorough explanation of MAC layers was given in the Network Latency section above, and thus only minimal additions to that description are given here. The highest throughput in a network is achievable when there are no contentions and minimal protocol overhead. Results in [8] showed that highest throughput for a given latency was achieved with a static access protocol, G-TDMA, where the bandwidth per node (in a network with heterogeneous traffic) was empirically adjusted for maximum throughput. This held in all but the highest traffic conditions, and even in the high traffic conditions, only dynamic access protocols became superior while random access protocols couldn't even support the traffic. Furthermore, the highest supportable traffic for a given power found in [8] was achieved by dynamic access protocols (static access protocols were not tested) for all but the extremely low traffic conditions. Under extremely low traffic conditions, a random access protocol, PRMA, used a lower total power. Overall, as throughput increases, for any of the protocols chosen, the power requirements will increase. Additionally, as throughput increases, the power overhead needed by any of the protocols increases. By choosing a protocol wisely, the amount of power taken by the overhead compared to the amount of power taken by transmitting actual data can be minimized.

These results as well as the description given in the Network Latency section above, demonstrate that for low traffic conditions, a random access MAC protocol is preferable to a schedule-based protocol to give the highest throughput for the lowest power. For moderate to high traffic conditions, a schedule-based MAC is preferable. The reason for this is that as traffic load increases, schedule-based protocols ensure no (or reduced) contention and thus no (or little) need for retransmission, whereas random access protocols do not provide this guarantee. Additionally, if collision avoidance is implemented as a part of the random access protocol, this will drastically increase the overhead causing it to never be the most power conservative, as is the case with 802.11.

Ultimately, if the traffic pattern is known ahead of time, a static access protocol should always be used for highest throughput and lowest power operation. If the traffic pattern is not known, but is expected to be high and uniform, a static access protocol may still be preferable. If the traffic is not known but expected to be high and non-uniform, a dynamic access protocol, such as RRR, may be the best choice. Finally, if the traffic is not known but expected to be low and non-uniform, a random access protocol is probably the best choice. If no assumptions are to be made about the traffic or a general solution is sought, a hybrid protocol, such as SRUC, that changes operation from random access to dynamic access, will give the best operation across the spectrum of possible traffic patterns.

Other: A description of the latency versus throughput trade was given above in the Network Latency section. Since throughput is directly dependent on the bandwidth of the channel, bandwidth and throughput are directly and linearly related. If the throughput of a channel is defined in relation to how much actual data is transmitted, as opposed to just how many packets are transmitted, then any overhead incurred by security provisions will degrade

throughput. Thus security and throughput are inversely related. It seems also that as network fairness is increased, i.e. a more schedule-based protocol is used, the throughput will also increase. This is proven experimentally in the results of [8], [7] and [6]. Since reliability and fairness are directly related, the fairness / throughput relation implies that reliability and throughput are also directly related.

4.3 Methodology for selecting a sensor networks

One of the objectives of the SDAC LDRD is to provide an initial methodology for exploring the space between the four mission areas (Military Operations in Urban Terrain, Intelligence Community Operations, Mobile Forces and Fixed Site Security, and Safe and Secure Borders) and the selection of a certain sensor network(s). This missing link is meant to help identify when, how, and what the role of sensor networks might be in certain real-world applications and scenarios, and how the applications themselves direct what sensor network architectures are used.

Creating a methodology for representing the relationship between the elements of a sensor network and the increasing application space for this technology can be based on the general idea of decomposition. By applying this basic concept we can take the higher-level concepts of wireless sensor networks and decompose it into a finite collection of entities identified as main components for wireless sensor networks. This same process can be applied to the four mission areas to generate a collection of abstract application concepts. These abstract applications will be based on the capabilities needed in the mission areas. This chapter provides an overview of the proposed methodology that can assist in answering this question.

4.3.1 Decomposing sensor networks

The key is to find some type of concrete criteria or overlapping specification space between the two very broadly defined fields. This specification space can be defined by the mission areas and met by the sensor networks and thus create workable solutions. Since the terms sensor networks and mission areas raise myriad discussions from numerous perspectives and audiences, the first important observation is that there are only a finite number of classes of discussion for each topic, and so a framework for researching them as a whole should be possible to establish. The primary complicating factor is that the two areas are extremely multi-disciplinary. Scoping is necessary, however, and the purpose of this chapter is to provide a proposed methodology for conducting this as yet ill-formed research.

Determining the points of overlap between the fields is important, and one method of doing this is to decompose each of these areas to identify their primitive components. This is akin to creating a tree of research where the trunk is the field as a whole, and the leaves are individual concerns or implementations of individual components of the field. The drilling down into each area should identify two lists: first, a constraint list, which can be applied to each leaf of the topic tree, and second, the hierarchical structure of the topic tree itself.

Applying this methodology to the area of sensor networks, for example, one might identify the following components: system architecture, general purpose processing capability per node and inter-node, network topology, routing, security, throughput, lifetime, latency, time

synchronization, cost, communications, signal processing capability per node, size of hardware, size of software, etc… Each of these components can either be drilled into further or combined with others components to generate super-class components that are represented as some of the main branches (Routing and Topology) of the tree illustrated in Figure 14. Other components identified during this initial decomposition pass represent potential general constraints for sensor networks. These constraints reflect application or user defined concerns for issues associated with: signal-processing capability per node, size of hardware, size of software, throughput, latency, etc...

Applying this framework to the area of sensor communications provides a detailed decomposition containing routing, MAC, and physical communications layers, and in turn, routing is a class containing DSDV, AOVD, and TORA algorithms. Security, latency, and lifetime may be primitive specifications that can be applied to each leaf of the topic tree so comparisons between implementations can be meaningfully determined. For example, AOVD might give a lower lifetime system than DSDV, but it will also give a lower latency system. This will create a vector of constraints corresponding to each leaf of the sensor network tree as illustrated in Figure 14. These constraints are represented as external numbers on to the right side of the tree leaf structure, which are applied to the leaf elements (Net, Cluster, Bus, etc.). This approach to sensor network provides unique collection of course and fine grain capabilities that will ultimately help in matching implementation possibilities to application requirements.

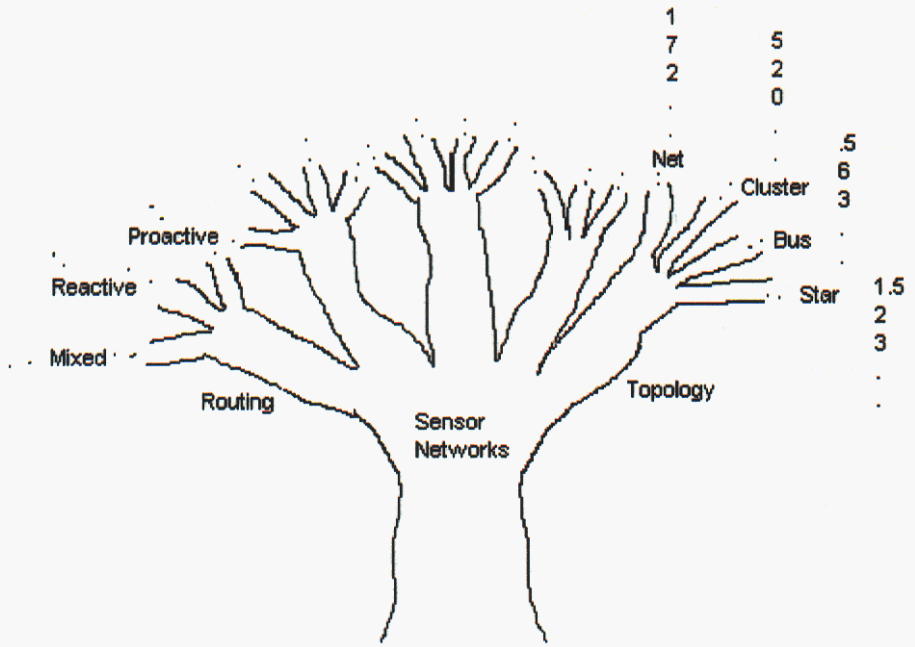


Figure 14:General sensor tree containing multiple levels of sensor network capabilities

4.3.2 Decomposing mission space

Drilling down into the mission areas will also be important. It should be noted, however, that it is unlikely that the individual mission areas are the primary branches of this topic tree. The overall goal is to provide a methodology to create working and useful systems of sensor networks, making each application space are more relevant than the larger mission areas. The mission areas will simply provide constraints on the applications, such as the density of sensors needed or the need for a wireless or wired implementation. Research into the mission areas will thus help supply a constraint and scenario list, whereas research into the applications will help identify components of applications necessary for specific types of scenarios.

In order to determine the applications, the mission areas must be broken down into scenarios for which sensor networks could be useful, so the construction of this topic tree will start with the twigs, and work towards the trunk. One possible super-set of primary applications is data collection, statistical data summarization, event detection, and tracking, which are based on the initial domain concern of the MOUT and Border areas (see Chapter 3 for details) for the mission areas. These are represented as the larger braches of the application tree in Figure 15. The scenarios are drilled into further to determine the enabling concepts or components needed to accomplish the scenario. In Figure 15 these concepts or components include the usage of technologies for tacking (i.e., agent-based systems, statistical algorithms, global/local data fusion) and data collection (i.e., 2 pairs of methods for data collection continuous or event-driven and centralized or decentralized). These components may be the leaves of the application tree, and taking these components, the scenario, and the mission area into consideration, specifications required of sensor network solutions can be determined. The combining of these three different items can be used to generate a constraint vector that can be assigned to each component or scenario.

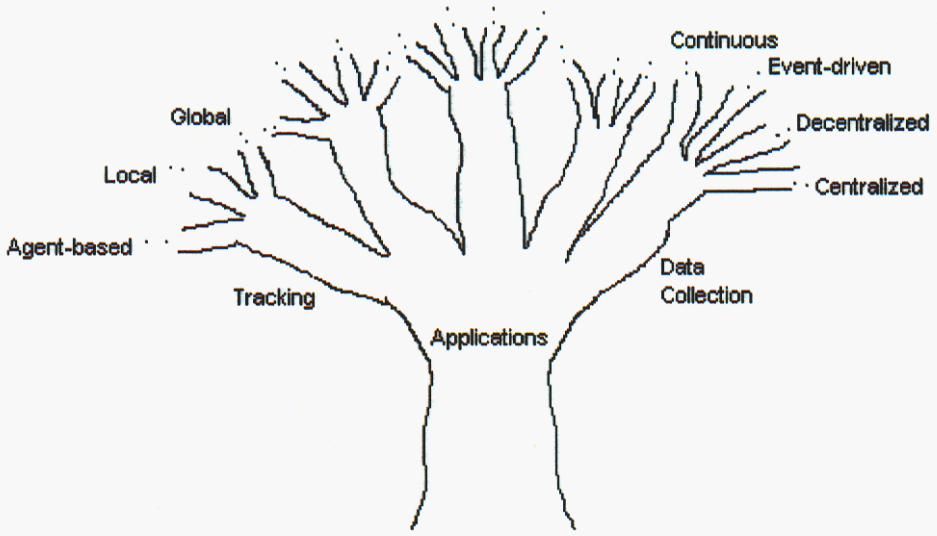


Figure 15: Application tree

4.3.3 Analyses of tree methodology with constraints

Once a list of constraints and the leaves of each topic tree have been identified, a few possible analyses fall out. First, a weighted constraint list must be applied to each leaf individually, and this list will become a vector capturing the relevant qualities of the leaf. Although the constraints will likely be inter-dependent, a unique constraint vector should be able to be determined. In order to find an optimal existing sensor network solution to solve a particular problem, the constraint vectors of the leaves on a scenario tree could be matched to the constraint vectors of the leaves on an existing sensor network tree. There are many possible ways to do this matching, and only one will be presented here.

First, a function to evaluate the matching of each application constraint to each sensor network constraint must be determined. If the application and sensor network constraints are identical, a perfect score, P, should result. If the application requirement is greater than the sensor network capability, a score less than P down to some minimum value, min, should result. If the application requirement is less than the sensor network capability, a score greater than P up to some maximum value, max, should result. The reason for limiting the scores between some max and min value is to eliminate the possibility of an outlying constraint mismatch eclipsing the application to sensor network matching. If there was no max, for example, one very highly ranked sensor network constraint/capability would cause that total constraint vector to be favored over others even if all other elements in the vector were less than their corresponding application requirements. These observations point towards using a function on the difference between application requirement, a, and sensor network capability, n (i.e., something of the form $f(a-n)$).

One possible function, f, with these properties is given by:

$$\frac{1}{f(x)+\max}+\frac{s}{f(x)+\min}=\frac{x}{c}$$

where the constant s is given by:

$$s=-\frac{P+\min}{P+\max}$$

which comes from solving for s when $x=0$ and $f(x)=P$. The constant c is used to stretch $f(x)$ horizontally in order to give the most appropriate matching function. This value can be determined empirically. Solving for $f(x)$ gives:

$$f(x)=\frac{\frac{x}{c}(\max+\min)+s+1-\sqrt{s^2+2s\left(1+\frac{x}{c}(\max-\min)\right)+\left(1-\frac{x}{c}(\max-\min)\right)^2}}{2\frac{x}{c}}$$

(At x=0, this function approaches P.) In order for $f(x)$ to be real valued, the discriminant in this function must be greater than 0 for all x. Solving the discriminant for x gives:

$$x = \frac{\max - \min - s(\max - \min) \pm 2\sqrt{-s(\max - \min)^2}}{(\max - \min)^2}$$

In other words, the solution to this equation for x cannot be a real valued number. Looking at the square root in this equation, we therefore find that:

$$s > 0$$

Assuming $\max > P > 0$, this requires that:

$$\min < -P$$

This is the only restriction on if a function $f(x)$ of this form is used. For max = 20, P=10, and min=-20, the plot shown in Figure 16 shows $f(x)$ for c=1, 10, 50, and 100.

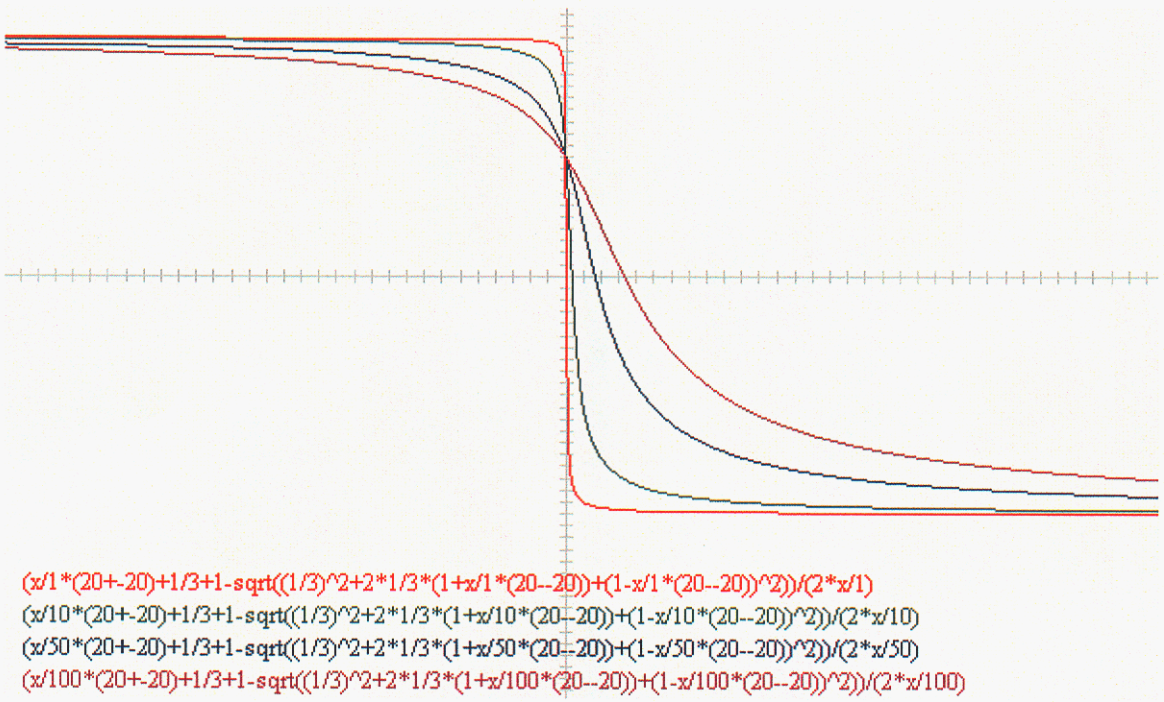


Figure 16: Example Score Generator Function

As can be seen, for very negative values of a-n, i.e. when the sensor network's capability far exceeds the application's requirement, this function asymptotes to max=20. For very positive values of a-n, i.e. when the sensor network's capability is far less than the application's

requirement, this function asymptotes to min=-20. For a-n=0, i.e. when the sensor network's capability exactly matches the application's requirement, this function returns the score P=10.

Another possible function for which there is no restriction on min, max, or P can be given by:

$$f(x)=\frac{\max-\min}{\pi}\,ArcTan\left(Tan\left(\frac{\pi\left(2P-(\max+\min)\right)}{2(\max-\min)}\right)-\frac{x}{c}\right)+\frac{\max+\min}{2}$$

For max=20, min=-20, and P=10, the plot illustrated in Figure 17 shows $f(x)$ for c = 1, 5, and 10.

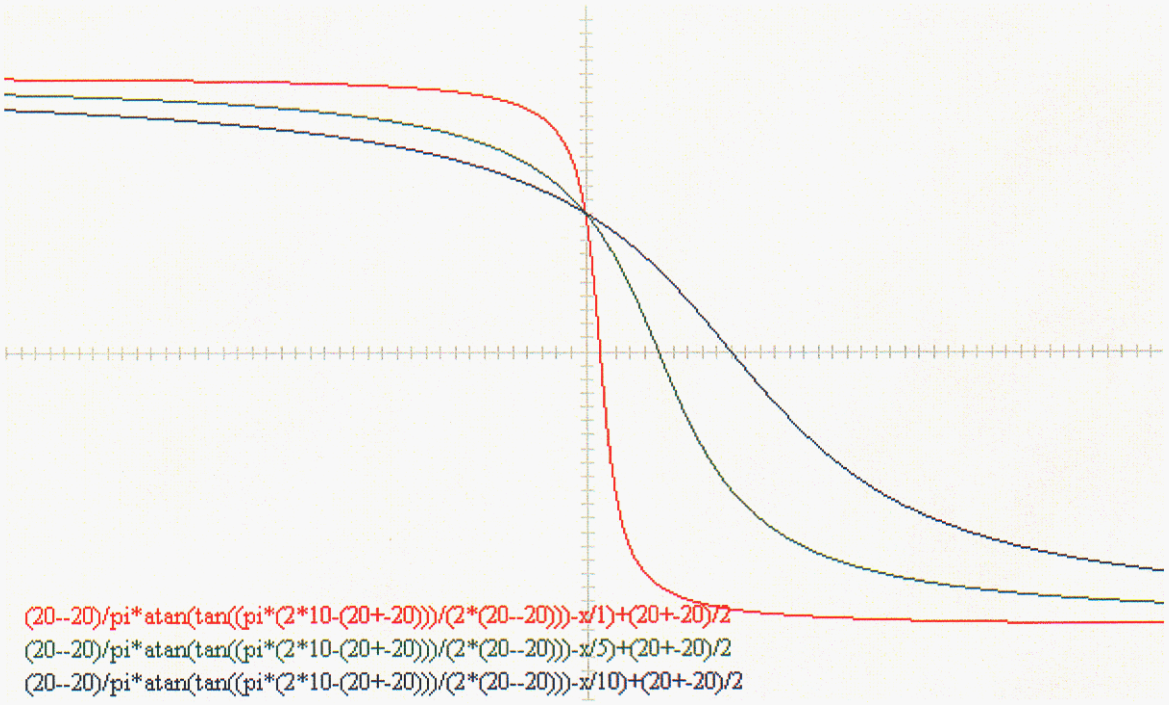


Figure 17: Another example Score Generator Function

Qualitatively, the difference between the two possible matching function developed is that the first one approaches its asymptotes much faster than the second. In other words, for sensor network capability deviations away from application requirements, the first function will score small deviations more extremely than the second function. This can be compensated for by the constant c, however, so the second function may be better since it does not restrict min, max, or P.

We may also want to specify a greatest value on a constraint that is unacceptable. For example, we may want a capability of 2 on bandwidth to give a score of 0 instead of a capability of 0 giving a score of 0. We can solve for the proper scaling factor, c, which will give us this property. Since the second matching function was preferable due to its unrestrictive nature, this

is the only function for which the proper scaling will be determined. Setting $f(x)=0$ and solving for c we find:

$$c=\frac{x_i}{Tan\left(\frac{\pi(2P-(\max+\min))}{2(\max-\min)}\right)+Tan\left(\frac{\pi(\max+\min)}{2(\max-\min)}\right)}$$

where x_i is the value for which the ith constraint score should be 0. The only restrictions implied by any of this development are that $\max > P > 0 > \min$, where these inequalities are strict.

Once the individual elements of the constraint vectors are passed through this function, the matching of the total vectors must be made. Again, there are several potential ways to perform a matching. An element-wise matching has thus far been developed, but a holistic approach, such as taking a normalized dot product, would also be possible. Only one potential element-wise approach will be developed here. In order to find the matching score for the entire system, Q, the weighted average of all the scores of all of the constraints is computed.

$$Q=\frac{\sum_{i=1}^{N} w_i f(a_i-n_i)}{\sum_{i=1}^{N} w_i}$$

where w_i is the weight given to the i^{th} constraint and N is the total number of constraints. The weight is a decided importance of a particular constraint. For example, if security is more important than network bandwidth for a particular application, security would be given a higher weight than network bandwidth. The weight may not be negative, but for cases when a higher constraint is less desirable, x_i, the value at which a constraint is scored 0, will be higher than the application requirement, a_i. This would be the case with power consumption, for example. A weight of 0 would imply that one does not care what a particular constraint value may be. (The file "Score Calculator.xls" can be used to examine properties and calculate examples of this scoring formula.)

Besides simply scoring existing solutions against potential applications and specific scenarios, the constraint vectors can be used in other ways. It may be possible to synthesize a sensor network to optimally match an application specification if existing systems don't suffice. This would be a difficult problem to give a closed form algorithm for, but it is likely that a system designer experienced in sensor network design would be able to take a constraint list and determine what components would be needed to implement a solution. It is possible that one way to help the designer do this would be to gather together all of the existing leaves of different sensor network systems that optimally match the application specifications (via the matching formula) into one system, but problems might arise as to integrating these disparate components. More research is needed to determine what possibilities exist along these lines.

One other aspect that can be examined by quantifying the constraints is a tradeoff analysis. It is likely that the primitive specifications will be dependent on each other, and it may

be possible to determine multi-variable functions for their relations. For example, increasing network security would probably decrease the lifetime of a system for a fixed size battery because it takes more communication power to have more secure transmissions. However, changing security requirements may have no affect on changing a specification such as the need for acoustic sensors, so there will likely be multiple functions that need to be determined. Designers could use these equations as a tool to help them design more reasonably achievable solutions.

Using this quantitative and tree like framework, all of the proposed LDRD problems can be solved. This framework takes into consideration that there is no single sensor network that is optimal for all situations. It will also help identify the best possible existing solutions to be used in a particular application. Since choosing a sensor network for an application seems to be akin to choosing a car for a person, there may not be any perfect solution to the problem, but there may rather be a set of solutions that would all suffice, and a matching formula can help identify this group. Finally, this framework could quantitatively aid designers in identifying the necessary components when designing new sensor networks. If the LDRD follows this framework, the result will be a top-down summary of many aspects of sensor networks and depending on time, the level of detail of the study can be extended downward as far as reasonable. This ability to extend or retract the amount of material that the study covers as permitted by time is very beneficial since it is currently unclear as to how much research will be needed or wanted at any specific level of detail.

4.4 Chapter 4 References

[1] Feng, J. and Potkonjak, M., "Power Minimization by Separation of Control and Data Radios".

[2] Walrod, J., *Military Sensor Networks symposium*, Technology Training Corporation, 2003.

[3] Yuan, M. and Qu, G., "Design Space Exploration for Energy-Efficient Secure Sensor Network", *Proceedings of the IEEE International Conference on Application-Specific Systems, Architectures, and Processors* (ASAP'02).

[4] Rex Min and Anantha Chandrakasan, "Energy-Efficient Communication for Ad-Hoc Wireless Sensor Networks", *35th Asilomar Conference on Signals, Systems, and Computers*, vol. 1, November 2001, pp. 139-143. (Invited Paper)

[5] [Wireless Networking for Ad-Hoc Sensor Fields, Wireless Sensor Networks and Their Tactical Applications, UCLA Extension, February 27, 2003]

[6] H. Peyravi, D. Wieser. "Simulation Modeling and Comparison of Multiple Access Protocols".

[7] Linnenbank, G.R.L. A power dissipation comparison of the R-TDMA and the Slotted-Aloha wireless MAC protocols, Moby Dick technical report, (1997) http://wwwhome.cs.utwente.nl/~havinga/mdpapdir.html.

[8] Chen, J.-C., Sivalingam, K.M. Agrawal, P., and Kishore,S., " A Comparison of MAC Protocols for Wireless Local Networks Based on Battery Power Consumption" in IEEE International Conference on Computer Communications (INFOCOM), (San Francisco, CA), pp. 150 - 157, Mar. 1998.

[9] Broch, Josh; Maltz, David; Johnson, David; Hu, Yih-Chun; Jetcheva, Jorjeta. "A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols".

5 SDAC Proposed Architecture

<< Text is this chapter is directly taken from Jesse Davis, Ron Kyker, Nina Berry, "A System Level Hardware Architecture for a Distributed Sensor Network Node", SAND2003-8209. See the actual report for complete document. >>

5.1 Modular Architecture Motivation

A sensor node in an ad-hoc distributed network that incorporates distributed computing must have the following hardware capabilities: sensing, node-to-node communication, processing, and data storage.[13] Any other hardware capabilities, such as satellite communication, the ability to add new sensors, or the ability to act as an access point to the network, should be able to be easily integrated into the system.[21] In this way, the system should be flexible and extensible to accommodate a variety of possible mission scenarios. The design of the system should also be easily upgradeable without necessitating a complete system redesign since sensors, processors, and communication links are continually improving. This adaptation is of primary importance because it will allow the exploration of which hardware architectures, routing algorithms, distributed computing algorithms, etc., are best suited for distributed sensor network applications. The perspective of this document is that an architecture developed for extensibility will allow the development and deployment of applications for both today and the future.

Low power consumption is another essential goal of the system. Since processing and sensing technologies are developing far faster than energy delivery technologies, the power consumption of the system acts as a bottleneck on its real world applicability. To make the situation worse, faster and more capable processors generally consume more power than their predecessors, further decreasing the lifetime of finite energy availability systems (such as battery operated systems)[14]. Many power conscious architectures have been researched on both the network and node level (see the DARPA PAC/C[33],[36] and SensIT[24] projects), and many power reduction techniques have been developed as a result. For example, task and instruction time extension via dynamic voltage and frequency scaling can conserve up to 40% of the power consumed by a processor.[6],[25],[31],[32] Other techniques, such as power adjustable signal processing[10],[35], power efficient multi-hop routing protocols[5],[30], power conservative MAC protocols[40], light weight[20] and power-aware[25] operating systems, and power aware compilers[2] have also been developed.[4]

One area of power conservation in sensor nodes that has not been adequately exploited is application specific computing. The central idea of application specific computing is that processing speed can be increased and power consumption can be decreased through specialization of hardware sub-components.[11],[12],[19],[38] Highly reconfigurable processors waste substantial amounts of energy in circuits that remain inactive but cannot be powered down. Conversely, application specific hardware allows all inactive sections of the system to be completely powered down thus decreasing the overall power consumption of the system.[15],[16],[17] This concept has been primarily confined to FPGA development[1],[18], but the idea is applicable to larger scale systems as well.

The price of using application specific hardware is that the individual processing engines will not be readily reconfigurable to perform different types of tasks. This problem can be easily remedied by keeping a more flexible general purpose processor in the overall system.[15],[24] The application specific processors can be used as satellite computational units servicing only their specific responsibilities and allowing the general purpose processor to either go to sleep or carry out other functions in parallel. This task separation has been demonstrated by the Sandia HERD program to allocate the duty of network routing to a low performance, power conservative processor while allowing the central processor of the system to remain in a low power mode. Research in ad-hoc multi-hop routing protocols has shown that depending on network topology, a large portion of messages that a node receives will be intended for other nodes in the network. Allowing the central processor to remain powered down while the satellite processor handles the network routing reduces the power consumption of the system.

The movement of data in a system is another energy and time intensive operation. At the network level, this observation led to the idea of developing distributed rather than centralized sensing. Instead of sending sensor data collected at each node to be analyzed on a centralized computer, distributing the data processing over a local set of nodes greatly reduces the amount of communications traffic. For wireless distributed sensor systems, the energy required per bit of computation is generally at least 100 times less than the energy required per bit of wireless communication[14],[29], so this localization leads to lower power consumption systems. At the processor level, the MIT uAMPS project[23] found that the continual need to move data in and out of memory and through different computational structures can consume a considerable amount of processing energy. Simply rearranging the software in a more data management efficient way can reduce processor power consumption significantly.[37] At the system level, a primary source of power consumption comes from driving large, high capacitance system-wide buses.[17] Since these bus structures are necessary, the best approach to reducing power consumption is to use them as infrequently as possible or encode the data that is put on them to minimize transitions.[4]

From this non-exhaustive, brief listing of the engineering concerns and potential solutions at numerous levels of distributed sensor system hardware and software architecture, it should be clear that there are many interesting problems that have yet to be solved in this field. The specific problem that this document seeks to address has already been described, but in order to better scope the remainder of this document, a specific class of distributed sensor network systems will be targeted. The primary application of the architecture to be proposed is in event-driven wireless distributed sensor networks. Though the proposed architecture is certainly applicable elsewhere, this class is a primary target for the design.

5.2 Proposed modular node design

This background research and motivation supports the idea of using a modular decentralized architecture for wireless distributed sensor networks that incorporate collaborative distributed computing. Not only will this modularity allow extensibility and upgradeability, but it will also decrease the total power consumption of the system. This decrease in power consumption comes from a decrease in data movement to a centralized location, and the ability to use application specific hardware instead of a single power intensive general purpose processor so inactive modules can be powered down. Furthermore, this modularity leads to a

more robust system since if one of the modules fails, the other parts of the system will still function. A validation of the fully modular approach is given by its current and on-going development by a joint effort of USC/ISI, MIT, Berkeley, and Raytheon, but as of January, 2003, no results have been demonstrated. Many of the principles of distributed computing have been applied to networks of nodes but not to the internal node structure itself. Several partially modular examples include the Sandia HERD nodes as explained above, the Berkeley Wireless Research Center picoRadio test beds[27], the Berkeley MICA motes[8], and the Rockwell WINS and Infocube platforms.[3] None of these systems allow the flexibility to easily add or change modules, however, and the data collection and event detection are still centralized on power hungry processors such as StrongARMs. The SDAC architecture proposed in Figure 18 below is fully modular and satisfies many more of the desired system attributes than previous systems.

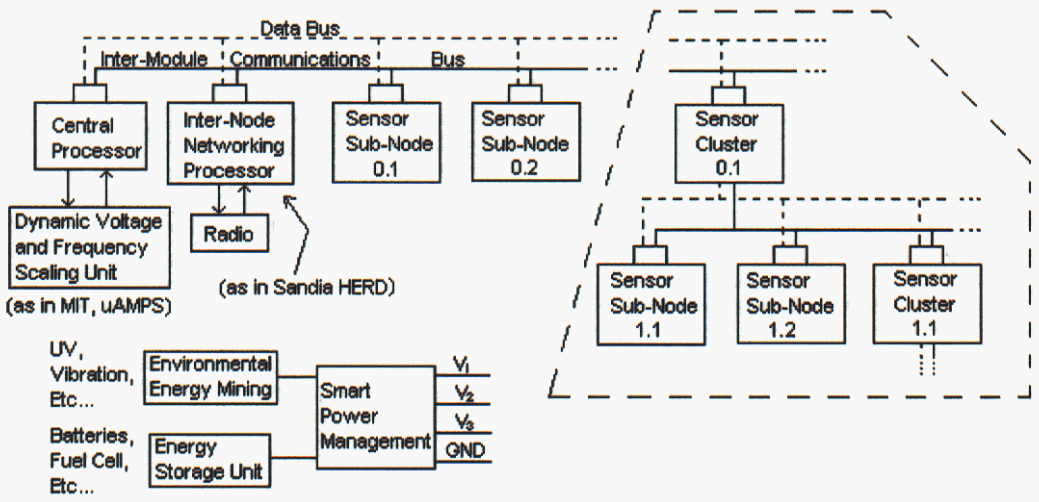


Figure 18: Node System Architecture

In the architecture shown in Figure 18, each of the modules attached to the central buses act as stand alone components. Each of the sensor modules will have their own data pre-processing component (either small general purpose processing units or application specific FPGAs) and data storage (either internal or external to the data pre-processor depending on memory requirements) as shown in Figure 19. This allows the high power general-purpose processor module to remain in a low power sleep mode for the majority of the nodes' operation. The general purpose processor will only be woken up when an external request for data processing is made by either another node in the wireless network or another module on the intra-node network. At this point, the general-purpose processor will collect the pre-processed data from each of the sensor modules and perform any further computation necessary. The wireless networking processor module will handle all routing and network message handling, as in Sandia's HERD, and may employ a two radio scheme for additional energy savings. There are a variety of schemes that employ two radios to separate data and control [14] or wake-up and communication [34] or allow a node to participate in multiple cluster relationships as in Sensoria's WINS platform. Using two radios has been shown to give both energy savings and logistical benefits.

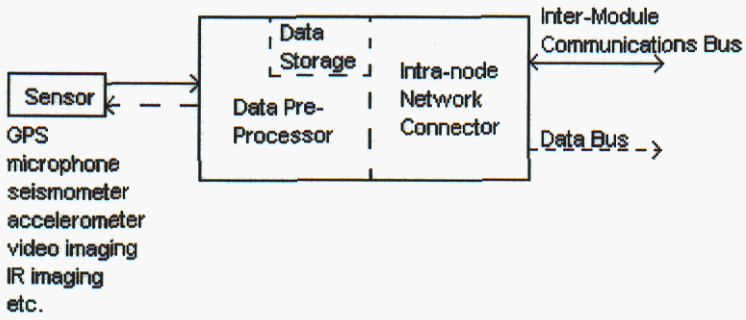


Figure 19: Module Architecture

The purpose of the sensor module data pre-processor will be two-fold. It will acquire and process the raw sensor data into a standard format, and it will act as a possible event detector. The processing of the raw data is necessary in the case of a temperature sensor, for example, in order to take the voltage measured over a thermistor, translate it into a temperature, trigger events requiring higher level processing, and package the collected data into a standard format that each module must adhere to and understand. Typically, only this very preliminary processing will be performed; any higher level transforms or computation on the data will take place on a more capable general purpose processor module when necessary. (It is an open question as to what level of processing should take place on the pre-processor, and what level on a general purpose processor module. For example, would a motion detection operation using an imaging sensor be a pre-processing event detection or a general purpose processing computation? Decreasing the amount of data passed over the bus, increases the complexity and power consumption of individual modules, so a tradeoff analysis must be conducted. See Mathematical Analysis section below.) The possible event detection service that the pre-processor performs will take a first pass look at the data from the sensor to determine if an event has occurred. This first pass look will likely take the form of threshold monitoring, envelop detection, or something similarly undemanding. When an event occurs, the sensor module will send a request to a more capable processing module (or a more capable processor on another sensor module) for verification. The processing module will gather the sensor's buffered data, fuse it with other relevant sensor data if necessary, and analyze it using higher-level, computationally intensive algorithms in order to determine if an event has actually occurred. If the processor verifies the module's detection, the processor will then pass along the processed situational information to other wireless nodes and start a distributed computation thread in the network to classify and track the event.

In order to make this architecture viable, each of the modules on the intra-node network will have to communicate using a common protocol. The bus protocol must have a few key attributes: it must have a low communication overhead, be secure (see more about security below), be simple to use and implement, not require complex routing or messaging routines, be multi-master capable, be extensible, and not require many channels. Since this architecture is designed for event-driven sensing, there is no master controller in the system. Instead, any module that detects an event becomes an effective master until its inter-module communication needs have been satisfied at which point the bus is released. With these requirements in mind, two potential candidates for bus protocols are encrypted versions of I2C[26] or USB OTG (On-The-Go)[41]. As inspired by the IEEE 1451 specification[7],[22],[39] and the MIT Media Lab Snap! Project[9], each module must also have a separate hardware or software intra-node (inter-module)

networking section. This additional section will be able to power on or off the module back-end and act as a gateway to the intra-node network. (In Figure 1, this interfacing section is denoted by a small box connecting the buses to the components, and in Figure 2 it is labeled as "Intra-Node Network Connector".) This Intra-Node Network Connector (INNC) protects each module from the heterogeneity of any of the other sensor or processor modules on the bus, and allows for the extensibility and reconfigurability of the system. Using this architecture, the modules could even be hot-swappable. The INNC provides a decentralized control backbone to the modular system. It controls the waking up, powering down, synchronization, and inter-module communication tasks necessary for and between separate module back-ends. They are the only sections of the node system, which need to be continuously on and ready to receive interrupts. This requires them to be very low power, and to have the capability to go into interrupt ready sleep modes to conserve energy.

There are two immediate potential extensions to this architecture in order to accommodate additional sensor capabilities. First, clusters of modules could become composite module sub-systems or meta-modules. An example of this idea would be an environmental cluster. This environmental cluster would be comprised of other, lower level sensor modules, such as temperature, humidity, or UV, and would act as a type of meta-sensor. This composite sensor would require an additional gateway module (as shown in Figure 1) in order to transparently interact with the existing modules on the intra-node bus, and act as a data fusion processor if appropriate and necessary. The second extension of this architecture would be the incorporation of a data bus for high bandwidth intensive sensors. Though this certainly would extract a power cost, it may be worthwhile depending on how processing is distributed among the modules. While data from a temperature module could easily be transmitted on a low bandwidth communications bus, this would limit the amount of data that could feasibly be sent to a processor module for analysis, and it would also potentially clutter this communications and control channel preventing other modules from access. With the incorporation of a separate data bus, these problems would be mitigated, and allow for complex sensors such as imaging modules to pass their data around the system more easily. This data bus could use higher speed data transfer protocols and have multiple channels in order to speed the data transfer and reduce collision between modules.

Aside from the already described benefits, this architecture also allows for extremely straightforward, modular packaging. Sensor nodes could be assembled by simply stacking together the necessary sensors, processors, communications unit, and power supply modules. One possible packaging scheme is displayed in Figure 20. This particular scheme has the advantage of putting only a single dimension size constraint on any of the modules, and so can easily adapt to shrinking module sizes, but the multi-layered aspect of its exterior may not be as robust or secure as may be needed.

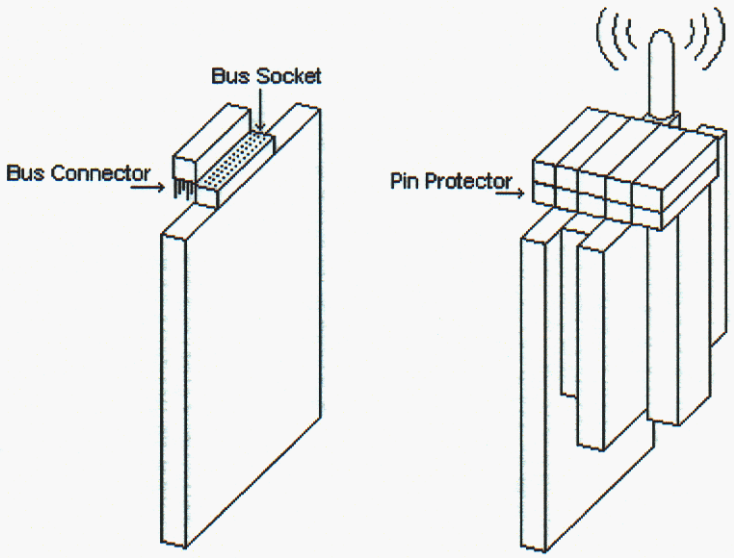


Figure 20: Example Packaging Scheme for Modular System Architecture

Another packaging possibility that is more robust but doesn't allow for shrinking module sizes as easily because it limits two dimensions of each module is shown in Figure 21.

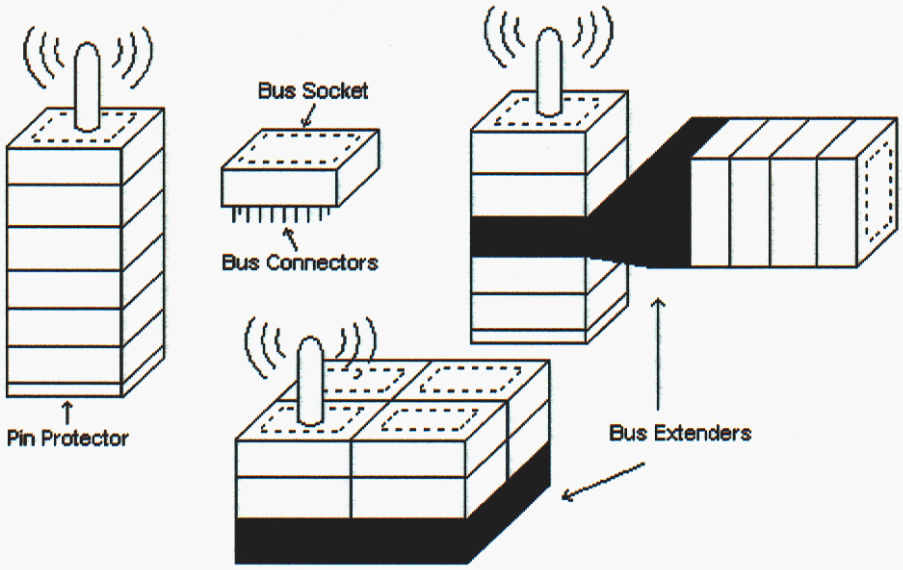


Figure 21: Another Example Packaging Scheme for Modular System Architecture

It would also be possible to fabricate an ultra miniaturized version of this architecture via system-on-a-chip and multi-chip-module technologies. This would shrink the packaging considerably as well as making the system more covert and robust.

For all applications, there should be an importance emphasis on the issue of security. This security would include the ability of the system to resist both hostile attack and unwanted intrusion or eavesdropping. Furthermore, the security would have to exist on multiple levels ranging from the sensors themselves, to the node hardware and software, to the wireless network as a whole. Wireless ad-hoc protocols and communications have been given almost the sole

focus of security research in sensor networks, and hardware and software security on the individual nodes have been all but ignored. A detailed discussion on software hardware and networking security issues is given in Chapter 6 Sensor Node and SDAC Security Considerations.

5.3 Compare centralized and modular architectures

Since the architecture proposed in this chapter is significantly different than previous, more centralized architectures, the benefits and drawbacks associated with each design should be examined. A centralized architecture is one in which the sensors and radio interface directly with a single general purpose processor that performs all of the data collection, computation, and communication of the system - an architecture similar to traditional, standard desktop computers. Table 9 provides a brief comparison of the two architectures, it is by no means meant to be exhaustive, but it should be used at least as a starting point for comparison. Table 9 might also help create specifications for other architectures to be proposed.

Table 9: Centralized vs Modular Architectures

Comparison Factor	Centralized Architecture	Modular / Decentralized Architecture
Power Consumption	Higher power consumption due to centralized data collection and the resulting inability to power down inactive sections. (Assumes micro-processor centralized system.)	Lower power consumption due to decentralized data collection and the resulting ability to power down inactive modules.[See note 1]
Speed	Application dependent. The centralized architecture may create a bottleneck at the processor and does not allow for any parallel computing. However, the design can be optimized to increase speed.	Application dependent. The decentralized architecture removes any major bottlenecks and allows for parallel and application specific computing on each module. However, data transfers over the inter-module bus may reduce overall speed.
Security	Centralized architecture provides adversaries with a single point of attack. However, software is less accessible than hardware, so the system structure is less accessible.[See note 2]	Decentralized architecture distributes the mechanisms for security making defeat more difficult. However, system structure is more accessible since the structure is more hardware based.[See note 2]
Extensibility	Limited extensibility. The architecture is mostly static, and only minimal hardware re-configurability would be possible.	Much broader extensibility. Extensibility fundamentally limited only by the number of bits in the address of each module.[See note 3]
Upgradeability	Not as easily upgradeable. Upgrading any hardware component of the system is a more involved process, requiring at least as much effort as for the decentralized architecture, and may require a complete architectural redesign. Software upgradeability is at best as easy as for	More easily upgradeable. Since the inter-module communication will be standardized, upgrading any hardware component simply involves building it into a new module. Software upgradeability is at least as easy as for centralized architecture.

	modularized architecture.	
Application Fit	Each unique application will require the design of an entirely new system. Centralized design is customized to a smaller application space.	Each unique application will use different modules but keep the same general system architecture. This vastly reduces the time, money, and effort required to implement new solutions. Modularized design is applicable to a wider application space.
Initial Implementation	Easier. The main initial complications come only from the number of tasks the central processor will have to service, the task scheduling, and interrupt servicing of multiple critical data input streams.	More involved. Time and effort must be expended to thoroughly design the overall architecture, and develop standards for module interactions. Decentralization will lead to more difficult inter-module communications debugging.
Future Implementation	More involved. Since the system is customized, future implementations require more effort to design. There is limited reusability of previous components when future applications deviate significantly from prior ones.	Easier. Once the building blocks are in place, each future implementation requires simply putting the right pieces together and programming their interaction. Reusability of previous work is more likely when future applications deviate significantly from prior ones.
Short Term Cost	Probably less expensive. The lower part count drives the cost lower, but the possible necessity for a more powerful central processor to handle all of the operations of the system might balance this.	Probably more expensive. The higher part count drives the cost higher, but the possibility of getting a less powerful general purpose processor (since it wont need to handle as much traffic) might balance this.
Long Term Cost	More expensive. Since the architecture is not as extensible or upgradeable, only problem or mission area specific solutions can be engineered. Any future development in a wider application range will start from scratch and hence require much more effort and cost.	Less expensive. Since the architecture is widely extensible and upgradeable, future development will have an already built infrastructure from which to work hence requiring less effort and cost. Mission space applicability is much broader.
Robustness	Less robust. Highly sensitive to single point failure. Any robustness will be due to complicated software reducing the ease of implementation.	More robust. Less sensitive to single point failure due to the modularity. Robustness is inherently built into the system architecture.
Size	Smaller. The layout can be optimized over the whole system.	Larger. Only the layout of individual modules can be optimized. Also a standardization of module packaging and connectors will likely result in some wasted space.[See note 4]
Part Count	Lower part count since the data storage and computation is centralized to a single processor.	Higher part count since data storage and some level of computation is decentralized onto individual modules.

Note 1: Even though there are more parts in the decentralized architecture, a brief survey of processors will show that more power will be consumed by one large processor than by several smaller processors.
Note 2: Assume that tamper protection is implemented on both systems, and all means of security possible are employed on the individual nodes. The largest security concern with distributed sensor networks is not in the individual nodes, but in the network communications. The operational power of a sensor network is mainly in the network, not in the individual nodes.
Note 3: Since I^2C is a candidate for the intra-node communication, it should be noted that the official I^2C specification allows either 7 or 10 bit addresses. This will allow up to 127 or 1023 modules with one additional broadcast address. An actual implementation would be limited to far fewer modules by various logistical and engineering issues including the increasing bus capacitance as more modules are added.
Note 4: A single Multi-Chip-Module or System-on-a-Chip implementation of the decentralized architecture could be made for optimal space conservation. This would incur higher costs depending on production volume.

From Table 9, it can be seen that there are tradeoffs associated with both types of architectures. There is no single best solution to every distributed sensing scenario – every mission application should be analyzed to find the best programmatic fit. It should be noted here that it would also be possible to combine a centralized and decentralized architecture into a hybrid system where only certain portions are modularized. There is a potential to design this type of compromise system to gain as many good aspects of both systems as possible while limiting the number of negative characteristics inherited from either. In this way, it may be possible to make the architecture itself a tunable characteristic to optimally satisfy application constraints.

5.4 Chapter 5 References

[1] Abnous, Arthur. Low-Power Domain-Specific Processors for Digital Signal Processing. PhD Thesis University of California at Berkeley.
http://bwrc.eecs.berkeley.edu/Publications/2001/Theses/L-pwr_domain-spec_processors/Abnous/thesis.pdf

[2] Acevedo, Oscar and Jimenez, Manuel. A Survey of Software Optimization Techniques for Low-Power Consumption. University of Puerto Rico.
http://mayaweb.upr.clu.edu/crc/crc2002/papers/Acevedo_Oscar.pdf

[3] Asada, G.; Dong, M.; Lin, T.S.; Newberg, F.; Pottie, G.; and Kaiser, W.J. Wireless Integrated Network Sensors: Low Power Systems on a Chip.
http://www.janet.ucla.edu/WINS/download_publications/esscirc98.pdf

[4] Benini, Luca and De Micheli, Giovanni. System-Level Power Optimization: Techniques and Tools.
http://iacoma.cs.uiuc.edu/CS497/LP7a.pdf

[5] Broch, Josh; Maltz, David; Johnson, David; Hu, Yih-Chun; and Jetcheva, Jorjeta. Multi-Hop Wireless Ad Hoc Network Routing Protocols. Carnegie Mellon University.
http://citeseer.nj.nec.com/cache/papers/cs/882/http:zSzzSzwww.monarch.cs.cmu.eduzSzmonarch-paperszSzmobicom98.pdf/broch98performance.pdf

[6] Chandrakasan, Anantha; Min, Rex; Bhardwaj, Manish; Cho, Seong-Hwan; and Wang, Alice. Power Aware Wireless Microsensor Systems.
http://www-mtl.mit.edu/research/icsystems/uamps/pubs/Chandrakasan_A4.pdf

[7] CogniSense. Building Plug-and-Play Networked Smart Transducers.
http://ieee1451.nist.gov/edc-pap.pdf

[8] Crossbow. MICA Wireless Measurement System: Datasheet.
http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/MICA.pdf

[9] Debski, Matthew. The Snap! Toolkit for Developing Sensor Networks and its Applications to Cross-Country Skiing. Masters Thesis, MIT.
http://www.media.mit.edu/pia/Research/ESP/DebskiThesis/thesis.html

[10] Dong, Michaeil; Yung, K. Geoffrey; and Kaiser, William. Low Power Signal Processing Architectures for Network Microsensors.
http://www.janet.ucla.edu/WINS/download_publications/islped97.pdf

[11] Mangione-Smith, William; Ghang, Phil Seong; Nazareth, Sean; Lettieri, Paul; Boring, Walt; and Jain, Rajeev. A Low Power Architecture for Wireless Multimedia Systems: Lessons Learned From Building A Power Hog.
http://www.janet.ucla.edu/glomo/finalreport/final_report_documents/islped96.pdf

[12] Fei, Yunsi and Jha, Niraj. Functional Partitioning for Low Power Distributed Systems of Systems-on-a-chip.
http://www.ee.princeton.edu/~yfei/pdf_files/vlsi02.pdf

[13] Feng, Jessica; Koushanfar, Farinaz; and Potkonjak, Miodrag. System-Architectures for Sensor Networks Issues, Alternatives, and Directions.
http://www.cs.ucla.edu/~jessicaf/paper/ICCD_SN.pdf

[14] Feng, Jessica and Potkonjak, Miodrag. Power Minimization by Separation of Control and Data Radios.
http://dsp.jpl.nasa.gov/cas/short/feng.pdf

[15] Havinga, Paul and Smit, Gerard. Octopus – an energy-efficient architecture for wireless multimedia systems.
http://wwwhome.cs.utwente.nl/~havinga/papers/octoProRISC99.pdf

[16] Havinga, Paul and Smit, Gerard. Minimizing Energy Consumption for Wireless Computers in Moby Dick.
http://wwwhome.cs.utwente.nl/~havinga/papers/energy.MD.icpwc97.pdf

[17] Havinga, Paul and Smit, Gerard. Design techniques for low power systems.
http://wwwhome.cs.utwente.nl/~havinga/papers/energy.design.techniques.jsa.pdf

[18] Hwang, Enoch; Vahid, Frank; and Hsu, Yu-Chin. FSMD Functional Partitioning for Low Power.
http://jamaica.ee.pitt.edu/Archives/ProceedingArchives/Date/Date99/papers/1999/date99/pdffiles/01b_1.pdf

[19] Henkel, Jorg. A Low Power Hardware/Software Partitioning Approach for Core-based Embedded Systems.
http://www.sigda.org/Archives/ProceedingArchives/Dac/Dac99/papers/1999/dac99/pdffiles/08_1.pdf

[20] Hill, Jason; Szewczyk, Robert; Woo, Alec; Hollar, Seth; Culler, David; and Pister Kristofer. System Architecture Directions for Networked Sensors.
http://tinyos.millennium.berkeley.edu/papers/tos.pdf

[21] Hollar, Seth. COTS Dust. Bachelors Thesis, MIT.
http://www-bsac.eecs.berkeley.edu/~shollar/shollar_thesis.pdf

[22] MicroChip AN214. The PICmicro MCU as an IEEE 1451.2 Compatible Smart Transducer Interface Module (STIM).
http://www.microchip.com/download/appnote/pic16/00214a.pdf

[23] Min, Rex; Bhardwaj, Manish; Cho, Seong-Hwan; Sinha, Amit; Shih, Eugene; Wang, Alice; and Chandrakasan, Anantha. An Architecture for a Power-Aware Distributed Microsensor Node.
http://www.mit.edu/~rmin/research/min-sips00.pdf

[24] Parker, Robert. Distributed Sensors Group: Goals, Metrics, and Challenges. PAC/C DARPA presentation, USC. November 1-3, 2000.
http://pads.east.isi.edu/presentations/pi_2000_10/parker_challenges.pdf

[25] Pereira, Chritiano; Gupta, Rajesh; and Srivastava, Mani. PASA: A Software Architecture for Building Power Aware Embedded Systems.
http://dsp.jpl.nasa.gov/cas/full/gupta.pdf

[26] Philips Semiconductor. The I^2C-bus Specification.
http://www.semiconductors.philips.com/acrobat/various/I2C_BUS_SPECIFICATION_3.pdf

[27] Rabaey, J.; Ammer, J.; Karalar, T.; Li, S.; Otis, B.; Sheets, M.; and Tuan, T. PicoRadios for Wireless Sensor Networks: The Next Challenge in Ultra-Low-Power Design.
http://bwrc.eecs.berkeley.edu/Publications/2002/presentations/isscc2002/ISSCCslides.pdf

[28] Rabaey, J. Issues in Low Power Design – Architecture and System Level.
http://bwrc.eecs.berkeley.edu/People/Faculty/jan/presentations/Lausanne/lecture3.pdf

[29] Raghunathan, V.; Schurgers, C.; Park, S.; Strivastava, M. Energy-Aware Wireless Microsensor Networks.
http://www.parc.xerox.com/zhao/stanford-cs428/readings/Physical/raghunathan_sp_02.pdf

[30] Rahul, Shah and Rabaey, Jan. Energy Aware Routing for Low Energy Ad Hoc Sensor Networks. University of California at Berkeley.

http://bwrc.eecs.berkeley.edu/Publications/2002/presentations/WCNC2002/wcnc.rahul.pdf

[31] Saewong, Saowanee and Rajkumar, Ragunathan. Optimal Static Voltage-Scaling for Real-Time Systems.
http://www.cs.mcu.edu/afs/cs/project/rtml-2/Papers/dvs_rtss.ps.gz

[32] Saewong, Saowanee and Rajkumar, Ragunathan. Practical Voltage-Scaling for Real-Time Systems.
http://www.cs.mcu.edu/afs/cs/project/rtml-2/Papers/dvs_rtas.ps.gz

[33] Schott, Brian; Parker, Bob; Chien, Charles; Srivastava, Mani; and Gupta, Rajesh. Power Aware Distributed Systems. Joint PAC/C DARPA presentation, USC, Rockwell, UCLA, and UC Irvine. November 1-3, 2000.
http://pads.east.isi.edu/presentations/pi_2000_05/pacc_pads_may00.pdf

[34] Schurgers, Curt; Tsiatsis, Vlasios; and Srivastava, Mani. STEM: Topology Management for Energy Efficient Sensor Networks.
http://www.ee.ucla.edu/~curts/papers/Aerospace02.pdf

[35] Sinha, Amit and Chandrakasan, Anantha. Operating System and Algorithmic Techniques for Energy Scalable Wireless Sensor Networks. MIT
http://www-mtl.mit.edu/research/icsystems/uamps/uAMPS-1/presentations/sinha_mdm2001Talk.pdf

[36] Srivastava, Mani and Chien Charles. PADA: Power Aware Distributed Systems, Architecture Approaches. Joint PAC/C DARPA presentation, USC and Rockwell.
http://pads.east.isi.edu/presentations/review_2001_06/pads_arch_isi.pdf

[37] Tiwari, V.; Malik, S.; Wolfe, A., and Lee, T.C. Instruction Level Power Analysis and Optimization of Software.
ftp://ftp.ee.princeton.edu/pub/vivek/jdsp96.ps

[38] Wang, Alice and Chandrakasan, Anantha. Energy Efficient System Partitioning for Distributed Wireless Sensor Networks.
http://www-mtl.mit.edu/research/icsystems/uamps/pubs/aliwang_icassp2001.pdf

[39] Woods, Stan et. al. IEEE-P1451.2 Smart Transducer Interface Module.
http://ieee1451.nist.gov/senoct11-2col.pdf

[40] Ye, Wei; Heidmann, John; and Estrin, Deborah. An Energy-Efficient MAC Protocol for Wireless Sensor Networks.
http://www.isi.edu/~weiye/pub/smac_report.pdf

[41] On-The-Go Supplement to the USB 2.0 Specification.
http://www.usb.org/developers/onthego/otg1_0.pdf

5.5 Sensor Node and SDAC Security Considerations

The threats to wireless sensor networks (WSN) extend beyond traditional computer systems due to the radio frequency issues and the need to place the sensors in hostile or unmonitored environments. This chapter provides insight into the general issues associated with sensor networks security from communication and networking issues to hardware and software architecture tampering.

5.6 General Security in Sensor Networks

Wireless sensor networks differ from other forms of distributed systems in a very important way, and therefore many fundamental solutions in network security fail to apply and must be discarded. The resource-starved nature of sensor networks creates a new and paramount challenge for their security. Since sensors generally have very little computational power, public-key cryptography is considered excessively expensive and most symmetric-key ciphers can be used in a limited scope. Sensors have a minimal amount of memory, which limits the amount of state information we can maintain within security mechanisms. Communication bandwidth presents an even greater problem, where each transmitted bit generally consumes as much power as many program instructions. Therefore, any data expansion (increasing message size, etc.) added with the inclusion of security provisions within a sensor network comes at extreme cost to power consumption [1]. Every increase in power consumption within a sensor network is a decrease in the lifetime of a sensor node, and security must therefore be designed with power in mind. Despite its increase in resource consumption, security is a very necessary feature to be added to any sensor network, especially those deployed in possibly malicious areas. Therefore, care must be taken in identifying and addressing vulnerable areas of a sensor network in order to find the maximum increase in security while providing the least amount of power consumption increase.

A thorough analysis of the general security problems within sensor networks will provide the motivation and background needed to address problems within the SDAC architecture. The following outline contains selected topics in sensor network security with direct application and importance to the MOUT (Military Operation in Urban Terrain) or Border applications of SDAC. They follow from the assumption that there is no way of ensuring physical security on the individual nodes. In the SDAC architecture, it will be desirable to include the base security enhancements as outlined below, in addition to those specifically outlined in Section 5.7. It is important to note, however, that it may not always be feasible to apply all security provisions to all aspects of node functionality given the extremely limited resources of the architecture. Therefore, various applicable security problems will be identified as well as what provisions are needed to solve them; additionally, it must be analyzed how these provisions can be implemented and applied in an efficient and effective manner.

5.6.1 Technology Exposure

Technology exposure is perhaps the most critical concern facing sensor networks deployed in potentially sensitive areas. It provides the ability for an adversary to gain specific knowledge about and functionality with a given sensor node such that it can use for an unintended (and perhaps malicious) purpose.

Exposure of information is a physical, implementation specific concern. The question revolves around the fact that, once a sensor network is deployed in the field, what functionality will be available (exposed) to a random entity that happens upon a device? More specifically, will a possibly malicious entity be able to accomplish anything significant without specific knowledge about device functionality? Will there be any possible information leakage via an insecure *side-channel*? In reality it must be assumed that there are no physical protections on a node and that each node is completely exposed to physical attack and interception. Before further analysis can be made, it is important to define *physical* attack against a node as well as insecure side-channels:

Side-Channel Attack:
A side-channel attack consists of performing an attack against a non-standard entry point on a device. In the context of a sensor node, the normal entry point would be via the wireless communication module through which all inter-node communication takes place. A side-channel is any other possible entry point on a system other than traditionally specified, which would indicate any other entry point on our node other than the wireless communication channel.

Bus Sniffing:
Bus sniffing is a side-channel attack wherein an attacker reads data bits as they pass along a system bus. This can be accomplished in varied difficulty, and the degree of difficulty depends on the complexity of the bus. This attack has been used extensively in the past to figure out bus communication protocols as well as extracting specific data from the bus (such as cryptographic primitives, etc.). Equipment needed consists of a custom made tap-board that will extract data from the bus lines, and some sort of analyzer for processing the data. This is a highly specific attack and requires a great deal of hardware knowledge.

Differential Power Analysis:
Differential power analysis (DPA) is a side channel attack against any electronic system. By monitoring a channel that utilizes system power, such as the connector from a node battery to the main processing unit, an adversary can determine a great deal of information about a system. It can be readily identified that when a system is using more power it is likely performing some heavy computational task such as encryption or key generation. By examining the electromagnetic emission of various components of a node, one can achieve the same desired affect as physical contact.

Utilizing some or all of the above methods, an attacker can gain access to a node on varying levels. For instance, with knowledge of the bus transmission protocol, one could

carefully construct messages sent within the node such that they are able to modify the behavior of the node. DPA could possibly allow one to extract certain properties of a node, such as its cryptographic keys. These are all considerations when designing the physical architecture of a node. However, it is important when addressing security outside of the physical layer to assume that a node is fully exposed to possible adversaries. Given this, it is important to construct a base of security that allows for an acceptable level of protection in the presence of an exposed and possibly intercepted node within the network.

5.6.2 Member Enforcement

One of the most basic concepts in sensor security is the enforcement of network membership. A member of a network is generally considered a trusted entity and is allowed normal access to network functions and resources. The primary concern is in keeping a malicious entity from joining the network and performing some "bad" operation such as flooding the nodes with erroneous data, perhaps to the extent that normal operation is no longer possible. In addition, it is necessary to ensure that a malicious base station is not able to subvert and take over the original sensor network's traffic. Fundamentally, it is important to ensure that a sensor network performs in a safe and secure manner in the presence of one or more possibly malicious outside sensors.

5.6.3 Data Authenticity

It is important to maintain some mechanism for ensuring the authenticity of data being transferred across the network. Data must not be random noise, but some meaningful sensor data or network management information coming from a trusted source. There are five main types of traffic traveling throughout a sensor network: sensor data, routing and infrastructure management information, node management information, noise and a fusion of the above.

In general, sensor data is uni-directional in that it will be traveling from a node and routing through to the base station, or possibly interpreted (*fused*) at an intermediary node. Routing and infrastructure management information will travel in many directions, and is used to modify the network structure and routing tables in case of node loss, and may or may not involve the base station. Node management information can be thought of as uni-directional and moving "downstream" from the base station to its children nodes, performing some type of administrative task for the network.

Noise is not necessarily acceptable network traffic, but is indeed an element of concern in real-world deployment. It is important to discern normal information from noise, and ensure the presence of noise does not alter valid data in any harmful way. In regards to sensor readings, it could become critical to verify their origin. This will provide a framework in which non-member nodes cannot insert arbitrary information into the network.

Another problem is the case when a node is tricked into producing and propagating through the network some type of erroneous reading. This is an indirect attack and its solution is non-trivial. For example, take the scenario when sensors are deployed across a border to monitor travel in and out of a country. One sensor suddenly looks like it has many entities in movement about it, causing alarms that a large flow of people is moving across a region. The attack

concerns the case when, in reality, one or two people are repeatedly activating the sensor in order to draw attention from another location which would normally trigger concern. Mechanism utilizing various sensor types can be used to prevent against this method of attack.

Network management information should be subjected to some form of authentication from node to node to ensure that the ad-hoc managed structure of the sensor network is maintained. A denial-of-service (DoS) attack is possible if the routing tables contained on certain nodes are modified such that network traffic follows a circular path, never being routed to a base station. Only trusted nodes should be allowed to modify the routing structure of others, and therefore their network management messages should be strongly authenticated. Authentication can be approached by various methods. The network can be setup such that each node has a unique identification that carries throughout the network, or authenticates via cryptographic methods. Additionally, authentication can be performed through a third party, or perhaps designate some nodes as a semi-authority. Node management traffic propagating from the host down to the nodes should always be authenticated, as it has the power to modify the behavior and properties of individual nodes. If more than a single base station is present then the means for authenticating each should be present in each individual network node.

5.6.4 Timestamps

An important security concern is in identifying anomalous behavior present on a network. Authentication provides a strong level of protection against erroneous membership and data, but does not prevent a denial-of-service attack. In reality there are two types of DoS attacks: (1) when data is never routed to the base station, and (2) when network traffic is so flooded that it cannot successfully propagate to its desired endpoint. Timestamps give an accurate representation of the lifecycle of data as it traverses the network, and can be used to identify problems in specific areas of the network topology. Intelligent filtering must be used to determine what an acceptable traversal time is, and subsequent modification of data routing will avoid and perhaps remove possibly compromised nodes.

5.6.5 Fault-Tolerance

In the event of node failure there must be concrete means of recovery for the network. The following scenario provides a concrete example: take a self-organized network of nodes with each following a standard routing algorithm. If a single node fails or is compromised and the routing tables fail to update in a timely fashion, the resulting network behavior can be used to the advantage of an attacker. The first concern lies with the traffic that is lost to that node or waiting for the node to respond; secondly, it is possible for a new malicious node to come online during the time delay of the network that effectively emulates the original node. The latter problem is prevented by use of strong cryptographic authentication. Timeouts should also be placed on message transmission such that a node failure will be almost immediately recognized and time/resources will not be wasted on transmitting data to the absent or possibly malicious sector of the network. Therefore, a routing structure should be self-modifying and persistent in order to be fault-tolerant enough to discover possible network anomalies.

5.6.6 Routing

One of the most heavily addressed issues in sensor network security is routing. A secure routing protocol can create a high level of operation security in a sensor network. Theoretically, a secure routing protocol should guarantee the integrity, authenticity, and availability of messages in the presence of a malicious entity [1]. It is very difficult to protect against inside attacks by members of the network, and as a result provisions must be included to prevent improper admission to the network by untrusted nodes. Protection from eavesdropping and data replay must be handled within other security provisions as they are better addressed in the application and link layers (of the network stack). Due to the fact that most routing protocols in sensor networks are designed to be as simple as possible, they are more susceptible to attack than general ad-hoc routing protocols. The following list of known attacks against routing ad-hoc schemes, and will allow for design of a secure, robust routing scheme for SDAC [1].

5.6.6.0 Spoofed, altered or replayed routing information

The most direct attack against sensor routing targets the routing information exchanged between sensor nodes. If an adversary can spoof, alter or replay information they could possibly create routing loops or attract/repel network traffic, extend or shorten source routes, generate false errors, partition the network or increase latency [1]. All of these can be catastrophic to the sensor network as they will greatly increase power consumption or cause the network to function in an unintended fashion.

5.6.6.1 Selective forwarding

Often it is assumed that a node will forward received messages when necessary to the best of its ability. Selective forwarding is an attack wherein a malicious entity (which may or may not be a node) selectively chooses which traffic to forward. An extreme case of this is when a node forwards no traffic and acts like a *black hole* within the sensor network. This attack is unlikely, however, as neighboring nodes will view the malicious node as broken and reroute traffic around it. It is much more damaging when a node decides to hold onto traffic from select nodes, rendering them useless within the network. This sort of attack likely seems only feasible by an insider; however, an outsider can perform this attack by intercepting traffic over the radio and jamming the channel between the nodes, acting as a third node in the situation. It is much more likely that an attacker will try and insert themselves as members of the network to perform this attack, as the jamming and intercepting of signals is considered very difficult.

5.6.6.2 Sinkhole attacks

The goal of a sinkhole attack is for an adversary to lure as much traffic as they can from a set of nodes through a compromised node, creating their own central routing and monitoring point on the radio network. This is generally accomplished by making a compromised node (or third-party within radio range) look like the highest probable next target in the routing algorithm. For example, a compromised node could lie about the quality of its radio link to the base station to lure traffic in its direction. If an attacker had a higher-power machine he/she could actually provide a powerful link to the base station and advertise it to the individual nodes in an attempt

to lure their traffic through the malicious channel [1]. A sinkhole attack makes the attack in 1.6.2 trivial, as they can selectively modify and/or deny traffic as it is routed through the sinkhole.

5.6.6.3 Sybil attacks

In a Sybil attack, a malicious node creates multiple "identities" for itself on the sensor network. Other sensor nodes have no idea that the multiple identities they are aware of are actually a single physical node. This greatly decreases the fault-tolerant nature of multipath routing as there is both a greater chance that traffic will be routed to a malicious node, and less diversity between physical nodes on the network. On another occasion, a malicious node may create multiple geographic identities for itself in case it is discovered. That is, it will store multiple coordinate systems for itself if it is discovered such that the malicious node can appear to be in multiple locations or possibly multiple nodes. [1]

5.6.6.4 Wormholes

Wormholes are a method of achieving a sinkhole attack from within the network. During a wormhole attack, an adversary tunnels messages received in one location of the sensor network over an alternative link and replays them in a different location. For example, a single node is situated between two other nodes and is forwarding messages between the two of them. Now, imagine that the two nodes are actually communicating with two geographically distant malicious nodes with a discreet communication side-channel. The two malicious nodes can lie about their relative distance to each other to ensure that communication passes through them, and all traffic from two geographic locales will travel through the two malicious nodes. An adversary situated close to a base station can disrupt normal routing by placing a wormhole such that it can convince nodes that are multiple hops from the base station that they are much closer. The resultant effect is a sinkhole on the sensor network. [1]

5.6.6.5 HELLO flood

The HELLO flood was first introduced in [1], and is based on the following idea: many routing protocols require nodes to broadcast HELLO packets to announce their presence to nodes within radio range. Nodes receiving the HELLO packets generally assume (not always correctly) that they are within normal radio range of the sender. Take, for instance, the case when an outside attacker has a powerful machine capable of broadcasting strong radio signals. They could then send a HELLO packet to each machine on the network broadcasting that they are just one hop from the base station. Each node would, in turn, send packets to the adversary to forward to the base station. In most cases, these packets would be lost to the channel and the resultant network would be completely ineffective. Even worse, if a node were to figure out the ruse and re-route to a closer node, the next node may be still broadcasting to empty air.

5.6.6.6 Acknowledgement spoofing

Generally, routing algorithms used in sensor networks rely on acknowledgements at the link layer. Given the wireless broadcast medium, an adversary can intercept messages with relative ease and spoof acknowledgements for overheard packets transmitted to neighboring nodes. In doing so, an attacker can convince a sending node that a weak link is strong or that a dead node is still available. Since packets sent in either of these situations should be lost, the adversary can mount a selective forwarding attack using acknowledgement spoofing. [1]

The above attacks give a general identification of routing security problems faced by a sensor network. Once again, it is important to stress that a secure routing scheme cannot protect against insider attack, and serious consideration must be given to preventing unauthorized network membership.

5.6.7 Power Considerations

When adding security mechanisms to a sensor network the most important affect is increased power overhead. Additional computational and data complexity to modern computing devices with their vast amount of resources is generally not of concern, but the contrary is evident within sensor networks. Any enhancement of security will increase the resource usage of a system proportional to the amount of security added, and careful consideration must be given to the tradeoff of security versus resource consumption. The most expensive operation on a sensor node is radio communication, where the cost per bit is vastly increased over regular computation. It is important to note that the encryption of a data segment will not increase the size of the data being transmitted, and therefore additional security will not increase communication overhead directly. Indirectly, the need to maintain and update cryptographic primitives and perform authentication between nodes will increase the amount of necessary network traffic, and therefore increase the power consumption of the average node. It remains to be analyzed exactly what the optimal solution is in regards to key maintenance and authentication versus power consumption. Computational power requirements are generally much less than those of communication, but must still be considered if a high amount of computational resources are required for cryptographic operations. Therefore, it is necessary to find a cryptographic algorithm that meets the requirements for minimal computational (and indirectly, power) overhead within a sensor node. The choice of cryptographic algorithm and implementation depend on the sensor architecture used, as various architectures may have separate specifications regarding available resources.

5.7 SDAC Specific Security Issues

SDAC, though subjected to the aforementioned security provisions, contains a unique set of problems due to its modular architecture. It may be irresponsible to look at only the above problems of general sensor networks assuming that they cover the gamut of possible situations after deployment. Additional properties of SDAC must be identified that can have impact on its security in the field. It is convenient to break up security into three separate domains based on its respective fundamental functionality, with some overflow from one into the other as well as analyze each situation in regards to our power and computational limitations.

5.7.1 Hardware

The desired environment for deployment of the SDAC sensor network is, in essence, highly malicious. Both nation-borders and urban warfare environments (MOUT) are rife with adversaries that will benefit from acquiring the knowledge and means to subvert any protections provided by the sensor network. It will likely be impossible to protect against an adversary learning the network infrastructure and thus gaining the ability to subvert sensor via avoiding their geographic locale. This is a deployment aspect of SDAC and not within the scope of security. Rather, concern lies with an adversary gaining specific knowledge about the sensor nodes such that they can modify or use them for malicious purpose. Earlier this concept was introduced as "technology exposure," where it is desirable for SDAC to be engineered in such a way that this exposure is minimized.

Given that the hardware design of SDAC relies on modularity (which provide a number of benefits in a sensor environment) it is susceptible to a greater chance of side-channel attack and exposure of key architectural and data components.

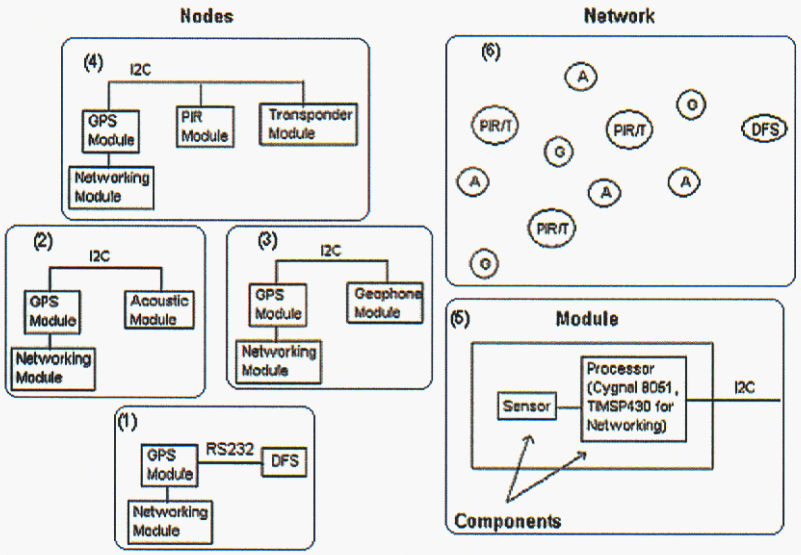


Figure 22: Hardware Level Subsections of SDAC

The above diagram in Figure 22 illustrates the individual architectural subsections of the SDAC architecture. Each component is subject to individual security analysis, which will help develop optimal solutions in regards to resource consumption versus strength. Examine the module-level diagram illustrated in Figure 23.

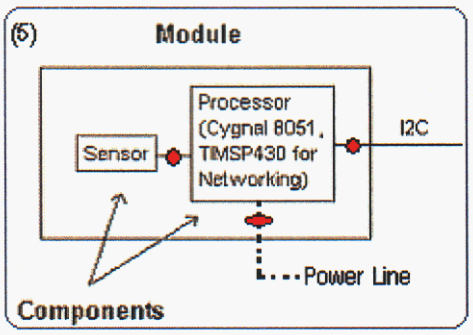


Figure 23: Individual Module Points-of-Attack

From this three main points of attack are identified: the sensor connection to the Cygnal 8051, the power line from the 8051, and the outgoing I²C bus. Additionally, though not shown above, are the ports that perform the programming of the processor. The sensor connection to the main processor is designated as a point-of-attack due to the fact that it can be manipulated to produce erroneous sensor readings that work in the favor of an adversary. It is infeasible to protect this data line via encryption (as it has no processing power), and the only possible existing solution is to integrate the sensor into the processor packaging such that the line of communication is not physically exposed.

The external power line from the 8051 to the power source exposes individual modules to an advanced attack wherein one analyzes the power consumption of the device in order to extract information. This is an extremely difficult problem to solve, and requires one to attempt to mask the statistical power consumption over various program segments and cryptographic operation as identified in [2]. One positive note is that this attack is non-trivial and requires a high level of knowledge by an attacker and specialized equipment (meaning that it is not easily implemented in the field). However, it is important to assume that an attacker is easily as knowledgeable as the system engineers and has the technological resources required. Therefore this attack it identified as a major exposure of information by the device. There are some provisions that can be developed for a device to reduce its susceptibility to such an attack, and are identified in [3,4,5,6,7]. The outgoing I²C bus is a major focal area of attack, but as it generally applies to a higher abstraction layer it will be covered at a later time.

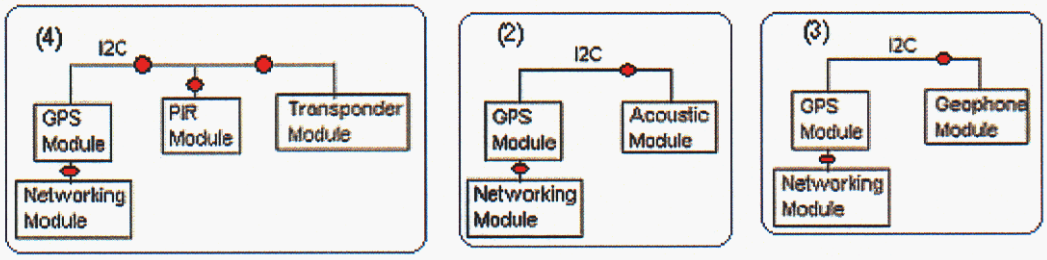


Figure 24: Nodal Points-of-Attack

Individual node types are covered within subsections 1-4 in. Of these, (1) is considered to be in the possession of a trusted entity as it is the control point of the overall sensor network and

not subjected to the same scrutiny (in regards to security) as the others. Figure 24 shows nodal types that arouse concern and the focal points for attack. It should be identified that the attacks occur on the Inter-IC Bus (I^2C), which is the main communication channel between the individual modules. Using an attack commonly called "bus sniffing" one can easily (assuming access to necessary hardware) discern the bus communication protocol as well as the contents of any data being transferred. Bus sniffing requires that an attacker construct some sort of a *tapboard* that will read the signals from the individual bus lines, and then process them in such a way as to extract useful information. A successful method of performing this attack was exhibited on the Microsoft Xbox™ in [8], where the shared secret key for encryption of hardware communication (as well as other information) was extracted from a single unprotected bus line. This method of attack leads to a plethora of security vulnerabilities such as data and key extraction, replay, man-in-the-middle attacks and malicious module insertion. It should be obvious the implications of data and key extraction in regards to the security of the sensor network, especially if a shared, static key is used across all nodes. Simple knowledge of the bus communications protocol and perhaps the secret key (if used) makes it possible for an attacker to create a malicious node module which can provide them with a means of joining the sensor network. For example, a simple attack would be to construct a faux sensor module that acted like a trusted module but rather repackaged and communicated readings and inter-node information to a malicious entity (essentially acting as a pseudo base station with much less functionality). Such a node, given proper information, would appear as a trusted sensor node. Another example occurs when an erroneous module is inserted into a node that simply produces incorrect readings in order to trick the base station into incorrect action. This could have dire effects on the quality of the sensor network, and could compromise its integrity as a whole.

Additionally it is important to note that the exposed pins on the processing board will impose a serious security threat. Simply stated, they should be removed in such a way that an attacker entity cannot use them. This means covering them in some material whose removal will break the board, or simply breaking the pins in such a way that their signals cannot be read and/or modified.

5.7.2 Software

It is vitally important that SDAC be concerned with the security of its running software for trusted operation. The extraction and subsequent modification of software within the nodes would have disastrous consequences in regards to the trust model of SDAC. With the addition of security enhancements to limit technology exposure on the hardware level, it should be sufficiently difficult to determine the functional purpose of software as well as the data contained within. However, if an attacker were able to extract the software from the processing component of each module, then they could theoretically recreate a node with their custom (and likely malicious) attributes. To protect against the extraction of software, it will be necessary that node-level programs be stored in protected flash memory such that reading its contents is difficult if not impossible. Mechanisms should also be included for monitoring the execution integrity of the software. This can be accomplished via a concept known as execution tracing outlined in [9]. The fundamental idea is that a method is used to compare the execution path of the current code with that of one that is expected normal code behavior; if they are not equivalent the code is labeled as tampered and execution is refused. This provides no means of recovery

from tampering, but prevents successful operation of a node in the event of tampering. It will remain important to assume that this will add significant computing overhead to the module processor, and correct amount of code to monitor in order to achieve a desirable level of security without sacrificing too much on power consumption must be determined.

5.7.3 Wireless Communications

Node-to-node communication is perhaps the most diverse area in regards to possible points of attack. It is simplest to say that the wireless channel is extremely vulnerable and difficult to protect. However, there are specific attacks that are at least feasible to protect against within the SDAC sensor environment. For example, all node-to-node communication can be snooped on in plain form as it is transmitted via radio link. However, not all traffic need be protected and a certain level of information leakage can be accepted. That is, only data and critical network administration information should be protected against snooping by a malicious third party. This reduces the amount of overhead required by security mechanisms as they are not always necessary. The practical solution needed in order to protect radio communications is the application of a fast, symmetric block encryption cipher. In our case, AES (Advanced Encryption Standard) should be used as it is optimized for small memory and computational power. See Appendix, Section 8.2 for a detailed overview of the AES algorithm.

The symmetric nature of the AES algorithm allows for utilization of half its functionality on any given node. That is, if on the Data Fusion System (DFS) or root node, any data sent to all child nodes we need only use decryption (as it has a larger memory and computational footprint than encryption). This is due to the fact that if the decryption function is applied to a data segment followed by the application of the encryption function, the result will be the original plaintext value as if the inverse method were applied. This method requires the same key be used for decryption and encryption. Given that the DFS or root node should have some extended computational resources beyond the individual nodes, the more intensive decryption operation is used on it exclusively.

In addition to data protection, the use of AES allows for authentication of individual network nodes as members of the sensor network. This is due to the fact that only authorized members will have access to the proper encryption/decryption key (theoretically) and only authorized members will be able to perform encrypted communication. Using Figure 25 as an example, if node A is a member of the network and wants to initiate communication with node B, whose membership is unknown, they can perform the following. Node A requests authentication from node B. Node B knows the authentication message, which could be a random seed similar to a key that is distributed by the root node. Node B then encrypts the authentication message with the secret key and sends it to node A. Node A then decrypts the messages and checks it against their authentication message. From that point it is a simple pass/fail.

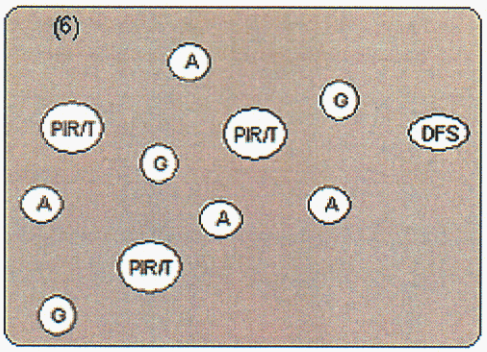


Figure 25: Network Points-of-Attack

Concerning routing (as was introduced in section 1.6), possible solutions can now be identified for some of the presented attacks. Many outsider attacks against SDAC routing can be prevented by simple link-layer encryption and authentication in the same fashion as was used to protect wireless data. A Sybil attack (1.6.4) is no longer a threat because nodes will not accept even a single identity for the malicious entity; it is authenticated each time communication occurs. Most selective forwarding (1.6.2) and sinkhole attacks (1.6.3) cannot occur as the adversary is prevented from joining the network topology. Link-layer acknowledgements (1.6.7) are also authenticated such that they cannot be spoofed. Encryption and authentication will not solve all routing attacks such as wormholes and HELLO floods. Additionally, there are no provisions to protect against a malicious entity that is already a trusted member of the network (or has achieved that position through careful reverse engineering of a node). More advanced solutions do provide us with a positive solution, however. For example, the best defense against a HELLO flood is to verify the bi-directional characteristic of a link before taking action on a message received. Geographic routing is a positive direction for protection against wormhole and sinkhole attacks. For further information the reader is referred to [1].

A more technical analysis should be given to the actual radio broadcast method used, as some modulation schemes provide greater security than others. Concern is focused at Layer 1 (Physical Layer) of the network stack, with a focus on detection, interception and jamming (resistance against) characteristics of the different modulation schemes. It is out of the scope of this document to cover all modulation schemes, so coverage is restricted to the best candidates in regards to security. The two best methods to use are Direct Sequence Spread Spectrum (DSSS) and Frequency Hopping Spread Spectrum (FHSS). For more on other modulation schemes the reader is referred to any digital communications textbook. DSSS spreads a given transmission signal over an allowed band, and is modulated via a random binary string called the spreading code. Both the transmitting and receiving ends must have the same spreading code. Due to the fact that DSSS spreads a signal over a wide band, it can recover fast from narrowband interference. From a security standpoint, DSSS is very desirable in that it generally appears as noise to traditional radio signals and is hard to detect/intercept. FHSS hops in a pseudorandom sequence between frequency sub-channels on which it transmits short bursts of data over a period of time before moving on. Senders and receivers must both know the pseudorandom sequence and be synchronized with each other for successful data transmission. Given the fact that the used frequency is always shifting in a pseudorandom fashion, FHSS is rather insusceptible to interference and interception. This joined with the fact that an attacker must jam

and entire band to successfully break communications makes FHSS the more secure scheme of the two. An added benefit of FHSS is that there are fewer collisions with nearby networks operating on the same band (which in the current application would mean that SDAC *A* would not interfere significantly with SDAC *B* in close proximity, if they were to be kept as separate networks). An interesting and highly desirable solution is in Ultra-Wideband (UWB) transmission. UWB does not use modulation for data transmission, but rather uses impulses that carry data in their timing or presence. UWB is perhaps the most secure in regards to signal detection, but is still in development and research phases and is therefore not feasible until it becomes standardized.

If the application of error correction/detection schemes within the communication layer of SDAC is necessary, it will add an additionally beneficial level of security varying with the rate of the encoder. That is, for every m data bits being transmitted we will also have n coded bits being transmitted (where n/m is the rate of the encoder) which will add additional complexity to the data stream. It will be difficult, given that an attacker is not synchronized with the transmitting entity, to distinguish between coded and data bits. This is similar to encryption, but would occur in addition to encrypting the data. It is important to note that adding n coded bits for each m data bit to the wireless communication channel will increase the necessary power consumption (based purely on transmission) by a factor of $(n + m)/m$.

5.7.4 Power Considerations

SDAC sensors must perform for long time periods on as little battery power as possible, the additional power consumption required of any security provision is of the utmost importance. All the aforementioned security provisions are minimal in that they consume the least amount of power while providing the highest level of security when compared with their peers. Of course, stronger security enhancements such as public-key cryptography can be pursued, but are likely unfeasible due to the high increase in power requirement.

5.8 Chapter 5 References

[1] C. Karlof and D. Wagner. "Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures". http://citeseer.nj.nec.com/576488.html

[2] P. Kocher, J. Jaffe and B. Jun. *Differential Power Analysis*. Lecture Notes in Computer Science, 1666: 388-397, 1999. http:// citeseer.nj.nec.com/kocher99differential.html

[3] C. K. Koc and C. Paar (Eds.). *Proceedings of Cryptographic Hardware and Embedded Systems (CHES)*. Lecture Notes in Computer Science, 1717. 1999.

[4] C. K. Koc and C. Paar (Eds.). *Proceedings of Cryptographic Hardware and Embedded Systems (CHES)*. Lecture Notes in Computer Science, 1965. 2000.

[5] C. K. Koc, D. Naccache and C. Paar (Eds.). *Proceedings of Cryptographic Hardware and Embedded Systems (CHES)*. Lecture Notes in Computer Science, 2162. 2001.

[6] B. S. Kaliski Jr., C. K. Koc and C. Paar. *Proceedings of Cryptographic Hardware and Embedded Systems (CHES).* Lecture Notes in Computer Science, 2523. 2002.

[7] C. D. Walter, C. K. Koc and C. Paar. *Proceedings of Cryptographic Hardware and Embedded Systems (CHES).* Lecture Notes in Computer Science, 2779. 2003.

[8] A. Huang. "Hacking the Xbox". No Starch Press, 2003.

[9] G. Vigna. *Cryptographic traces for mobile agents.* Lecture Notes in Computer Science, 1419:137-149, 1998. http:// citeseer.nj.nec.com/vigna98cryptographic.html

[10] *Advanced Encryption Standard (AES).* National Institute of Standards and Technology: FIPS 197, 2001. http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf

6 SDAC Demonstration System

The development of a demonstration system was part of the three main objectives of this LDRD. Due to the brief development cycle the team and program managers established some initial constraints to scope the demo system development. These constraints included: (1) using established hardware; (2) consulting with experts on validity of demo scenario; (3) conceptual demonstrate that reflect the SDAC vision of smart wireless sensor networks. This chapter covers the details of the SDAC demo system and other exploratory system development, like cameras.

6.1 SDAC network for MOUT

A distributed wireless sensor network, based on the Sense, Decide, Act, Communicate (SDAC) framework, provides a method for collecting real-time knowledge on otherwise intractable systems. The distinction between knowledge collection and data collection is important because it implies the engineering of an *embedded intelligence*, *collaboration*, and *data reduction* in the network itself rather than on any centralized unit. This *decentralization* makes the network more robust against single point failures, and thus more adaptable to the quickly changing conditions faced in MOUT. Making the network "smart" also allows the possibility of complex and *multi-modal sensing* which can decrease false alarm rates while increasing confident dependency on proper operation. Finally, the removal of a central computational unit decreases the amount of wireless traffic necessary between the nodes, and this extends network lifetime with *lower power* operation while also decreasing the probability of wireless transmission detection by opposing forces.

The application of SDAC sensor system will be derived from some of the concepts (italicized phases) provided in the prior paragraph. The actual scenario mission space was determined during a group meeting in NM with other Sandians, who felt MOUT provided the most interesting potential for a conceptual demonstration. The scenario was based on survey information of the MOUT domain, ongoing war situation in the Middle East, discussions with DOD personal during Washington DC meetings, and discussions with knowledgeable Sandians in the area of MOUT. This section provides details about the MOUT scenario, a brief evaluation of potential hardware and highlights of the SDAC conceptual issues used in this demonstration.

6.1.1 Overview of demo scenario

One of the primary issues that face the military in MOUT situations is the inability to identify an entity as friend, foe, and non-combatant (IFFN). MOUT environments are inherently complex and densely populated with sight and communications obscuring structures. Along with the additional complications of dynamic terrain and difficult mobility problems associated with MOUT, securing battlefields and assuring enemy retreat poses a challenging problem. This lack of battlespace awareness, at all levels of control, can lead to the development of dangerous operational chaos. The close proximity of forces and extremely short engagement ranges in this already problematic arena creates the further hazard of not being able to quickly and easily distinguish between troops from friendly and opposing units and civilian non-combatants typical in urban areas. This confusion increases the possibility of fratricide and civilian casualty, both unacceptable consequences of any military operation.

The IFFN demonstration will consists of an array of SDAC sensors spread over an area to detect the proximity of any movement within the field of sensors. The sensor array will contain three categories of sensor (1) passive infrared (PIR -120 Degrees), (2) acoustic and (3) geophone. Data from the other two sensors will be fused on the nearest PIR sensor to reduce the number of false positive indications of movement in the sensor field. The correlation of events coming from these orthogonal sensing sources is imperative in order to reduce the false alarm rate of the network. If the acoustic sensor detects a distant gunshot but the PIR sensor detects nothing, for example, no proximity event will be registered. If an event is detected and validated on any particular node, the node will query the presence with an RF transmission. Each friendly asset, both human and material, will carry an RF transponder tag designed to receive and respond to queries from the nodes in the network. This query and response between the network and friendly asset will categorically identify (e.g. tank, personnel, etc…) the asset to the network, and along with GPS coordinates provided by another sensor on each node, allow the network to localize friendly assets. If a node detects a presence that does not respond to its RF query, the presence will be assumed to be non-friendly. Since non-friendly could imply either foe or non-combatant, there is ambiguity as to the identification of the non-friendly presence. In future developments of the system, an imaging device may be integrated onto a specialized node in order to allow a user to visually identify the presence.

The localization information of both friendly and non-friendly presence will be passed to a laptop (or command center) where assets can be displayed and tracked on a GIS map. The external collection of data from this network will not have knowledge of the details of the network implementation, and so the network itself effectively becomes the sensor. The demonstration will incorporate aspects of data fusion, data reduction, event detection and validation, and multi-modal sensing in order to provide a situational awareness (SA) and identification of friend, foe, and non-combatant (IFFN) aid in a military operations in urban terrain (MOUT) environment.

6.1.2 Survey of potential hardware

The SDAC conceptual demonstration would require existing hardware due to time limitations on this 9-month LDRD. While the program managers made the final selection of the actual hardware platform used in the demonstration, the LDRD team provided a comparison of positives and negatives of some potential hardware systems. The combinations of systems being considered were Hybrid Emergency Radiation Detector (HERD) with Cygnal 8051, Crossbow Motes, iPaq with Crossbow Motes, Dust Inc. Motes, and HERD with Rabbit. We also considered the following other wireless sensor systems: Rockwell WINS – no longer obtainable, expensive; and Sensoria – very expensive and not for sale at this time.

HERD w/ Rabbit

High Performance, High Power. In-network intelligence can be built in, complex sensing (e.g. imaging) possible, computation could be distributed with some effort.

Cost = ~$400/unit, programmers and software development environment are already owned

Pros

- Sandia owned
- Units are flexible, customization easier than other platforms
- Units have ample memory and processing power
- Units can be made more power conscious

- Flexible platform to suit our needs
- Units are inexpensive
- Parts of the system software we would need already exists

CONS

- No external support required
- Internal support is highly limited
- Lacks documentation at this time
- Never been either fully or field tested

PDAs (Zaurus or iPaq?) with Crossbow Mote

Attached to lower performance devices like Estrin's iPaq/Crossbow platform?

Cost = ~$200, (already have several)

Pros

- They run Linux and are well supported
- They are very flexible
- They have a lot of memory and processing power
- They are designed around StrongARM processors
- They are operationally the highest performance system
- They have a display built in

Cons

- They are architecturally uninteresting
- They are very high power
- Networking would be 802.11 based or else we would have to build our own entire stack from physical layer up
- Pre-existing systems, application use only

HERD w/ Cygnal 8051

Low Performance, Low Power. Limited in-network intelligence, complex sensing (e.g. imaging) very difficult or impossible, computation would likely be mostly centralized

Cost = ~$400/unit

Pros

- They are very low power
- Sandia owned
- Units are flexible, customization easier than other platforms
- Units can be made more power conscious
- Flexible platform to suit our needs
- Units are inexpensive
- Parts of the system software we would need already exists

Cons

- They have limited memory (2k RAM) or processing capability
- No external support required
- Internal support is highly limited
- Lacks documentation at this time
- Never been either fully or field tested

Dust Inc. Motes

Cost = $15k for software development environment, 5 gateways, 3 programmers, evaluation kit, configuration utility, and 100 motes.

Pros

- They are a better version of the crossbow motes
- They are very architecturally interesting
- Dust Inc. will support their products well
- They have some built in routing protocols
- Very cost effective, and could be used for future development
- Low power

Cons

- First customers to receive the SDK
- Limited memory (2k RAM) and processing power
- The OS is closed source
- They are based on a distributed data base query system making any other actions, e.g. in-network processing, difficult to program
- They are designed to be a centralized system, not to do any in-network computation

Crossbow Motes

Cost = $1500 for 4 motes, 3 sensor boards, 1 programming board.

Pros

- They are well-known platforms that have been used widely
- We are close to Berkeley where we may be able to find support for them
- They are low power and architecturally nice
- The OS is open source

Cons

- They are not very flexible or easily customizable
- They have limited memory (4k RAM) and processing power
- Crossbow itself does not provide much support, and Berkeley support is not guaranteed
- They are designed to be a centralized system, not to do any in-network computation

The program managers selected the HERD units as the demonstration platform, due to a few different variables. The main factor was the prior knowledge of the LDRD team with the HERD unit. The team also saw this as a chance to extend the HERD platform and test it out in a domain for which it was not initially envisioned.

6.2 SDAC demo hardware and software architecture

<< This section is based on text taken directly from Douglas Stark and Jesse Davis, "Friendly Object Tracking and Foreign Object Detection and Localization", SAND2003-8736C. >>

The Data Fusion System (DFS) is the user interface to the SDAC system. The DFS collects the type and location of each node in the network and displays that information on a Geographical Information Systems (GIS) viewer, which shows the location of nodes against geographical features such as roads, waterways, and terrain. Figure 26 provides a picture of the initial GIS viewer with two little nodes centered around the green area on the screen.

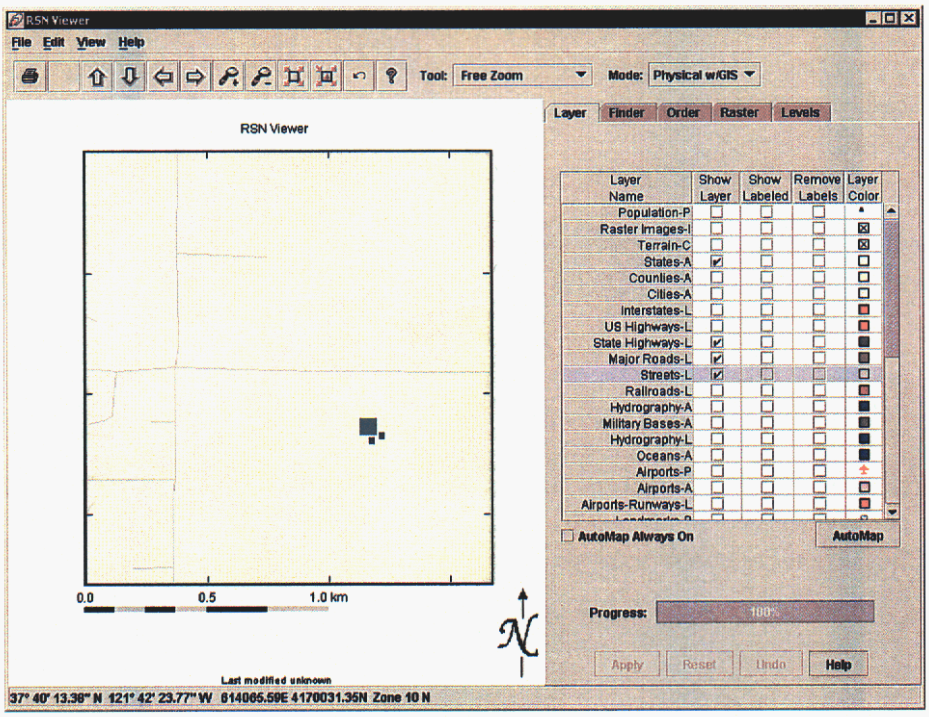


Figure 26: GIS viewer showing a PIR, microphone, and geophone node and roads.

Events are displayed on the GIS viewer in real-time at the location they were detected. Figure 27 illustrates the blue icons on the GIS viewer indicating events generated by friendly objects, while red icons represent events generated by foreign objects. The icons persist on the GIS viewer as long as the object continues to generate events.

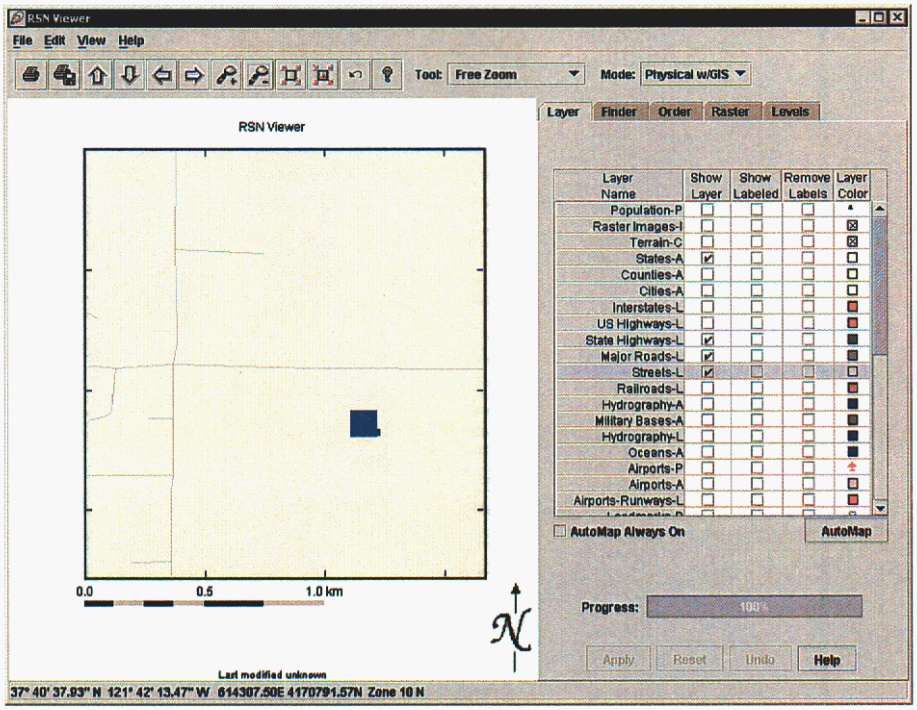


Figure 27: GIS viewer showing a friendly event.

The DFS stores events, which contain the ID numbers of friendly objects that were present during an event, in a database. The DFS software can search the database and provide tracking information about friendly objects, which can then be rendered on the GIS viewer to show the path of a friendly object as it moved through the network.

The SDAC network contains four types of nodes, shown in Figure 28: microphone, geophone, passive infrared and transponder (PIR/T), and a gateway node that provides a connection to the DFS. Another part of the system, operating on a different frequency and independent of the network, is a reply transponder located on each friendly object.

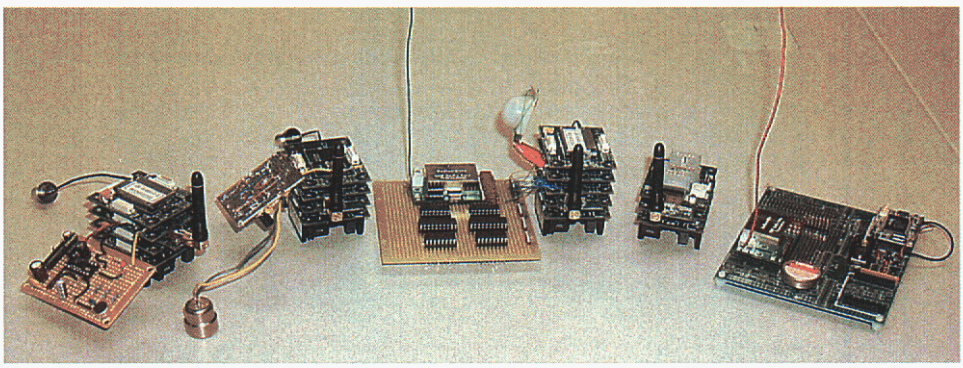


Figure 28: A microphone, geophone, PIR/T, and gateway node and reply transponder (left to right).

The system begins operation by establishing an ad-hoc wireless network among the nodes. Once established, the network detects sensor events and performs sensor fusion on the events. Since the PIR sensor has a finite range of only 10 to 30 feet, and the geophone and microphone sensors have ranges dependent on the magnitude of the events they are sensing, the PIR/T nodes were chosen as the sensor fusion nodes of the system. The PIR/T nodes store their own GPS-time stamped sensor events as well as receive wireless communications containing GPS-time stamped sensor events from neighboring microphone and geophone nodes. The PIR/T node applies sensor fusion rules to the events as they are received to determine if the sensor events constitute a verified event. A verified event is defined as a PIR event and a microphone event or a PIR event and a geophone event that occur within one second of each other. This fusion of two orthogonal phenomena decreases the probability of false alarm in the detection of events. When a verified event occurs, the PIR/T node broadcasts a transponder query to its local area and collects friendly object replies from the object-worn reply transponders. The time and location of the verified event, defined as the location of the PIR/T node and the time of the PIR event, and the ID numbers of the friendly objects that reply are packaged into a message and sent to the DFS, which displays the events in real-time on the GIS viewer.

6.2.1 Module Descriptions

Each type of node in the SDAC network contains a different compliment of modules; Figure 29 shows the seven different modules. Microphone and geophone nodes contain a controller, a wireless networking module, a microphone or geophone sensor module, and a power supply module. The PIR/T node is similar to the microphone and geophone modules, but instead includes a PIR sensor module and adds a transponder module.

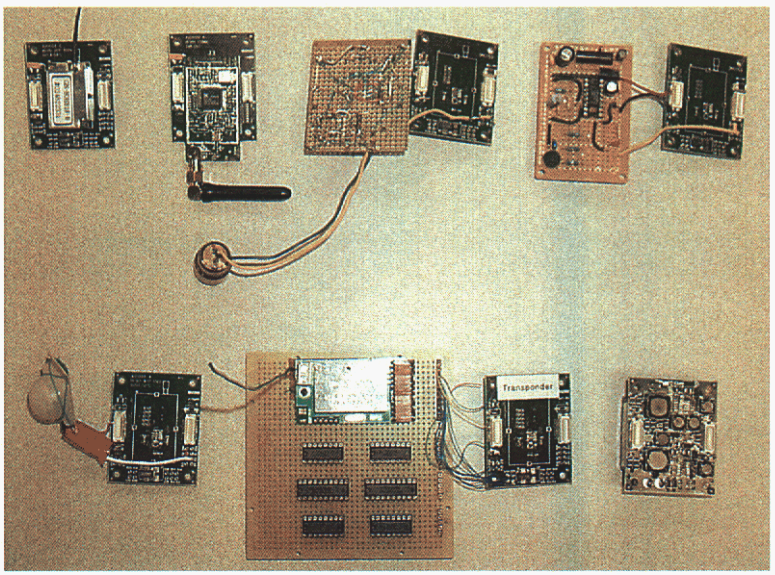


Figure 29: Controller, networking, geophone, microphone, PIR, transponder, and power supply modules (left to right, top to bottom).

On start-up, the controller module performs an automatic discovery of the other modules present in its node. Once the controller has established which modules are present in the node, it

automatically sets its own corresponding mode of operation so that the node can begin functioning appropriately.

The controller module serves as a central processor for the node. It receives events from sensor modules, determines if a verified event has occurred, sends query requests to the transponder, receives query replies from transponder modules, and operates a GPS receiver. The GPS receiver (a Furuno GN-80) is used to provide the controller with location data and to allow time synchronization among all the nodes in the network without the complication of a wireless communication-based time synchronization protocol. The controller obtains the time from the GPS receiver and distributes time synchronization messages to the other modules in the node. This allows sensor events to be time-stamped on the sensor modules as soon as they occur. In order to maintain network-wide synchronization in the presence of clock drift, the controller modules obtain new time and location data from the GPS receiver every five minutes.

The controller firmware is organized into software modules. The core of the firmware is five software modules and their APIs. The serial module handles communications with the networking module. The real-time clock module operates the real-time clock and provides an API for setting and reading the time. The I^2C driver handles communications with the sensor and transponder modules in the node. The GPS module operates the GPS receiver and performs time synchronization. Finally, the event decision software module handles sensor events, determines when a verified event has occurred, and handles transponder queries. Figure 30 shows how the software modules communicate.

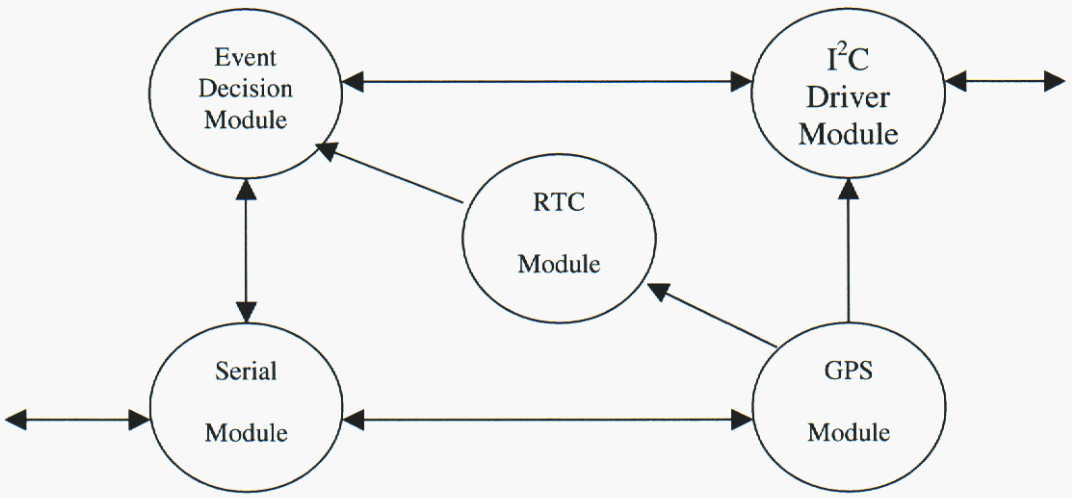


Figure 30: The controller software modules.

There are two conduits for information to move into and out of the controller module. The serial software module allows communication with the networking module. It sends events from other nodes to the event decision software module and handles requests for information from the DFS, such as requests for the node's location. The I^2C driver receives events from other modules in the node and sends the events to the event decision software module. The event decision software module in turn uses the I^2C driver to send query requests to the transponder. It also

sends verified event messages to the serial software module, which sends them to the networking module to send them to the DFS. The GPS module is the only module that sets the real time clock. It also uses the I^2C driver to send time synchronization messages to the other modules in the node. The event decision software module reads the real time clock when it sends a query request to the transponder module.

The event decision software module is one of the most complicated software modules in the controller. It has the task of storing events, checking for verified events, querying the transponder, generating service requests, and sending verified event messages to the DFS. In order to accomplish this task, the event decision software module was organized into a state machine shown in Figure 31.

The event decision software has three states. In the default WAITING state, the software simply stores events and waits for the events to combine to make a verified event. In this case, a timeout of "0" is used to signify that the software should never timeout while waiting for an event. Once a verified event is detected, the software issues a query request to the transponder, sets a 1.5 second timeout, and enters the QUERY state. In the QUERY state, the software looks for a query response message from the transponder. If sensor events are received in this state, the events are stored and the timeout is recalculated. When a query response is received, the software combines the verified event and the query response into a verified event message and sends the message to the DFS. If the software times out while waiting for a query response from the transponder, the state machine enters the SERVICE state. This indicates that the transponder failed to reply within its 1.5-second time limit. In this situation, the software sends a service request to the DFS to notify the user of a potential problem with the node.

The networking module handles all aspects of the wireless communication in the network. It contains a Xemics XE1202 900 MHz, 76.8kbps radio and runs a low-power MAC layer called Sensor MAC (SMAC) developed at UCLA[1]. SMAC time-synchronizes nodes and allocates time slots to each node, thus allowing nodes to duty cycle their radios. This greatly reduces power consumption, but has the consequence of restricting when and how often nodes can transmit. SMAC uses a neighbor list to track the nodes with which it can communicate. This allows for simple neighbor-to-neighbor communication and broadcast and unicast functionality. Routed communication is accomplished with a lightweight proprietary routing algorithm. The networking module encapsulates all wireless networking functionality. This keeps routed information from entering inter-module busses and reinforces the modularity of the system. Networking modules can also function as a gateway to the network for the DFS computer through a serial port and supporting hardware. The gateway node appears as a regular node to the rest of the network, but contains no sensors modules or controller.

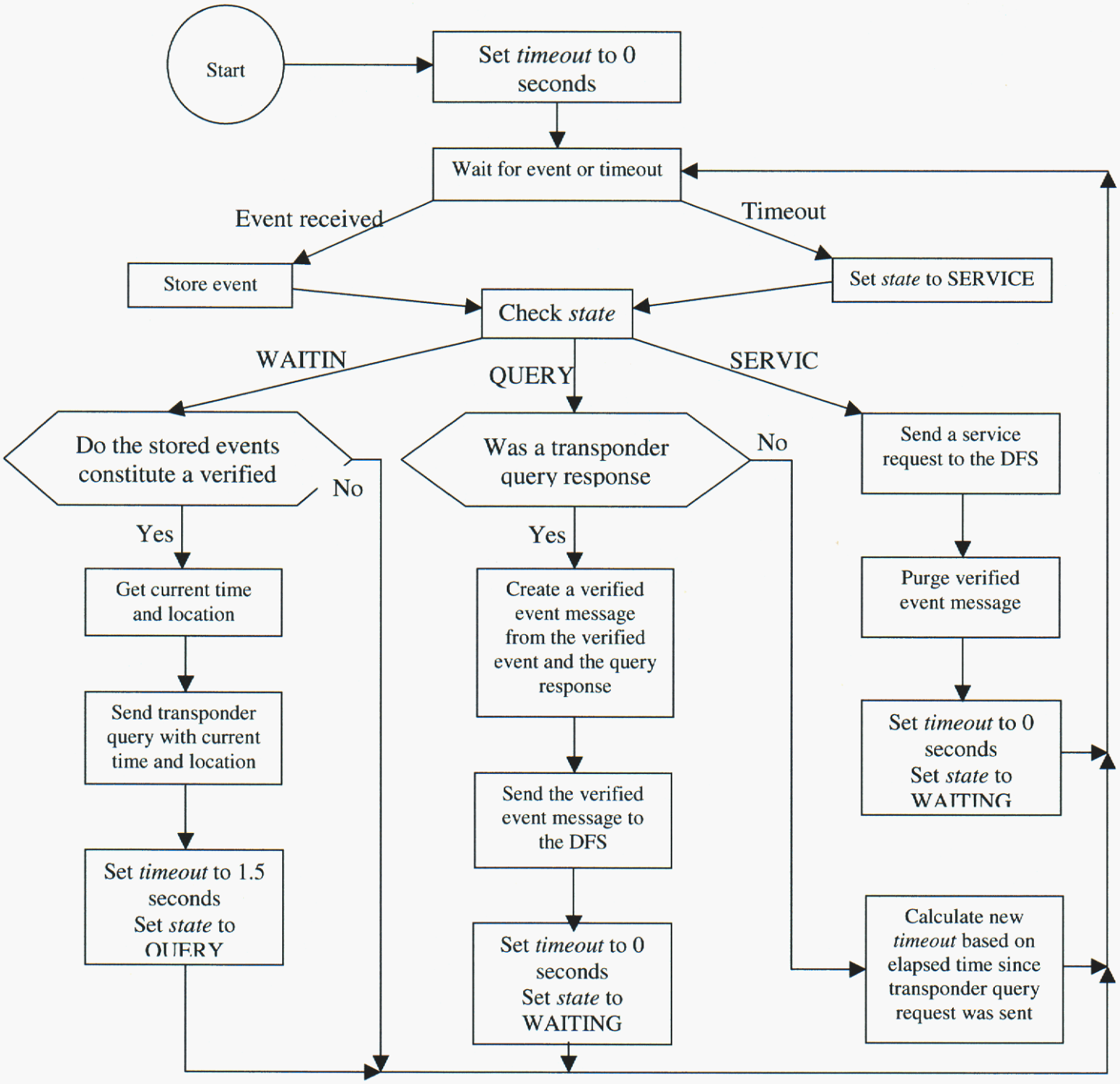


Figure 31: The event decision software module state machine.

Sensor modules (the geophone, microphone, and PIR modules) consist of sensor circuitry and a microcontroller. The sensor circuitry output is connected to a comparator on the microcontroller. A DAC on the microcontroller generates the reference voltage for the comparator. The reference level can be hard-coded, or the sensor module can auto-calibrate the reference level with a preprogrammed offset. If auto-calibration is selected, the processor samples the output of the sensor with an onboard ADC and calculates the average of the samples. The preprogrammed offset is subtracted from the average and the result is supplied to the DAC as the reference for the comparator. The microcontroller can be configured to interrupt on either positive or negative edges of the comparator output. This interrupt creates an event in the microcontroller software, which time-stamps the event with its internal real time clock, and sends the event to the controller module on its host node. If the node is a microphone or geophone node, the controller sends the event to the networking module. The networking module in turn unicasts messages to neighboring PIR/T nodes. If instead the node is a PIR/T node, the event is simply logged and a verified event check is performed. Following a sensor event, the sensor module disables the comparator interrupt for a programmable period to de-bounce sensor operation. This results in a programmable maximum event frequency for each sensor module.

The sensor module firmware uses parts of the controller firmware and maintains the same modular software architecture. The firmware, shown in Figure 32, is divided into the three modules: the I^2C driver, a sensor module, and a real time clock module. In the sensor module, the I^2C driver sets the real time clock when it receives a time synchronization message from the controller. The sensor software module detects sensor events, reads the real time clock and packages the event into a message. The I^2C driver sends the message to the controller. The only conduit for information to pass into or out of the sensor module is the I^2C driver.

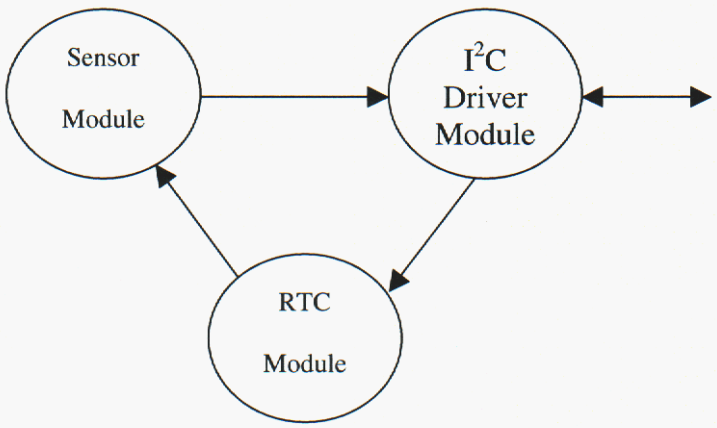


Figure 32: The sensor module software modules.

The transponder module is similar in functionality to a sensor module. It receives requests to broadcast a query from the controller and responds to the controller with a list of the friendly objects that reply to its broadcast. The reply list is empty if no friendly objects reply. The transponder module operates at 433 MHz, making its radio operation independent of the networking module. It broadcasts messages with a time and location stamp, and then listens for

one second for replies from reply transponders. The replies contain the query time and location along with an appended friendly ID. The transponder validates each reply against the time and location of its most recent broadcast, and if the time and location match, the ID of the replying friendly is added to a growing list of friendly objects. When the one second long period elapses, the transponder packages the list of friendly objects and the time and location of the query into a message and sends the message to the controller.

The reply transponder resides on friendly objects in the field. It listens for broadcasts from transponder modules. When the reply transponder receives a broadcast, it appends the friendly object's unique friendly ID to the original message and broadcasts this reply. The protocol and message structure used allow for future expansion of the information appended by the reply transponder. The system may thus eventually be used to collect information from friendly objects, such as physiological data or status, as they pass through the network.

The power supply module provides each module in the node with a 3.3V digital power supply and a 5V analog power supply. The power is drawn from two Lithium CR123 batteries in series. Two Linear Technology LTC3440 buck-boost converters create the 3.3V and 5V supplies from the six volts supplied by the batteries. The overall operational efficiency of the power supply board is about 80%. Filtering of the power rails is done both on the power supply module itself and locally on other modules in the node.

6.2.2 Platform Description

The sensor, controller, and transponder modules were designed around a Cygnal C8051F124. The C8051F124 was chosen because of its large amount of RAM (8 kB), large amount of ROM (128 kB), onboard analog components (two comparators, two DACs, and an eight channel ADC), low power consumption (.5 mA/MHz), and high speed (up to 50 MHz). This combination of features provided a processor that could run a real-time operating system (RTOS) and was configurable to meet the needs of a variety of modules while offering substantial computing power and maintaining relatively low power consumption. The networking module was designed around a TI MSP430F149. The MSP430F149 is an ultra-low power RISC microprocessor. Since the MSP430F149 has only 2 kB of RAM, there is not enough memory to run an RTOS on top of the MAC layer and routing algorithm. Consequently, the networking module uses a software scheduler and relies heavily on interrupts and timers.

The controller, sensor, and transponder modules incorporate an RTOS into their firmware. Micro-C OS-II (uC/OS-II) was chosen as the RTOS because of its low cost, flexibility, availability for many processors, and scalability. uC/OS-II is a multithreaded preemptive RTOS with many features well suited to embedded systems. The common environment makes firmware development easier and more consistent among different modules. Using an RTOS has maintenance benefits since it makes adding new functionality as simple as adding a new task. Interfacing with existing tasks is as simple as posting to existing queues and semaphores. The RTOS also helps to avoid redesigning software from the ground up for new modules by providing an existing software platform.

6.2.3 Implementation Issues

Several problems became evident while implementing this system. One of the most significant difficulties encountered while developing the firmware for the system was debugging the firmware while the modules were stacked into nodes. The hardware was designed in such a way that only one module in the stack could be debugged at a time. Furthermore, if the stack included modules made from the same hardware design (a PIR sensor module and transponder module for example), none of these modules could be debugged since they shared the same programming interface. Consequently, it was very difficult to debug the hardware in-circuit. In the future, a break-out board will allow the node to be assembled with the modules laid out flat. This will allow the modules to be debugged individually in-circuit.

The modular, stacked architecture was also difficult to probe and observe in action. Since the modules are stacked together very closely, it is difficult or impossible to access much of the hardware with oscilloscope probes. The break-out board will also help this situation because it will effectively eliminate the tight spaces created by the stacked modules. The shared I2C bus also introduced difficulties. Monitoring and analyzing the bus traffic proved nontrivial. The only available method was to watch the bus on an oscilloscope. While this was useful, an oscilloscope cannot store the bus activity for future analysis. The break-out board will also include a computer interface that will allow computer software to log and decode bus activity for future analysis.

An unexpected problem area was the power supply board and power distribution through the modules. The power supply was designed with adequate power filtering on the voltage rails, but the individual modules were not designed with local power filtering. This allowed noise from the modules to travel through the power rails onto other modules. This noise caused many problems, but was most evident in the analog sensor circuitry. In many situations, noise from the GPS receiver was actually being amplified in the microphone and geophone sensor circuitry, and causing the sensor module to register an event. In order to solve the noise problems, three solutions were implemented. First, LC filters were added to the networking boards. This greatly reduced the noise caused by turning the radio on and off. Second, larger bypass capacitors were added to the controller modules. The capacitors helped isolate the noisy GPS receiver from the rest of the system. Third, the analog sensor circuitry was moved to a power supply separate from the rest of the system. These three changes combined to greatly reduce the noise found on the power supply rails. The DC/DC converters on the power supply module also proved somewhat unreliable. While the cause is still under investigation, the end result was always the same: the converters would get very hot and cause the node to stop functioning. A more reliable power supply will be developed in the near future.

6.2.4 Results

In testing, the microphone sensor demonstrated the ability to detect loud voices at 10 m, the geophone was able to detect heavy footsteps at 5 m, and the PIR was able to detect a person at 10 m. Events with larger magnitudes, such as vehicles passing by or loud claps could be detected at greater distances. While the networking module radio can in theory transmit 300 m line-of-sight, they were limited to less than 100 m in practice. These facts combined to allow the system as built to detect people and vehicles as they passed through an area about 100 m^2 in size. In theory, the system could be expanded to include nearly 2^{16} nodes and cover hundreds of square

kilometers. In practice, however, the system is most likely limited to an area a few square kilometers in size due to the network and routing protocols used and the necessary density of the range-limited sensors.

6.3 SDAC Future Directions

6.3.1 SDAC Future Directions

Several task were identified, as important additions to the SDAC systems but were lower priority for the budget allocated. These are enhancements to the concept and are listed here as Future Directions. Alternate sensors added to the SDAC platform.

6.3.1.0 Tilt sensor

We have identified a small, low cost tilt sensor that would perform several useful functions including deployment, tamper detection, and movement detection. This sensor is useful to determine when deployment has been completed, (motion has stopped) as an indication that the system should enter an initialization phase used for enhancing network discovery, system configuration, and operation, saving power during pre-deployment. Note: a method of deployment detection is also required but this is another problem. Secondly the tilt sensor operates as a tamper detection circuit for either disabling or disarming the sensor. This is for use in hostile environments where the information stored within the node may need protection. Finally, the tilt sensor can be used as an assistance to the GPS to determine when GPS position may have moved and therefore an update is required. The net effect is that the position has not changed and therefore a lot of power can be saved by not powering up the GPS to re-acquire location. This would apply to non-hostile deployments where curious individuals may pick up a sensor or move it such as the HERD system deployed in a city.

6.3.1.1 Compass sensor

An inexpensive compass sensor has been identified that provides 45 degree orientation for those sensors that are directional such as PIR, Camera, Ultrasonic. With an integrated compass and three 120 degree PIR sensors(for 360 degree detection), one could establish not only the detection but the general direction of the detected motion. This would be useful when combined with GPS coordinates to validate and differentiate detections as well as identify friend or foe situations where friendly and non-friendlies are in the same area within transponder range.

6.3.1.2 Ultrasonic

An ultrasonic sensor was identified as a method to determine distance to a detected object. This is useful as an augmentation and sanity check for PIR as well as an enabling the capability for a smart mine or electronic fence. The concept here is this; as a non-friendly gets closer to the sensor, one could modify the response or alert to alarm the non-friendly to stay away or else a lethal or non-lethal deterrent will be activated.

6.3.1.3 Camera

We identified several potential imaging sensors, however the complexity with interfacing, buffering, and sending an image across the network would have consumed too much resources

and was therefore reduced in priority. This work however is continuing through another center. Imaging is particularly interesting because an image has a tremendous amount of spatially rich information. The problem however is that an imaging sensor is data intensive. One concept considered for data reduction is to extract as many of the details as possible from the image and send this descriptive information rather than the image. It is more desirable to try some out-of-the-box thinking about image processing rather that perform standard image processing compression techniques that are established concepts.

6.3.2 Other SDAC node modifications

- Photovoltaics (PV) cell based charging system –for SDAC systems of reasonably long life, standby power dominates the power requirement that in turn drives the physical size constraint. If a reasonable scavenged power source can be established, then life times become unlimited. Photovoltaic (PV) cells are the most readily available source.

- Uni-cast capability in the Network modules to only pass messages to relevant sets of nodes. E.g. Network modules will pass preliminary event messages only to PIR nodes in its neighbor list.

- Sensor power control – presently there is no way to power off redundant sensors. In the future one could imagine nodes configuring and powering off if the density is greater than required for a given mission.

- Develop diagnostics – The inherent added complexity of a modular SDAC system was unappreciated until the system was built. As a result, the need for a set of diagnostic tools became apparent late in the process. These include: a wireless network sniffer to diagnose traffic in the local area (this software has been written), a diagnostics breakout module so that modules can be made to flat so that they can be probed, and finally, an I2C bus diagnostics module. Building diagnostics as you go could be categorized as a lesson learned.

- Remote turn-on of sensor nodes - Sensor nodes in a wireless sensor network often extend their usable lifetime by taking advantage of low power sleep modes. One problem faced by sensor networks is how to wake up a node that is in a sleep mode. In general, only an event internal to the node can cause the node to resume normal operation. One possible solution is the use of a surface acoustic wave (SAW) correlator to receive a specifically coded RF signal. The SAW correlator turns the RF signal energy into an electrical pulse which can trigger an ultra-low power wake-up circuit internal to the node, thus causing the node to exit its sleep mode

6.4 Investigating Imaging in Distributed Sensor Networks

The unique characteristics of wireless sensor networks have the potential to revolutionize the way we sense the environment. Distributed sensors offer several advantages over the traditional centralized architecture including improved sensing resolution, robustness against failure, and increased adaptability. These advantages have made wireless sensor networks (WSNs) applicable to a diverse set of domains such as target tracking, environmental sensing, medical monitoring, machine diagnosis, and security systems. These systems consist of

scattered nodes that sense the environment, transform the data from sensors into information, and communicate with other nodes in their network.

The potential advantages of WSNs are dependent on the information we can extract from the network. Typical sensors used in research currently are one-dimensional sensors – temperature, passive infrared, geophone, acoustic, and etc. They work together across multiple nodes to describe events in the environment and report back to a user at a base station. Even though cameras provide potentially thousands of bytes of information about the environment, which are orthogonal to these other sensors, the integration of a camera has been hindered by the concern of power consumption and the restrictions in networking.

WSNs can be comprised of tens to thousands of unattended battery-powered nodes, which need to operate for extended periods of time. These power constraints limit the lifetime of a sensor unit and directly reduce radio transmissions, which represent the most power hungry function of a sensor unit and network. Local processing at the node level allows for a required bandwidth that is much less than bandwidth available. This has shaped the development of network protocols and data fusion algorithms, which sacrifice bandwidth for power conservation.

The advent of small low-power cameras typically called CMOS image sensors have made adding images to WSNs a possibility, allowing images to be captured with as little as 2 mJ. While many have recognized the benefits of obtaining a visual snapshot of the environment when and where there are events of interest, the design of the networking protocols in WSNs is not complementary to the needs of transmitting an image through a network. In order to effectively employ networks of visual sensors, one must devise a way to minimize power usage and network traffic while not losing any relevant data.

In Section 6.4.1, we will describe the available hardware and current networking stack and its constraints, and describe various means of creating a compact image representation.

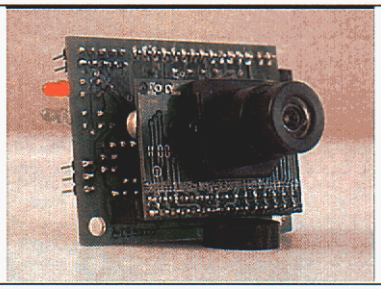
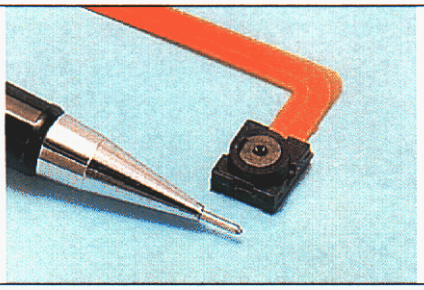
6.4.1 Hardware

We considered two CMOS Image Sensors – the OmniVision OV6630 and the Fujitsu MB86S02.

Table 10 highlighting some of the important features of both cameras. While the Fujitsu offers better power savings and a smaller form factor than the OmniVision, it also has less options in data format, image size, and video output. Because power and size are an overwhelming concern, we decided to interface the Fujitsu camera with the SDAC unit.

Table 10: Camera specifications

	OmniVision	Fujitsu

		
Pixel array number	7--- ,352x288, 176x144	352x288, 176x144
Video Output	4-bit, 8-bit, 16-bit Parallel	8-bit Parallel
Data Format	RGB, YCrCb, YUV	YCrCb, YUV
Command Interface	I2C bus	I2C bus
Pixel size	9μm x 8.2 μm	5.5 μm x 5.5 μm
Image Area	3.1 mm x 2.5 mm	1.96 mm x 1.61 mm
SN ratio	> 48 dB	45 dB
Active Power	< 66 mW	30 mW
Standby Power	< 33 μW	22.4 μW

For experimentation with possible techniques, the research discussed in this paper was done using the OmniVision camera. While the Fujitsu camera required a special board to interface directly with the camera, the Robotics Institute at Carnegie Mellon University had build a board to interface the OmniVision camera to a PC. This allowed for easy access to sample camera images for evaluation of different algorithms.

Like many of the small low-power sensor nodes being built, the HERD units have only sufficient memory to contain the operating system, program, a few packet sized buffer for the network communications, and some scratchpad space for data manipulation. It was not built to dedicate ~25 KB of memory for a 176x144 image. A camera board was developed with a dedicated processor to read an image into a FIFO buffer.

6.4.2 Networking

The current SDAC network stack is not optimized for transmitting images. We describe the implemented network stack, its weaknesses, and suggested modifications and additions to the network stack to ease image transmission.

6.4.2.0 The Current SDAC network stack

The current SDAC network stack is shown in Figure 33. The diagram shows the salient features of the network stack. The application layer sends and receives minimal data – only a notification of which sensor has detected an event. The rest of the network traffic comes from creating routes to and from the base station and enforcing reliable delivery of packets on neighboring nodes.

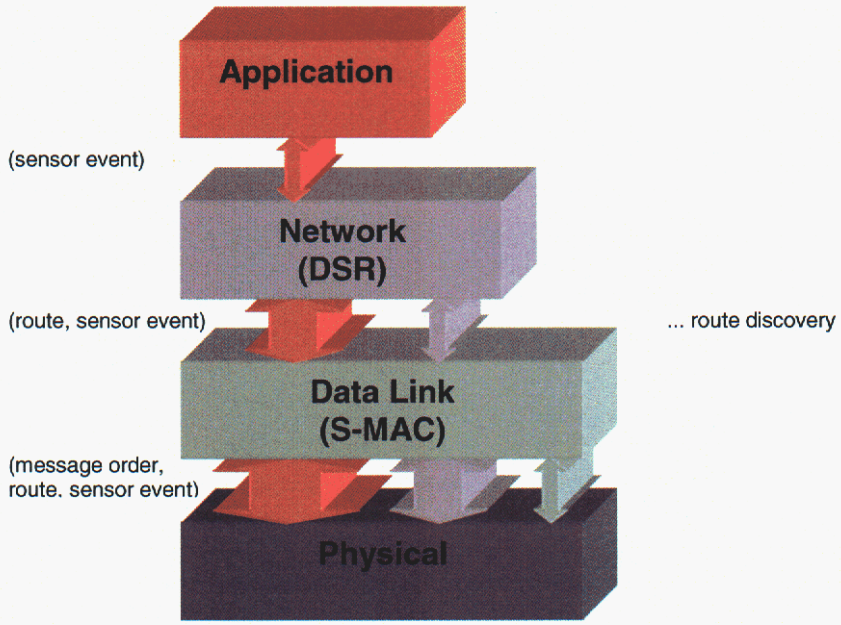


Figure 33: Current SDAC network stack. The application layer sends and receives only data about a sensor event. The underlying layer handles the transfer of packets between any two nodes in the network.

The number of bytes that need to be transmitted increase as the data traverses down the network layers. The network layer is implemented with the Dynamic Source Routing (DSR) protocol.[1] DSR is characterized by discovery routes on demand and attaches a variable length route header to each packet, describing the path it must traverse to get to its destination. The data link layer is implemented with the S-MAC protocol.[2] It is a hybrid MAC layer which takes on many of the aspects of the transport layer. While as in traditional MAC layers, there is a maximum packet size the MAC layer agrees to handle as a single entity, the S-MAC will further divide a packet into smaller subpackets for transmission and repackage it at the receiving end as shown in Figure 34. A MAC header is added to facilitate repackaging and media access with neighboring nodes. This protocol enforces reliable delivery at each node along the path. In addition to the extra bytes attached to each subpacket, each packet is accompanied by a RTS/CTS packet for securing channel access. Each subpacket also has an accompanying ACK packet to ensure correct transmission.

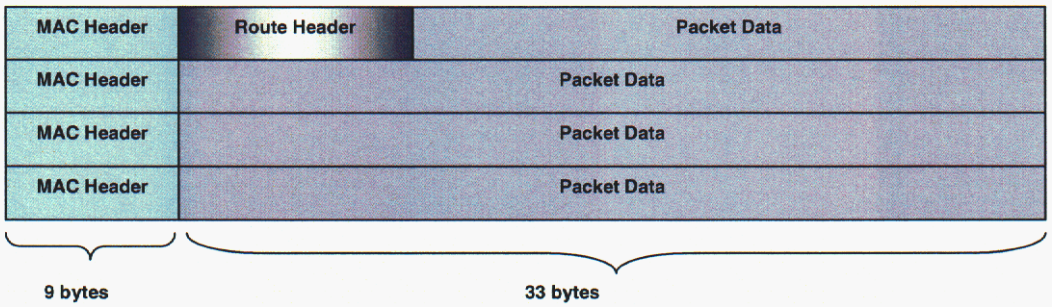


Figure 34: Breakdown of SDAC packets. In our implementation, we implemented a maximum packet size to be given the MAC layer of 132 bytes. The MAC layer breaks down the packet into 4 subpackets of 33 bytes, attaches a 9 byte header to each subpacket. The physical layer transmits treats each subpacket independently.

6.4.2.1 Suggested Changes

Currently, the needs of SDAC do not require a transport layer because the data the application layer sends at any particular time is much smaller than the maximum packet size the MAC layer defines. With the addition of a camera and the desire to transmit large amounts of interconnected data across the network, the addition of a transport layer to manage in-order reliable delivery of a stream of data would seem appealing.

It is important to note that S-MAC implements a sudo-transport layer to manage the transport of a few interconnected subpackets between neighboring nodes. Creating subpackets minimize the effect of the high packet error rate and the overhead from upper layers in the network stack. The S-MAC protocol would be an insufficient solution for image transmission because several kilobytes would be wrapped in a single packet and S-MAC does not support preempting the transmission of a single packet.

While the image data is interconnected, they are not interdependent as it is in other data such as a speech signal. Using Application Level Framing (ALF) [3] allows the application to utilize parts of large interconnected independent data without waiting for the complete transmission. This means the application can respond concurrently as the data is received in the face of lost transmission, which is important when we are dealing with a low-bandwidth system and high-bit error transmitters. The ALF would require an additional header, describing where the data fits into the application. Specifically, the header must contain an index into the position of the data.

Minimizing the overhead of a packet is critical for efficient use of the radio. Embedding the route in a variable length network header may be an acceptable solution when there is minimal traffic between nodes, but with a large amount of data through the same path, an alternative protocol would be preferable. One possible solution would be Ad-Hoc On-Demand Distance Vector Routing (AODV) [4] is one that would require only a small fixed header. AODV uses a similar method as DSR to discover routes on demand, but maintains next hop tables to eliminate the route requirement in the header.

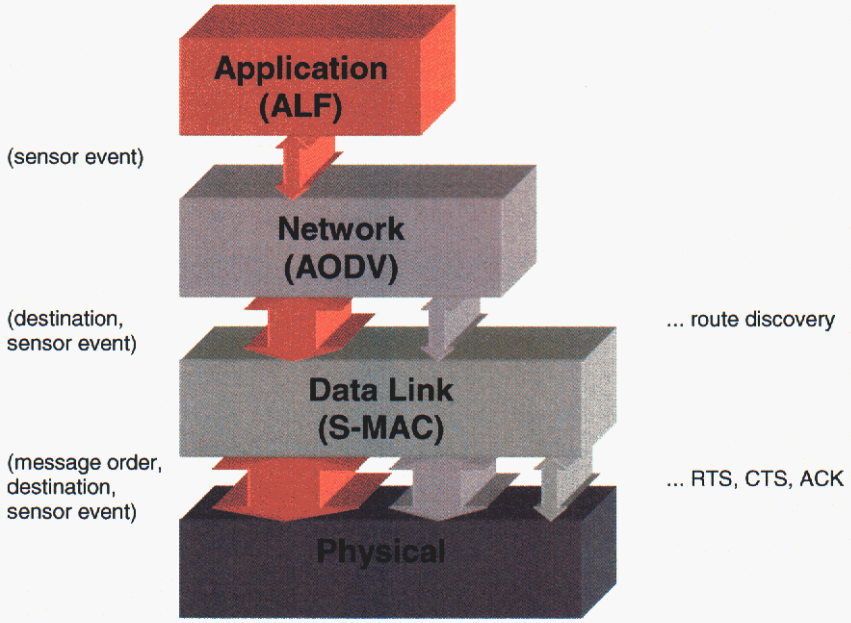


Figure 35: Recommended Network Stack. The stack is changed only at the Application and Network layers. The network layer was replaced with a different routing protocol to minimize the overhead of sending large amounts of data to the same source. The Application Layer adds the Application Layer Framework to allow for instant use of the partial out-of-order data from the lower layers.

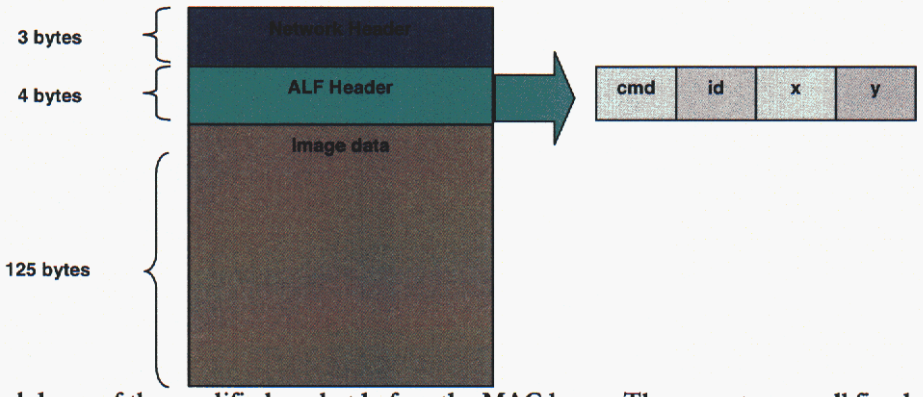


Figure 36: Breakdown of the modified packet before the MAC layer. There are two small fixed size headers for the Network and Application layer, allowing transmission to be scalable across different length routes. ALF resolves the packet data to its exact use in the application. The example header shown has 4 fields, describes which command should be executed, an identifier to resolve different images, and the (x,y) position of the first byte.

The changes to the network stack are summarized in Figure 35 and Figure 36. The new network stack features two major changes: the inclusion of ALF and the modification of our routing protocol. The ALF enables the application to instantly use partial out-of-order data from the lower layers of the network stack. Modifying our routing protocol to one that does not require the path in our header allows for a much more scalable sensor network, as well as minimizing overhead for paths that are traversed often. On average, for a network that is greater than 5 in diameter, our packet contains a smaller overhead from both the network header and the ALF header than before. The routing header will contain the final destination of the packet, and the ALF header will describe how the data fits into the application.

6.4.3 Compact Image Representation

Because radio transmissions consume so much power, it's important that every byte we decide to send contains as much information as possible. Because these images will be viewed by a human eye and will not be used for machine tasks, we can take advantage of the fact that the eye is more sensitive to certain types of information. We can also take use of the fact that some parts of the image may contain more relevant information than others.

6.4.3.0 Color Space

The eye is more sensitive to changes in luminance than chrominance. By using a color space like YUV or its phase shifted equivalent, YCrCb, shown in Figure 37, we can separate luminance and chrominance, and process and transmit only the luminance, reducing the number of bytes in half. Because we have separated the data that contains a lot of information from the data that contains little information, we are able to discard part of our data, reducing the number of bytes needed to represent our image without losing much information.

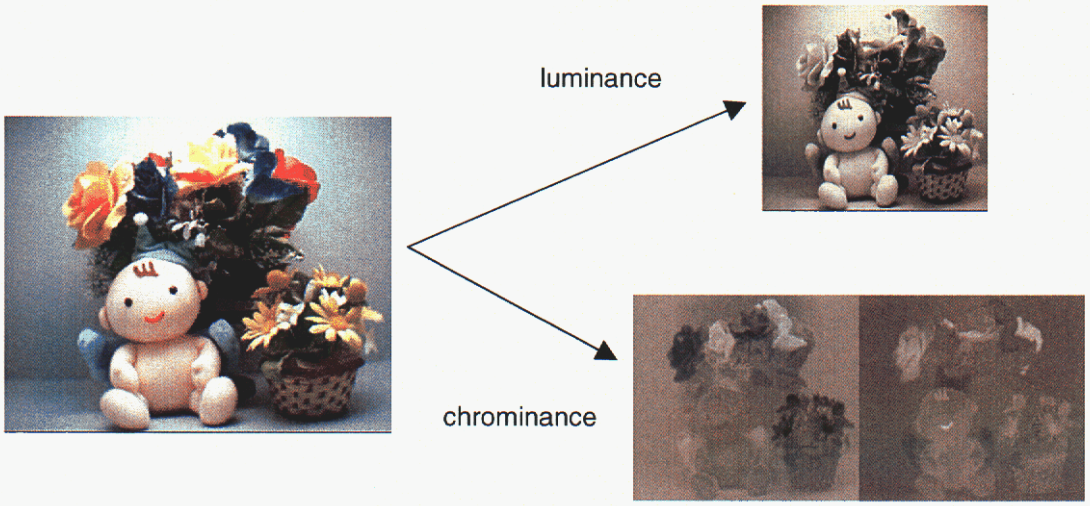


Figure 37: YUV Color Space.

6.4.3.1 Frequency vs Spatial Description

The human eye is also more sensitive to certain frequencies over others, most notably the lower frequencies over the higher frequencies. To take advantage of this, it is necessary to convert our standard spatial description of an image to a frequency description. Just like in the color space, it is important to transform the data where the more information-rich data is independent of the information-poor data.

By sending the lower frequencies first and waiting or never sending the higher frequencies, the user is given the most discernable image first. While there are many types of compression schemes, both designed specifically for images and not, the most promising ones were the JPEG [5] and JPEG2000 [6]. JPEG is discussed in detail because of the limited availability of JPEG2000 standards at the time. JPEG2000 should enable even better resolution images.

JPEG is a lossy algorithm that uses the Discrete Cosine Transform (DCT) to decompose the spatial components of an image into its frequency components, shown in Figure 38. The image is processed in blocks of an 8x8 matrix of pixels. For each block, the DCT is computed and stored in an 8x8 matrix of DCT coefficients. Each element in this matrix tells how much a particular frequency occurs in that spatial block. The 8x8 matrix of DCT coefficient is quantized to give more weight to certain frequencies that the human eye can discern and less to those it can't. Quantization results in an 8x8 matrix of mostly 0's at high frequencies. The degree of quantization is directly related to the degree of compression and the degree of information loss. The more an image is quantized, the smaller the number of bytes needed to describe the image and the more the degradation of the image. The final step of JPEG compression is to use a run-length encoding scheme like arithmetic or Huffman encoding to encode the matrices. Because most of the values in the matrices are zero, the encoding should reduce the number of bytes dramatically. To decompress the image for viewing, the run-length encoded bytes are decoded, and the Inverse DCT (IDCT) is computed for each 8x8 matrix block.

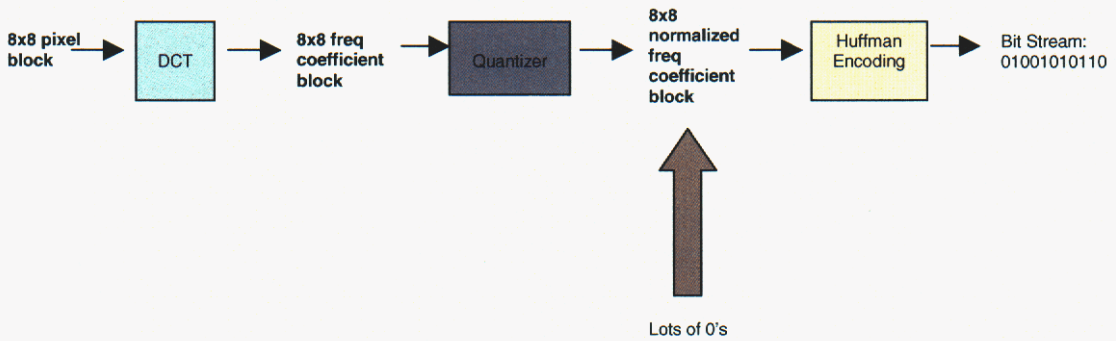


Figure 38: JPEG Algorithm. The image is broken into 8x8 pixel blocks. Each block is converted into the frequency domain using the Discrete Cosine Transform. The block is then quantized to emphasize the frequencies most sensitive the eye, and then encoding in run-length encoding.

An attractive quality of JPEG compression is its 8x8 matrix of DCT coefficient. Each element in the matrix encodes in its position and its value an independent contribution to the image.

Figure 39 and Figure 41 show original images received from an OmniVision image sensor. An image is transformed into its frequency components using the JPEG compression technique. The frequency components are then sent across the wireless network in small packets. Four packets would be sufficient to transmit the first DCT coefficient, given the mean value of each 8x8 pixel blocks. As we continue to transmit packets, more coefficients can be transmitted at the same time because at higher coefficients, most of the values are zero, as shown in Figure 40 and Figure 42.

Figure 39: Original Image

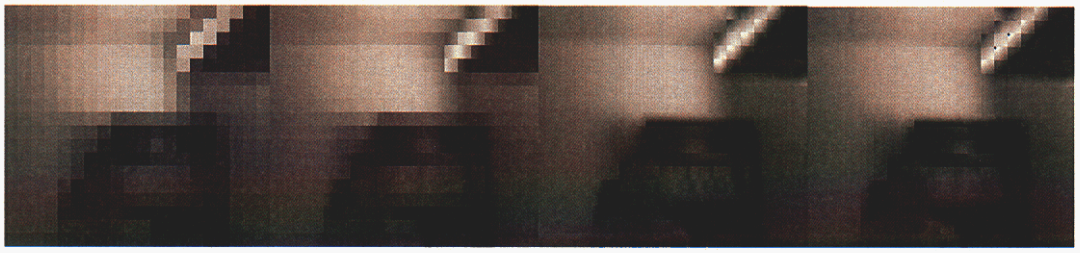


Figure 40: JPEG Compression. a) Image using only the first coefficient of the DCT matrix. Shows the mean value for each 8x8 pixel block. b) Image using the first two coefficients of the DCT matrix. c) Image using the first three coefficients of the DCT matrix. d) Image using the first four coefficients of the DCT matrix.

Figure 41: Another Sample Image.

Figure 42: JPEG Compression moving from left to right. a) Image using only the first coefficient of the DCT matrix. Shows the mean value for each 8x8 pixel block. b) Image using the first two coefficients of the DCT matrix. c) Image using the first three coefficients of the DCT matrix. d) Image using the first four coefficients of the DCT matrix.

6.4.3.2 User-feedback Compression

Many of the images the system will be dealing with will have only parts that contain information the user is interested in. The sample images, Figure 39 and Figure 41, show two

plausible images where most of the image contains the background, and object of interest, the face, is less than 25% of the total image.

Using ALF, we can immediately begin displaying parts of the image as the data arrives. In some images, after a few frequencies have been sent, a person is able to rule out regions where there is nothing worth seeing in more detail and discerning possible regions of interest. The user could give feedback to the sensor system by sending a short message describing the bounding box to the node that contains the image, and the node can respond by sending only information about the part of the image, as shown in Figure 43.

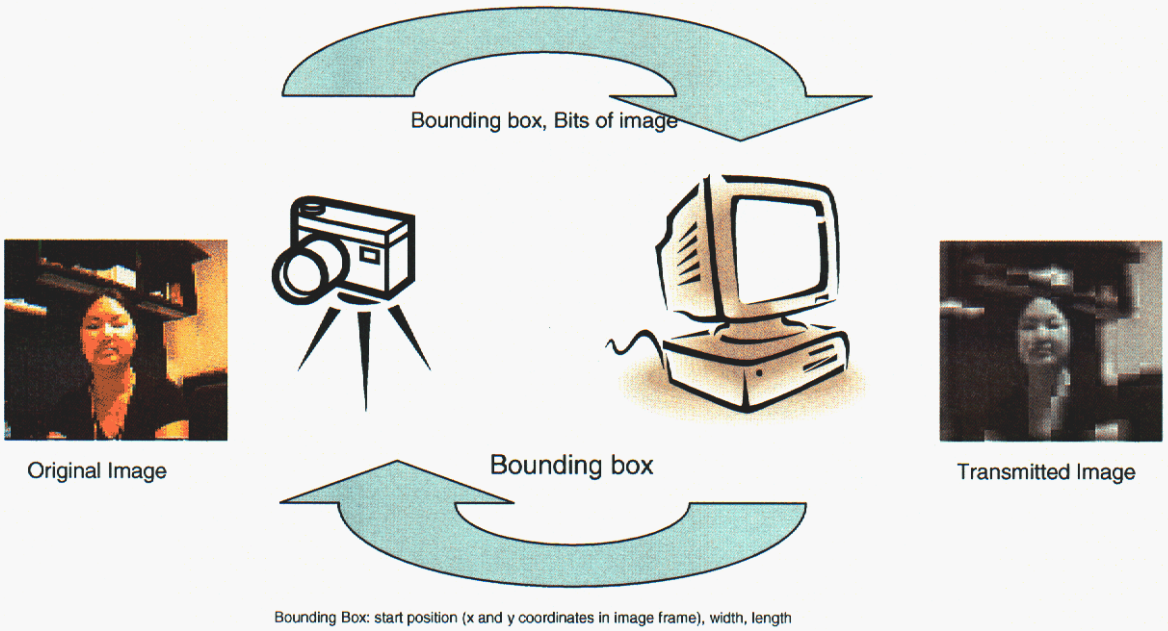


Figure 43: User-feedback scheme. The image is transmitted to the base station where the user is viewing the image and can respond by selecting a bounding box of the region of interest. When the sensor node receives a packet with the bounding box description, it will begin sending data on only the part of the image within the bounding box.

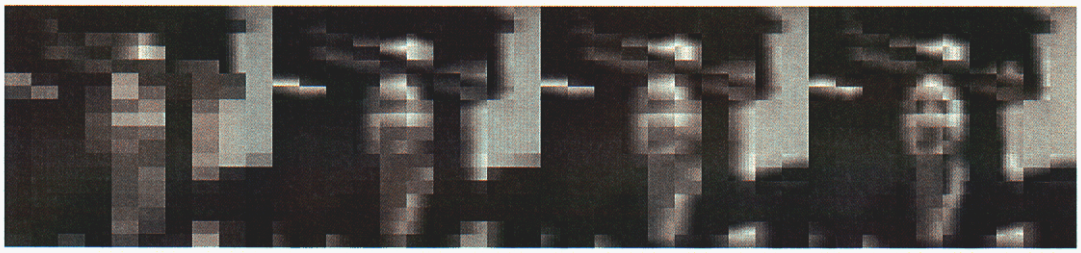


Figure 44: Standard Image through JPEG Compression. After a few frequencies, a face is discernable.

Figure 45: A Sample bounding box around face.

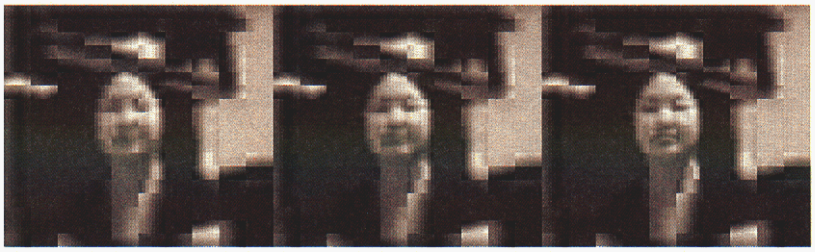


Figure 46: The subsequent transmissions improve only the area within the bounding box.

6.4.3.3 Intelligent Compression

The user-feedback mechanism is only helpful if the user can differentiate between objects of interest and background faster than the image is transmitted. This requires a slow transmission speed to prevent wasteful transmission. If the sensor node could perform the differentiation, the transmission speed would not need to be reduced. In addition, the regions of interest can be detected from the start, and bytes required to describe the background would never need to be sent.

For demonstration, we built a simple skin-color detector to detect faces. Because skin color across all nationalities exhibits the same hue, it makes it a good single-feature classifier. Still, a single-feature detector is unlikely to be robust enough for many situations. Using a combination of several features would give better face-detection results; we have focused on the effect of using a detector rather than the detector itself in this body of work. Figure 47 shows the results of the skin-color detector, and transmissions. The face is clearly discernable after only a few transmissions.

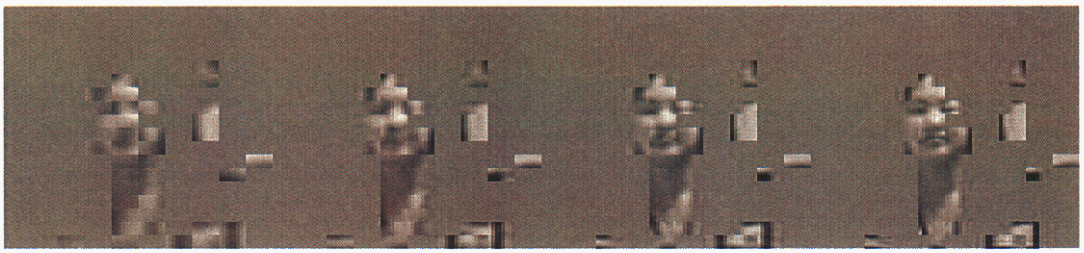


Figure 47: Images showing the progression using a skin-color face detector, to eliminate transmitting the background.

6.4.3.4 Analysis

These steps greatly reduce the amount of data that needs to be transmitted. The original image of ~25 KB would consume around .54 J if transmitted. This is equivalent to the energy consumed for transmitting 550 sensor events. By using the YUV Color Space and only transmitting the luminance, we reduce the amount of data to ~13 KB. JPEG compression further reduces the number of bytes by an order of magnitude. Figure 48 compares the time and power needed to transmit the image shown in Figure 44. In both cases, the smart compression provides the best performance. In both cases, the smart detector on the sensor node wins out over the simple straight forward transmissions of the image bytes and the user feedback. User feedback and the smart detector are close on power consummation; the smart detector only saving a few bytes by eliminating the background. Part of this is due to the simplicity of the detector. Implementing a smarter detector would result in even greater power savings.

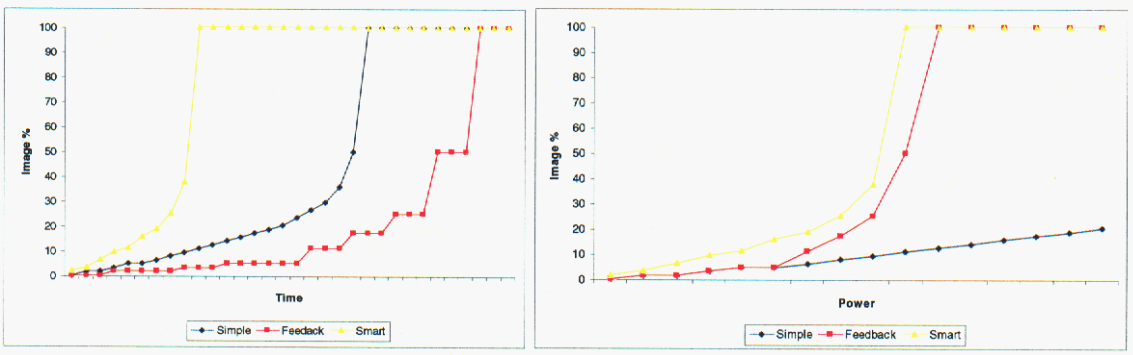


Figure 48: Comparison of different methods.

6.4.4 Summary

Instead of sending a large stream of data, we propose to incrementally improve the image, by sending the most general information before the details. To manage large amounts of data without the overhead of a transport level and the ability to utilize partial information, we suggested using Application Level Framing, a paradigm that enforces usable independent packets of information. Switching to a network layer that does not require the transmission of the route in every packet also would help minimize the overhead.

Incrementally sending information allows us to direct how the sensor network is improving the image, specifying areas in the image that we would like to know more detail about. One approach to deciding when to send data is to rely on an external source, such as a human or other sensor. While this is a useful feature for sensor networks, we would like to also process visual data at the nodes, and transmit only the data relevant to the task at hand. Or, we will accomplish this locally on the sensor node, using various feature detectors.

6.5 Chapter 6 References

[1] Johnson, David B., Maltz, David A., and Broch, Josh. DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks. In *Ad Hoc Networking*, C. Perkins, Ed. Addison-Wesley 2001, pp. 139-172.

[2] Ye, W., Heidemann, J., and Estin, D. An Energy-Efficient MAC Protocol for Wireless Sensor Networks. In *Proc. IEEE INFOCOM* (June 2002).

[3] Clark, David D. AND David L. Tennenhouse. Architectural Considerations for a New Generation of Protocols. In *Proc. ACM* (1990).

[4] Perkins, Charles E. and Royer, Elizabeth M. Ad-hoc On-Demand Distance Vector (AODV) Routing. In *Proc. 2nd IEEE Workshop on Mobile Computing Systems and Applications*, New Orleans, LA (Feb 1999)

[5] Lossless and near-lossless coding of continuous tone still images. ISO?IEC JTCI/SC29 WG1. (July 1997).

[6] Skoras A., Christopoulos C, and Ebrahimi T. The JPEG2002 still image compression standard. In *IEEE Signal Processing Magazine, Volume 18 Issue 5* (Sep 2001).

7 Conclusions

This chapter provides results and future research directions for the SDAC LDRD. On behalf of the ERI LDRD team, we would like to thank the Adv. Concept Group for giving us the opportunity to work on such an exciting project.

7.1 Results of SDAC LDRD

The SDAC LDRD was a 9 nine months effort on behalf of the Embedded Reasoning Institute research staff and internship program. This LDRD developed a body of knowledge that explored current and future trends in wireless sensor networks. The results of this body of work have been captured in this final report. The objective of this project was to explore the concept of using wireless smart sensor technology for a set of four mission areas in the War on Terrorism (WoT) domain. While the team later reduced the mission areas back to border protection and military operations for urban terrain (MOUT) the results more than adequately covered the bounds of the LDRD objectives. To achieve the overall objective the team surveyed existing technology, proposed a sensor architecture, and generated an initial methodology and metric model for understanding connections between the applications and sensor technology capabilities.

The SDAC demo system provides a way to detect and localize objects moving through an area and distinguishes friendly objects from foreign objects. The system demonstrates that currently available technology can be used to implement a fully functional sensor network capable of performing in-network sensor fusion. The modular architecture developed to create the system allows for a more extensible and upgradeable system than standard centralized systems. More research is needed to refine the system and further develop the modular architecture. In particular, careful attention needs to be paid to providing methods for debugging the system and isolating the power supplies of individual modules from interference caused by the rest of the node. Innovations in sensor technology, wireless communications, power-aware software techniques, and distributed computing will continue to drive future developments of advanced sensor networks like SDAC.

7.2 Future research directions

The SDAC LDRD represented fertile ground for exploring many different potential areas for future development and next generation concepts. This chapter provides a brief look at two-funded project that resulted from initial funded by the SDAC LDRD. The first project deals the concept of adding small cameras to the small SDAC sensor platform. While the ideas seemed practical shipping large image files via wireless communication proved to be an impossible challenge for the FY03 SDAC sensor platform. Researchers started exploring with ways to understand image data on the sensor unit itself. An overview of this project is discussed in Section 7.2.1 and is being led by Teresa Ko (8961) under funding from 15200. One of the biggest concepts to come out of this LDRD was the proposed architecture, which was unveiled earlier in the LDRD. Jesse Davis (8961) extend the idea of modularity as a key necessity to flexible and adaptable SDAC sensor systems to generate a successful proposal for build the next generation SDAC, known as Modular Architecture for Sensor Systems (MASS). An overview of the MASS project, which is supported by the CSRF, is presented in Section 7.2.2 and promises to provide sensor units that adapt to a variety of domain. Section 7.3 also provides a set of

potential project that would extend the current SDAC vision and direction. While the projects in this section are not currently funded, these ideas generated by Ron Kyker (8945) illustrated that the next generation sensor system is an on going development.

7.2.1 Feature-based Vision Data for Distributed Wireless Sensor

While it is possible to transmit an image across a wireless network, the power consumption and the latency of the system may be too high of a cost for the information gained. While researchers continue to investigate improvements in compression techniques the processing power, memory, and radio power may still be better used elsewhere. The benefit of including cameras in a sensor network is in their ability to give thousands of bytes of data at one instance. The question is how to best use obtainable imaging data within a constrained platform and networking situation. One option would be to send portions of the image across the wireless network to a user at a central location for human decision-making and additional processing. A preferred option, which fits into the SDAC sensor system paradigm, would be for the sensor node to process this raw data into information that a sensor network could use on its own.

The addition of cameras can facilitate a sensor network in distinguishing between different objects of interest (e.g., people, tanks), determining relative position/distance of objects, and predicting future areas of interest. These tasks can be directly applied to increase battlefield awareness in unknown terrain by tracking the enemy's movements and characterizing their behavior, their numbers, and their composition. Also, this work would add robustness and accuracy to monitoring of facilities or materials by removing the dependency of motion from intruders and locating their positions throughout the area, by working in conjunction with other sensors.

We propose to integrate the XScale based PASTA board, shown in Figure 49, created by USC/ISI with our current wireless sensor network node platform and a low-power image sensor. The PASTA board will provide additional computational power and memory to our current platform allowing us to process images onboard. With the addition of cameras and computational powerful processors in our wireless sensor network, the network can describe the environment at a resolution not previously available. This research will focus on extracting understandable and informative features from sensors and effective reasoning across node and sensors.

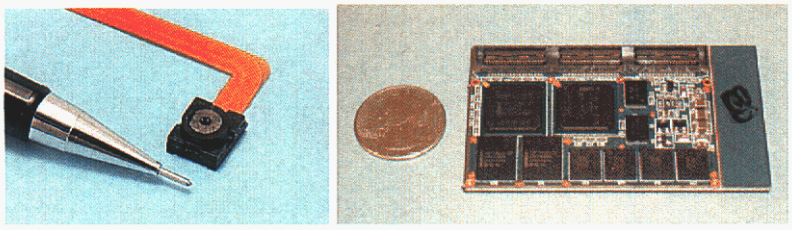


Figure 49: Additional hardware requirements for SDAC. The camera and XScale based board depicted will be integrated with the current HERD unit to provide more sensor information and processing power on the unit for intelligent data fusion.

Research in computer vision has explored many of the salient features of images and video, which help to describe an image and classify and track objects across different views and times.

Moving feature extraction and data fusion to the distributed domain raises many challenges not previously emphasis. While computer vision with multiple cameras has explored the exploitation of multiple views, images are typically collected at a centralized unit and results are computed with complete global knowledge. In the WSNs, there is no one unit with global knowledge of the information, so each node is required to make decisions about its environment with only partial knowledge. The questions we will address will ask what collection of features can be used to describe an objects of interest that fit in the constraints, how these features transcend across different images in time and position and orientation of the object, and how these features can be communicate across different nodes.

7.2.2 Modular architecture sensor systems

Wireless sensor networks are made up of individual wireless nodes each containing a mix of sensors, processors, power supplies, and wireless transceivers. Integrating these various resources into a unified hardware platform, as well as controlling this platform and fusing the on-node sensor data with a robust software framework is a complex task. Several mature programs, both internal to Sandia and throughout academia and industry, have tackled this multi-faceted node design issue. The resulting systems each have their own unique capabilities and limitations. The most notable Sandia programs are TALON, HERD, ISM/SMA, T1/T2, and SDAC (a.k.a. MicroTALON). TALON is a high speed, high bandwidth, target recognition sensor network, but it is high power and uses completely centralized data processing. HERD is a miniature, low power, distributed radiation detection network, but it allows limited, if any, application flexibility. ISM/SMA provides a flexible and robust system solution for high performance data processing applications, but it is limited to environments with access to wall power. Finally, the first prototype SDAC demo system is a distributed event detection network demonstrating simple in-network computation on low power hardware, but it is limited in its flexibility and high performance data processing capabilities.

There are two primary observations of previously developed systems, which drive the MASS project. First, application flexibility and power management of the individual nodes are significant design requirements. Since sensor networks are applicable to many different missions, and the windows of opportunity for developing new applications are generally quite short, the flexibility of the node platform is imperative. Instead of spending the extra time, money, and manpower to build single use systems for each mission, designing flexibility into a single node platform allows it to be reconfigured or reprogrammed quickly for many different applications. Additionally, in wireless nodes, power supplies, and thus node lifetimes, are severely limited. This power constraint makes mission efficient, low power operation vital as well.

The second observation is that each sensor network system developed thus far can demonstrate either mission efficiency or application flexibility, but none, either at Sandia or otherwise, has achieved both. Application flexibility is traditionally achieved via a highly adaptable, high performance central processor with several I/O interfaces, but these processors have high power consumption making them inefficient for many tasks. Mission efficiency is traditionally achieved via a low power, highly specialized central processor with only required peripheral support, but these processors are limited in their flexibility. Based on these traditional

methods, efficiency and flexibility have therefore been viewed as contrasting requirements. Building on advances in power aware microprocessors and a further understanding of wireless sensor network applications and requirements derived from previous programs, a new perspective is possible.

The MASS project suggests a novel design approach achieving mission efficiency and application flexibility in a single system. The key to achieving this goal is a modular, multi-processor hardware architecture operating under an intelligent distributed software control. Instead of relying on the traditional flexible node architecture, in which resource control is centralized on a single high power processor, each resource in the node will be built into physically separable modules with supporting resource-specific processors. On sensor modules, the module processors will perform data collection and preliminary data analysis tasks. On the wireless communication module, the module processor will manage power states of wireless transceivers and route network traffic. If a high performance processor is needed for complex data manipulation or data fusion, this processor will also be integrated into its own module, and its module processor will control its power states and I/O with the rest of the node. **Figure 50** illustrates the individual module-level structure envisioned. **Figure 51** provides a collective view of multiple modules making up a complete individual node-level architecture.

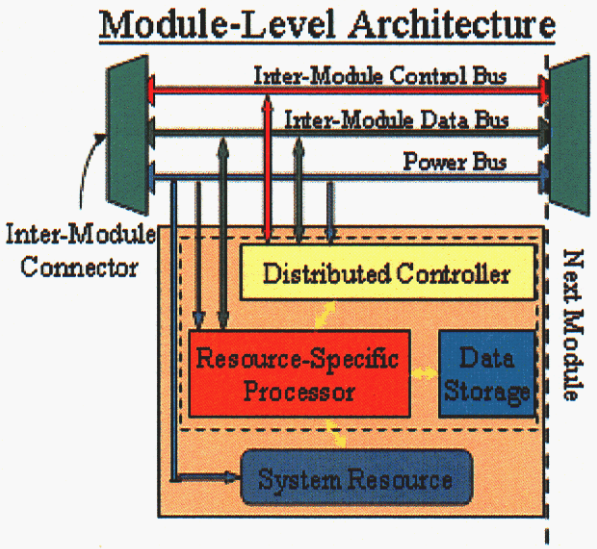


Figure 50: Individual module-level architecture

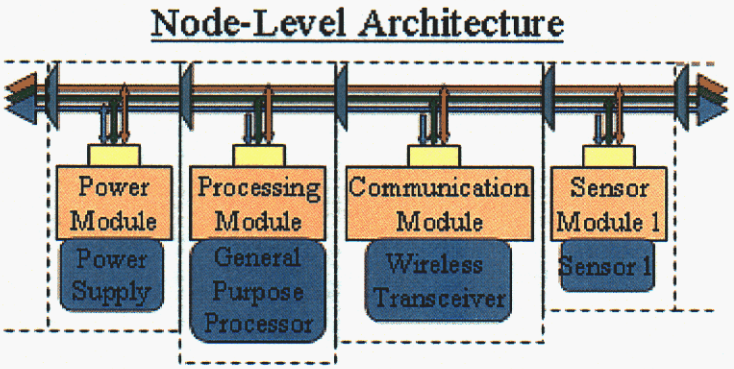


Figure 51: Individual node-level architecture

Since the module processors will control all node resources, there is no need for centralized control or centralized resource management. Furthermore, the modules will all be connected by a common bus giving them the ability to request services of each other. If a sensor module detects an event, for example, it can request validating data or tasks from other sensor modules or a high performance processor module. A module may also request data to be sent to or collected from another node through the wireless communications module to enable distributed network computation. This fine-grained, event-driven resource adaptation allows resources to remain in a low power sleep mode when they are unneeded. The modularity also allows resources to operate in parallel. In previous flexible systems, a high power central processor had to remain constantly powered in order to collect and analyze sensor data even when no events were occurring. The proposed architecture thus results in decreased node power consumption and an enhancement in mission efficiency.

In addition to these power and performance benefits, the modularity also gives the necessary flexibility, extensibility, and upgradeability to allow utilization of the architecture in multiple mission spaces. If a particular mission requires only a certain set of sensors and no high performance processing resource, for example, only each sensor module required and no high performance processor module will be integrated to configure the nodes for the mission. Previously, when flexible systems were adapted to individual missions, parts of the system that were unnecessary for the mission could not be removed and would add to the power overhead. An additional benefit of the proposed architecture is that upgrades to sensors or processors can be integrated into the system simply by building them into their own module. The architecture thus allows for a more mission-centric adaptation of hardware and software to specific application requirements.

7.3 SDAC Future Technologies

Alternately, there are technologies that are emerging that will play a significant role in SDAC like systems.

7.3.1 Fast non-volatile (unifying) memories

Ferroelectric, Magneto-resistive, and Ovonic Unified Memory are a few examples of emerging high speed non-volatile memories. These memory devices have the non-volatility, re-

writability, and density of flash memory with speed of SRAM in both read and write plus write cycles longer than flash. If the read/write cycle lifetimes eventually approach that of SRAM, one could imagine a unified memory type. No longer is there the need for separate ROM, RAM, EEPROM, and Flash memories. This one memory would do it all. In fact, the standby current could be practically zero so one could imagine ultra-low power operation where the memory is completely shut off while the processor is idle. While this capability does not seem to be significant at first glance, some new capabilities emerge that are very interesting. One such capability is the ability to make unused program memory space operate as RAM storage. This would be very useful in network applications where additional RAM could be used for buffers. One could imagine small operating systems written with the idea of trading program space for data space, similar to paging techniques used in larger machines. Further one now has the ability to have instant bootable computers. Computers could be designed such that they do not shut down but merely stop operation where they left off and start from the same place. In fact this will have some un-intended side effects such as memory de-allocation never happening or losing its previous contents on boot-up. One could imagine having to explicitly reset a processor or flush memory for security reasons. Another capability, energy scavenging machines would be able to crunch large problems by performing computations when they had collected enough energy to perform some computation, stop when power runs out and continue when power resumes.

7.3.2 Code vaults and context configurable software

With flash memory density increasing, one could imagine extremely large flash memories that are very small physically which could be turned on, accessed, and then shut down. This could act as a very large program or code vault provides essentially the same capability as a hard drive for PCs with the exception that they are low power, can be shut down, and aren't used for paging. Combined with fast non-volatile memories, one could have a system self configure on startup or dynamically based on the programmed mission, system modules installed, or context of a situation. One could even imagine dynamically changing the configuration in the field. People have envisioned multimodal sensing using a collection of configurable sensors capable of multiple modalities of sensing. The same concept could be applied to software where collections of processing capabilities are configurable from a much larger code vault. Perhaps at first this would be code for handling different sensors or sensing modes such as implementing a plug-and-play sensor system that would self configure its software based on the physically connected sensor hardware at startup.

One could take this concept further however and imagine field configurable sensors for different mission space. Suppose a sensor network has too many temperature sensors in a particular area but those same sensors have other capabilities. Nodes could be re-configured to provide alternate capability based on the need. This allows for a limited amount of operating memory to be used (always a fixed resource) but to have access to a vast array of software that could be executed. In this way, only the software that is necessary to perform the required task is loaded and executed, as opposed to a really large flexible program in memory that only 10% of the program gets executed. There are benefits in power savings because of the memory savings but also in new ways of executing and operating software. One could also imagine a design where the software loads modules as required. Say, one is uniquely identifying vehicles but on a rare occasion there is an instance of something not typically detected such as a human, animal, or other rare occurrence. Perhaps the signal analysis processing for these instances are very

different than of the normal occurrences. Or perhaps one only uses this situation when a target identification correlation is poor or inconclusive. In these cases, one could load in additional signal processing modules designed to handle other types of targets, attempt to identify the target, and then go back to normal configuration once the target is identified. The alternative is to have enough storage to cover all target set of detections or to only include those that are optimum, both of which are non-optimal.

Eternal power systems – as mentioned above, scavenged power provides the ability to operate in a physically small size for long life. For some systems, scavenged energy is not available. For these situations, some emerging options include radioactive batteries such as beta emitters.

7.3.3 Distributed heterogeneous processors

The current approach to many problems is to write a routine that controls each function of a system yet runs in a single process and maybe uses an RTOS such that each function becomes a separate process so as to create the appearance of multiprocessing. What would happen if the problem were scaled via processors instead of processes. Each processor would be custom programmed with a given task that it specializes in solving. In this way, each processor acts as an independent agent with it's own set of rules. This allows for selective and specialized processing.

7.3.4 Ultra-low power operating systems

There is a need for multithreaded applications in SDAC systems to handle complexity, maintainability, and competing resources. Current operating system techniques do not allow for the lowest operating power and in fact add overhead. It is possible to optimize this by executing only when processing is required through dynamic scheduling, resource balancing, and hardware based scheduling. The appearance of this in some of the latest Intel and Crusoe processors is called hyper threading and is similar in concept, essentially pushing threading into the hardware.

7.3.5 Ultra high-speed 8/16 bit processors

The 8 bit and definitely the 16 bit processor isn't dead. By optimizing architectures and making them low power with the latest processes, 8 bit processors could be running in the ghz regime. By providing scalability, some problems could be addressed by these processors that is currently being handled by larger and more power intensive processors with smaller scaleable logic thereby reducing overall power consumption. Imagine a processor that ran from 32khz to 1ghz, a scale of 4 orders of magnitude!

7.3.6 Wireless ad-hoc routing in hardware

Much of the resources of wireless communications in ad-hoc environments comes from the MAC layers and ad-hoc protocols. One could imagine ad-hoc protocols that adapt or are configurable based on the level of mobility, the amount and frequency of traffic, etc. Much of this could be moved to dedicated hardware for lower power and less memory requirement. One could imagine a set of three classes of routing algorithms that were intertwined and configurable such that the routing could be adapted to a particular environment or sub-environment with few switches. These three classes of algorithms would be designed to cover the majority of the classes of problems most people would be concerned with. Trade-offs could be used such as

mobility versus fixed, source routed versus temporal, etc. Perhaps the coded bank concept could be applied here to configure one of many algorithms.

8 Appendix

8.1 Additional existing sensor system evaluation

	Sensoria sGate	Ember
Network		
Routing algorithms	Multihop, self-organized	
Network Architectures		Decentralized multi-hop mesh topology. Requires one gateway
Robustness		Reconfigures routing as necessary
Hardware		
Node architecture	Modular: DSP processor for analog front end; 167MHz RISC processor for applications and networking; wireless and digital I/O module. Ethernet or RS-232 PC interface.	Communications processor with expansion bus and SPI connection for a host processor
Reconfigurability	15 digital I/O lines, four fully differential analog sensor inputs	Bus, standardized host processor communication, and API should provide for rapid prototyping
Upgradeability		The EmberNet Protocol Stack should mean any hardware changes on Ember's side would be transparent. Standardized interfaces should allow the consumer to easily upgrade any other their additions
Sensors	Integrated GPS; seismic and acoustic sensor modules available	
Software		
Architecture	Linux 2.4 Kernel, with API for applications	EmberNet Protocol Stack communication stack. Software routes incoming data or passes it along to a host processor. API provided for communication with a host processor.
Extensibility	Extensive API	Designed as a platform for wireless communication, so extensibility should be built in
Data processing	Distributed processing available	undefined
Power aware	Power management API included	
Intelligence	High speed processor should allow agent-based applications	undefined
Communication		
Wireless	2.4GHz RF, plus Bluetooth and 802.11b capabilities; 10 or 100mW transmit power	900MHz or 2.4GHz, 192kps
Range	25 – 100m indoor, 500m outdoor depending on antenna and transmit power	300m or 100m according to frequency
MAC		
Power		
Lifetime	12 hours on 7.2Ahr lead-acid battery	Requires DC power supply
Power consumption	600mA based on lifetime information	
Other		
Size	21cm x 15cm x 7cm	4cm x 7cm x 1cm

Cost		Development kit: $2500. Includes 6 nodes, gateway, and software
Application	User defined. Seemingly designed for fixed installation	Wireless sensing and control platform

8.2 The Advanced Encryption Standard (AES)

The Advanced Encryption Standard is the new encryption standard adopted by the National Institute of Standards and Technology (NIST) to replace the Data Encryption Standard (DES). AES is an algorithm named Rijndael developed by Joan Daemen and Vincent Rijmen and is outlined in [10]. The reader should refer to this reference for algorithm specifics, as they will not be outlined here and are mathematically intensive.

AES was chosen as a security mechanism within SDAC to allow for secure wireless communication as well as cryptographic authentication among nodes while maintaining a low level of power consumption. Its low power consumption is by design, as it is intended to run on 8-bit systems natively and uses a symmetric keying system. The use of the symmetric keying system as well as 8-bit arithmetic reduces the number of operations that need be performed on data before it can be used in its cryptographic form.

The SDAC implementation of AES was constructed in such a way to minimize power consumption and memory use during the individual stages of the algorithm. All operations are performed on a single 128-bit data segment through pointer arithmetic rather than static copying of smaller data regions into working forms that are then copied to the ciphertext result at completion. Additionally, round key generation is done in real-time and requires only a single 128-bit round key to be stored per round during encryption and decryption operations. Implementing the AES algorithm in this manner results in slower execution time, but a much smaller memory footprint in order to fit on the individual nodes.

In following with the modular theme of SDAC, AES was implemented in three modular forms: Encryption, Decryption and Both. In separating functionality the size of the implemented security mechanism is minimal unless the functionality of both encryption and decryption is necessary. This allows for inclusion of AES encryption on individual nodes and requires the least possible amount of memory. The DFS or root node contains the decryption library, which requires twice the memory of encryption. This is of little consequence as the memory constraints of the DFS are much more lenient than individual sensor nodes.

The libraries can be inserted into communication layers within individual node controllers, and all data passed to the network or I²C can be transparently passed through the encryption or decryption routines. The public interface for encryption is as follows:

```
int AES_encrypt(char* input, char* output, char* cipherKey);
```

This routine takes as input three 128-bit buffers passed by-reference. The input buffer contains data to be processed and should be padded to 128-bits with 0x00 if there is insufficient data to fill the buffer. The output buffer should be empty, yet allocated as 128-bits wide, and will contain the resultant ciphertext after encryption. The cipherKey buffer contains the static, shared key that is common throughout the sensor network. This can be changed at any given time, but data must be encrypted and decrypted with the same shared key.

The public interface for decryption is as follows:

```
int AES_decrypt(char* input, char* output, char* cipherKey);
```

Its input and output buffers follow the same constraints as above, but the input buffer should be in the form of ciphertext and if encrypted by the above AES_encrypt() routine will always be 128-bits wide. The output buffer will contain the resultant plaintext that was originally passed to the AES_encrypt() routine. The cipherKey parameter is the shared secret key across all nodes.

It is important to note that AES is a symmetric cipher, and will produce the same results no matter what order the functions are called in. That is, if the DFS wishes to send data to the nodes, which contain only the AES_encrypt() library, then it can apply the AES_decrypt() function on its plaintext data and the result will be ciphertext to be transmitted to the individual nodes. The nodes can then apply AES_encrypt() to the encrypted data stream and retrieve the resultant plaintext. This requires the same key be used for both operations. In doing this, the required memory for the libraries becomes minimal, as AES_encrypt() requires roughly half the memory of AES_decrypt().

The additional power requirements for AES are minimal in that it heavily relies on pointer arithmetic. Copying of data between memory locations is minimal, and all operations are native CPU operations such as Exclusive-OR (XOR) and addition. A concrete increase in power consumption is unknown, but due to the fact that wireless data transmission overhead is zero (due to the lack of increase in transmitted bits) the increase in computational resources required is minimal.

8.3 Sensor system evaluation

Systems contact information

1. HERD: Doug Stark (925 294 3898), Ron Kyker (925 294 3065)
2. EmberNet: info@ember.com
3. ISM from 8200: Ron Kyker (925 294 3065)
4. Crossbow Motes: Bob Bingwell (408 965 3332), info@xbow.com
5. Dust, Inc. Motes: Kris Pister (510 643 9268), pister@eecs.berkeley.edu
6. Sensoria sGate: Frederic Newberg (310 641 1331 x 211), information@sensoria.com
7. Rockwell Scientific HiDra: Max Pedyash, (805 373 4110)
8. Darpa Self-Healing Minefield: Dr. Thomas Altshuler (703 696 0222), SHM@darpa.mil
9. JPL/NASA Sensorwebs: Kevin A. Delin, Kevin.A.Delin@jpl.nasa.gov
10. Graviton: technologies@graviton.com
11. Seekernet: Robert Twitchell (678 662 3819), info@seekernetinc.com
12. Pacific Northwest National Labs: Jim Skorpik, jim.skorpik@pnl.gov
13. Steel Rattler, Steel Eagle: g.prado@sentech-acoustic.com, (781 279 9871)
14. USC/ISI: Deborah Estrin, destrin@cs.ucla.edu
15. Army Research Lab: Jon Eicke (301 394 5000), jeicke@arl.army.mil

Distribution

1	MS0839	Gerry Yonas	16000
1	MS0839	Ron Pate	16000
1	MS0839	Curtis Johnson	16000
1	MS 0865	Regan Stinnett	01903
1	MS0839	Tim Moy	16000
1	MS9003	Ken Washington	8900
1	MS9003	James Handrock	8960
1	MS9003	Chuck Hartwig	8940
1	MS9011	Barry Hess	8941
1	MS9019	Brian Maxwell	8945
1	MS9019	Steve Carpenter	8945
1	MS9037	Jim Berry	8947
1	MS9012	Steve Gray	8949
1	MS9915	Mike Koszykowski	8961
1	MS9217	Steve Thomas	8962
1	MS9012	Jerry Friesen	8963
1	MS9012	Mike Hardwick	8964
1	MS1188	Steve Tucker	15311
3	MS9915	Nina Berry	8961
1	MS9915	Jesse Davis	8961
1	MS9915	Teresa Ko	8961
1	MS9036	Ron Kyker	8245
1	MS9913	Doug Stark	8245
1	MS9036	Greg Cardineli	8245
1	MS9007	Doug Henson	8200

1	MS 0323	D. Chavez, LDRD Office, 1011
3	MS9018	Central Technical Files 8945-1
1	MS0899	Technical Library 9616
1	MS9021	Classification Office, 8511, for Technical Library, MS 0899,9616 for DOE/OSTI via URL