Spring 2008

Energy efficient wireless sensor network protocols for monitoring and prognostics of large scale systems

James W. Fonda
Missouri University of Science and Technology, fonda@mst.edu

ENERGY EFFICIENT WIRELESS SENSOR NETWORK PROTOCOLS

FOR

MONITORING AND PROGNOSTICS OF LARGE SCALE SYSTEMS

by

JAMES WILLIAM FONDA

A DISSERTATION

Presented to the Faculty of the Graduate School of the

UNIVERSITY OF MISSOURI-ROLLA

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

IN

ELECTRICAL ENGINEERING

2008

Approved by

Jagannathan Sarangapani, Advisor
Steve E. Watkins, Advisor
R. Joe Stanley
Yahong Rosa Zheng
Can Saygin

PUBLICATION DISSERTATION OPTION

This thesis consists of the following five articles that have been submitted for publication as follows:

Pages 14 to 47 were submitted to the IOP *Journal of Smart Materials and Structures*

Pages 48 to 92 were submitted to *International Journal of Distributed Sensor Networks*

Pages 93 to 127 is intended for *IEEE Transactions on Wireless Communications*

Pages 128 to 176 were submitted to *IEEE Transactions on Controls Technology*

Pages 177 to 195 (Appendix) in press for *SPIE 15th Annual International Symposium on Smart Structures/NDE: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2008, vol. 6932.*

Other Pertinent Publications produced during the course of this work include:

[1] J. W. Fonda, M. J. Zawodniok, S. Jagannathan, S. E. Watkins, "Development and Implementation of Optimized Energy-Delay Sub-network Routing Protocol for Wireless Sensor Networks," *Proc. of the IEEE Int. Symp. on Intelligent Control*, Munich, Germany, pp. 119-124, 2006.

[2] J. W. Fonda, M. J. Zawodniok, J. Sarangapani, and S. E. Watkins, "Adaptive Distributed Fair Scheduling and Its Implementation in Wireless Sensor Networks," *Proc. of the IEEE Conf. on Systems, Man and Cybernetics*, Taipei, Taiwan, pp. 3382-3387, 2006.

[3] J. W. Fonda, Watkins, S. E., "Health Monitoring of a Truss Bridge using Adaptive Identification," *Proc of the IEEE Conf. on Intelligent Transportation Systems Conf. (ITSC)*, pp.944-949, 2007.

[4] K. Mitchell, S. E. Watkins, J. W. Fonda, and J. Sarangapani, "Embeddable Modular Hardware for Multi-Functional Sensor Networks," *Smart Mater. Struct*., vol. 16, no. 5, pp. N27-N34, 2007.

[5] S. E. Watkins, J. W. Fonda, and A. Nanni, "Assessment of an Instrumented Reinforced-Concrete Bridge with Fiber-Reinforced-Polymer Strengthening," *Opt. Eng.,* vol. 46, no. 5, 051010 2007.

[6] S. E. Watkins, T. M. Swift, and J. W. Fonda, "Development of Autonomous Triggering Instrumentation," *Smart Structures and Materials 2008: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems,* SPIE Proc, vol. 6932, San Diego, CA, March 2008, In Press, 2008.

ABSTRACT

In this work, energy-efficient protocols for wireless sensor networks (WSN) with applications to prognostics are investigated. Both analytical methods and verification are shown for the proposed methods via either hardware experiments or simulation. This work is presented in five papers. Energy-efficiency methods for WSN include distributed algorithms for i) optimal routing, ii) adaptive scheduling, iii) adaptive transmission power and data-rate control. In the first paper a reactive optimized energy-delay sub-networking routing (OEDSR) is developed and implemented on custom motes. OEDSR provides optimal routing through consideration of available energy, delay, and distance to the destination. Results show an improvement in energy usage and extension of network lifetime. In the second paper an adaptive distributed fair scheduling protocol for multi-channel networks (MC-ADFS) is developed that provides scheduling support for WSN via management of bandwidth capacity based on priority. Results for MC-ADFS demonstrate performance increases in end-to-end delay, bit-rate, and fairness. In the third paper an adaptive distributed rate and power control (ADRPC) and optimal tracking control through the generalized Hamilton-Jacobi-Bellman are developed. Lyapunov methods are used to show the stability and performance of these protocols and a comparison between these are also introduced. Finally, in the forth paper a robust observer and a prognostics scheme is developed. This robust observer is utilized as an online approximator for detecting faults in large scale systems. Additionally, a prognostics method is developed providing estimates of future values of critical physical parameters resulting in estimation of the time-to-failure (TTF) of components.

ACKNOWLEDGMENTS

I would like to thank my wife Minna, my parents David and Donna, and the rest of my family. I also would like to acknowledge the help from my advisors, Jagannathan Sarangapani and Steve E. Watkins, who made this work possible and the rest of my committee for reviewing my work. Also, special thanks go to Maciej J. Zawodniok, Travis Dierks, Kainan Cha, Ahmet Soylemezoglu, Jeffery Birt, and Qinmin Yang for their advice and hours of discussion. Finally, I would like to thank the National Science Foundation, GAANN Fellowship Program from Dept of Education, the Air Force Research Laboratory, the Missouri University of Science and Technology (MST), formerly University of Missouri-Rolla, Intelligent Maintenance Systems I/UCRC, and the Electrical and Computer Engineering Department for their support of this work through funding. Finally, thanks to the many other members of the faculty and staff of the Electrical and Computer Engineering Department for their friendship and support in my years of residence.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

PAPER III

PAPER IV

APPENDIX

LIST OF TABLES

1. INTRODUCTION

Wireless sensors networks (WSN) provide advantages over traditional monitoring and prognostics technologies, due to their distributed architecture, for large scale systems. Application spaces include agricultural, industrial, safety, infrastructure, military, and residential deployments. However, major challenges remain in the development of communication protocols to enable distributed monitoring and prognostics schemes. Additionally, few protocols have been implemented and widely deployed to date. In particular, associated drawbacks with these networks include durability issues such as battery power and data intensive processing decreasing network life-time. To enhance network lifetime energy-efficiency through energy-efficient protocols are required.

Development of efficient protocols and deployment methods allows WSN's to be applied to a variety of practical problems. Energy-efficient networks provide durability to applications by satisfying needs in deployments, signal processing, and distributed computing. Signal processing and distributed computing using WSN hardware provide opportunities to integrate intelligence into a network. For instance, such capabilities can support maintenance applications to predict component lifetime, estimate physical parameters, and provide fault detection.

Energy-efficient networking topics include routing, scheduling, transmission power, and data-rate adaptation. Also, prognostics topics include parameter estimation and fault detection. Additionally, there is a cross-over of research interests between monitoring and prognostics due to the supporting role that WSN's play in the deployment of prognostics applications. Through cooperative deployment WSN and prognostics methods can enable large-scale system monitoring and maintenance.

Figure 1.1 illustrates two aspects for deployment of WSN's and describes the linkage existing between topics. Large-scale monitoring and prognostics is reliant on the use, and development, of embedded hardware that can host both applications; however, the use of WSN's provide the embedded hardware needed to enable deployment of prognostics methods alongside the network.

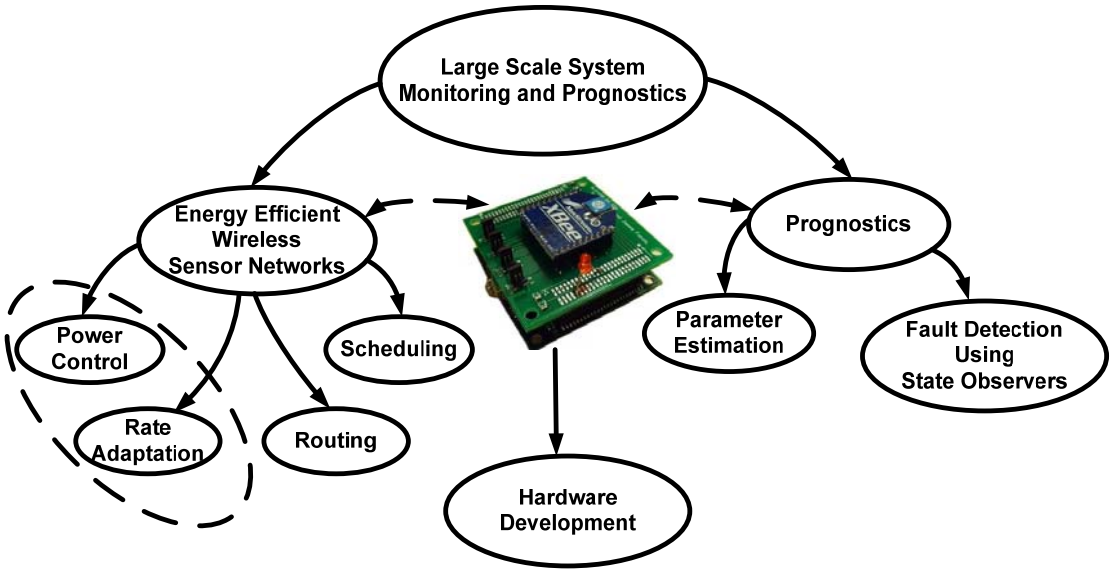

Fig. 1.1 Interaction of WSN and distributed prognostics

1.1 ENERGY EFFICIENT PROTOCOL DEVELOPMENT AND DEPLOYMENT

WSN's have the ability to transform networking applications and to create new uses for embedded technology. In many applications the parameters of network durability are key issues, namely the energy efficiency, network lifetime, and quality-of-service (QoS). Analytical methods to assure QoS while meeting the efficiency needs of an application are also important. To address these needs consideration must be given in the design of routing, scheduling, and control of the physical layer.

The application space encompassed by WSN's and prognostics provides large-scale monitoring, intelligent processing, and information flow due to the low-cost and low-power platforms used in WSN. These platforms provide adequate processing and enhance

applications through distributed processing. The processing capabilities enable QoS and energy-aware networks increasing lifetime, decreasing network maintenance, and enabling visibility for process variables. Energy efficiency has been investigated in the context of routing, scheduling, and physical hardware interactions with QoS. A balance is required between network performance and QoS to a level of user perceived optimality.

In many networks simplistic routing cannot satisfy the needs of QoS and resource management. Simplistic routing mechanisms such as flooding [1, 2] and gossiping [3, 4] can provide data flow; however they do not provide resource awareness. Therefore, while providing data paths their resource wastage and lack of QoS awareness are undesirable. Past innovations in routing methods have included resource awareness [5, 6], hierarchal protocols such as LEACH [6], PEGASIS [7], TEEN and APTEEN [8] and data centric methods such as SPIN [9], and directed diffusion [10].

Resource-aware protocols take into account energy, location, or other physical parameters in order to provide improved routing. Hierarchal methods normally provide clustering and possibly data aggregation to maintain performance. Finally, data-centric methods provide data consolidation and aggregation to minimize traffic while maintaining the data space being conveyed. Aggregation methods include duplicate suppression as well as functional methods such as minimum, maximum, average, variance, or statistical data metrics. However, to provide energy efficiency for routing use of a distributed architecture is desirable. Once routing is completed packet, or bandwidth, scheduling is required to provide fair bandwidth utilization on the network.

Scheduling methods range from the simplest to implement to analytically guaranteed methods. Basic methods such as first-in-first-out (FIFO) methods and round-robin (time-

slotting) methods provide scheduling with little processing overhead or implementation issues. FIFO methods operate with no preference, or priority, and do not guarantee performance under load. FIFO queuing can suffer from process starvation, or unfair resource scheduling. In contrast, round-robin methods do not suffer from process starvation, however there is no priority given to a particular process, or flow. Methods such as weighted round-robin [11] have been proposed and help to alleviate some of these issues. Methods such as strict fair queuing (SFQ) [12] provide weighted selection of packets based on the acceptable end-to-end (E2E) delay. SFQ methods provide a mechanism for weighting of packets based on the desired QoS and priority of the packet; however, SFQ methods do not provide adaptation to changing conditions. Extension of SFQ to provide this adaptability resulted in adaptive distributed fair scheduling (ADFS) [13]. The author has ported ADFS to WSN platforms previously [14]. Use of the ADFS methodology provides the needed adaptation to network resources and provides additional performance gains over FIFO and SFQ methods. Comparisons of the FIFO, SFQ, and ADFS methods can be found in [14]. Channel scheduling is also of interest as utilization of more than a single channel can effectively increase the total bandwidth of a network. Additionally, in bandwidth-constrained networks using the 802.15.4 standard channel scheduling provides the advantage of increasing available bandwidth resources.

To minimize interference, transmission power control at each node provides regulation of signal-to-interference ratio (SIR) across the network. Additionally, use of the minimum transmission power to satisfy the SIR for the desired data-rate provides energy savings and reduces interference levels. Also, the use of a minimal power level provides energy efficiency to the network. Therefore, rigorous work on distributed power

control (DPC) has been performed for cellular networks [15-17]. Also, several DPC schemes [18-22] were developed for wireless ad-hoc networks with constantly changing topologies due to mobility and link failures. Power control has also been investigated using rule-based methods and via linear control methods [23]. While these methods provide control of the power, only through joint control of power and data rate can the needed cooperative control be accomplished.

Joint control of the power and rate provides linkage between the physical layer and the data-rate QoS. To provide a desired rate the SIR level of a link must be maintained via control of the transmission power. To ensure QoS for a wireless channel the power level and data-rate must be controlled [24, 25]. Methods of joint power and rate control have included stochastic [26], game-theory [27], linear programming [28], neuro-fuzzy [29], optimization methods [30], optimal control methods [31], linear quadratic control [23], and adaptive methods [23]. Adaptive and robust methods previously are presented in [23]; however they fall short of providing analytical stability analysis and performance guarantees. Additionally, the methods proposed in [23] do not adequately consider QoS through selection of rates and its linkage to SIR control.

1.2 DISTRIBUTED MONITORING AND PROGNOSTICS

Diagnostics and prognostics can change the methods that are used to maintain large-scale systems. Classical methods reliant on human observation and intervention are slow and cumbersome, while wired and wireless networks can enable network-centric diagnostics and prognostics methods to be deployed. Manual monitoring requires much overhead and incurs lag time until files are received and archived. Networked systems provide the needed speed and ease of file management with wireless systems providing

easier and cost-effective deployments. For many applications the overall technology and deployment paradigms for monitoring has shifted towards low-cost, low-power, and small devices using mesh networks, commonly known as WSN's.

Prognostics are used to predict the remaining usefulness of systems and components. Classical prognostic and diagnostic methods have been reliant on human operators and observations. As mentioned earlier, these methods can be slow and tedious and do not react quickly enough for all failures. Other prognostic methods are based on model-based [32-33] and stochastic approaches [34, 35]; however, they do not provide the desired performance in all environments. Therefore analytically solutions are desirable. Parameter estimation has been performed via adaptive techniques [36]. The adaptive methods provide estimation of physical parameters of systems and can provide feedback or maintenance decisions. However, the use of adaptive methods is a diagnostic approach and is not appropriate for prediction of failures. To provide prognostics estimates of the time-to-failure (TTF) based on the physical parameters is required.

1.3 ORGANIZATION OF THE DISSERTATION

Work presented here encompasses WSN developments for practical deployments. Analytical methods and practical experimentation for WSN are explored to show capabilities and limitations. Along with theoretical contributions, development of hardware for support of WSN is shown in the appendix. In the appendix a discussion design criteria and subsequent design processes are given for a WSN platform that was developed for infrastructure and machine monitoring .

The dissertation is divided into four major papers, or chapters. The first three papers discuss the implementation of optimal energy-delay-sub-networking routing (OEDSR)

protocol, adaptive distributed fair scheduling scheme in the presence of multiple channels (MC-ADFS) and joint power and rate control respectively. The final chapter addresses deployment of these protocols for civil infrastructure monitoring and prognostics through the development of both a robust observer and a prognostics prediction method.

Innovations in the energy efficiency for routing methods were investigated. Application of OEDSR [37] and a comparison to AODV [38] was performed on Missouri University of Science and Technology (MST) motes [39]. The motes used in both the routing and scheduling, discussed next, testbeds were predecessors to the F1 mote developed in the appendix. In this study verification of the performance of the protocols under real-world conditions was performed. Energy savings as much as ten times was possible during the routing phase using OEDSR. This energy savings is due to the improved search method during routing. The method also provides a reduction in routing traffic and interference among nodes.

Bandwidth scheduling was addressed. The adaptive distributed fair scheduling (ADFS) method was investigated and compared to traditional first-in-first-out (FIFO) methods. Additionally, the ADFS method was extended for multi-channel networks resulting in MC-ADFS. Implementation results for MC-ADFS show as much as a 16 percent improvement over non-adaptive DFS methods.

Additionally, joint control of the data rate and transmission power was investigated to provide QoS and energy efficiency for wireless networks. The data rate and transmission power are physically coupled through the channel capacity and the requirement on the signal-to-interference ratio (SIR). Therefore, the channel capacity is managed by the transmission power in order to achieve a desired data rate on a network link. To provide

rate and power control an adaptive control method is proposed. Additionally, a neural network (NN) based optimal control method using the generalized Hamilton-Jacobi-Bellman (GHJB) equation was applied. Results show that desired data rates can be achieved with more accuracy and QoS parameters can be met.

Finally, methods for application support of deployed WSN were investigated. A robust NN observer with robust integral sign of the error (RISE) feedback was developed. The NN/RISE method provides robustness to the standard NN observer with respect to noise and disturbances. For many applications this robustness can be a practical advantage for embedded systems. Also, in this section of the work a method for prognostics based on prediction of parameter values was developed. Using adaptive control methods that provide physical parameter predictions a solution was developed to predict future values of a parameter with a focus on predicting a time-to-failure (TTF) value. The emphasis on the TTF was sought in order to provide prognostics capabilities to current methods. The standard parameter update law was manipulated to provide this result. The robust NN/RISE observer provides as much as a 25 percent increase in performance. Additionally, the use of TTF estimation shows that prediction of failure levels for parameters can be accurately performed.

1.4 CURRENT FOCUS AND FUTURE WORK

Future work should include methods to combine metrics for both routing and scheduling to provide a more integrated networking protocol. This combined metric would provide increased performance and efficiency by improving interaction between packet routing and scheduling. This interaction could provide improved metrics for evaluation of network performance that include both energy and fairness on the network.

For prognostics methods more work into analytical methods should be investigated. During literature review many rule-based methods were found and closed-loop analytical methods would provide continued innovation in the field.

Future work for the prognostics methods should include continued work into improving methods for prediction of physical parameters as in Paper 4. By increasing the accuracy of such methods and extending the horizon for prediction prognostic methods can be improved. Additionally, other applications can serve as hardware testbeds for future prognostics work. One such work was WSN deployment to support diagnostics on pneumatic systems. The pneumatic system is shown in Fig. 1.2 and includes a PLC, pneumatic actuator, and associated supporting hardware. A mote was used to monitor normally-available signals that are on the PLC such as command signals and limit switches. The MST instrumentation sensor node (ISN) was used to monitor the signals from the pneumatics system. A rule-based method was implemented to provide diagnostics capability to the pneumatics system and is capable of diagnosis of stuck actuator, bad limit switches, and leaks in supply lines. Results for this method showed near 100% accuracy for the types of failures considered. Other methods have been implemented on WSN hardware that extends signal processing past rule-based methods. One such method is the Mahalanobis-Taguchi strategy [40, 41], which provides a statistical method for detection of fault states and classification of root-case analysis. Extension of this work to include prognostics capabilities can lead to the prediction of pneumatic cylinder lifetimes. Validation of these methods could be done through experimentation on the existing testbed. One such application could be estimation of friction values in the pneumatic cylinder for prognosis of failure.

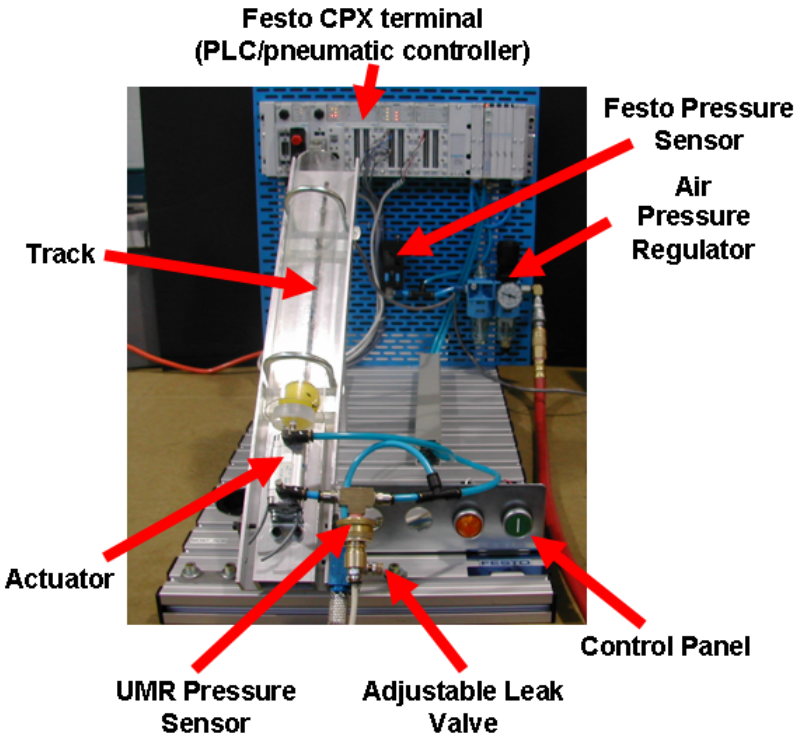

Fig. 1.2 Example application

Bride monitoring is another application using WSN hardware. This application space can be extended to investigate the long-term monitoring and prognostics needs. In Paper 4 we present methods to improve robustness of observers to noise and other uncertainties; however, long-term deployment of such methods should be studied and performance evaluated. This type of long term study can be done on a structure such as the Missouri University of Science and Technology (MST) Smart Composite Bridge (SCB) as seen in Fig. 1.3, or on other laboratory structures where fatigue testing is being performed. Performance of prognostics methods could then be validated through hardware experimentation.

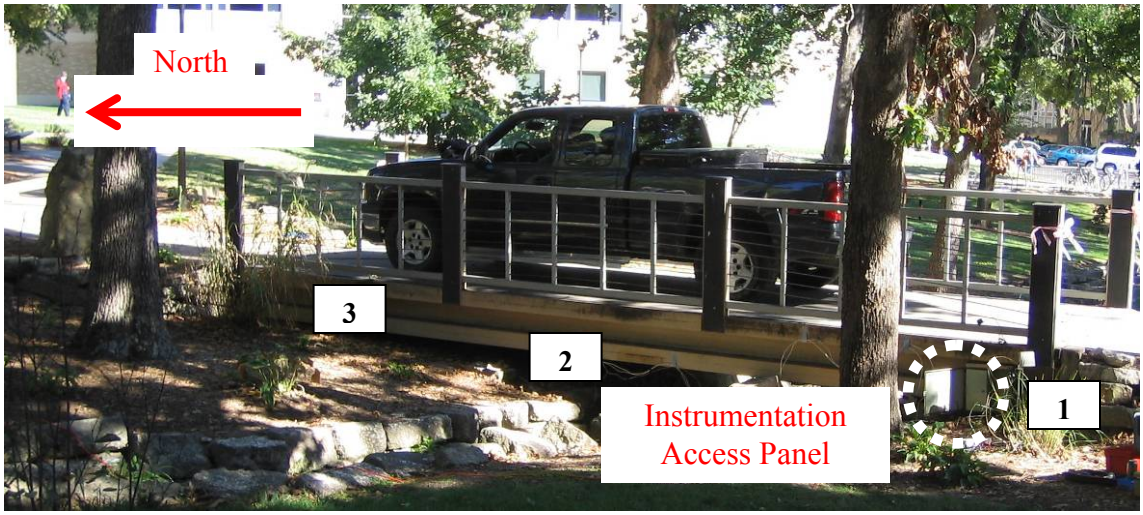

Fig. 1.3 Photo of the MST SCB

1.5 REFERENCES

[1] H.K. Pung, J. Song, and L. Jacob, "Fast and efficient flooding based QoS routing algorithm," *Proc. of the IEEE Int. Conf. on Computer Communications and Networks,* pp. 298-303, 1999.

[2] R. Farivar, M. Fazeli, and S. G. Miremadi, "Directed flooding: a fault-tolerant routing protocol for wireless sensor networks," *Proc. of the IEEE Systems Communications,* pp. 395-399, 2005.

[3] L. Xiang-Yang, K. Moaveninejad, and O. Frieder, "Regional gossip routing for wireless ad hoc networks," *Proc. of the IEEE Int. Conf. on Local Computer Networks,* pp. 274-275, 2003.

[4] N. Gupta, and D. Manjunath, "Gossiping in multihop radio networks," *Proc. of the IEEE Int. Conf. on Personal Wireless Communication*, pp.78-82, 1999.

[5] J. Zhang, Z. Yang, B. H. Cheng and P. K. McKinley, "Adding safeness to dynamic adaptive techniques," *Proc. of the Int. Conf. on Software Engineering (ISCE)*, pp. 371-380 2006.

[6] W. R. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," *Proc. of the IEEE Hawaii Int. Conf. on System Systems Science (HICSS),* vol. 2, 10 pp., 2000.

[7] S. Lindsey and C. Raghavendra, "PEGASIS: Power-efficient gathering in sensor information systems," *Proc. of the IEEE Aerospace Conf.,* vol. 3, pp. 1125-1130, 2002.

[8] A. Manjeshwar, and D.P. Agrawal, "APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks," *Proc. of the IEEE Int. Parallel and Distributed Processing Symp. (IPDPS),* pp. 195-202, 2002.

[9] W. R. Heinzelman, J. Kulik and H. Balakrishnan, "Adaptive protocol for information dissemination in wireless sensor networks," *Proc. of the ACM/IEEE Int. Conf. on Mobile Computing and Networking (MobiCom)*, Seattle, WA, pp. 174-185, 1999.

[10] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann and F. Silva, "Directed diffusion for wireless sensor networking,", *IEEE/ACM Transactions Networking*, pp. 2-16, Feb. 2003.

[11] S. Ramadhadran and J. Pasquale, "The stratified round robin scheduler: Design, analysis, and implementation," *IEEE/ACM Trans. on Networking*, vol. 14, no. 6, pp. 1362-1373, 2006.

[12] P. Goyal, H.M. Vin, and H. Cheng, "Start-time fair queuing: A Scheduling Algorithm For Integrated Services Packet Switching Networks," *IEEE/ACM Trans. on Networking*, vol. 5, pp. 690–704, 1997.

[13] N. Regatte and S. Jagannathan, "Adaptive and distributed fair scheduling scheme for wireless adhoc networks," *Proc. of the Word Wireless Congress*, pp. 101-106, May 2004.

[14] J. W. Fonda, M. J. Zawodniok, J. Sarangapani, and S. E. Watkins, "Adaptive distributed fair scheduling and its implementation in wireless sensor networks," *Proc. of the IEEE Conf. on Systems, Man and Cybernetics*, Taipei, Taiwan, pp. 3382-3387, 2006.

[15] N. Bambos, S. Chen and G. J. Pottie, "Channel access algorithms with active link protection for wireless communication networks with power control," *IEEE/ACM Trans. on Networking*, vol. 8, no. 5, pp.583-597, October 2000.

[16] S. Dontula and S. Jagannathan, "Active link protection for wireless peer-to-peer and cellular networks with power control," *Proc. of the World Wireless Congress*, pp. 612-617, May 2004.

[17] S. Jagannathan, A. T. Chronopoulos and S. Ponipireddy, "Distributed power control in wireless communication systems," *Proc. of the IEEE Int. Conf. on Computer Communications and Networks*, pp. 493-496, Nov. 2002.

[18] S-J. Park and R. Sivakumar, "Quantitative analysis of transmission power control in wireless ad-hoc networks," *Proc. of the IEEE Int. Conf. on Parallel Processing Workshops (ICPPW)*, pp. 56-63, 2002.

[19] E-S. Jung and N. H. Vaidya "A power control MAC protocol for ad hoc networks," *Proc. of the ACM/IEEE Int. Conf. on Mobile Computing and Networking (MobiCom)*, pp. 36-47, 2002.

[20] J. Gomez, A.T.Campbell, M.Naghshineh, and C.Bisdikian, "Conserving Transmission Power in Wireless Ad Hoc Networks," *Proc. of the IEEE Int. Conf. on Network Protocols (ICNP)*, pp. 24-34, 2001.

[21] P. Karn, "MACA – A new channel access method for packet radio," *Proc. ARRL Computer Networking Conf.*, pp. 134-140, 1990.

[22] M.B.Pursley, H.B. Russell, and J.S.Wysocarski, "Energy efficient transmission and routing protocols for wireless multiple-hop networks and spread-spectrum radios," *Proc. of the IEEE/AFCEA Conf. on Information Systems for Enhanced Public Safety and Security (EUROCOMM)*, Munich, Germany, pp. 1-5, 2000.

[23] A. Subramanian and A. H. Sayed, "Joint rate and power control algorithms for wireless networks," *IEEE Trans. on Signal Processing*, vol. 53, no. 11, pp. 4204-4214, 2005.

[24] K.T. Kornegay, Q. Gang, and M. Potkonjak, "Quality of service and system design," *Proc. of the IEEE Computer Society Workshop on VLSI*, Orlando, FL, pp. 112-117, 1999.

[25] E. Babulak, "User perception of quality of service provision," *Proc. of the IEEE Int. Conf. on Industrial Technology*, vol. 2, pp. 1022-1025, 2003.

[26] L. T. Ong, A. Olfat, and M. Shikh-Habaei, "Joint SNR target and rate adaptation based on bit error rate estimation," *Proc. of Int. Symp. on Wireless Communication Systems*, pp. 470-474, 2005.

[27] Y.E. Sagduyu, and A. Ephriemides, "Power and rate adaptation as stochastic games for random access," *Proc. of the IEEE Conf. on Decision and Control (CDC)*, vol. 4, pp. 4202-4207, 2003.

[28] A. K. Karmokar, D. V Djonin, and V. K. Bhargava, "Delay constrained rate and power adaptation over correlated fading channels," *Proc. of the IEEE Global Communications Conf. (GLOBECOM)*, vol. 6, pp. 3448-3453, 2004.

[29] C-H. Jiang, and J-K. Lain, "Adaptive neuro-fuzzy power control and rate adaptation for multirate CDMA radio systems," *Proc. of the IEEE Int. Conf. on Networking, Sensing, and Control*, pp. 1307-1312, 2004.

[30] M. Shikh-Bahaei, M. Mouna-Kingue, and G. Charbit, "Joint optimization of outer-loop power control and rate adaptation over fading channels," *Proc. of the IEEE Vehicular Technology Conf. (VTC-Fall)*, vol. 3, pp. 2205-2209, 2004.

[31] S. A. Jafar, and A. Goldsmith, "Optimal rate and power adaptation for multirate CDMA," *Proc. of the IEEE Vehicular Technology Conf. (VTS-Fall)*, Boston, MA, vol. 3, pp. 994-1000, 2000.

[32] G. Vachtsevanos, and P. Wang, "Fault prognosis using dynamic wavelet neural networks," *Proc. of the IEEE Systems Readiness Technology Conf.(AUTOTESTCON),* pp. 857-870, 2001.

[33] L. C. Jaw, "Neural networks for model-based prognostics," *Proc. of the IEEE Aerospace Conf.*, vol.3, pp.21-28, 1999.

[34] A. Ray and S. Tangirala "Stochastic modeling of fatigue crack dynamics for on-line failure prognostics", *IEEE Trans. on Control Systems Technology*, vol. 4, no. 4, pp. 443-51, July 1996.

[35] S. J. Engel, B. J. Gilmartin, K. Bongort, and A. Hess, "Prognostics, the real issues involved with predicting life remaining," *Proc. of the IEEE Aerospace Conf.*, vol.6, pp. 457-469, 2000.

[36] J. W. Fonda, Watkins, S. E., "Health monitoring of a truss bridge using adaptive identification," *Proc of the IEEE Conf. on Intelligent Transportation Systems Conf. (ITSC)*, pp. 944-949, 2007.

[37] J. W. Fonda, M. J. Zawodniok, S. Jagannathan, S. E. Watkins*, "*Development and implementation of optimized energy-delay sub-network routing protocol for wireless sensor networks," *Proc. of the IEEE Int. Symp. on Intelligent Control*, Munich, Germany, pp. 119-124, 2006.

[38] C. E. Perkins, E. Belding and S. Das, "Ad-hoc on-demand distance vector (AODV) routing," *IETF MANET Working Group*, available online: www.ietf.org/internet-drafts/draft-ietf-manet-aodv-14.txt, Feb. 2003.

[39] K. Mitchell, S. E. Watkins, J. W. Fonda, and J. Sarangapani, "Embeddable modular hardware for multi-functional sensor networks," *Smart Mater. Struct*., vol. 16, no. 5, pp. N27-N34, 2007.

[40] R. B. Chinnam, B. Rai, N. Singh, "Tool-condition Monitoring from Degradation Signals Using Mahalanobis-Taguchi System Analysis," *ASI's Annual Symp. of Robust Engineering*, pp.343-351, 2004.

[41] G. Taguchi and R. Jugulum, The Mahalanobis-Taguchi Strategy – A Pattern Technology System, John Wiley & Sons Inc., 2002.

PAPER 1

Optimized Energy-Delay Sub-Network Routing Protocol Development and Implementation for Wireless Sensor Networks

James W. Fonda[1,2,*], Maciej Zawodniok[1], S. Jagannathan[1], and Steve E. Watkins[2]
Embedded Systems and Networking Laboratory[1]
Applied Optics Laboratory[2]
Department of Electrical and Computer Engineering
Missouri University of Science and Technology, Rolla, Missouri 65409-0040
jfonda@ieee.org, mjzx9c@mst.edu, sarangap@mst.edu, and steve.e.watkins@ieee.org

Abstract— The development and the implementation issues of a reactive optimized energy-delay sub-network routing (OEDSR) protocol for wireless sensor networks (WSN) are introduced and its performance is contrasted with the popular Ad-hoc On-demand Distance Vector (AODV) routing protocol. Analytical results illustrate the performance of the proposed OEDSR protocol; while experimental results utilizing a hardware testbed under various scenarios demonstrate improvements in energy efficiency of the OEDSR protocol. A hardware platform constructed at the University of Missouri-Rolla (UMR), now the Missouri University of Science and Technology (MST), based on the Generation 4 Smart Sensor Node (G4-SSN) prototyping platform is also described. Performance improvements are shown in terms of end-to-end (E2E) delay, throughput, route-setup time, drop rates, and energy usage is given for three topologies, including a mobile topology. Additionally, results from the hardware testbed provide valuable lessons for network deployments. Under testing OEDSR provides a factor of ten improvement in the energy used in the routing session and extends network lifetime compared to AODV. Depletion experiments show that the time till the first node failure is extended by a factor of three with the network depleting and network lifetime is extended by 6.7%.

Index Terms—**Wireless sensor network, self organization, clustering, energy-delay, sub-network, energy efficiency, AODV, OEDSR.**

I. INTRODUCTION

Energy efficient network protocols are an integral part of constructing a practical wireless sensor network (WSN) for deployment [1, 3-17]. Available routing protocols for sensor networks can be classified as data centric, location-based, quality-of-service (QoS) aware and hierarchical protocols. Data centric protocols [17] such as SPIN [4], directed diffusion [5], and GRAB [6,7] consolidate redundant data while routing. Location-based

Research supported in part by Dept. of Education GAANN Fellowship, Air Force Research Laboratory Grant (FA8650-04-C-704) through Center for Aerospace Manufacturing Technologies and Intelligent Systems Center.

routing protocols such as GPSR [11], GEAR [12] and TTDD [13] require GPS information to determine an optimal path so that the flooding of routing-related control packets is not necessary. On the other hand, QoS aware protocols such as SPEED [14], address requirements such as energy efficiency, reliability, and real-time requirements. Finally, the hierarchical protocols such as LEACH [3], TEEN[8], APTEEN [9], and PEGASIS [10] form clusters and use cluster heads (CHs) in order to minimize the energy consumption for both processing and transmission of data. These network protocols have been evaluated in simulation; however, there are limited experimental results reported of performance on hardware.

Performance of wireless sensor network (WSN) protocols are traditionally evaluated through the use of network simulators such as NS2 [1, 3-6, 8, 9, 13], OPNET, PARSEC [7], and GloMoSim [2, 14, 18]. Simulations provide performance comparisons of competing protocols under ideal conditions; however, simulations often do not evaluate the protocol against realistic hardware constraints or dynamic environments with channel uncertainties. Typical applications for WSN deployments include forest fires, structural monitoring, and factories; and each of these environments present unique challenges for WSN. For example, industrial environments contain obstacles such as of machinery and goods that often move compromising line-of-sight communication and introducing multi-path interference. As a result, communication links and network topology vary requiring frequent retransmissions and rerouting. Such overhead can significantly reduce network performance or even render the network useless.

WSN support many applications including, but not limited to, civil, military, industrial, agricultural, and residential sensing. In civil infrastructure applications WSN

can be used to monitor and detect infrastructure health and safety measurands. WSN methods can be applied to traffic monitoring, structural health monitoring, and provide an augment to emergency communications. For applications listed above the use of distributed energy-efficient routing protocols are desirable due to the large-scale of the applications and the adaptable nature of WSN. Additionally, energy-efficient protocols allows for construction of networks that are reliable, low-cost, low-power, and that are energy aware. Figure 1 depicts an application example where a distributed WSN is applied to civil infrastructure monitoring. The figure highlights the use of WSN in the relay of data, the collection of traffic data, and also for the use of load testing and inspection methods.

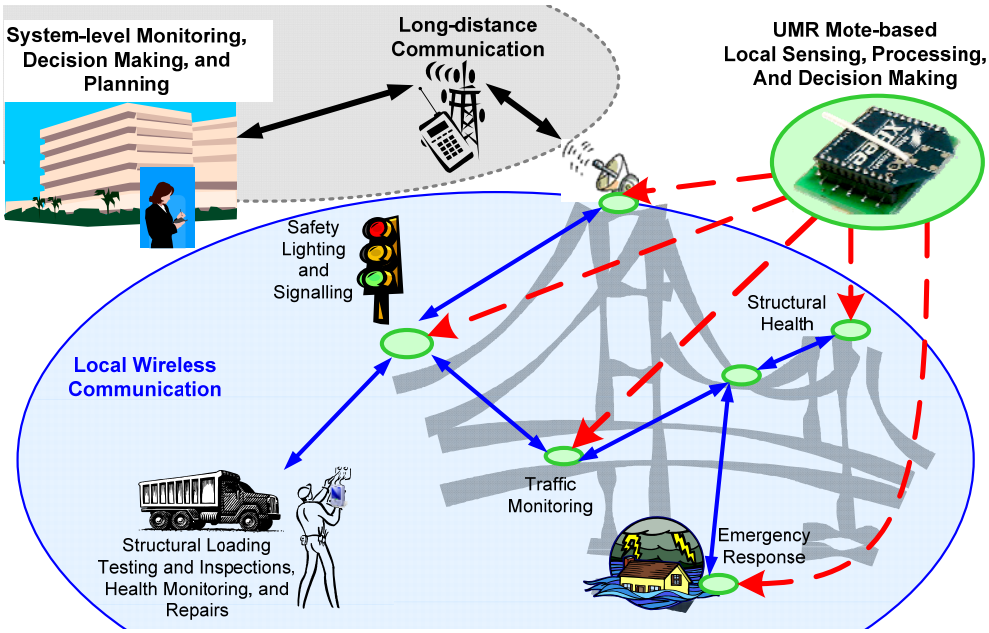

Fig. 1 Application example for WSN

This paper presents the optimized energy-delay sub-network routing (OEDSR) [18, 19]. The OEDSR protocol is used in WSN to provide optimal routing calculations in energy and delay dependent environments at the CH level. The OEDSR protocol is compared to the ad-hoc on-demand distance vector (AODV) routing scheme. Both

protocols are implemented on two hardware motes and their performance tested. A comparison of protocol performance is presented using throughput, end-to-end (E2E) delay, route-setup time, re-routing optimality, drop-rate, and dropped packets. Additionally, comparison of the energy consumption for routing is shown to illustrate the reduction in overhead in the OEDSR protocol. Key contributions include:

- Expansion of the OEDSR protocol to include detection and mitigation of link-failures at the point of failure, varying the retries allowing for interference mitigation, and RSSI filtering of packets to provide link stability,
- Comparison of the performance for the OEDSR protocol verses the AODV protocol in a hardware environment showing energy consumption during routing, route-setup times, and scalability, and
- Discussion of application constraints related to hardware, software, and environmental issues.

II. Optimized Energy-Delay Sub-Network Routing (OEDSR) Protocol

In OEDSR, sub-networks are formed around an event/fault and these sub-network member nodes wake up while elsewhere in the network nodes are left in sleep mode. An appropriate percentage (typically 20%) of nodes in the sub-network are elected as CHs based on a metric composed of available energy and relative location to an event [18,19] in each sub-network. Once the CHs are identified and the nodes are clustered relative to the distance from the CHs, the routing towards the base-station (BS) is initiated. First, the CH checks if the BS is within the communication range, if so data is sent directly to the BS. Otherwise, the data from the CHs in the sub-network are sent over a multi-hop route to the BS. The proposed routing algorithm is fully distributed requiring only local

information for constructing routes, and is dynamic by adapting to changing network conditions. The BS is assumed to have sufficient power to allow high power beacons from the BS to be sent to all nodes in the network to provide knowledge of the distance to the BS. Therefore, it is assumed that all nodes in the network know the distance to the BS at all times. Though OEDSR borrows concept of relay nodes from the energy-delay metric from OEDR [1], selection of relay nodes (RN) does not maximize the coverage of two hop neighbors. Here, the selection of a relay node is set to maximize the link cost factor which includes energy, end-to-end delay and distance from the BS to the RN.

In OEDSR routes are selected based on the Link Cost Factor (LCF) of nodes in the range of the CH or RN. Initially, a CH broadcasts HELLO packets to all nodes in range and receives ACK packets from all the relay candidates that are in the communication range. The ACK packets contain information such as the node ID, available energy, average queuing and processing delay at a node, and distance from the BS. The RNs that are farther away from BS than the current node do not respond to the HELLO packets. If one of the ACK packets was sent from the BS, then it is selected as a next hop node; thus ending route discovery procedure. Otherwise, the current node builds a list of potential RNs from the ACKs. Then it selects the optimal RN using the LCF parameter. The same procedure is carried out for all hops to the BS. The advantage of this routing method is reduction of the number of relay nodes that have to forward data in the network, and hence the scheme reduces overhead, minimizes number of hops, and communication due to flooding.

A. Optimum Relay-Node-Based Link Cost Factor

The LCF from a node to its next hop node '*i*' is given by (1) where D_i represent the delay that to reach the next hop node, x_i is the distance between the next hop node and the BS, and E_i is the energy remaining at the next hop node.

$$LCF_i = \frac{E_i}{D_i \cdot x_i} \tag{1}$$

In equation (1), consideration of the remaining energy at the next hop node increases network lifetime; the distance to the BS from the next hop node reduces the number of hops and end- to-end delay; and the delay incurred to reach the next hop node minimizes any channel fading problems and processing delay. These energy considerations do not include the transmission path energy since no assumptions for transmission power control are made. Therefore, the cost of path energy is considered constant for each potential link. When multiple RNs are available for routing of packets; the most optimal RN is selected based on the maximum LCF. Hence, the proposed OEDSR protocol is a high-performance on-demand routing protocol. For detailed discussion of OEDSR refer to [18]. The route selection is illustrated through the following example.

B. OEDSR Routing Algorithm through Example

Consider the topology shown in Fig. 2, where the link cost factors are taken into consideration to route packets to the BS.

OEDSR Steps:

i) Start with an empty relay list for the source node *n*: *Relay(n)={ }*. Here node n_4 and n_7 are CHs.

ii) First, CH n_4 checks which other nodes are in transmission range. In this case, CH n_4 is in range of nodes n_1, n_2, n_3, n_5, n_8, n_9, n_{12}, and n_{10}.

iii) The nodes n_1, n_2, and n_3 are eliminated as potential RNs since their distance to the BS is greater that of CH n_4 to the BS.

iv) Now, all the nodes in the range of CH n_4 transmit acknowledgement (ACK) packets and CH n_4 makes a list of possible RNs, i.e. n_5, n_8, n_9, n_{12}, and n_{10}.

v) CH n_4 then sends this list to CH n_7 which checks if it is range with any of the nodes in the list.

vi) Nodes n_9, n_{10}, and n_{12} are selected as the potential common RNs since they are in range with both CH n_4 and n_7. .

vii) The link cost factors for n_9, n_{10}, and n_{12} are calculated.

viii) The node with the maximum value of LCF is selected as the RN and assigned to *Relay(n)*. In this case, $Relay(n)=\{n_{12}\}$.

ix) Now node n_{12} checks if it is in direct range with the BS, and if it is, then it directly routes packets to the BS.

x) Otherwise, n_{12} is assigned as the RN, and all the nodes that are in range with node n_{12} and whose distance to the BS is less than its distance to the BS are taken into consideration, i.e. nodes n_{13}, n_{16}, n_{19}, and n_{17}.

xi) The LCF is calculated for n_{13}, n_{16}, n_{19}, n_{14}, and n_{17} and the node with the maximum LCF is selected as the next RN. In this case $Relay(n) = \{n_{19}\}$.

xii) Next the RN n_{19} checks if it is in range with the BS. If it is, then it directly routes the information to the BS. In this case, n_{19} is in direct range, so the packets are sent to the BS directly.

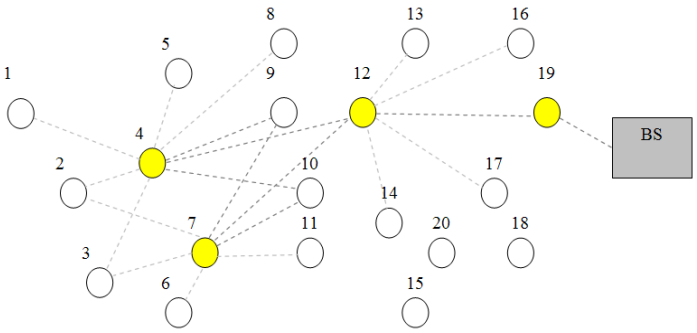

Fig. 2 Relay node selection

III. Optimality Analysis for OEDSR

The optimality and functionality of the OEDSR protocol is analyzed for the following assumptions and theorems.

Assumption 1: All nodes know the position of the BS irrespective of whether they are in motion or not. When the BS is mobile, it sends out location information to all the nodes in the network.

Assumption 2: Sufficient number of nodes exists for adequate node density and redundancy to provide at least one next-hop node towards the BS.

Assumption 3: Clustering and the use of low-power and sleep modes in conjunction with the energy balancing of OEDSR will minimize or eliminate routing holes. Energy consumption is balanced in clusters.

Note, in the cases where routing holes may form, a mitigation method could be added to compensate [11]. One option would be to use a simple flooding procedure to escape the routing hole and then continue with OEDSR procedures. Another option would be to introduce a feedback mechanism to help divert the route away from potential holes.

Theorem 1: The link cost factor-based routing generates viable RNs to the BS.

Proof: Consider the following two cases

Case I: When the CHs are one hop away from the BS, the CH selects the BS directly. In this case, there is only one path from the CH to the BS. Hence, OEDSR algorithm is not needed.

Case II: When the CHs have more than one node to relay information, the OEDSR algorithm selection criteria are taken into account. In Fig. 3, there are two CHs, CH_1 and CH_2. Each CH sends signals to all the other nodes in the network that are in range. Here, CH_1 first sends out signals to n_1, n_3, n_4, *and* n_5 and makes a list of information about the potential RN. The list is then forwarded to CH_2. CH_2 checks if it is in range with any of the nodes in the list. Here, n_4 and $n_5$are selected as potential common RNs. A single node must be selected from both n_4 and n_5 based on the OEDSR link cost factor. The cost to reach n from CH is given by (2). So based on the OEDSR link cost factor, n_4 is selected as the RN for the first hop. Next, n_4 sends signals to all the nodes it has in range, and selects a node as RN using the link cost factor. The same procedure is carried on till the packets are sent to the BS.

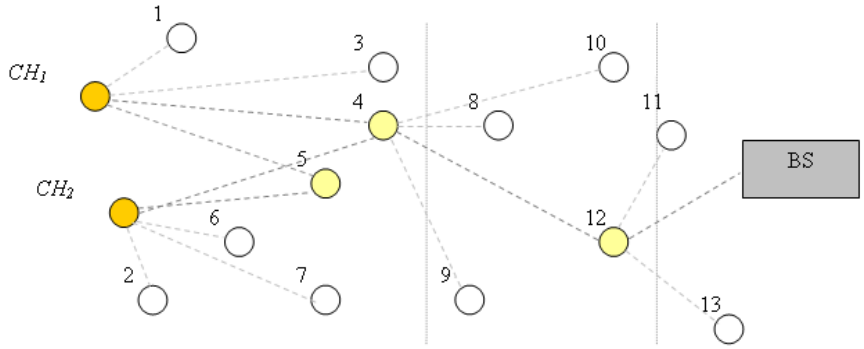

Fig. 3 Link cost calculation

Lemma 1: The intermediate nodes on the optimal path are selected as RNs by the previous nodes on the path.

Proof: A node is selected as a RN only if it has the highest link cost factor and is in range with the previous node on the path.

Lemma 2: A node can correctly compute the optimal path (with lower end to end delay and maximum available energy) for the entire network topology.

Proof: When selecting the candidate RNs to the CHs, the distance from the candidate RN to the BS is less than the distance from the CH to the BS. When calculating the link cost factor, available energy is divided by distance and average end-to-end delay to ensure that the selected nodes are in range with the CHs and close to the BS. The number of multi-point RNs in the network is minimized.

Theorem 2: OEDSR protocol results in an optimal route (the path with the maximum energy, minimum average end-to-end delay and minimum distance from the BS) between the CHs and any source destination.

Proof: Directly follows from Lemma 1 and 2 and Theorem 1.

IV. Overview of AODV and Comparison to the OEDSR routing protocol

The AODV protocol was selected for comparison to OEDSR since both protocols are reactive in nature, and since AODV is a widely accepted routing protocol in the ad-hoc mobile network area. Since OEDSR operates at the CH level and relies on another clustering agent such as SOS [18], it is then similar to mobile ad-hoc networking operations. While WSN mainly are stationary there is mobility in WSN systems depending on the applications.

The AODV protocol employs flooding by a series of route requests (RREQ), acknowledgements (ACK), and route-replies (RREP) to discover the network and to plot an appropriate route. In contrast, OEDSR operates by querying one-hop neighbors,

evaluating options for the next link and then selecting the most optimal route available. The functional differences in AODV and OEDSR are demonstrated through example.

A. Comparison of AODV and OEDSR through Example

In Fig. 4 an example of AODV routing is demonstrated. Node N_2, which is a CH, begins a routing session by sending a RREQ via broadcast mode. Each node receiving a RREQ broadcast sends an ACK packet (ACK packets are not shown in the figure). Then, each node that received the RREQ sends its own RREQ on broadcast continuing the pattern across the network. N_2 is the CH and is receiving data from a source node that is omitted from the graphic. As a CH, N_2, generates a RREQ and floods the network with this message. Each node receiving this message sends back an ACK (not shown in the graphic) allows the CH to construct a routing table. The RREQ propagates outward in all directions to all nodes in the network until it reaches the destination, in this case the BS. Once the BS receives a RREQ it sends a RREP to the first node from which it received the RREQ. The RREP is sent back along the known path where each node selected is along the path of least delay seen during the routing session, equating to a "first-come-first-select" paradigm.

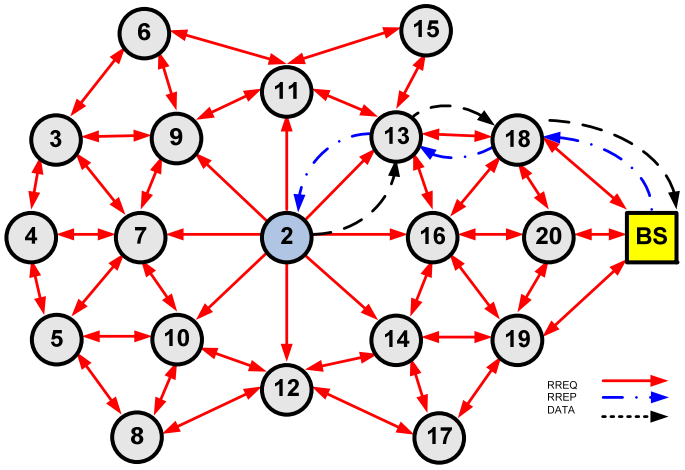

Fig.4 Central CH topology routing activity example for AODV

By comparison, OEDSR reduces the amount of traffic generated during the routing stage. As shown in Fig. 5, the CH sends a HELLO packet in broadcast mode to find next hop neighbors. Each node that receives the HELLO packet and is closer to the BS will respond with an ACK packet that includes the energy of that node, the distance of that node to the BS, and the delay estimated to the next hop node. The CH will wait for a short period to collect ACKs from its neighbors and then the CH compares LCF values to find a node that maximizes (1). Afterwards the CH sends a SELECT packet to the best node selecting it as a RN; the new RN will then repeat the HELLO/ACK/SELECT cycle in its local area. This procedure continues until the BS is reached. Once the BS is found packet flow is initiated between the CH and the BS. In contrast to flooding used by AODV, in OEDSR, only nodes that are known to be closer to the BS will continue the routing cycle. Additionally, the routing cycle is repeated only by the selected RN. Consequently, the amount of generated traffic is reduced in OEDSR over AODV.

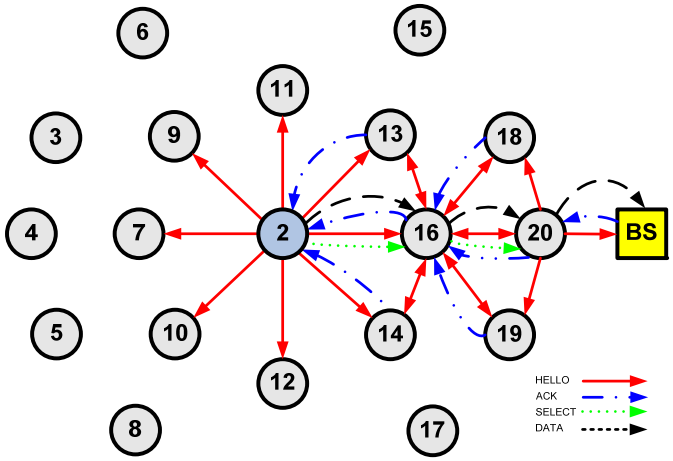

Fig. 5 Central CH topology routing activity example for OEDSR

B. Advantages of OEDSR over AODV

OEDSR exhibits some advantages over AODV. AODV inherently provides a path with a short delay and a low amount of interference since the path that responds the

quickest is selected. However, this provides no guarantees on route optimality. Consider a node that is on the optimal path and experiences a longer delay due to processing and queuing. The additional delay in forwarding RREQ will cause another node to be selected. Also, a sub-optimal route that ignores distance can establish a route through nodes that do not provide distance gain toward the BS.

Furthermore, OEDSR considers available energy in the network nodes in contrast to AODV. In this case AODV may select a node with low available energy thus causing it to quickly deplete. Consequently, it will be impossible to monitor the corresponding portion of a network as the depletion of the node may lead to partitioning of network. Finally, link-failure handling differs in the two protocols in that OEDSR heals routes at the failure point while AODV starts over from the original source. This difference in link-failure mitigation can lead to lost packets with key information about the monitored system that are queued up along the route and lead to longer route-healing times.

V. Hardware Implementation Description

A comparison of OEDSR and AODV capabilities in reference to sensing, processing, and networking support is shown using the same hardware platforms. The same hardware was used for both the sensing nodes and the CH's of the network. Additionally, the software architecture as implemented for testing is discussed with respect to packet types, control flow, and memory implications. Finally, hardware issues are discussed with regard to network density, RSSI filtering, and transmission retries.

A. Hardware Description and Limitations

Implementation of any algorithm is constrained by the capabilities of the selected hardware. For WSN, low-power, small form-factor, and high-speed processing hardware is desired. Low-power consumption was given the highest priority for the testbed. Silicon

Laboratories® 8051 variant hardware was selected for its ability to provide high-speed 8-bit processing, low-power consumption, and interface compatibility with peripheral hardware components. Limitations of the 8051 variant family are limited memory space and a maximum processing speed. Consequently, local data processing requirements are not be too great.

B. Hardware and Software Architecture

Two platforms are considered with commercially available software. The software imposes additional considerations.

1) Sensor Node – Instrumentation Sensor Node

The MST Instrumentation Sensor Node (ISN), as seen in Fig. 6(a), is used for interfacing to several types of sensors and as a data relay to CHs, i.e. it is a source of sensor traffic. The ISN provides sensor monitoring through use of a small and low-power device that can be controlled by CHs. The ISN can be instructed by control packets to transmit data in raw or pre-processed forms. Data pre-processing on the ISN allows for the opportunity to provide information flow on the network, implement adjustable digital filtering, or perform statistical studies at the sensor location. These capabilities provide the possibility of reduction of network traffic and embeddable sensors that provide intelligence to the application space.

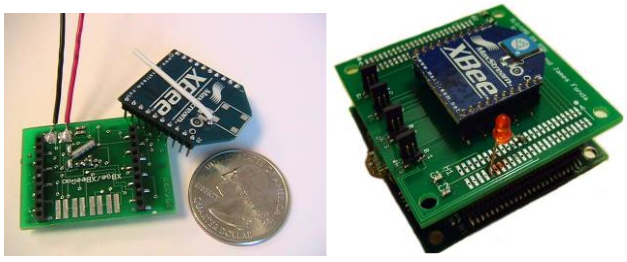

Fig. 6 (a) Instrumentation Sensor Node (ISN), (b) G4-SSN

2) Cluster Head and Relay Nodes

The Generation-4 Smart Sensor Node (G4-SSN) [20], as seen in Fig. 6(b), was originally developed at MST and subsequently updated at St. Louis University (SLU). It was chosen as the CH. The G4-SSN provides abilities for sensing and processing. The former include strain gauges, accelerometers, thermocouples, and general A/D sensing. The later include analog filtering, CF memory interfacing, and 8-bit data processing at a maximum of 100 MIPS. The G4-SSN provides memory and speed advantages over the ISN that make it a suitable choice for implementation as a CH or a RN. Future work is being undertaken to develop an improved platform that is more powerful than an ISN and smaller in size than a G4-SSN [21].

3) Comparison of ISN and G4-SSN Capabilities

The ISN was designed to be a simplified sensor node with the abilities to sample, process and transmit data and has a limited ability to process data relative to the G4-SSN. The abilities of the G4-SSN and ISN are shown in Table I and compared to a commercially available platform. The table shows that the G4-SSN has approximately 4-times the processing speed available relative to the ISN. Memory constraints are also compared between the platforms, with the G4-SSN having more available code space and RAM. These hardware differences show the design criteria for the ISN to be a ‘sample-and-send’ sensor node. In comparison, the G4-SSN can be used for networking functionality and other tasks that require more memory and processing capabilities.

TABLE I COMPARISON OF G4-SSN AND ISN CAPABILITIES

	Ic @ 3.3V [mA]	Flash Memory [bytes]	RAM [bytes]	ADC Sampling Rate [kHz]	Form-Factor	MIPS
G4	35	128k	8448	100 @ 10/12-bit	100-pin LQFP	100
ISN	7	16k	1280	200 @ 10-bit	32-pin LQFP	25
X-Bow	8	128k	4096	15 @ 10-bit	64-pin TQFP	8

4) Software Architecture for OEDSR

The software architecture utilized to implement OEDSR and AODV on the MST/SLU 8051 motes is shown in Fig. 7. A multi-level structure was employed to implement the protocols that allowed separation of software and hardware components. Platform portability and increased ability to modularize the network stack for future additions is accommodated.

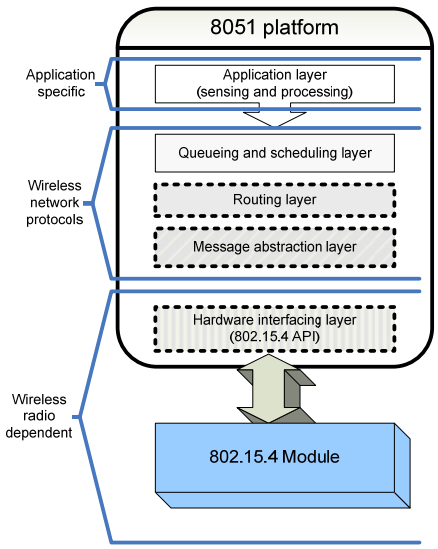

Fig. 7 Software architecture of OEDSR implementation

The three-tier structure provides flexibility to the radio and application design. The wireless radio dependent components are interfaced with networking layers using a message abstraction layer providing generic access to the physical and link level

parameters, for example transmission power level and RSSI indicator. Consequently, cross-layer protocols such as OEDSR can be easily implemented.

The main components of the software architecture consists of

Physical interface between 8051 and 802.15.4 module – in this setup a standard serial interface connects the processor with the radio module

Abstraction layer – provides generic access to the physical and link layers

Routing layer – contains routing implementations

Queuing layer– a simple drop-tail queuing policy is employed, and

Application layer –measurement and processing of sensor data

C. OEDSR Routing Implementation

The implementation of the OEDSR protocol is now discussed. The implementation of the control flow and memory requirements is shown to illustrate the capabilities of the protocol.

1) OEDSR Traffic Cases

Figure 8 presents a block diagram of the routing control flow implementation. The handling of the received message starts at the RX block, where the type of the packet is determined. Next, packet processing depends on the packet type. Once the packet has been handled the control flow returns to the idles state, or RX mode, and awaits a new packet.

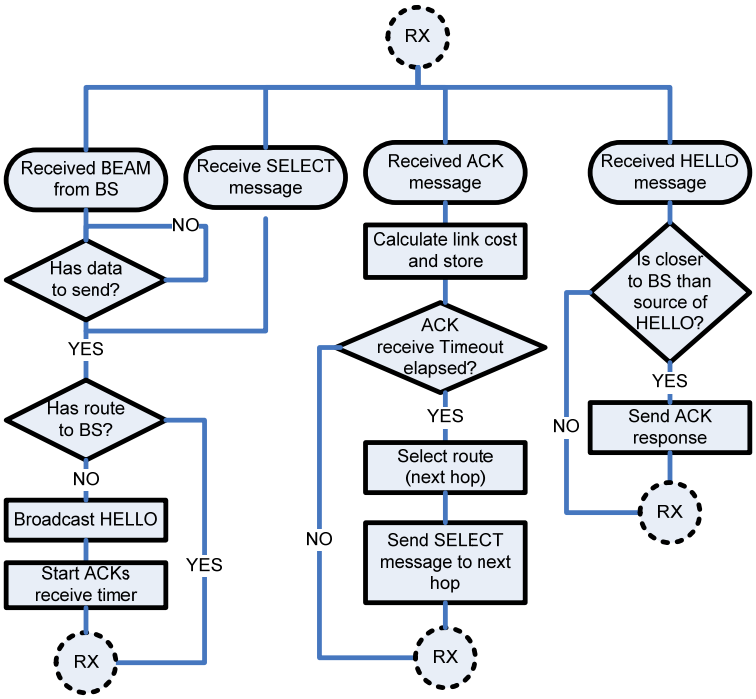

Fig. 8 Control flow scheme for OEDSR routing implementation

D. AODV Implementation description

The implementation of the AODV routing protocol is now discussed. An overview of the structure of the protocol including handling of traffic cases and memory requirements are shown.

1) AODV Traffic Cases

Implementation of the AODV protocol control flow is shown in Fig. 9. Upon reception of a packet the AODV protocol determines the type and follows the pre-described action. After packet processing it also returns to an idle state similar to the diagram in Fig. 8.

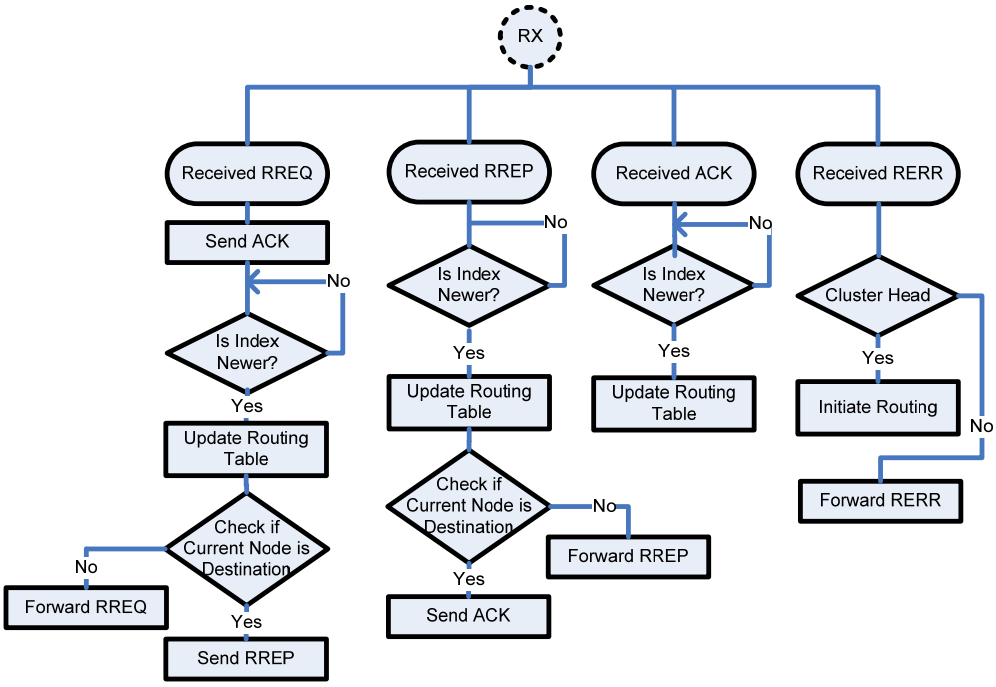

Fig. 9 Control flow scheme for AODV routing implementation

E. Hardware Implementation Lessons Learned

The use of a hardware testbed provided useful knowledge of deployment issues that are not always encountered during simulation studies. The importance of memory limits, network density, RSSI filtering, channel conditions, and other environmental factors are highlighted.

1) Memory Limitations

Memory requirements for network and data applications must be calculated. Basic requirements include buffer space for ADC's and UARTs, network queues, routing tables with supporting variables, and application-specific buffers and variables. For networking applications, the main impact on memory capacity is due to queuing of packet flows originating from, or passing-through, a particular node. As the number of sensor nodes increases either in a cluster, or in the CH population, queuing of the flows becomes more memory intensive.

Each protocol has different memory requirements. In the case of AODV only the two 2-byte addresses must be stored for each entry in the routing table; one address for the node in the direction of the particular destination and one byte for the destination node. In the case of OEDSR there are several entries for each possible link on the route to calculate the LCF. For each entry to the routing table include the distance, energy, and delay. Both the distance and energy are 4-byte values and the delay is a 2-byte value. Therefore each table entry under OEDSR must hold the 1-byte address, 4-byte distance and energy value, and the 2-byte delay value. In our test 15 entries where reserved in the routing tables for a total of 150-bytes. Also, there were ~50B of other routing variables. For AODV the RAM requirement is only 60-bytes. Additionally, the RAM used for the UART, ADC, networking and temporary variables are 400B, 100B, 200B and ~300B respectively.

In case of ISN, the memory requirements will leave no space for queuing. However, the G4-SSN has approximately 7k for use as queue or application space. The G4-SSN could hold about 70 packets for the case of network queuing. However, if this queue space is shared among multiple flows, the available capacity per flow will reduce to only a few vacant packet slots.

2) Network Density

Density of the network, measured as number of nodes per square foot (meter), was found to have an impact on performance of the routing protocol in the implementation. If node density is sufficiently high interference can cause the radio channel to be unreliable and increase route-setup times. In addition, the ability of a node to transmit/receive packets can be compromised; therefore, care must be given to distribution of the active nodes in the network and node sleep cycles to limit active nodes. Observations showed

that when nodes are closely packed during testing interference from neighboring can cause link failures.

3) RSSI Filtering

The use of RSSI as an indication of link reliability was also considered. A soft-limit was placed on the RSSI for accepting routing messages; thus rejecting packets with RSSI levels below an assigned threshold; these rejected packets would have a greater chance of corruption, or potentially be packets that occasionally pass through a weak link. Therefore, the RSSI filter prevents excessive packet drops and helps to reduce packet retransmissions for "borderline" stable links.

Consider a node receiving ACK packets. If the packet's RSSI value is above the limit, the transmitting node will be considered as a primary candidate for route selection. The original OEDSR selection criterion applies for such a node. However, when a packet is received with the RSSI below the soft threshold it will be considered only as a fallback choice when no node with RSSI above threshold is found.

4) Channel Condition

Poor channel condition such as a noisy or obstacle rich environment can hinder the overall performance of the network protocol. A spectral scanner measured interference levels for the testing area and the nodes were set for the lowest interference channel. . Specifically, the cross-over between 802.15.4 channels and 802.11x channels was of concern. Both technologies share the 2.4 GHz ISM band. The 802.11x channel uses a wider channel bandwidth and can interfere with two or more 802.15.4 channels. The performance statistically degraded when channels with high levels of noise and interference were used for testing. For mobile scenarios or dynamic environments, an initial scan may not be sufficient and a dynamic approach may be needed.

5) Transmission Retries

Experiments were conducted to evaluate performance for a different number of packet retransmissions. These experiments were aimed at alleviating problems with unreliable channels as discussed in previous section. Performance gains were achieved by adjusting the maximal number of retries for packet transmission allowing temporary interferences to be overcome. Performance improvements were seen in the route-setup times and reliability of the routing for OEDSR and AODV. On the other hand, an excessive number of retries can lead to selection of unreliable routes that result in frequent retransmission leading to reduced network capacity. The number of transmission retries was varied to investigate the performance variation and results are shown in the next section.

6) Route Repair

Increasing the number of retries can improve link reliability and network performance; however, the routing protocol still has to address the problem of link or node failures. For example, when the energy of a node depletes a section of the route will be inoperable regardless of the number of maximum retransmissions. In this work, an explicit and low latency method to heal broken routes is introduced for OEDSR. The nodes use MAC layer acknowledgements to packet transmissions for detection of dropped link. When the predefined number of packets are dropped due to lack of acknowledgements the link is considered broken. The node can then detect the lost link and initiate by beginning routing at the point of failure. In contrast, AODV protocol requires the source node to be informed about the lost link, and to again restart routing from the source node. As a result OEDSR recovers faster than AODV due to its reduced re-routing procedure.

VI. Experimental Setup and Results

Hardware experiments were performed with two objectives: (1) to evaluate performance for both AODV and OEDSR; and (2) to identify practical issues to be considered for protocol redesign and implementation. For the former, the performance of the protocols is evaluated in terms of throughput, E2E delay, route-setup time, drop rate, jitter, and quality of routing. The latter will investigate if typical assumptions used in simulation, and theoretical analyses, are true in hardware implementations. Scenarios considered include:

- Single round comparison of energy overhead using twenty nodes
    - Central CH and grid topologies
    - Experiments conducted with different number of transmission retries
- Network of ten nodes ran to network failure
    - Network ran until network failure due to energy depletion or partitioning of the network
- Network with mobile CH

A network of twenty nodes was assembled for testing. The nodes use 802.15.4 modules transmitting at a 250 kbps RF data rate. The ISN is used to generate CBR traffic and provide a data source. CH's and RN's are implemented using the G4-SSN. The CH provides the routing capabilities by choosing the RN for routing of traffic toward the BS. The nodes' processor interfaces to the 802.15.4 module at 115,200 bps. The ISN, CH, and RN's are equipped with low-power 1-mW 802.15.4 modules; while the BS is equipped with a high transmission power 100-mW 802.15.4 module to increase its range

for beam signals. CBR rates of 3700 bps were initiated at the ISN's for packet flow to the BS.

A. Description of the Experimental Scenario

Nodes were placed in the specified topology and signaled to run networking tasks initiated by the BS. The nodes are network configurable to set their distance, energy, routing protocol, and may be sent messages to emulate maintenance. Maintenance signals include messages to recharge a node, reset nodes, and configure analog-to-digital converter (ADC) parameters. Both central CH and grid topologies were considered for testing. The central CH topology demonstrates a "worst-case" scenario when compared to the grid topology. Under the central CH topology half of the network will be in-active with OEDSR while all nodes can be active using AODV.

1) Scenario 1: Single Round Tests with Single Cluster Head

Experimental results were obtained for a network of twenty nodes with a CH near the center of the network. Single round results are for forty seconds using both topologies. Additionally, the number of MAC retries was varied.

The central CH (CCH) topology is shown in Fig. 4 and Fig. 5. Results in Table II and Table III show that AODV consumes roughly ten times more energy for the six and nine retry cases. Differences are also seen in the route-setup times and the amount of energy used in routing. For OEDSR, the average route-setup times are shorter than for AODV since the flooding used by AODV causes multiple collisions and may result in failed routing session. In contrast, collisions are rare for OEDSR since it transmits only a limited number of messages when compared to AODV. The achieved throughput is close to the generated data rate for both routing protocols. The small variation is attributed to varying number of hops in different experiments and random channel uncertainties that

result in packet losses. In summary, the added overhead of OEDSR is outweighed by the higher energy efficiency over the AODV.

The maximum number of transmission retries has an impact on protocols performance. In case of a maximum of three retries, both protocols had difficulties in setting up the route. In contrast to AODV, the OEDSR was not able to establish route within the forty seconds of the experiments since it requires a reliable communication to finish the routing. Interference can easily disrupt communication of a critical message and lead to a failed routing session due to channel uncertainties with that lead to collisions and dropped packets. The higher number of retransmissions increases the probability of successful transmission, thus improving performance. As a result, the AODV protocol was able to establish a route using three retries since the use of flooding with multiple re-broadcasts increased probability of the request getting through. Similar positive effects were observed when the number of retries was increased for both protocols. Also, there is a decrease in route-setup time as the number of retries increases due to less dropped packets.

TABLE II COMPARISON OF AODV AND OEDSR PERFORMANCE FOR CENTRAL CH TOPOLOGY

	AODV	OEDSR	AODV	OEDSR	AODV	OEDSR
Number of Retries	3 Retries		6 Retries		9 Retries	
Avg. Throughput [bps]	2809.30	--	3211.60	3102.00	3017.70	3010.40
Avg. E2E Delay [sec]	0.267	--	0.228	0.246	0.255	0.274
Avg. Drop Rate [packets/sec]	0.061	--	0.011	0.103	0.022	0.172
% Energy for Routing	5.663	--	4.455	0.427	4.823	0.435
Avg. Jitter [sec]	0.046	--	0.022	0.043	0.044	0.075
Avg. Variance of Jitter	0.002	--	0.001	0.003	0.003	0.006
Avg. Route Setup Time [sec]	3.244	--	3.794	1.550	2.322	1.597

The grid topology, shown in Fig. 10, was used to evaluate the protocols in case of network using a larger number of hops between the source and destination. As in the case of CCH topology, the OEDSR results show a ten times reduction in the energy consumption. Also, the route setup time decreases with number of retries since more packets are successfully transmitted. The drop rates, jitter, and E2E delay were found to be similar; this is due to the similar number of average hops in the topology. It was observed that the E2E delay was similar on average between both protocols. This result was due to the number of hops being similar on average between the two protocols over the entire testing set.

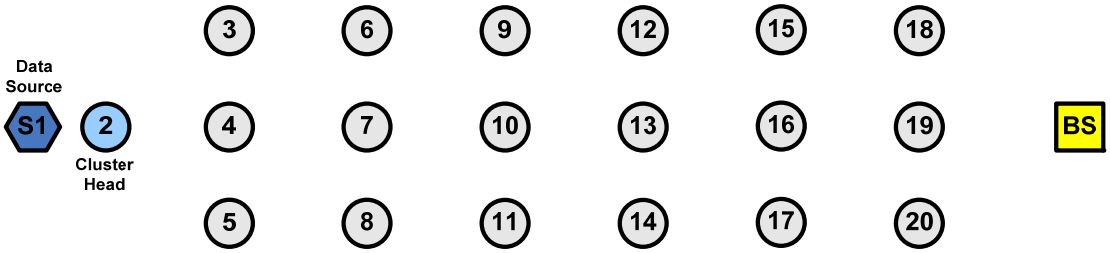

Fig. 10 Grid topology and single CH

TABLE III COMPARISON OF AODV AND OEDSR PERFORMANCE FOR GRID TOPOLOGY

	AODV	OEDSR	AODV	OEDSR	AODV	OEDSR
Number of Retries	3 Retries		6 Retries		9 Retries	
Avg. Throughput [bps]	1785.30	--	3387.60	3035.60	3642.50	2811.00
Avg. E2E Delay [sec]	0.151	--	0.229	0.244	0.249	0.263
Avg. Drop Rate [packets/sec]	0.043	--	0.013	0.028	0.016	0.022
Percentage Energy for Routing	26.609	--	6.188	0.529	4.000	0.515
Avg. Jitter [sec]	0.027	--	0.040	0.032	0.110	0.043
Avg. Variance of Jitter	0.002	--	0.001	0.001	0.004	0.001
Avg. Route Setup Time [sec]	11.253	--	5.441	2.809	1.338	1.947

2) Network with Single Cluster Heads Ran to Depletion

Network experiments were also performed to investigate how energy depletion of nodes affects network coverage and node-lifetime. In this scenario, the depleted nodes cause a lack of sensor visibility in the corresponding area leading to missed events. In these experiments, a particular amount of energy was assigned to each of the nodes. The energy was reduced for each packet that was processed by the node. The amount of reduction was linearly related to the length of the packets. Fig. 11 illustrates the network topology that was used for the experiments. Node time-of-death (TOD) values and address were observed. The experiments were run until the network becomes partitioned and no more valid routes were possible.

Depletion results show that OEDSR provides longer 100% network coverage as well as longer total network lifetime before partitioning. Once the network becomes partitioned the CHs in the detached part of network will not be able to transmit to the BS. Figure 12 shows failure times of the first five nodes for both OEDSR and AODV routing protocols. The figure shows that the energy aware routing of OEDSR more evenly distributes energy consumption. Consequently, OEDSR allows for a longer time before the first node failure, as a result, partitioning of the network begins three times later in contrast to AODV. Also, the nodes for the OEDSR scenario become depleted together as a network since the energy usage is evenly distributed. Moreover, the network with OEDSR will operate 6.7% longer than with AODV since OEDSR reduces energy consumption during routing session.

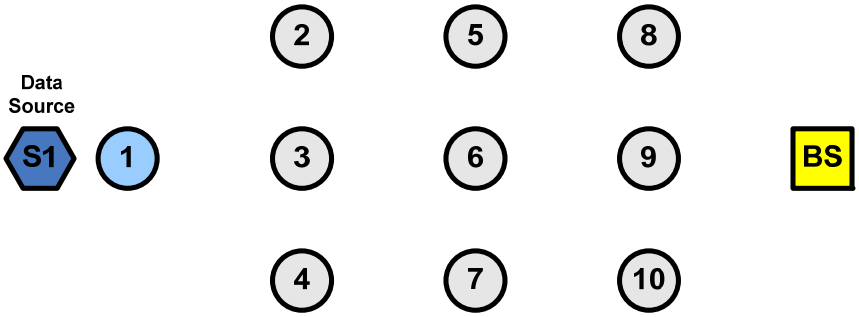

Fig. 11 Depletion experiment topology

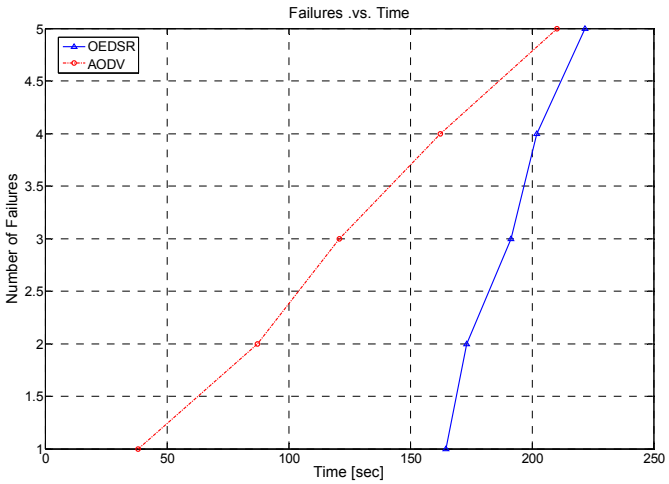

Fig. 12 Number of node failures verses time for OEDSR and AODV

3) Network with Single Cluster Heads and Node Mobility

Node mobility experiments demonstrated the ability of the routing protocols to tolerate nodes moving through the network area. This scenario could be viewed as a moving vehicle with a sensor cluster onboard transmitting data through the stationary peer-nodes. Mobility experiments were performed by moving CH_2 through the network at a constant speed. These experiments were performed with the grid topology as shown in Fig. 10. It took CH_2 thirty seconds to move between each layer of the grid topology.

Figs. 13 and 14 illustrate the throughput (bps) for five trials of OEDSR and AODV. Results show that both routing protocols provide adequate coverage and service to the mobile nodes. Both protocols are able to sustain connectivity during testing. The major differences in the results show that OEDSR changes routes on a periodic basis to balance

the energy usage among the relay nodes. However, due to the route change OEDSR also has the opportunity to select a more optimal route and provide a higher throughput during some rounds. Throughput decreases with distance to BS since number of required retransmissions increases with distance; thus reducing channel capacity. However, the OEDSR protocol periodically reroutes allowing for a selection of more favorable path where the link quality increases and higher throughput is possible.

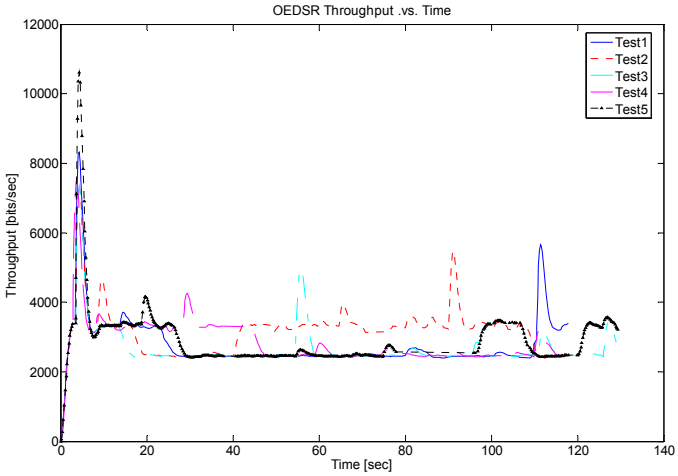

Fig. 13 OEDSR Throughput during mobility conditions for OEDSR and AODV

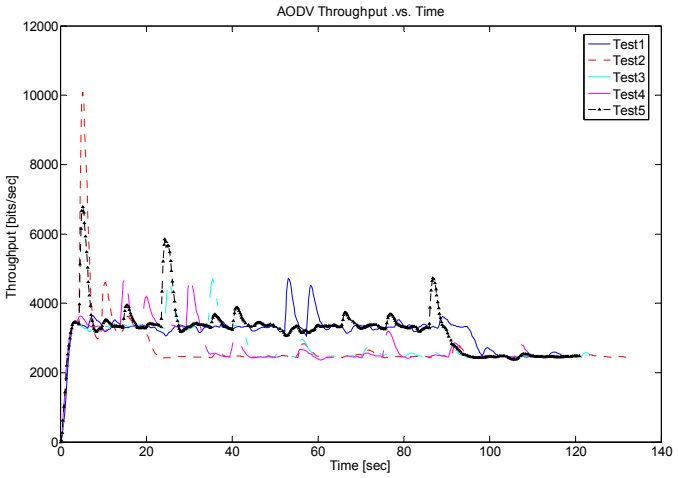

Fig. 14 AODV throughputs during mobility conditions for OEDSR and AODV

VII. Discussion of Results

Experiments using the AODV and OEDSR protocols on MST/SLU motes demonstrate the performance of both protocols under realistic conditions. Through the use of experiments presented above the advantages, disadvantages and performance of the protocols are discussed. Single round comparison of the protocols show energy consumption is an average of ten times lower for OEDSR than for AODV. Additionally, comparisons for depletion of the network energy over node lifetime shows that OEDSR effectively increases the network longevity and allows for the nodes to be on the network longer before the first node TOD. Also, experiments show the ability of both protocols to operate satisfactorily during node mobility.

Intrinsic differences to the protocols include that AODV does not employ any waiting periods while waiting for ACK packets. Instead, AODV takes routing nodes on a "first-come-first-select" paradigm and does not guarantee that the most optimal node is selected or does it help mitigate packet collisions during routing. In comparison, OEDSR has a built in window during routing allowing several nodes to offer viable routes. Additionally, during this window only nodes that are closer to the destination respond to HELLO packets, introducing less channel congestion during routing.

OEDSR provides advantages in terms of its ability to provide optimal routing using energy, delay, and distance as a routing metric. AODV selects only the nodes that respond the quickest but can travel a longer distance geographically, while OEDSR guarantees that the route taken will be as short as possible due to the inclusion of distance in the routing metric. OEDSR provides even energy dissipation due to its use of rounds, and its ability to select nodes that have more residual energy. Additionally, OEDSR re-routes from the point of failure, in contrast to AODV. This ability allows for packets that

are queued up on the current route not to be dropped and provides quicker re-routing in larger networks. In comparison, AODV must begin the route again at the source node after a link-failure. However, the complexity of OEDSR is slightly more than AODV due to the need for metric evaluation, but this is acceptable with the added energy performance.

Hardware implementation highlighted key issues in deployment and configuration. Insufficient memory limits the ability to host queues and implement larger routing tables. Other issues are related to environmental conditions such as network density, RSSI filtering, channel conditions, and transmission retries, which are interrelated. For instance, network density has a direct effect on selection of a RSSI level for a "good" link; while it also impacts the channel conditions since a densely packed network will provide more interference from node-to-node. Transmission retries is also an environmental factor and was changed manually to characterize its impact. Certainly, a minimum number of retries are needed under different environmental conditions, however an adaptive method for controlling the maximum number of retries should be explored in future work. Also, adaptive methods could be applied for selection of the RSSI level for "good", or stable, links.

VIII. Conclusions

In this work an extended study of the OEDSR protocol has been shown in the hardware environment. Comparison of the AODV protocol was also shown to demonstrate the performance gains with use of a reactive protocol that optimizes distance, available energy, and delay. For OEDSR route selection is based on a metric given by the ratio of energy available and delay multiplied with distance to provide the

link cost factor for optimal routing. By comparison, AODV provides a reactive routing protocol as well; however there are not provisions to ensure optimal routing is achieved.

The OEDSR protocol computes the energy available and average E2E delay values of the links and this information along with the distance from the base station determines the best RN. While ensuring that the path from the CH to the BS is free from loops, it also ensures that the selected route is both energy efficient, and has a minimal E2E delay. AODV however routes through RN's that respond the quickest during the routing session providing a potentially low E2E-delay path, but does not ensure that the route is short, or has enough energy to support the packet flow. Additionally, the lifetime of the network is maximized with OEDSR since the energy is taken into account while selecting nodes forming a route. Due to the consideration of energy level in the routing protocol, there is also a balancing of energy consumption across the network.

Testing of the OEDSR and AODV routing protocols on MST/SLU motes revealed implementation and hardware issues. Selection of hardware must consider the complexity and memory footprint of the proposed methods. For example, the ISN nodes were designed to minimize physical size of the node and reduce energy consumption. However, the selected processor lacks sufficient RAM to provide routing and queuing functionality on the network as compared to the G4-SSN. Therefore, minimum hardware requirements in terms of memory size, processing power, energy consumption, physical size, and the corresponding tradeoffs must be explored before a protocol is implemented on a particular platform.

Energy consumption during routing with OEDSR was improved by a factor of ten for the two topologies considered. In addition, OEDSR balanced energy consumption which

extended the network lifetime and coverage area. Specifically, the first node TOD is seen at a factor of three longer with OEDSR than with AODV. The extended time before failure means that the network has complete coverage over the deployment area for three times as long as the same network running AODV. Also, there is a 6.7% increase in network lifetime, and a tighter clustering of TOD's for the nodes under OEDSR. The TOD clustering demonstrates the ability of OEDSR to distribute the energy consumption and to deplete the nodes at closer to the same time letting the network fail more as a unit. Improvements in the failure mode of the network provide increased reliability and ensure that portions of the network are not partitioned off early in the network lifetime.

ACKNOWLEDGMENT

Thanks go to Dr. Kyle Mitchell of St. Louis University, St. Louis, MO. His continued support and advice on the G4-SSN system and hardware implementation issues are much appreciated.

REFERENCES

[1] N. Regatte and S. Jagannathan, "Adaptive and distributed fair scheduling scheme for wireless adhoc networks," Proc. of the Word Wireless Congress, pp. 101-106, May 2004.

[2] About GloMoSim*, Global Mobile Information Systems Simulation Library*, available online: http://pcl.cs.ucla.edu/projects/glomosim/

[3] W. R. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," *Proc. of the IEEE Hawaii Int. Conf. on System Systems Science(HICSS),* vol. 2, 10 pp., 2000.

[4] W. R. Heinzelman, J. Kulik and H. Balakrishnan, "Adaptive protocol for information dissemination in wireless sensor networks," *Proc. of the ACM/IEEE Int. Conf. on Mobile Computing and Networking (MobiCom)*, Seattle, WA, pp. 174-185, 1999.

[5] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann and F. Silva, "Directed diffusion for wireless sensor networking," *IEEE/ACM Trans. on Networking*, vol. 11, no. 1, pp. 2-16, 2003.

[6] F. Ye, G. Zhong, S. Lu and L. Zhang, "GRAdient Broadcast: A robust data delivery protocol for large scale sensor networks," *ACM Trans. on Wireless Networks*, vol. 11, no. 3, pp. 285-298, 2005.

[7] F. Ye, G. Zhong, S. Lu and L. Zhang, "PEAS: A robust energy conserving protocol for long-lived sensor networks," *Proc. of the IEEE Int. Conf. on Distributed Computing Systems (ICDCS)*, pp. 28-37, 2003.

[8] A. Manjeshwar and D. P. Agrawal. "TEEN: A routing protocol for enhanced efficiency in wireless sensor networks," *Proc. of the IEEE Int. Parallel and Distributed Processing Symp. (IPDPS)*, pp. 2009-2015, 2001.

[9] A. Manjeshwar, and D.P. Agrawal, "APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks," *Proc. of the IEEE Int. Parallel and Distributed Processing Symp. (IPDPS),* pp. 195-202, 2002.

[10] S. Lindsey and C. Raghavendra, "PEGASIS: Power-efficient gathering in sensor information systems," *Proc. of the IEEE Aerospace Conf.,* vol. 3, pp. 1125-1130, 2002.

[11] B. Karp and H. T. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," *Proc. of the ACM/IEEE Int. Conf. on Mobile Computing and Networking (MobiCom)*, pp. 243–254, 2000.

[12] Y. Yu, R. Govindan, and D. Estrin, "Geographical and energy aware routing: A recursive data dissemination protocol for wireless sensor networks," UCLA Computer Science Department Technical Report, UCLA-CSDTR-01-0023, May, 2001.

[13] H. Luo, F. Ye, J. Cheng, S. Lu and L. Zhang, "TTDD: A two-tier data dissemination model for large-scale wireless sensor networks," *Proc. of the ACM/IEEE Int. Conf. on Mobile Computing and Networking (MobiCom),* pp. 148-159, Sept. 2002.

[14] T. He, J. A. Stankovic, C. Lu and T. Abdulzaher, "SPEED: A stateless protocol for real-time communication in sensor networks," *Proc. of Int. Conf. on Distributed Computing Systems*, pp.46-55, May 2003.

[15] C. E. Perkins, E. Belding and S. Das, "Ad-hoc on-demand distance vector (AODV) routing," *IETF MANET Working Group*, available online: www.ietf.org/internet-drafts/draft-ietf-manet-aodv-14.txt, Feb. 2003.

[16] D. Johnson, D. Maltz and Y. Hu, "The dynamic source routing protocol for mobile ad hoc network (DSR)," *IETF MANET Working Group*, available online: draft-ietf-manet-dsr-09.txt, April 2003.

[17] M. Esler, J. Hightower, T. Anderson, and G. Borriello "Next century challenges: data-centric networking for invisible computing: The Portolano project at the University of Washington," *Proc. of the ACM/IEEE Int. Conf. on Mobile Computing and Networking (MobiCom),* pp. 256-262, 1999.

[18] S. Ratnaraj, S. Jagannathan and V. Rao, "OEDSR: Optimal energy delay subnet routing protocol for wireless sensor networks," *Proc. of the IEEE Conf. on Sensing, Networking, and Control (ICNSC)*, pp. 330-335, April 2006.

[19] J. W. Fonda, M. J. Zawodniok, S. Jagannathan, S. E. Watkins*, "*Development and implementation of optimized energy-delay sub-network routing protocol for wireless sensor networks," *Proc. of the IEEE Int. Symp. on Intelligent Control*, Munich, Germany, pp. 119-124, 2006.

[20] K. Mitchell, S. E. Watkins, J. W. Fonda, and J. Sarangapani, "Embeddable modular hardware for multi-functional sensor networks," *Smart Mater. Struct*., vol. 16, no. 5, pp. N27-N34, 2007.

[21] J. W. Fonda, S. E. Watkins, S. Jagannathan, and M. Zawodniok, "Embeddable sensor mote for structural monitoring," *Smart Structures and Materials 2008: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems,* SPIE Proc, vol. 6932, San Diego, CA, March 2008, In Press, 2008.

PAPER 2

Adaptive Distributed Fair Scheduling for Multiple Channels in Wireless Sensor Networks

James W. Fonda[1,2,*], Maciej Zawodniok[1], S. Jagannathan[1], and Steve E. Watkins[2]
Embedded Systems and Networking Laboratory[1]
Applied Optics Laboratory[2]
Department of Electrical and Computer Engineering
Missouri University of Science and Technology, Rolla, Missouri 65409-0040
jfonda@ieee.org, mjzx9c@mst.edu, sarangap@mst.edu, and steve.e.watkins@ieee.org

Abstract—A novel adaptive and distributed fair scheduling (ADFS) scheme for wireless sensor networks (WSN) in the presence of multiple channels (MC-ADFS) is developed. The proposed MC-ADFS increases available network capacity and focuses on quality-of-service (QoS) issues. When nodes access a shared channel, the proposed MC-ADFS allocates the channel bandwidth proportionally to the weight defined by priority of the packet's flow. The packets are dynamically assigned to channels based on the packet weight and current utilization of the channel. The dynamic assignment of channels is facilitated by use of receiver-based allocation using alternative routes to forward packets towards the BS. Moreover, MC-ADFS allows for dynamic allocation of network resources with little added overhead. Packet weights are initially assigned using user specified QoS criteria, then they are subsequently updated as a function of delay, enqueued (is this a word? Why not just "queued") packets, flow arrival rate, and the previous packet weight. The back-off interval is also altered using the weight adaptation. The weight update and the back-off interval selection ensure that global fairness is attained even with variable service rates. The MC-ADFS performance is demonstrated through hardware experiments and shows improved performance in terms of fairness index, flow rate, and end-to-end delay (E2E) compared to first-in-first-out (FIFO) and distributed fair scheduling (DFS) scheduling methods.

Index Terms—**Fairness, Adaptive-fair-scheduling, Weight-adaptation, Quality-of-Service, Embedded System.**

I. Introduction

Introduction of the IEEE 802.15.4 standard has accelerated the application of wireless sensor networks (WSN) for industrial environments. Use of small, low-power, radio enabled networks provide observability through a cost effective and deployable platform.

Research supported in part by Dept. of Education GAANN Fellowship, NSF I/UCRC on Intelligent Maintenance Systems, Air Force Research Laboratory Grant (FA8650-04-C-704) and Intelligent Systems Center.

Research into WSN has shown the ability to provide dynamic routing [1], intelligent processing of data, and information flow from deployed sensors and other algorithms [2, 3]. Limited bandwidth is a significant constraint in WSN, thus effective, and fair, management of radio resources is crucial to guaranteeing the quality of service (QoS). The primary focus of this work is to address challenges in scheduling packets, or flows, when nodes reside in an adaptive multi–channel network and secondly to show performance of a hardware implementation of a fair scheduling network protocol [4] using a WSN test-bed. Fair scheduling methods are ones that use finish-times to decide which packet, or flow, needs priority at a given time interval.

In sensing applications, communication protocols have to guarantee that information from each source is delivered to the destination. Moreover, important flows should be given proportionally higher priority in terms of throughput and end-to-end (E2E) delay to meet QoS criteria. Hence, a fair scheduling has to be employed in WSN to provide proper packet flow for a given application. In the literature, many algorithms and protocols regarding QoS metrics are found; however few address hardware constraints [1, 4-11]. In the deployment of WSN hardware constraints are important in reference to limitations in the memory, processing speed, and network bandwidth and their impact on the implementation of complex algorithms. To provide network capacity and effective distributed processing for WSN a paradigm of hardware and software co-design is desirable.

A number of fair scheduling schemes exist in the literature including centralized [5-8] and distributed [9, 10] approaches. Also, work on achieving fairness using distributed medium access control (MAC) protocols for wireless networks has been done[10].

Recent literature address methods based on round-robin scheduling [12], energy efficiency [13-16], distributed schedulers [17], priority based scheduling [18], channel state [19], and weighted fair queuing scheduling [20-22] for WSN. Round-robin methods allow for balanced processing of the network queues, but in comparison to other methods do not allow for arrival time control. Energy efficient methods often center on network lifetime studies and the balancing of node usage. In the data-centric and priority-based networks, much emphasis is given to QoS parameters. Additionally, the evaluation of hardware constraints is covered with respect to the complexity of the algorithm [23-25], however they do not address the ideas of hardware and software co-design.

One work [9] proposes a DFS protocol for wireless LANs. DFS [9] allocates bandwidth proportional to the weights of the flows set by the network administrator. DFS performs a fair allocation of bandwidth resources using a self-clocked fair queuing algorithm. However, this protocol may not be suitable for multi-hop networks with dynamic channel state and topology. When the network state changes, for example nodes move changing the physical topology; DFS ensures fair allocation of resources among the hops. However, it does not address the variation in end-to-end performance when the number of hops varies. For example, DFS results in large delay variations, or jitter, in reception of packets at the destination. Consequently, use of fixed weights does not guarantee weighted fairness in a dynamic environment. Finally, selection of initial weights is not addressed in DFS as the weights are not typically updated. Unless weights are selected appropriately, fairness cannot be guaranteed even for wireless networks using stationary nodes.

In general, the existing fair scheduling schemes determine appropriate weights in order to meet the desired QoS criteria. In most schemes weights are assigned and are not dynamically updated when network conditions change, and therefore do not provide the advantage observed in an adaptive distributed fair scheduling (ADFS) enabled network [5-10]. In the proposed method, scheduling is based on QoS using a self-clocked fair scheduling system with adaptive weights. The method contrasts with other works in that the adaptive weights are used with a distributed framework allowing for effective fair scheduling with little packet overhead. Additionally, all packets are modified downstream eliminating the need for feedback packets for implementation of the method. These points allow for energy-efficient method that simultaneously accomplishes fair scheduling and satisfies QoS metrics.

This paper primarily focuses on development of the multi-channel ADFS or MC-ADFS. This work differs from others in that the scheduling method takes into account the state of the wireless channel while implementing an adaptive fair scheduling method. Additionally, an implementation of the MC-ADFS protocol [4] is shown in hardware to benchmark performance of the scheduling algorithm. The ADFS protocol was initially developed at University of Missouri-Rolla (UMR), known currently as Missouri University of Science and Technology (MST), for ad-hoc networks [4]. An ADFS-like protocol is developed for WSN in the presence of multiple channels and its implementation on a hardware platform is presented. The development of ADFS using multiple channels is considerably different as compared to development for a single channel; therefore, novel analytical work is included in this paper for inclusion of multiple channels. For the case of WSN, unlike ad-hoc networks, scheduling is required

both at the node level within the cluster and at the cluster head (CH) level which resembles a multi hop ad-hoc network. However, for WSN, the node density is greater and results in collisions causing significant delays.

Applications of the MC-ADFS protocol, as presented here, are intended for industrial monitoring and control WSN. Networked control systems operating in real, or near-real, time regimes require a higher QoS to guarantee that delays are adequate for the application. Control applications have unique requirements on QoS, thus the proposed MC-ADFS for WSN guarantees network performance due to its adaptive methods. With MC-ADFS, applications requiring either aggregated or non-aggregated data can be serviced reliably. MC-ADFS's dynamic allocation of network resources allows sensors to forward large volumes of packets while maintaining proper delay and throughput to other classes of packet flows.

Selection of fair bandwidth allocation for the packet flows is based on user-defined QoS. Hardware testing results include a performance comparison of MC-ADFS, a simple FIFO and the DFS protocols. Additional simulations studies results are available in [4], and previous hardware work shown in [26, 27]. Experimental results using the MST/SLU G4-SSN mote were performed using multiple clusters of five nodes. Multi-cluster scenarios were used to illustrate performance gains of the ADFS protocol over other schemes in terms of throughput, delay, and fairness index (FI).

In this work we provide contribution through:

- Multi-channel mechanism for ADFS
- Hardware implementation of the MC-ADFS protocol
- Dission of challenges for implementation of MC-ADFS including:

- Memory limitations
- Low processing power, and
- Selection of priority sensor flows.

II. BACKGROUND

In WSN, fair scheduling methods often are implemented through distributed algorithms to achieve performance; affecting the global fairness of the flows on the network. Consider a WSN node that maintains several input queues, belonging to several packet flows, to be transmitted onto an output link where a fair queuing algorithm is used to determine which flow to serve next. Moreover, consider several such WSN nodes in a given region both at the cluster member and CH levels accessing a shared wireless channel in the CSMA/CA paradigm. In this paradigm, the back-off interval is used to determine which WSN node should be granted access to the shared channel. Therefore, a fair scheduling protocol for WSN must be able to implement the fair queuing algorithm along with a fair back-off interval scheme to satisfy the required fairness criterion and achieve global fairness.

The notion of fairness must be guaranteed among contending flows even in a multi-channel environment. Multi-channel methods should be computationally distributed in nature. Thus, solutions for distributed fair scheduling for WSN must coordinate local interactions to achieve global performance, and must also achieve these goals within the constraints imposed by the hardware platform. Therefore, any fair scheduling algorithm proposed for a multi-hop WSN must consider the following design criteria: (adding bullets to this list may be more effective)

Centralized vs. Distributed Approaches: Distributed fair scheduling algorithm for WSN is preferred over a centralized scheme.

Fairness Metric: Selection of an appropriate fairness metric is important from a design perspective. It should address the fair allocation of service proportional to weights selected by user-defined QoS metrics.

Scalability: The scheduling scheme should deploy well in WSN with dynamic topology and link failures.

Efficiency of the Protocol*:* Since a trade-off exists between throughput and fairness, fair scheduling should render reasonable throughput to all flows.

Persistency of Quality-of-Service: Fair scheduling should meet QoS of all flows during topology changes and dynamic channel states.

Intuitively, allocation of the output link bandwidth is fair to all the flows if equal bandwidth on each active channel is allocated over every time interval. This concept generalizes to weighted fairness where the bandwidth is allocated in proportion to the weights associated with each of the flows. Formally, if ϕ_f is the weight of flow f on a channel c and $W_f(t_1, t_2)$ is the aggregate service (in bits) received by it in the interval $[t_1, t_2]$, then an allocation is fair when both flows f and m are backlogged and satisfies (1).

$$\frac{W_f(t_1, t_2)}{\phi_f} - \frac{W_m(t_1, t_2)}{\phi_m} = 0 . \tag{1}$$

Clearly, this is an idealized definition of fairness, as it assumes that flows can be served in infinitesimally divisible units of time. The objective of fair packet scheduling algorithms is to ensure that

$$\left|\frac{W_f(t_1,t_2)}{\phi_f}-\frac{W_m(t_1,t_2)}{\phi_m}\right|\le\varepsilon \tag{2}$$

where ε is a small positive number and |.| is the absolute value. Alternatively, for fairness, it has been shown in [5] for distributed fair scheduling of computer networks that if a packet scheduling algorithm guarantees

$$\left|\frac{W_f(t_1,t_2)}{\phi_f}-\frac{W_m(t_1,t_2)}{\phi_m}\right|\le H(f,m) \tag{3}$$

for all intervals $[t_1,t_2]$ where

$$H(f,m)\ge\frac{1}{2}\left(\frac{l_f^{\max}}{\phi_f}+\frac{l_m^{\max}}{\phi_m}\right)$$

and $H(f,m)$ is a function of the properties of flows f and m, while $l_f^{\max}$ and $l_m^{\max}$ denote the maximum packet lengths of flows f and m, respectively. Several fair, yet centralized, scheduling algorithms that achieve a value of $H(f,m)$ close to the lower bound have been proposed in the literature [5, 8, 11]. In this work the definition of fairness for a wireless network is modified to include multiple channels.

The ADFS protocol [4] for multiple channels is developed here and is referred to as MC-ADFS. To accommodate inclusion of multiple channels the aggregate service must be calculated over all channels and for a given flow, or set of flows, passing through an ADFS node. In (4) a new method of calculation of the aggregate service using multiple channels is given as

$$W_f=\sum_{c=0}^{C_{\max}}w_{f,c} \tag{4}$$

where f is a flow-id, c is the RF channel. Through summation of the aggregate service over all the channels, an evaluation of a flow in a per-channel basis can be performed. For this development, W_f is assumed to be an aggregate service for a flow over a multi-channel sensor network at a given ADFS node. Hence, an ADFS node with multi-channel capability is referred to as MC-ADFS node. Although MC-ADFS appears to be a simple extension of ADFS for ad-hoc networks [4], it differs significantly as inclusion of a multi-channel environment changes the analytical proofs as introduced in this paper.

Note that, service rate of flow-controlled, broadcast medium and wireless links may fluctuate over time. Two service models, fluctuation constrained (FC) and exponential bounded fluctuation (EBF) service model, which are suitable for modeling many variable rate servers, have been introduced in [28] for computer networks. Similarly, variable rate service models for WSN have to be defined to incorporate the channel and contention based protocols. An FC service model for WSN operating on a channel c in the interval $[t_1, t_2]$ has two parameters, average rate $\lambda(t_1, t_2, c)$ bps and variations parameter $\psi(\lambda, c)$ given by $\psi(\lambda, c) = \chi(\lambda, c) + \delta(\lambda, c) + \varpi(\lambda, c)$, where $\chi(\lambda, c)$ is the reduction in wireless channel capacity due to channel-uncertainties, $\varpi(\lambda, c)$ is the variation due to backoff interval and $\delta(\lambda, c)$ is the burstiness in bits on the channel c.

A. FC service model

A WSN node is said to follow the Fluctuation Constrained (FC) service model [5] on a channel c, if for all intervals $[t_1, t_2]$ in a busy period of the node, the work done on channel c, denoted by $W(t_1, t_2)$, satisfies

$$W(t_1, t_2) \geq \lambda(t_1, t_2, c)(t_2 - t_1) - \psi(\lambda, c)$$

where channel c has parameters $(\lambda(t_1, t_2, c), \psi(\lambda, c))$.

The EBF service model is a stochastic relaxation of FC service model. Intuitively, the probability of work done by a wireless sensor node following EBF service model deviating from the average rate by more than γ, decreases exponentially with γ.

B. EBF service model

A WSN node follows the Exponentially Bounded Fluctuation (EBF) service model [5] on channel c if for all intervals $[t_1, t_2]$ in a busy period of the node, the work done, denoted by $W(t_1, t_2)$, satisfies

$$P(W(t_1, t_2) < \lambda(t_1, t_2, c)(t_2 - t_1) - \psi(\lambda, c) - \gamma) \leq Be^{-\omega\gamma}$$

where channel c has parameters $(\lambda(t_1, t_2, c), B, \omega, \psi(\lambda, c))$,

Our proposed scheme works well even with WSN nodes having variable rates. For this development, a variable rate WSN node is defined as an ADFS WSN node if the transmission occurs in a single channel and as a MC-ADFS node if the transmission occurs using multiple channels. The proposed MC-ADFS scheme satisfies the fairness criterion given in (3). The description of the proposed scheme and the MAC protocol is given next.

III. Multi-channel Adaptive and Distributed Fair Scheduling (MC-ADFS) Protocol

The main goal of the proposed MC-ADFS protocol is to achieve fairness in WSN over multiple channels, The protocol must accommodate the presence of dynamic channel states that affects available bandwidth. Channel dynamics include channel uncertainties, for example shadowing and multi-path fading, and weight adaptation is

used to compensate for these changing channel states. ADFS employs an adaptive scheduling algorithm to provide fairness among local queues and a MAC protocol to provide fair channel access via dynamic selection of back-off interval. MC-ADFS performance was evaluated in the NS-2 simulator [4].

A. Protocol Implementation

To achieve fairness in the presence of multiple channels at the scheduling level, the proposed MC-ADFS protocol still implements the start-time fair queuing (SFQ) [4, 5] scheme, defined as follows:

i) On arrival, the j^{th} packet of flow f, defined as p_f^j and having length l_{fj} and weight ϕ_{fj}, is stamped with start tag $S(p_f^j)$, defined as

$$S(p_f^j) = \max\left\{ v(A(p_f^j)), F(p_f^{j-1}) \right\} \qquad j \geq 1 \tag{5}$$

where $F(p_f^j)$, the finish tag of packet p_f^j, is defined as

$$F(p_f^j) = S(p_f^j) + \frac{l_{fj}}{\phi_{fj}} \qquad j \geq 1 \tag{6}$$

where $F(p_f^0) = 0$.

ii) Initially, the virtual time, $v(\circ)$, at a given wireless sensor node is set to zero. During transmission, the WSN node's virtual time at time *t*, *v(t)*, is set equal to the start tag of the packet being transmitted at time *t*. At the end of a transmission, *v(t)* is set to the maximum of finish tag assigned to any packets that have been transmitted by time *t*.

iii) Packets are transmitted in the increasing order of the start tags.

1) Dynamic Weight Adaptation

To account for dynamics in traffic demands and channel state for a given channel selection that affects fairness and E2E delay, the packet weights for the flows are updated dynamically. The actual weight for the i^{th} flow, j^{th} packet denoted by $\hat{\phi}_{ij}$, is updated as

$$\hat{\phi}_{ij}(k+1) = \alpha \hat{\phi}_{ij}(k) + \beta E_{ij} \tag{7}$$

where $\hat{\phi}_{ij}(k)$ is the previous weight of the packet, α and β are design constants, where $\{\alpha\ ,\ \beta\} \in [-1,1]$, and the network state, E_{ij}, is defined as:

$$E_{ij} = e_{ij,queue} + \frac{1}{e_{ij,delay}} \tag{8}$$

where $e_{ij,queue}$ is the error between the expected length of the queue and the actual size of the queue and $e_{ij,delay}$ is the error between the expected delay and the delay experienced by the packet at the current time. According to (7) and (8), as queues grow, the packet weights will be increased to clear the backlog. Moreover, the weights of the packets with large E2E delays (delays closer to the expected delay) will be increased (due to smaller values of $e_{ij,delay}$) in order to speed up their transmission. However, packets experiencing delays greater than the expected delay will be dropped. Note that the term, E_{ij}, is a bounded value since the queue length and delay are finite values at each node. Equation (7) takes into account the channel state and buffer availability at each node via weight adaptation.

To calculate the back-off interval and to implement the scheduling scheme, the updated weights at each WSN node have to be transmitted in the data frame of the MAC

protocol. Changes are made to the data packet header to accommodate the current weight of the packet. Whenever a packet is received, the current weight is used to update the weights dynamically using (7) and the weight field in the packet header is replaced with the updated weight.

2) MAC Protocol - Dynamic Back-off Intervals

The proposed MC-ADFS protocol uses the CSMA/CA scheme similar to the IEEE 802.11 protocol. When multiple nodes of a WSN both at node and CHlevel compete to access the shared channel the selection of the back-off interval plays a critical role in deciding which node is granted access to the channel. In order to achieve global fairness, the nodes must access the channel in a fair manner.

The proposed MC-ADFS scheme is implemented at the MAC level to provide access of the shared channel by adjustment of the dynamic back-off intervals. MC-ADFS calculates the back-off interval relative to the weight of the packet. Since the weights are updated using (7), the back-off interval is also updated at each WSN node. Back-off interval, BI_{ij}, for i^{th} flow j^{th} packet with packet length l_{ij} and weight ϕ_{ij} is defined as

$$BI_{ij} = \left\lfloor \rho * SF * \frac{l_{ij}}{\phi_{ij}} \right\rfloor$$

where SF is a scaling factor and ρ is a random variable with mean 1. A collision handling mechanism is incorporated similar to [9], resulting in fair allocation of the bandwidth. Next the fairness guarantee is presented with relevant theorems.

B. Fairness Guarantee

To prove that MC-ADFS is fair, the bound on $\left|\frac{W_f(t_1,t_2)}{\phi_f}-\frac{W_m(t_1,t_2)}{\phi_m}\right|$ must be obtained for a sufficiently long interval $[t_1,t_2]$ over which both flows, f and m, are backlogged. In contrast to standard ADFS [4], the MC-ADFS scheme has to consider allocation of the channel and bandwidth. MC-ADFS assumes receiver-based allocation of channels to the nodes where nodes receive packets on an assigned channel. The receiver accepts incoming packets on the assigned channel and queues the packet for scheduling. Next, for the duration of the packet transmission to the receiver the transmitter channel is switched to that of the receiver for the next hop. The receiver-based allocation scheme assures a balanced allocation of channels to the receivers in communication range.

The allocation phase is assumed to precede the transmission of packet flow. Additionally, a multi-path routing scheme is employed, where the destination can be reached through several alternative paths. The next hop for each path can have different assigned channels allowing parallel transmission by the relay nodes along the alternative routes. Finally, an assumption is made that the base–station (BS) has multiple interfaces for concurrent reception on different channels allowing for higher network throughput.

The main challenge in developing MC-ADFS scheme is in balancing the load between alternative channels/paths to provide higher performance and fair allocation of the resources proportional to flow weights to meet the desirable QoS.

Periodically, an MC-ADFS node will communicate its service load to neighbors in order to provide feedback to the load balancing mechanism. When a packet is scheduled a node then selects the next hop among the available alternate paths based on the current

feedback information. The load balancing feature of MC-ADFS uses the current feedback information to select the next hop node with the lowest service load ensuring that no individual MC-ADFS node is overloaded; this ensures that the QoS of the network traffic is maximized.

Assumptions:

i) Assignment of nodes to orthogonal channels is performed during route discovery, before the scheduling of packets begins. The assignment of channels is receiver-based, in other words, based on the nodes.

ii) Each node on a route is capable of transmitting on multiple channels. The routing protocol generates multiple routes from a source to the destination.

For the scheduling of packets in the presence of multiple channels ADFS will dynamically decide which relay node and channel, will be used to transmit the packet. A simple example scenario with multi-channel ADFS is shown in Fig. 1. The alternative paths are shown with percentage of packets passed through each relay node. Node 1 generates packets and can route those packets on one of two channels. When the link to node 2 has too much bandwidth allocated to the flows passing through it, then node 1 can transmit traffic through node 3 on the second channel. Node 4 receives on the first channel since channels are assigned using receiver based allocation.

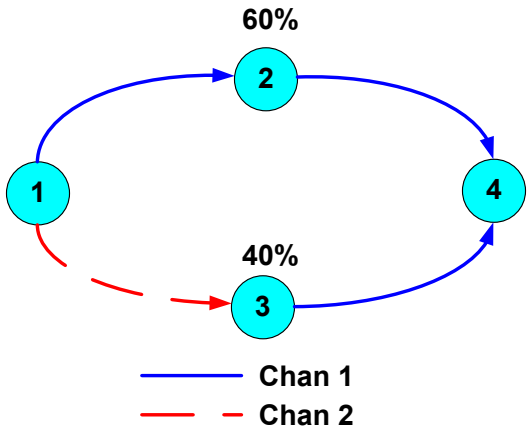

Fig 1 Example of multi-channel ADFS

The MC-ADFS schedules channel resources on a packet-by-packet basis using the alternative routes available from the proactive routing information. As flows are added to the network the MC-ADFS will balance the load based on available channel and relay-node capacity. When the packets are sent through channels, the relay nodes will evaluate the sum of the weights of transmitted packets for each channel. Next, the feedback will be sent to the transmitting nodes, then, the transmitting nodes allocate the new packets to the least utilized channel/relay-node provided the channel state is conducive for transmission. Note that when a different channel is utilized, E2E delays can increase or decrease. The proposed MC-ADFS scheme will incorporate these as the weights are tuned using delay.

Lemma 1*:* The MC-ADFS node will fairly service all flows for a channel provided

$$\sum_{n \in Q} \phi_{n,l} \leq 1$$

To provide a complete model of the wireless channel as related to the upper bound of the service the channel variations parameter from the FC and EBF service models. This additional term is then shown to be included in the upper bound of the service in any interval $[t_1, t_2]$. The proof is similar to the one shown above using Lemmas 1,2,3,4 and

Theorem 1. In Lemma 4 we begin with the addition of the maximum variation term given as $\psi_f^{\max}(\lambda, c)$.

In order to proceed, the following assumption is needed.

Assumption: *In order to arrive at a fair scheduling scheme, we assume that there exists a weight vector* ϕ_{ij} *for a* i^{th} *flow,* j^{th} *packet, at each WSN node* l *as*

$$\phi_{ij} = \begin{bmatrix} \phi_{ijl} \\ . \\ . \\ . \\ \phi_{ijm} \end{bmatrix}$$

Remark 1: *In fact, the weight update (7) ensures that the actual weight for the packet at each WSN node whether it is at the cluster or CH level converges close to its target value.*

Remark 2: ϕ_{ij} *is finite for each flow at each WSN node.*

Let $\tilde{\phi}_{ij}$ be the weight error defined as $\tilde{\phi}_{ij}$, which is given by

$$\tilde{\phi}_{ij} = \phi_{ij} - \hat{\phi}_{ij} \tag{9}$$

Lemma 2: *If the weights are updated using (7) for a sufficiently long interval* $[t_1, t_2]$, *then the weight error* $\tilde{\phi}_{ij}(k+1)$ *is bounded, provided* $|\alpha| < 1$.

Proof: Using (7) and (9) the weight estimation error is expressed as

$$\tilde{\phi}_{ij}(k+1) = \alpha \tilde{\phi}_{ij}(k) + (1-\alpha)\phi_{ij} - \beta E_{ij}. \tag{10}$$

Choose a Lyapunov function

$V = \tilde{\phi}_{ij}^2(k)$.

The first difference of the Lyapunov equation is obtained as

$\Delta V = V(k+1) - V(k)$,or

$$\Delta V = \tilde{\phi}_{ij}^2(k+1) - \tilde{\phi}_{ij}^2(k) \tag{11}$$

Substituting (10) in (11) to get

$$\Delta V = [\alpha \tilde{\phi}_{ij}(k) + (1-\alpha)\phi_{ij} - \beta E_{ij}]^2 - \tilde{\phi}_{ij}^2(k). \tag{12}$$

Equation (12) can be rewritten as

$$\Delta V = \alpha^2 \tilde{\phi}_{ij}^2(k) + (1-\alpha)^2 \phi_{ij}^2 + \beta^2 E_{ij}^2 + 2\alpha \tilde{\phi}_{ij}(1-\alpha)\phi_{ij} - 2(1-\alpha)\phi_{ij}\beta E_{ij}$$

$$-2\alpha \tilde{\phi}_{ij}\beta E_{ij} - \tilde{\phi}_{ij}^2(k). \tag{13}$$

Equation (13) can be simplified as

$$\Delta V = -(1-\alpha^2)\tilde{\phi}_{ij}^2 + (1-\alpha)^2 \phi_{ij}^2 + \beta^2 E_{ij}^2 + 2\alpha \tilde{\phi}_{ij}(1-\alpha)\phi_{ij} - 2(1-\alpha)\phi_{ij}\beta E_{ij} - 2\alpha \tilde{\phi}_{ij}\beta E_{ij},$$

This further implies that

$$|\Delta V| \le -(1-\alpha^2)|\tilde{\phi}_{ij}|^2 + 2\alpha|\tilde{\phi}_{ij}|\, a + b,$$

where

$$a = |[(1-\alpha)\phi_{ij} - \beta E_{ij}]|$$

and

$$b = |(1-\alpha)^2 \phi_{ij}^2 + \beta^2 E_{ij}^2 - 2(1-\alpha)\phi_{ij}\beta E_{ij}|.$$

$$|\Delta V| \le -(1-\alpha^2)\left[|\tilde{\phi}_{ij}|^2 - \frac{2\alpha}{(1-\alpha^2)}|\tilde{\phi}_{ij}|\, a - \frac{b}{(1-\alpha^2)}\right].$$

$|\Delta V| \leq 0$ implies that

$|\tilde{\phi}_{ij}| \geq B_{ij,\phi}$.

where $B_{ij,\phi}$ is defined as the upper bound on the weight error and it is given by

$$B_{ij,\phi} = \frac{\alpha a + \sqrt{\alpha^2 a^2 + b(1-\alpha^2)}}{(1-\alpha^2)}. \quad (14)$$

For $|\tilde{\phi}_{ij}| \geq B_{ij,\phi}$ $\Delta V < 0$. Therefore (14) can be treated as the upper bound on the weight error. Since $|\tilde{\phi}_{ij}| \geq B_{ij,\phi}$, from equation (9) we get

$$\hat{\phi}_{ij} \leq \sigma \phi_{ij}$$

for some σ.

Note: *From now the weight of a packet of flow f* at node *l is denoted as* $\phi_{f,l}$, *and it is given by* $\phi_{f,l} = \sigma_f \phi_f$

Lemma 3: *The actual weights* $\hat{\phi}_{ij}$ *at each node using (7) converge close to their target values in a finite time.*

Proof: Since $|\alpha| < 1$, define $\tilde{\phi}_{ij}(k) = x(k)$, then (10) can be expressed as

$$x(k+1) = c\, x(k) + d\, u(k) \quad (15)$$

where $c = \alpha$, $d = [(1-\alpha) \quad -\beta]$, and $u(k) = \begin{bmatrix} \phi_{ij} \\ E_{ij} \end{bmatrix}$

Equation (15) is a stable linear system [29], driven by a bounded input $u(k)$ (see Remark 2). According to linear system theory [29], $x(k)$converges close to its target value in a finite time.

Lemma 4: *If flow, f, is backlogged throughout the interval $[t_1, t_2]$, then in a MC-ADFS* wireless sensor node

$$\phi_{f,l}.(v_2 - v_1) - l_f^{\max} - \psi_f^{\max}(\lambda, c) \le W_f(t_1, t_2), \tag{16}$$

where $v_1 = v(t_1)$ *and* $v_2 = v(t_2)$. *The additional term* $\psi_f^{\max}(\lambda, c)$ is included to account for the channel number and channel state in a multi-channel network.

Proof: The proof follows on similar lines to ad-hoc networks [4] which in turn are based on computer networks [4].

Since $W_f(t_1, t_2) \ge 0$, if $\phi_{f,l}(v_2 - v_1) - l_f^{\max} - \psi_f^{\max}(\lambda, c) \le 0$, (16) holds trivially. Therefore, consider the case where $\phi_{f,l}(v_2 - v_1) - l_f^{\max} - \psi_f^{\max}(\lambda, c) > 0$, i.e., $v_2 > v_1 + \dfrac{l_f^{\max} + \psi_f^{\max}(\lambda, c)}{\phi_{f,l}}$. Let packet p_f^k be the first packet of flow f that receives service in the open interval (v_1, v_2). To observe that such a packet exists, consider the following two cases:

Case 1: Packet p_f^n such that $S(p_f^n) < v_1$ and $F(p_f^n) > v_1$ exists: Since flow f is backlogged in $[t_1, t_2]$, we conclude $v(A(p_f^{n+1})) \le v_1$. From (5) and (6), we get:

$$S(p_f^{n+1}) = F(p_f^n)$$

Since $F(p_f^n) \le S(p_f^n) + \dfrac{l_f^{\max} + \psi_f^{\max}(\lambda, c)}{\phi_{f,l}}$ and $S(p_f^n) < v_1$, we get:

$$S(p_f^{n+1}) < v_1 + \frac{l_f^{\max} + \psi_f^{\max}(\lambda, c)}{\phi_{f,l}} \tag{17}$$

$$< v_2 \tag{18}$$

Since $S(p_f^{n+1}) = F(p_f^n) > v_1$, using (18), we conclude $S(p_f^{n+1}) \in (v_1, v_2)$.

Case 2: Packet p_f^n such that $S(p_f^n) = v_1$ exists: p_f^n may finish service at time $t < t_1$ or $t \geq t_1$. In either case, since the flow f is backlogged in $[t_1, t_2]$, $v(A(p_f^{n+1})) \leq v_1$.

Hence, $S(p_f^{n+1}) = F(p_f^n)$.

Since $F(p_f^n) \leq S(p_f^n) + \frac{l_f^{\max} + \psi_f^{\max}(\lambda, c)}{\phi_{f,l}}$ and $S(p_f^n) < v_1$, we get:

$$S(p_f^{n+1}) \leq v_1 + \frac{l_f^{\max} + \psi_f^{\max}(\lambda, c)}{\phi_{f,l}} \tag{19}$$

$$< v_2 \tag{20}$$

Since $S(p_f^{n+1}) = F(p_f^n) > v_1$, using (20), we conclude $S(p_f^{n+1}) \in (v_1, v_2)$.

Since either of the two cases always holds, we conclude that packet p_f^k such that $S(p_f^k) \in (v_1, v_2)$ exists. Furthermore from (17) and (19), we get:

$$S(p_f^k) \leq v_1 + \frac{l_f^{\max} + \psi_f^{\max}(\lambda, c)}{\phi_{f,l}} \tag{21}$$

Let p_f^{k+m} be the last packet to receive service in the virtual time interval (v_1, v_2). Hence,

$$F(p_f^{k+m}) \geq v_2 \tag{22}$$

From (21) and (22), we conclude:

$$F(p_f^{k+m}) - S(p_f^k) \geq (v_2 - v_1) - \frac{l_f^{\max} + \psi_f^{\max}(\lambda, c)}{\phi_{f,l}} \tag{23}$$

But since flow f is backlogged in the interval (v_1, v_2), from (5) and (6) we know:

$$F\left(p_f^{k+m}\right) = S\left(p_f^k\right) + \sum_{n=0}^{n=m} \frac{l_f^{k+n} + \psi_f^{k+n}(\lambda, c)}{\phi_{f,l}} F\left(p_f^{k+m}\right) -$$
$$S\left(p_f^k\right) = \sum_{n=0}^{n=m} \frac{l_f^{k+n} + \psi_f^{k+n}(\lambda, c)}{\phi_{f,l}} \quad (24)$$

Hence, from (23) and (24), we get:

$$\sum_{n=0}^{n=m} \frac{l_f^{k+n} + \psi_f^{k+n}(\lambda, c)}{\phi_{f,l}} \geq \left(v_2 - v_1\right) - \frac{l_f^{\max} + \psi_f^{\max}(\lambda, c)}{\phi_{f,l}}$$

$$\sum_{n=0}^{n=m} l_f^{k+n} + \psi_f^{k+n}(\lambda, c) \geq \phi_{f,l}\left(v_2 - v_1\right) - l_f^{\max} - \psi_f^{\max}(\lambda, c)$$

Since $S\left(p_f^{n+1}\right) < v_2$, packet p_f^{k+m} is guaranteed to have been transmitted by t_2.

Hence, $W_f(t_1, t_2) \geq \sum_{n=0}^{n=m} l_f^{k+n} + \psi_f^{k+n}(\lambda, c)$ and the lemma follows.

Lemma 5: *In a MC-ADFS-based WSN node, during any interval* $[t_1, t_2]$

$$W_f(t_1, t_2) \leq \phi_{f,l}(v_2 - v_1) + l_f^{\max} + \psi_f^{\max}(\lambda, c), \quad (25)$$

where $v_1 = v(t_1)$ *and* $v_2 = v(t_2)$.

Proof: The proof follows on similar lines as that of wireless ad-hoc network [4] which is based on computer networks [5].

From the definition of MC-ADFS, the set of flow f packets served in the interval $[v_1, v_2]$ have service tag at least v_1 and at most v_2. Hence, the set can be partitioned into two sets

- Set D consisting of packets that have service tag at least v_1 and finish time at most v_2. Formally,

$$D = \left\{k \middle| v_1 \leq S\left(p_f^k\right) \leq v_2 \wedge F\left(p_f^k\right) \leq v_2\right\}$$

From (5) and (6), we conclude

$$\sum_{k \in d} l_f^k \leq \phi_{f,l}(v_2 - v_1) \tag{26}$$

- Set E consisting of packets that have service tag at most v_2 and finish time greater than v_2. Formally,

$$E = \left\{k \middle| v_1 \leq S\left(p_f^k\right) \leq v_2 \wedge F\left(p_f^k\right) > v_2\right\}$$

clearly, at most one packet can belong to this set. Hence,

$$\sum_{k \in E} l_f^k \leq l_f^{\max} \tag{27}$$

From (26) and (27) we conclude that (25) holds. Since unfairness between two flows in any interval is maximum when one flow receives maximum possible service and the other the minimum service. Theorem 1 follows directly from Lemma 1, 2, 3 and 4.

Theorem 1: *For any interval* $[t_1, t_2]$ *in which flows* f *and* m *are backlogged during the entire interval, the difference in the service received by two flows at a MC-ADFS WSN node is given as*

$$\left|\frac{W_f(t_1,t_2)}{\phi_{f,l}} - \frac{W_m(t_1,t_2)}{\phi_{m,l}}\right| \leq \frac{l_f^{\max} + \psi_f^{\max}(\lambda,c)}{\phi_{f,l}} + \frac{l_m^{\max} + \psi_m^{\max}(\lambda,c)}{\phi_{m,l}}$$

Remark 3: *If* $E_{ij} = 0$ *at each node and no channel variations, then the proposed MC-ADFS will become a DFS scheme [9]. Here a single channel is utilized at the ADFS node.*

Remark 4: *No assumption on the service rate of the wireless node was made to establish Theorem 1. Hence, Theorem 1 holds regardless of the service rate of the WSN node. This demonstrates that MC-ADFS achieves fair allocation of bandwidth and thus meets a fundamental requirement of fair scheduling algorithms for integrated services networks.*

Remark 5: *The addition of the wireless variations parameters* $\psi_f^{\max}(\lambda, c)$ *and* $\psi_m^{\max}(\lambda, c)$ *for flows f and m respectively take into account the state of the wireless channel over each time interval. By taking the maximum value of the channel variation the maximum aggregate service delay is included in the bound of the service error. The new bound accommodates the variation in a wireless channel and provides innovation in that the ADFS method is extended to accommodate channel conditions.*

C. Throughput Guarantee

Theorems 2 and 3 establish the throughput guaranteed to a flow by a MC-ADFS FC and EBF service model, respectively. These theorems are extended versions of throughput and delay guarantees of computer networks except as applied to WSN. Consequently, the throughput and end-to-end delay bounds will be a function of the service rate of the node, channel state and backoff interval in contrast to the case of computer networks where channel state and backoff interval do not exists.

Theorem 2: *If Q is the set of flows served by a MC-ADFS node following FC service model with parameters* $(\lambda(t_1, t_2), \psi(\lambda), c)$, *and* $\sum_{n \in Q} \phi_{n,l} \le \lambda(t_1, t_2, c)$, *then for all intervals* $[t_1, t_2]$ *in which flow f is backlogged throughout the interval,* $W_f(t_1, t_2)$ *is given as*

$$W_f(t_1, t_2) \ge \phi_{f,l}(t_2 - t_1) - \phi_{f,l} \frac{\sum_{n \in Q} l_n^{\max}}{\lambda(t_1, t_2, c)} - \phi_{f,l} \frac{\psi(\lambda, c)}{\lambda(t_1, t_2, c)} - l_f^{\max} - \psi_f^{\max}(\lambda, c)$$

Proof: The proof follows in similar lines to ad-hoc networks [4] which are again based on computer networks [5].

Let $v_1 = v(t_1)$ and let $\hat{L}(v_1, v_2)$ denote the aggregate length of packets served by the wireless node in the virtual time interval $[v_1, v_2]$. Then, from Lemma 4 we conclude

$$\hat{L}(v_1, v_2) \leq \sum_{n \in Q} \phi_{n,l}(v_2 - v_1) + \sum_{n \in Q} l_n^{\max}$$

Since $\sum_{n \in Q} \phi_{n,l} \leq \lambda(t_1, t_2, c)$,

$$\hat{L}(v_1, v_2) \leq \lambda(t_1, t_2, c)(v_2 - v_1) + \sum_{n \in Q} l_n^{\max} \tag{28}$$

Define v_2 as:

$$v_2 = v_1 + t_2 - t_1 - \frac{\sum_{n \in Q} l_n^{\max}}{\lambda(t_1, t_2, c)} - \frac{\psi(\lambda, c)}{\lambda(t_1, t_2, c)} \tag{29}$$

Then from (28) we conclude:

$$\hat{L}(v_1, v_2) \leq \lambda(t_1, t_2, c)\left(v_1 + t_2 - t_1 - \frac{\sum_{n \in Q} l_n^{\max}}{\lambda(t_1, t_2, c)} - \frac{\psi(\lambda, c)}{\lambda(t_1, t_2, c)} - v_1 \right) + \sum_{n \in Q} l_n^{\max}$$

$$\leq \lambda(t_1, t_2, c)(t_2 - t_1) - \psi(\lambda, c)$$

Let $\hat{t}_2$ be such that $v(\hat{t}_2) = v_2$. Also let $\mathrm{T}(w)$ be the time taken by the WSN node to serve packets with aggregate length w in its busy period.

Then,

$$\hat{t}_2 \leq t_1 + \mathrm{T}(\hat{L}(v_1, v_2))$$

$$\leq t_1 + \mathrm{T}(\lambda(t_1, t_2, c)(t_2 - t_1) - \psi(\lambda, c)) \tag{30}$$

from the definition of FC service model, we get:

$$\mathrm{T}(w) \le \frac{w}{\lambda(t_1, t_2, c)} + \frac{\psi(\lambda, c)}{\lambda(t_1, t_2, c)} \tag{31}$$

From (30) and (31) we get:

$$\hat{t}_2 \le t_1 + \frac{\lambda(t_1, t_2, c)(t_2 - t_1) - \psi(\lambda, c)}{\lambda(t_1, t_2, c)} + \frac{\psi(\lambda, c)}{\lambda(t_1, t_2, c)}$$

$$\le t_2$$

From Lemma 4 we know that:

$$W_f(t_1, \hat{t}_2) \ge \phi_{f,l}(v_2 - v_1) - l_f^{\max} - \psi_f^{\max}(\lambda, c)$$

Since $\hat{t}_2 \le t_2$, using (29) we get:

$$W_f(t_1, t_2) \ge \phi_{f,l}(t_2 - t_1) - \phi_{f,l} \frac{\sum_{n \in Q} l_n^{\max}}{\lambda(t_1, t_2, c)} - \phi_{f,l} \frac{\psi(\lambda, c)}{\lambda(t_1, t_2, c)} - l_f^{\max} - \psi_f^{\max}(\lambda, c) \tag{32}$$

Remark 6: *Since $\psi(\lambda, c)$ is a function of changes in bandwidth resulting from, channel number and its state, and backoff interval during contention, (32) clearly indicates that the proposed MC-ADFS protocol throughput depends upon channel number, radio channel state and backoff intervals in contrast to [1].*

Theorem 3: *If Q is the set of flows served by a MC-ADFS node following EBF service model with parameters $(\lambda(t_1, t_2, c), B, \omega, \psi(\lambda, c))$, $\gamma \ge 0$, and $\sum_{n \in Q} \phi_{n,l} \le \lambda(t_1, t_2, c)$, then for all intervals $[t_1, t_2]$ in which flow f is backlogged throughout the interval, $W_f(t_1, t_2)$ is given as*

$$P\left(W_f(t_1,t_2) < \phi_{f,l}(t_2-t_1) - \phi_{f,l}\frac{\sum_{n\in Q} l_n^{\max}}{\lambda(t_1,t_2,c)} - \phi_{f,l}\frac{\psi(\lambda,c)}{\lambda(t_1,t_2,c)} - \phi_{f,l}\frac{\gamma}{\lambda(t_1,t_2,c)} - l_f^{\max} - \psi_f^{\max}(\lambda,c) \right) \le Be^{-\omega\gamma}$$

Remark 7: *Since the term, $\psi(\lambda,c)$ depends upon wireless channel and its state, and backoff intervals due to contention, Theorems 2 and 3 clearly demonstrates that the MC-ADFS protocol throughput in an WSN is affected by the radio channel state variations, channel number and backoff intervals, unlike in [5].*

D. Delay Guarantee

Generally, a network can provide a bound on delay only if its capacity is not exceeded. The weight $\phi_{f,l}$ can also mean the rate assigned to a packet of flow f at node l. Let rate function for flow f at virtual time v, denoted by $R_f(v)$, be defined as the rate assigned to the packet that has start tag less than v and finish tag greater than v. Formally,

$$R_f(v) = \begin{cases} \phi_{f,l} & \text{if } \exists j \ni \left(S(p_f^j) \le v < F(p_f^j)\right) \\ 0 & \text{otherwise} \end{cases}$$

Let Q be the set of flows served by the node. Then a FC or EBF node with average rate $\lambda(t_1,t_2,c)$, is defined to have exceeded its capacity at virtual time v if $\sum_{n\in Q} R_n(v) > \lambda(t_1,t_2,c)$. If the capacity of a SFQ–based node is not exceeded, then it guarantees a deadline to a packet based on its *expected arrival time*. Expected arrival time of packet P_f^j, denoted by $T_a\left(P_f^j, \phi_{f,j}\right)$, is defined as

$$\max\left\{ A\left(P_f^j\right), T_a\left(P_f^{j-1}, \phi_{f,j-1}\right) + \frac{l_f^{j-1} + \psi_f^{j-1}(\lambda,c)}{\phi_{f,j-1}} \right\} \quad j \ge 1 \qquad (33)$$

where $T_a\left(P_f^0, \phi_{f,0}\right) = -\infty$. A deadline guarantee based on expected arrival time has been referred to as *delay guarantee.* Theorem 3 and 4 establish the delay guarantee for FC and EBF service models, respectively. The proofs follow similar steps as that of computer networks [5].

Theorem 4: *If Q is the set of flows served by a MC-ADFS node following FC service model with parameters* $(\lambda(t_1,t_2,c),\psi(\lambda),c)$, *and* $\sum_{n\in Q} R_n(v) \le \lambda(t_1,t_2,c)$ *for all* v, *then the departure time of packet* P_f^j at the node, denoted by $T_d\left(P_f^j\right)$, is given by

$$T_d\left(P_f^j\right) \le T_a\left(P_f^j, \phi_{f,j}\right) + \sum_{n\in Q\wedge n\neq f} \frac{l_n^{\max}}{\lambda(t_1,t_2,c)} + \frac{l_f^j + \psi_f^j(\lambda,c)}{\lambda(t_1,t_2,c)} + \frac{\psi(\lambda,c)}{\lambda(t_1,t_2,c)}$$

Proof: Let H be defined as follows:

$H = \left\{m \middle| m > 0 \wedge S\left(P_f^m\right) = v\left(A\left(P_f^m\right)\right)\right\}$

Let $k \le j$ be largest integer in H. Also, let $v_1 = v\left(A\left(P_f^k\right)\right)$ and $v_2 = S\left(P_f^k\right)$. Observe that as the node virtual time is set to the maximum finish tag assigned to any packet at the end of a busy period, packet P_f^k and P_f^j are served in the same busy period of a wireless node. From the definition of MC-ADFS, the set of flow f packets served in the interval $[v_1, v_2]$ have start tag at least v_1 and at most v_2. Hence, the set can be partitioned into two sets

- This set consists of packets that have start tag at least v_1 and finish tag at most v_2. Formally the set of packets of flow n, denoted by D_n, in this set is

$$D_n = \left\{m \middle| v_1 \le S\left(p_n^m\right) \le v_2 \wedge F\left(p_n^m\right) \le v_2\right\}$$

Then, from the definition of $R_n(v)$ and $F(P_n^m)$, we know that the cumulative length of such flow n packets served by wireless node in the virtual time interval $[v_1, v_2]$, denoted by $C_n(v_1, v_2)$, is given as

$$C_n(v_1, v_2) \leq \int_{v_1}^{v_2} R_n(v)dv$$

Hence, aggregate length of packets in this set, $\sum_{n \in Q} C_n(v_1, v_2)$, is given as

$$\sum_{n \in Q} C_n(v_1, v_2) \leq \sum_{n \in Q} \int_{v_1}^{v_2} R_n(v)dv$$

$$\leq \int_{v_1}^{v_2} \lambda(t_1, t_2, c)dv$$

$$\leq \lambda(t_1, t_2, c).(v_2 - v_1)$$

But since $v_2 = S(P_f^k)$, from the definition of k , $v_2 - v_1 = \sum_{n=0}^{n=j-k-1} \frac{l_f^{k+n} + \psi_f^{k+n}(\lambda, c)}{\phi_{f,k+n}}$. Hence,

$$\sum_{n \in Q} C_n(v_1, v_2) \leq \lambda(t_1, t_2, c) \sum_{n=0}^{n=j-k-1} \frac{l_f^{k+n} + \psi_f^{k+n}(\lambda, c)}{\phi_{f,k+n}}$$

- This set consists of packets that have start tag at most v_2 and finish tag greater than v_2. Formally, the set of packets of flow n, denoted by E_n, in this set is

$$E_n = \{m | v_1 \leq S(p_n^m) \leq v_2 \wedge F(p_n^m) > v_2\}$$

Clearly, at most one packet of flow n can belong to this set. Furthermore, $E_f = \{j\}$. Hence, the maximum aggregate length of packets in this set is

$$\sum_{n \in Q \wedge n \neq f} l_n^{\max} + l_f^j$$

Hence, the aggregate length of packets served by the wireless node in the interval $[v_1, v_2]$, denoted by $\hat{L}(v_1, v_2)$, is

$$\hat{L}(v_1, v_2) \le \lambda(t_1, t_2, c) \sum_{n=0}^{n=j-k-1} \frac{l_f^{k+n} + \psi_f^{k+n}(\lambda, c)}{\phi_{f,k+n}} + \sum_{n \in Q \wedge n \ne f} l_n^{\max} + l_f^j$$

Let $\mathrm{T}(w)$ be the time taken by the wireless node to serve packets with aggregate length w in its busy period. From the definition of FC service model, we get

$$\mathrm{T}(w) \le \frac{w}{\lambda(t_1, t_2, c)} + \frac{\psi(\lambda, c)}{\lambda(t_1, t_2, c)}$$

Since packet P_f^j departs at system virtual time v_2 and all the packets served in the time interval $[v_1, v_2]$ are served in the same busy period of the wireless node, we get

$$A(p_f^k) + \mathrm{T}(\hat{L}(v_1, v_2)) \ge T_d(P_f^j)$$

$$A(p_f^k) + \sum_{n=0}^{n=j-k-1} \frac{l_f^{k+n}}{\phi_{f,k+n}} + \sum_{n \in Q \wedge n \ne f} \frac{l_n^{\max} + \psi_n^{\max}(\lambda, c)}{\lambda(t_1, t_2, c)} + \frac{l_f^j + \psi_f^j(\lambda, c)}{\lambda(t_1, t_2, c)} + \frac{\psi(\lambda, c)}{\lambda(t_1, t_2, c)} \ge T_d(P_f^j)$$

From (33) we get

$$T_a(P_f^j, \phi_{f,j}) + \sum_{n \in Q \wedge n \ne f} \frac{l_n^{\max} + \psi_n^{\max}(\lambda, c)}{\lambda(t_1, t_2, c)} + \frac{l_f^j + \psi_f^j(\lambda, c)}{\lambda(t_1, t_2, c)} + \frac{\psi(\lambda, c)}{\lambda(t_1, t_2, c)} \ge T_d(P_f^j) \qquad (34)$$

Remark 8: *The departure time from (34) at an MC-ADFS WSN node depends upon the following: the first term indicates the arrival time, second term reflects the consequence of packet-by-packet rather than bit-by-bit transmission, third term is the transmission time of a packet based on its length, and the final term accounts for the channel number and its state, and back off interval.*

Theorem 5: *If Q is the set of flows served by a MC-ADFS WSN node following EBF service model with parameters $(\lambda(t_1,t_2,c), B, \omega, \psi(\lambda,c))$, $\gamma \geq 0$, and $\sum_{n\in Q} R_n(v) \leq \lambda(t_1,t_2,c)$ for all v then the departure time of packet P_f^j at the node, denoted by $T_d(P_f^j)$, is given by*

$$P\left(T_d(P_f^j) \leq T_a(P_f^j,\phi_{f,j}) + \sum_{n\in Q\wedge n\neq f} \frac{l_n^{\max} + +\psi_n^{\max}(\lambda,c)}{\lambda(t_1,t_2,c)} + \frac{l_f^j + \psi_f^j(\lambda,c)}{\lambda(t_1,t_2,c)} + \frac{\psi(\lambda,c)}{\lambda(t_1,t_2,c)} + \frac{\gamma}{\lambda(t_1,t_2,c)}\right) \geq 1 - Be^{-\omega\gamma} \quad (35)$$

Theorems 4 and 5 can be used to determine delay guarantee even when a WSN node has flows with different priorities and services them in the priority order.

Remark 9: *Similar to Remark 8, the departure time from (35) at a MC-ADFS WSN node depends upon the arrival time, consequence of packet-by-packet rather than bit-by-bit transmission, transmission time of a packet based on its length, channel state and backoff interval, and the deviation from the average rate given by γ.*

Theorem 6: *The end-to-end delay denoted by $T_{EED}(P_f^j)$, is given by*

$$T_{EED}(P_f^j) = \sum_{i=1}^{m} \left(T_{d,i}(P_f^j) - T_{a,i}(P_f^j, \phi_{f,j})\right) + T_{prop}(P_f^j)$$

where $T_{d,i}(P_f^j)$ and $T_{a,i}(P_f^j, \phi_{f,j})$ are the departure time and expected arrival time of packet P_f^j at hop, i, in the multi-hop network. T_{prop} is the total propagation delay experienced by the packet, from source to the destination.

Remark 10: *As expected, the end-to-end delay is a function of packet length, channel number and its state, and backoff interval of the CSMA/CA protocol. By contrast, in*

wired computer networks, the end-to-end delay is a function of packet length alone. Similarly, for ADFS node with a single channel, the channel number is not a variable.

E. Overhead Analysis

An analysis was also performed to estimate the packet overhead for the hardware implementation of ADFS on a single channel WSN. The typical message includes the packet data field and a timestamp for the corresponding event. The timestamp is also utilized by the ADFS scheme thus not requiring additional fields. The only added overhead in ADFS protocol is due to inclusion of the current weight value in the header of each packet. Besides the ADFS incurred overhead, the implementation adds overhead due to the 802.15.4 message format and radio module interface. In this implementation, the ADFS overhead consists of one extra byte (octet) per packet and seven more bytes (octets) to the overhead. The time-stamp adds further 4 bytes (32-bit). Channel (is a word missing – channel what) also adds some temporal overhead but (what) does not incur overhead to the storage and transmission of packets. Additionally, switching time is generally small and normally negligible. Consequently, a total bit overhead for 100 byte message is 11% and 12% for the non-ADFS and MC-ADFS schemes respectively. In conclusion, the MC-ADFS protocol bit overhead is negligible compared to the benefits of the fairness guarantee and improved QoS if considered against just DFS alone.

F. Initial Weights Selection

Initial weights are selected according to QoS classes that are user, or administrator, defined. The flows within the same class are assigned the same initial weight value. For instance, in some simulations [4] the flows are classified into 4 different classes. Alternatively, for some experimental scenarios with n flows, all the flows are assumed to

belong to the same class and therefore the initial weight assigned to each flow is *1/n* treating all flows equally. In both cases, the initial weights are selected in a manner such that overall sum of weights of all the flows is equal to one. Let flows of the ranges of 0-0.2, 0.2-0.5, and 0.5-1.0 be the low, medium, and high priority, respectively. Moreover, the proposed ADFS protocol uses dynamic weight adaptation to react to the varying channel state and network topology. The overall goal is to achieve a per flow service that is proportional to the initial weights of the flows.

G. Performance Evaluation Metric

Evaluation of the performance is carried out using the fairness index (FI), shown in (37), which illustrates the weighted fairness among the flows is maintained regardless of the network dynamics. The FI [11] of the network is calculated using (37) and is used as a metric to further evaluate the performance of the ADFS protocol.

$$FI = \left(\sum_f \frac{T_f}{\phi_f} \right)^2 \Bigg/ \eta * \sum_f \left(\frac{T_f}{\phi_f} \right)^2 \tag{37}$$

where T_f is the throughput of flow f, ϕ_f is the initial weight of flow f, and η is the number of flows.

IV. HARDWARE IMPLEMENTATION DESCRIPTION

In this section, an overview of the hardware implementation of the ADFS scheduling protocol is given. This example was chosen since MC-ADFS and ADFS differ only in the channel selection and their performance in the scheduling is similar. Use of the MST/SLU motes [26, 27, 30] for development of sensing, processing, and networking is presented. A description of capabilities, limitations, and support for networking

applications is also presented. Also, in this section an overview of the software architecture is given with respect to the ADFS protocol and its hardware requirements.

A. Hardware Description

Hardware for implementation of the ADFS was selected to be energy-efficient, performance-oriented, and of small form-factor. Use of Silicon Laboratories® 8051 variant hardware was selected for its ability to provide high-speed 8-bit processing, low-power consumption, and easy interfacing to peripherals. A Maxstream XBee™ RF module was also employed for this work. Additionally, use of external RAM (XRAM), UART interfacing, and A/D sensing allow the microcontroller to perform the tasks needed for a WSN platform. Next, a discussion of the hardware capabilities and limitations will be given.

1) Considerations and Limitations

Protocol implementation is constrained by the limitations of the selected hardware platform. The use of a specific hardware platform must be weighed against the precision, speed, and criticality of an algorithm's implementation. For this protocol, low-power consumption was given the highest priority. However, the desire for low-power constrains the processor architectures that can be deployed. The selection of the Silicon Laboratories® 8051 variants was based on these criteria. Limitations for the implementation that are incurred through the use of the 8051 variant family are a small memory space and a maximum processing speed. In the next section, a description of the specifications for the hardware platform is given.

B. Architecture of the Hardware System and Software

This section outlines the hardware and software components of the ADFS implementation and the software implementation is also discussed. Software architecture, control-flow, system architecture, and hardware implications are also shown.

Fig. 2 G4-SSN node

The Generation-4 Smart Sensor Node (G4-SSN), seen in Fig. 2, was originally developed at MST and subsequently updated at St. Louis University (SLU) [30]. The G4-SSN has various abilities for sensing and processing. The former include strain gauges, accelerometers, thermocouples, and general A/D sensing. The later include analog filtering, compact flash (CF) memory interfacing, and 8-bit data processing at a maximum of 100 MIPS.

ADFS requires that the sensor nodes are synchronized and for this reason testing was performed on the G4-SSN real-time-clock (RTC). Experimentation consisted of a statistical analysis of the RTC accuracy. Results show that extended use of the RTC without re-synchronization can result in time drift which must be quantified to determine the confidence of time measurements of the RTC. A 32.768 kHz quartz crystal is used to supply a timer on the 8051 and is used as the time-base of the RTC. The RTC was allowed to run for 10 minutes and the drift results are shown in Table I.

TABLE I G4-SSN RTC DRIFT EXPERIMENTAL RESULTS

Test Time [min]	Δt [sec]	Variance [μ-sec]	Mean [sec]	STD [sec]	Error [sec]
10	0.05	50.45	0.0504	0.007	0.515

As seen in the table the RTC has a drift error of approximately a half-second over ten minutes, this translates to 3.5 seconds per hour (6x0.5 is not 3.5 s, change one or the other). In the context of this application, a X-s drift per X second is acceptable as the RTC is synchronized with the BS every 30 seconds. Synchronization is facilitated via a BEAM packet as used in the OEDSR [27] protocol that has been modified to carry the network time to the nodes.

Hardware implementation of the protocols reveals several differences as opposed to the use of simulators such as NS2 or GloMoSim. Differences using embedded hardware are highlighted by the constraints in memory, processing power, and available energy. While simulations are useful for evaluating proposed network protocols they often lack the realism of a dynamic environment that can reveal performance limitations not considered in simulation environments. In this section a discussion of the considerations and problems encountered during hardware implementation are discussed.

Limits on memory in embedded hardware affect networking schemes in several key places. Memory limits the amount of available space for queuing of packets, space for routing tables, space for other networking applications, and space for application memory. Use of WSN platforms to process sensor data to form information flow requires memory resources normally used to either hold or route packets for transmission to the BS.

The amount of memory allocated for aspects such as queuing, routing, operation variables, and applications is dependent on the amount of available memory on the WSN platform. For the G4-SSN there is 8448 bytes of memory available. The maximum network traffic able to be queued is of interest. For this application, 5 queues of 10 packets could be held using the processors on-board RAM resulting in a usage of 5kB. Only around 3.5kB are left for applications and operational variables. Operational variables include ADC buffers, UART buffers, and other needed variables for process execution.

V. Experimental results and discussion

ADFS experiments where performed using up to four clusters of five nodes. Each cluster has a designated CH and each node generates CBR traffic with rate equal to 1000 bps. Additionally, OEDSR routing [27] is used to provide routing services from the CH's to the BS. Figure 3 illustrates the experimental topology for 4 clusters. Each cluster's members join the network in five-second intervals and initial weights were selected such that each source has an equal priority. Experiments for each network topology were repeated three times and results were averaged. Experiments for FIFO, DFS, and ADFS enabled networks were performed for comparison.

Experiments were also performed for a single cluster containing nodes with differing priorities. For two of the six nodes in the cluster the weight of the packets was doubled to facilitate priority flows.

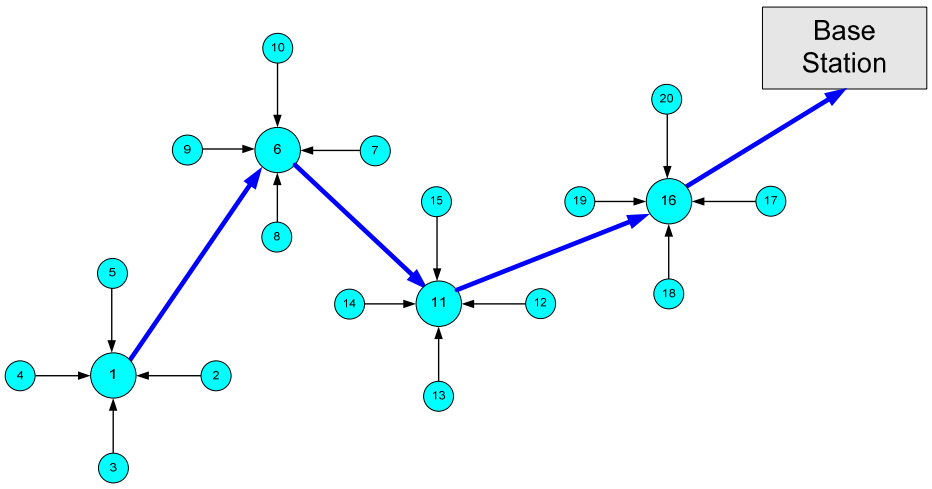

Fig. 3 Network topology for testing

These experiments were repeated for FIFO, DFS, and ADFS protocols using the multi-cluster topology for comparison and the results are shown in Table II. The DFS scheduling protocol was facilitated by disabling the adaptive weight updates in the ADFS method and using the initial packet weights as static values.

TABLE II RESULTS FOR EVENLY WEIGHTED FLOWS IN MULTIPLE CLUSTERS

	Average Throughput [bps]	Average E2E Delay [sec]	Average Dropped Packets	Average Drop Rate [packets/sec]	Average FI	Increase over FIFO	Increase over DFS
1 Cluster							
FIFO	291.440	2.212	149.330	1.383	0.978	X	X
DFS	820.350	0.734	81.083	0.751	0.994	1.6 %	X
ADFS	891.840	0.653	88.333	0.818	0.996	1.8 %	0.2 %
2 Cluster							
FIFO	134.780	13.255	116.170	1.189	0.827	X	X
DFS	416.890	2.027	100.960	1.058	0.985	19.1 %	X
ADFS	540.870	1.229	138.960	1.458	0.988	19.5 %	0.3 %
3 Cluster							
FIFO	115.680	24.403	101.070	1.072	0.577	X	X
DFS	417.820	2.487	84.364	1.040	0.746	29.3 %	X
ADFS	536.880	1.497	107.850	1.325	0.777	34.7 %	4.2 %
4 Cluster							
FIFO	121.830	17.988	90.155	0.926	0.331	X	X
DFS	270.900	11.769	71.830	0.942	0.523	58.0 %	X
ADFS	390.000	3.853	107.440	1.505	0.608	83.7 %	16.3 %

Results show that the ADFS enabled network provides improvements in the average throughput, delay, and FI among flows since it is adopts weights based on the network state and topology. The weights are increased for the flows that have to pass over more hops or through congested relay nodes; while in case of flows that are close to the BS and do not experience congestion the weights are reduced. Consequently, the ADFS compensates for the hindered flows resulting in improved overall fairness. The FIFO and DFS networks have a lower performance in average throughput and E2E delay when compared to the ADFS network. As the number of clusters was increased it was observed that the E2E delay scales more linearly than with either DFS or FIFO. The linearly increasing E2E delay seen in the ADFS method is due to a controlled queue size and E2E delay since the weight adaptation pushes older packets through the network first. The weight adaptation provides an automatic control into the system allowing the network to react to changing conditions. Additionally, the protocol scales with the size of the network allowing the introduction and removal of nodes without reducing network efficiency. In all there is as much as a 16.3 % increase in fairness in the larger network cases over the DFS using ADFS. If the results are viewed in comparison with the FIFO method a total of 83.7% improvement is seen. This clearly shows that the added overhead of 1-byte (octet) is acceptable to ensure the ability of the network to provide fair scheduling of resources to all flows.

Results show that for the larger topologies that improvements are greater. This improvement is attributed to the ability of ADFS to maintain flows that are dropped or neglected by both the DFS and FIFO methods. As the trunk of the network grows the sources farthest from the BS are neglected due to buffer overflows in the FIFO method

and the lack of weight adaptation in DFS resulting in lower service for those flows. The ADFS method provides even data rates and lower neglect of the sources that are further away from the BS. Also, the E2E delay in larger topologies shows that the FIFO method does not provide good delay characteristics. The three cluster case shows that there is a large delay while the four cluster case shows a lower average delay. This difference is due to how the performance is calculated. During the three cluster cases there are more packets delivered from the third cluster, while in the four cluster case the third and fourth clusters drop more packets that then do not enter the performance calculations. While the DFS method does provide moderate improvement in the E2E delay, it is unable to enhance the rate of delivery for the furthest nodes due to the static weights.

Figure 4 shows an example of network activity for a single cluster FIFO experiment. The plots illustrate that the bandwidth allocation is not even and variation between the flows is high. This variation in the flows is due to the first-come-first-serve nature of a FIFO implementation. The FIFO does not discriminate packets based flow-id or packet type, therefore the packets are serviced as they come in. Also, since there is no separation into queues there is no way to drop packets from flows in a controlled manner. Packets in the FIFO method are dropped by an overflow of the ring-buffer such that the oldest packet is dropped regardless of its priority, or type. For the DFS case, shown in Fig. 5, flows are more even and approach the desired rate. In comparison, activity for an ADFS trial, shown in Fig. 6, shows better throughput over the trial while still providing more even data rates. The improvements of DFS to the FIFO method are due to the ability of the DFS to evaluate if a packet is experiencing too much delay or that its queue is becoming overly large. The additional improvements seen by the ADFS protocol

compared to DFS are due to the ability of ADFS to dynamically update the packet weights based on the actual experienced delay and queue size allowing under serviced flows to “catch-up” while flows that have been serviced are made to wait.

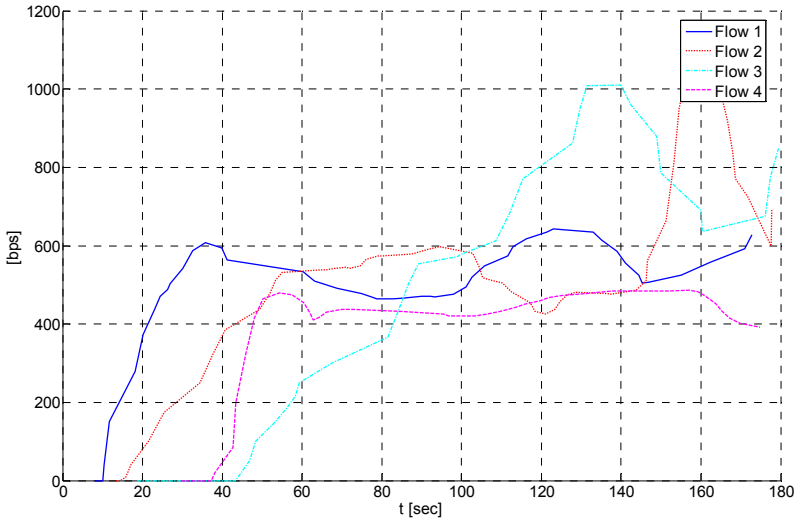

Fig. 4 Network flows for single cluster and FIFO

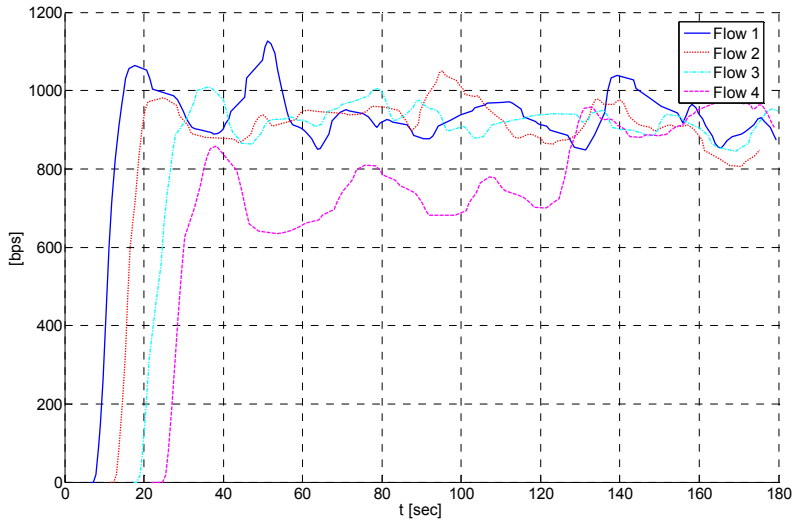

Fig. 5 Network flows for a single cluster and DFS

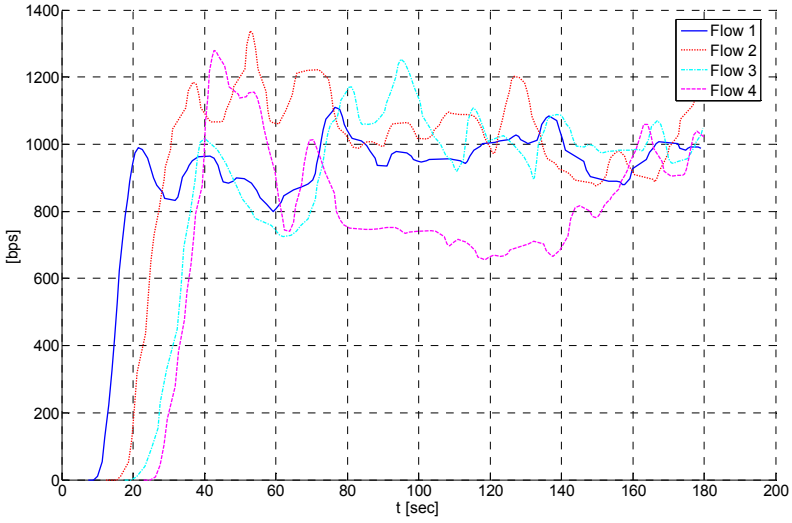

Fig. 6 Network flows for a single cluster and ADFS

Experimental results for networks with varying initial weights are shown next. Table III. ADFS provides improved fairness while increasing average throughput and decreasing E2E delay on a per-flow basis in the presence of priority flows. The ability of ADFS to adapt packet weights allows the network to react to the dynamic nature of the channel state and other conditions in the network. The adaptation of the weights ensures that the E2E delay and the queue size remain at an appropriate level for each flow and at each relay node. This stability in the delay and queuing allows for the improvements in the variation of throughput and the E2E delay values.

TABLE III RESULTS UNEVEN FLOW-WEIGHTING

	Average Throughput [bps]	Average E2E Delay [sec]	Average Dropped Packets	Average Drop Rate [packets/sec]	Average FI	Increase over FIFO	Increase over DFS
1 Cluster							
FIFO	230.270	2.900	139.170	1.357	0.867	X	X
DFS	662.760	0.985	92.000	0.899	0.909	4.9%	X
ADFS	810.350	0.719	107.330	1.052	0.918	5.9%	1 %

VI. Conclusions

The proposed protocol, MC-ADFS, introduces a new metric for the upper bound of the service for a flow and the upper bound for the error in service for two contending flows. Introduction of MC-ADFS for WSN allows for increased capabilities and transmission capacity allowing dynamic scheduling of packets over multiple channels to and facilitating increased network capacity, reduction of congestion, and efficient packet scheduling.

Additionally, testing results for the ADFS implementation on the MST/SLU motes show that the performance of such methods on WSN hardware is viable. Performance of the DFS and ADFS protocol as compared to FIFO scheduling shows that DFS and ADFS methods are able to maintain higher and more even transfer rates, between similarly weighted sources. The increases fairness allows for assurance of the QoS packet flows to the BS. ADFS provides decreased delays and higher average throughput translating to even usage of the channel compared with FIFO or DFS. Additionally, the E2E delay of ADFS is more linearly scalable to the size of the network as compared to FIFO or DFS methods allowing variation of the network size.

Analytical assurance of the QoS level provides confidence that the rate of packet transfer is guaranteed for each source and due to the distributed adaptive nature of ADFS allows implementation with low packet overhead. When real-time performance at a specified data rate is vital to application feasibility, such as the deployment of networked controls systems, ADFS can be useful to a WSN. In a real-time or near real-time system, both the throughput and delay of the sampled data are crucial to application stablity.

ADFS allows for satisfying these types of requirements while remaining scalable and sustainable.

ACKNOWLEDGMENT

The authors appreciate the continued advice and support of Dr. Kyle Mitchell of St. Louis University, St. Louis, MO on the G4-SSN system and hardware implementation issues.

REFERENCES

[1] S. Ratnaraj, S. Jagannathan and V. Rao, "OEDSR: Optimal energy delay subnet routing protocol for wireless sensor networks," *Proc. of the IEEE Conf. on Sensing, Networking, and Control (ICNSC)*, pp. 330-335, April 2006.

[2] S. Sahni and X. Xu, "Algorithms for wireless sensor networks," *Int. Journal of Distributed Sensor Networks*, vol. 1, pp. 35-56, 2005.

[3] N. Jain, D. P. Argawal, "current trends in wireless sensor network design" , *Int. Journal of Distributed Sensor Networks*, vol. 1, pp. 101-122, 2005.

[4] N. Regatte and S. Jagannathan, "Adaptive and distributed fair scheduling scheme for wireless adhoc networks," *Proc. of the Word Wireless Congress*, pp. 101-106, May 2004.

[5] P. Goyal, H.M. Vin, and H. Cheng, "Start-time fair queuing: A Scheduling Algorithm For Integrated Services Packet Switching Networks," *IEEE/ACM Trans. on Networking*, vol. 5, pp. 690–704, 1997.

[6] S.J. Golestani, "A self-clocked fair queuing scheme for broadband applications," *Proc. of the IEEE Conf. Networking for Global Communications (INFOCOM)*, Toronto, Canada, vol. 2, pp. 636–646, 1994.

[7] A. Demers, S. Keshav, and S. Shenker, "Analysis and simulation of a fair queuing algorithm," *Proc. of the ACM Conf. on Special Interest Group on Data Communications (SIGCOMM)*, vol. 19, no. 12, pp. 1-12, 1989.

[8] J. C. R. Bennett and H. Zhang, "WF2Q: Worst-case fair weighted fair queuing," *Proc. of the IEEE Joint Conf. of the IEEE Computer Societies (INFOCOM)*, vol. 1, pp. 120–128, 1996.

[9] N.H. Vaidya, P. Bahl, and S. Gupta, "Distributed fair scheduling in a wireless LAN," *IEEE Trans. on Mobile Computing,* vol. 4, no. 6, pp. 616-629, 2005.

[10] H. Luo, P. Medvedev, J. Cheng, and S. Lu, "A self-coordinating approach to distributed fair queuing in adhoc wireless networks," *Proc. of the IEEE Conf. Networking for Global Communications (INFOCOM)*, vol. 3, pp. 1370–1379, 2001.

[11] R. Jain, G. Babic, B. Nagendra, and C. Lam, "Fairness, call establishment latency and other performance metrics," Tech. Rep. *ATM_Forum*/96-1173, August 1996.

[12] S. Ramadhadran and J. Pasquale, "The stratified round robin scheduler: Design, analysis, and implementation," *IEEE/ACM Trans. on Networking*, vol. 14, no. 6, pp. 1362-1373, 2006.

[13] Z. Xilang and X. Cheng-Zhong, "Energy-efficient wireless packet scheduling with quality of service control," *IEEE Trans. on Mobile Computing*, vol. 6, no. 10, pp. 1158-1170, 2007.

[14] X. Qin, M. Alghamdi, M. Nijim, Z. Zong, and K. Bellam, "Scheduling of periodic packets in energy-aware wireless networks," *Proc. of the IEEE Int. Conf. on Performance, Computing, and Communications (IPCCC)*, New Orleans, LA, pp. 210-217, 2007.

[15] S. Hong, D. Kim, and J. Kim, "Battery aware real time task scheduling in wireless sensor networks," *Proc. of IEEE Inter. Conf. on Embedded and Real-Time Computing System and Applications*, pp. 269-272, 2005.

[16] Y. Yao and G. B. Giannakis, "Energy efficient scheduling for wireless sensor networks," *IEEE Trans. on Communications*, vol. 53, no. 8, pp. 1333-1342, 2005.

[17] A. Ghosh and T. Givargis, "LORD: a localized, reactive, and distributed protocol for node scheduling in wireless sensor networks," *Proc. of Design, Automation, and Test in Europe*, vol. 1, pp. 190-195, 2005.

[18] X. Yin, H. Chen, Y. Shen, and W. Qi, "A priority-based packet scheduling method in wireless sensor networks," *IEEE Int. Conf. on Information Acquisition*, pp. 627-631, 2006.

[19] P. Sharma, "Channel-state based scheduling in wireless sensor networks for reliable transmission," *Proc. IEEE Int. Parallel and Distributed Processing Symp.*, 6 pp., April 4-8, 2005.

[20] A. Sahoo and D. Manjunath, "Revisiting WFQ: minimum packet lengths tighten delay and fairness bounds," *IEEE Comm. Letters*, vol. 11, no. 4, pp. 366-368, 2007.

[21] M . Alam, A. Hamid, and C. S. Hong, "QoS-aware fair scheduling in multihop wireless ad hoc networks," *Proc. IEEE Int. Conf. on Adv. Communications Technology (ICACT)*, vol. 2, 5 pp. 2006.

[22] F. Liu, Y. Huang, Y. Ma, and N. Yi, "A new priority calculation method for sorted-priority fair queuing," *Proc. of the IEEE Conf. on Consumer Communications and Networking (CCNC)*, pp. 110-115, 2004.

[23] E.M. Popovici, "Coding and Cryptography for resource constrained wireless sensor networks: A hardware-software co-design approach," *Proc. of the IEEE Inter. Semiconductor Conf.*, vol. 1, pp. 19-27, 2006.

[24] K. Virk, J. Madsen, A. V. Lorentzen, M. Leopold, P. Bonnet, and M. Hansen, "Design of a development platform for HW/SW codesign of wireless integrated sensor nodes," *Proc. of the IEEE Euromicro Conf. on Digital Systems Design (DSD)*, pp. 254-260, 2005.

[25] F. Zhao, J. Liu, J. Reich, M. Chu, and J. Liu, "Programming embedded networked sensor systems," *Proc. of the IEEE/ACM/IFIP Inter. Conf. on Hardware/Software Codesign and System Synthesis*, p. 52, 2003.

[26] K. Lee, "Performance bounds in communication networks with variable-rate links," *Proc. of the ACM Conf. on Special Interest Group on Data Communications (SIGCOMM)*, 1995, pp. 126–136, 1995.

[27] J. W. Fonda, M. J. Zawodniok, J. Sarangapani, and S. E. Watkins, "Adaptive distributed fair scheduling and its implementation in wireless sensor networks," *Proc. of the IEEE Conf. on Systems, Man and Cybernetics*, Taipei, Taiwan, pp. 3382-3387, 2006.

[28] J. W. Fonda, M. J. Zawodniok, S. Jagannathan, S. E. Watkins, "Development and implementation of optimized energy-delay sub-network routing protocol for wireless sensor networks," *Proc. of the IEEE Int. Symp. on Intelligent Control*, Munich, Germany, pp. 119-124, 2006.

[29] William L. Brogan, *Modern Control Theory*, 3rd Edition, Prentice Hall, New Jersey, 1990.

[30] K. Mitchell, S. E. Watkins, J. W. Fonda, and J. Sarangapani, "Embeddable modular hardware for multi-functional sensor networks," *Smart Mater. Struct*., vol. 16, no. 5, pp. N27-N34, 2007.

PAPER 3

Joint Adaptive Distributed Rate and Transmission Power Control Scheme for Wireless Sensor and Ad-hoc Networks

James W. Fonda*[1,2], S. Jagannathan[1], and Steve E. Watkins[2]
Embedded Systems and Networking Laboratory[1]
Applied Optics Laboratory[2]
Department of Electrical and Computer Engineering
Missouri University of Science and Technology, Rolla, Missouri 65409-0040
jfonda@ieee.org, sarangap@mst.edu, and steve.e.watkins@ieee.org

Abstract—A novel adaptive distributed rate and power control (ADRPC) protocol is introduced for wireless sensor (WSN) and ad-hoc networks. The proposed controller contrasts from others in the literature by providing simultaneous nonlinear compensation to the problem of transmission power and bit-rate adaptation and its coupling to the management of channel capacity. The protocol provides control of both signal-to-interference ratio (SIR) and quality-of-service (QoS) support to bit-rate adaptation. Bit-rate adaptation is performed by local estimation of congestion levels, which renders little packet overhead, on the network using Lyapunov-based adaptive control methods. Additionally, the problem of power and bit-rate adaptation is modeled as an affine nonlinear discrete-time system and an optimal scheme is utilized to control the rate and SIR levels. The optimal scheme provides the nearly-optimal control effort in order to provide an ideal transmission power and data rate for the current channel condition and congestion levels. Performance of the proposed control scheme is validated through simulation. Results include a probabilistic drop rate, random topologies, and increasing node populations. Also included, are performance comparisons between a continuous and discrete steps in the power settings to highlight differences in hardware implementations.

***Index Terms*—** Weight-adaptation, Quality-of-Service, Lyapunov, Adaptive Control, Rate adaptation, Power Control

I. INTRODUCTION

In wireless networking applications, the quality of service (QoS) issues are inter-related to channel dynamics and are important to network performance. To ensure QoS over a wireless channel for a particular bandwidth capacity, the transmission power and bit-rate must be controlled [1,2]. In many applications involving wireless sensor networks (WSN) and ad-hoc networks increasing energy-efficiency while maintaining QoS is

Research supported in part by Dept. of Education GAANN Fellowship, NSF I/UCRC on Intelligent Maintenance Systems, Air Force Research Laboratory Grant (FA8650-04-C-704) and Intelligent Systems Center.

desirable to extend battery life. In these networks, distributed control methods that provide analytically guaranteed performance while requiring little traffic overhead are desirable. In this paper, joint control of bit-rate and transmission power using Lyapunov-based decentralized adaptive and a decentralized optimal control schemes is proposed.

Transmission power control provides a reduction in power consumption, minimizes mutual interference, and maintains link capacity. Interference management at each node allows signal-to-interference ratio (SIR) to be satisfied for a specific data rate. Previously, rigorous work involving distributed power control (DPC) was performed for cellular networks [3-6]. Also, several DPC schemes [7-11] were developed for wireless ad-hoc networks where the topology is dynamic due to node mobility and communication link failures.

Unlike wired networks, radio channel uncertainties in a wireless network, such as path loss, shadowing, and Rayleigh fading, can attenuate the power of the transmitted signal causing variations in the SIR degrading performance. Low SIR levels results in higher bit error rates (BER), causing increased numbers of dropped bits and/or packets. Many DPC schemes [7-11] for ad-hoc networks assume that: i) only path loss is present, ii) no other channel uncertainty exists, and iii) the mutual interference among the users is held constant during users' power updates. Currently there is previous work into power control methods that account for channel uncertainties by Zawodniok et. al.[12].

In contrast, this work proposes a Lyapunov-based method for joint control of bit-rate and transmission power, known as the distributed or decentralized rate and power control (ADRPC) protocol by modeling the wireless sensor and ad-hoc network as an affine nonlinear discrete-time system. Lyapunov methods allow for nonlinear compensation for

the dynamic wireless channels and to satisfy user-defined QoS metrics for bit-rates. Joint control of bit-rate and transmission power provides QoS assurances to overcome channel uncertainties and congestion levels. Additionally, an optimal control strategy based on solution of the generalized Hamilton-Hamilton-Jacobi (GHJB) equation is proposed. Methods presented in [13] are extended to provide optimal tracking control to a general class of nonlinear wireless networking systems by specifically applying the near-optimal tracking controller to the joint bit-rate and SIR dynamics. This general class of wireless networking system includes not only wireless sensor and ad-hoc networks but also cellular networks.

Previously, methods of joint power and bit-rate control have included stochastic [14], game-theory [15], linear programming [16], neuro-fuzzy [17], optimization [18], optimal control [19], linear quadratic control [20], and adaptive schemes [20]. In [20], the SIR is determined by the rate; however the rate is not suitably controlled to provide QoS and is considered to be a stochastic input to the system. Additionally, congestion levels are assumed to be known a priori and are not estimated. To provide suitable QoS to a network, both the power and bit-rate must be controlled in a reliable and an analytically guaranteed manner.

In this work, the application of both adaptive and optimal tracking control methods to analytically guarantee performance of both the bit-rate and power control and to overcome limitations in previous works are presented. Application of an adaptive power control method to estimate channel conditions and a bit-rate controller that provides estimates of congestion levels allows the proposed methods to provide higher QoS over the entire network by limiting the number of retransmissions and thus reducing the

wastage of energy used for transmission. Additionally, desired SIR levels are chosen based on the desired bit-rate via Shannon's Capacity formula, thus assuring sufficient channel capacity for generated traffic. The use of control methods to estimate the changes in both the channel and congestion state provide innovation to the problem of joint power and bit-rate control. The proposed methods allow for a distributed controller requiring little network overhead to overcome channel variation and network congestion. Innovations include application of adaptive control methods for bit-rate utilizing estimations of the congestion levels. Additionally, the joint control method provides innovation in the form of a closed loop controller for the SIR and the bit-rate to meet user-defined QoS while compensating for variations in channel state and network congestion. Finally, application of GHJB control methods provides innovation through extension of the joint bit-rate and power control to near-optimal control. The use of optimal control methods provides minimization in the control effort in terms of transmission power and selection of appropriate bit rate to reach a particular level of performance. Additionally, innovation in the GHJB is found through presentation of theorems to support optimal tracking control.

II. Adaptive Controller Methodology

Adaptive estimation of the channel state and the network congestion facilitate control of both the bit-rate and transmission power. In this section, a description of desired bit-rate selection based on user QoS is presented. Additionally, the impact of the bit-rates on the transmission power control for selection of the desired SIR level is discussed. Next, both the bit-rate and SIR dynamics are presented with controller methodologies. Finally,

mathematical derivation of the controller performance provides analytical proof of the stability of the proposed methods.

A. Desired Rate Selection

To ensure proper QoS desired bit-rates must be based on users' perceptions of how the network should perform for a given application. In the scope of this work, the QoS is set by the time a user requires to send a particular quantity, i.e. the end-to-end (E2E) delay, of data. The desired bit-rate is calculated and then updated for each time interval to account for received and dropped bits. Calculation of the bit-rate over the first time interval is given by the total number of bits for transmission divided by the E2E-delay requirement. At each subsequent time interval, the desired bit-rate is updated by calculation of the number of bits left for transmission and the time remaining for delivery. Desired bit-rates at each time interval for the i^{th} node are selected as

$$r_{des,i}(l) = \frac{\chi_i(l)}{t_{f,i}} \tag{1}$$

where the number of bits to send is given by $\chi_i(l)$, and $t_{f,i}$ is the time remaining for the transmission. Desired bit-rates are then updated at each time interval to accommodate the QoS and account for any dropped bits, or lags, in the rate due to other variables. It should be noted in this work the desired rate selection is not complex. More sophisticated rate control methods could be used in place of (1) using the same control architecture.

B. Radio Channel with Uncertainties

Radio channels place fundamental limitations on wireless communication systems. The transmission path can vary from line-of-sight (LOS) to a path obstructed by buildings, mountains, and foliage. In wireless networks, channel uncertainties such as path loss, shadowing, and Rayleigh fading attenuates signal power and introduces

variations in the SIR at the receiver degrading performance of DPC schemes. The effect of these uncertainties is represented through a channel loss (gain) factor that typically multiplies with the transmitter power. Therefore, the channel loss or gain, g, can be expressed as [21, 22]

$$g = f(d, n, X, \zeta) = d^{-n} \cdot 10^{0.1\zeta} \cdot X^2 \tag{2}$$

where d^{-n} is the effect of path loss due to the distance, d, and the $10^{0.1\cdot\varsigma}$ term corresponds to the effect of shadowing. For Rayleigh fading the power attenuation is modeled as X^2, where X is a random variable with a Rayleigh distribution. Typically the channel gain, g, is a function of time and affects not only the transmission power, but also the channel bandwidth via Shannon's capacity relation. In other words, if the channel gain decreases then the capacity of the channel also reduces. A decrease in channel capacity can lead to added congestion unless the network traffic is reduced; therefore, to provide the needed channel capacity, and avoid congestion, the SIR must be chosen to compensate for the dynamic nature of the channel state over time.

C. Desired SIR Selection

The desired SIR at each time interval is based on the actual bit-rate that is being produced at each node ensuring that the SIR requirements at each node are satisfied. In contrast, desired SIR levels could be selected to satisfy the desired bit-rate, however this method leads to cases where the transmission power is greater than needed when a desired bit-rate is not reached, introducing additional interference. The desired SIR is based on Shannon's capacity formulation and is given by

$$\left[SIR_{des,i}(l)\right]_{dB} = \frac{E_b}{N_o} + 10\log_{10}\left(\frac{rate_i(l)}{BW}\right) \tag{3}$$

where the bandwidth efficiency, $\frac{E_b}{N_o}$, is chosen based on the desired bit-error-rate (BER) and modulation scheme. The desired bandwidth efficiency can be calculated or found in tables [23].

D. Adaptive Controller Formulation

Transfer of payload is controlled through the bit-rate adaptation by estimation of congestion levels in the network and user QoS metrics as described. The development of a Lyapunov-based nonlinear adaptive control of transmission rate and power is also shown. First, the rate dynamics and the bit-rate control method are given by representing the network as an affine nonlinear discrete-time system. Next, the SIR dynamics and the associated control method are described. Finally, a unified control scheme is presented and is shown to be asymptotically stable for ideal conditions and to have bounded error in the presence of disturbances.

1) Rate Dynamics and Control

Bit-rate impacts both the QoS and the congestion level of a wireless network. Distributed control of the bit-rate and estimation of the congestion level is facilitated through adaptive methods. Bit-rate dynamics are used as proposed in [20] and are given as

$$\begin{aligned} &r_i(k+1) = r_i(k) + \mu\left[d_i(k) - c_i(k) r_i(k)\right] = \\ &r_i(k)\left[1 - \mu c_i(k)\right] + \mu d_i(k) = r_i(k)\Phi_i(k) + \mu d_i(k) \end{aligned} \tag{4}$$

where μ is a positive step-size, $c_i(l)$ is the congestion level, and $\Phi_i(l) = \left[1 - \mu c_i(l)\right]$. In [20], the term $d_i(l)$ is the bit-rate control input and is assumed to be a random variable with mean m_d and variance σ_d^2. In contrast, the proposed method provides control of the bit-rate via $d_i(l)$ in conjunction with the transmission power levels, $p_i(l)$, to control the

SIR and bit-rates simultaneously. A joint control methodology is important since congestion and channel state effect the channel capacity. While congestion is a function of bit-rates generated at network nodes, it is also influenced by the channel gain due to an increase and decrease in channel capacity. In situations where the network is near capacity and the channel gain decreases congestion will increase. This relationship infers that SIR is a nonlinear function of the channel state and the bit-rates. In other words, the congestion level of the network is a function of the path loss, shadowing, and fading of the channel as well as the errors in the desired SIR levels, or mathematically $c_i(l) = f\left(d, n, X, \zeta, r_i(l) e_{SIR}(l)\right)$. Additionally, congestion levels are unknown making the bit-rate control problem more challenging. However, adaptive estimation of the congestion levels and subsequent selection of the bit-rate is proposed to overcome these issues.

The error dynamics for the transmission rate control subsystem are given as

$$e_{r,i}(l) = r_i(l) - r_{des}(l) \tag{5}$$

and the error in the next time step can be shown as

$$e_{r,i}(l+1) = r_i(l+1) - r_{des}(l+1) = r_i(l)\Phi_i(l) + \mu d_i(l) - r_{des}(l+1) \tag{6}$$

Introducing the control law of

$$d_i(l+1) = \frac{1}{\mu}\left[k_r e_r(l) + r_{des}(l+1) - r_i(l)\hat{\Phi}_i(l)\right] \tag{7}$$

where $\hat{\Phi}_i(l)$ is an unknown bounded parameter vector to be estimated due to congestion level, which is typically unknown. Using the error dynamics given in (6) and the control law in (7) the resulting error dynamics for the bit-rate are given by

$$e_{r,i}(l+1) = r_i(l)\Phi_i(l) + k_r e_r(l) + r_{des}(l+1) - r_i(l)\hat{\Phi}_i(l) - r_{des}(l+1) = r_i(l)\tilde{\Phi}_i(l) + k_r e_r(l) \quad (8)$$

where $\tilde{\Phi}_i(l)$ is the parameter estimation error for the rate control. With the addition of a parameter estimation error term that is given as $\varepsilon_r(l)$, which is considered bounded as $\|\varepsilon_r(l)\| \le \varepsilon_{r,\max}$ where $\varepsilon_{r,\max}$ is a known constant.

Next, two cases for control are shown. In the first case, an asymptotically stable system is demonstrated in ideal situations whereas in the second case boundedness can be shown in the presence of bounded estimation errors without the persistency of excitation (PE) condition.

Case 1: Φ_i and $B_i(l)$ are known. In this scenario, the feedback bit-rate control is selected as

$$u_i(l) = B_i^{-1}\left[k_r e_{r,i}(l) + r_{des,i}(l) - \Phi_i(l) r_i(l)\right] \quad (9)$$

where the error in the states is defined as in (5)and $B_i = \mu$. The resulting error dynamics are given as

$$e_{r,i}(l+1) = k_r e_{r,i}(l) \quad (10)$$

With appropriate selection of k_r, the Eigenvalues are placed within the unit disc, it is easy to show that the closed-loop system is asymptotically stable in the mean as given in Theorem 1.

Case 2: Φ_i and $B_i(l)$ are unknown. In this scenario, equation (4) can be expressed as

$$r_i(l+1) = \Phi_i(l) r_i(l) + B_i(l) u_i(l) \quad (11)$$

Selecting the feedback control as (6), where $\hat{\Phi}_i(l)$ is the estimate of $\Phi_i(l)$, the state error system is expressed as

$$e_{r,i}(l+1) = k_r e_{r,i}(l) + \tilde{\Phi}_i^T(l) r_i(l) + \varepsilon(l) \tag{12}$$

where $\tilde{\Phi}_i(l) = \Phi_i(l) - \hat{\Phi}_i(l)$ is the parameter estimation error. In this case, a standard adaptive parameter update law is chosen as

$$\hat{\Phi}_i(l+1) = \hat{\Phi}_i(l) + \sigma \psi_i(l) e_{r,i}^T(l+1) \tag{13}$$

where $\psi_i(l) = r_i(l)$.

Theorem 1: Given the rate dynamics above with uncertain levels of congestion, if the feedback bit-rate control is selected as (7), then the congestion estimation error along with the mean rate error converges to zero asymptotically, if the parameter updates are taken as (13) provided

$$\sigma \|\psi_i(l)\|^2 < 1 \tag{14}$$

$$k_{v\max} < \frac{1}{\sqrt{\delta}} \tag{15}$$

where $\delta = \frac{1}{1 - \sigma \|\psi_i(l)\|^2}$, and σ is the adaptation gain.

Proof: Define the Lyapunov function candidate

$$J_i = e_{r,i}^T(l) e_{r,i}(l) + \frac{1}{\sigma} \kappa \left[\tilde{\Phi}_i{}^T(l) \tilde{\Phi}_i(l) \right] \tag{16}$$

whose first difference is

$$\Delta J = \Delta J_1 + \Delta J_2 = e_{r,i}^T(l+1) e_{r,i}(l+1) - e_{r,i}^T(l) e_{r,i}(l) + \frac{1}{\sigma} \kappa \left[\tilde{\Phi}_i{}^T(l+1) \tilde{\Phi}_i(l+1) - \tilde{\Phi}_i{}^T(l) \tilde{\Phi}_i(l) \right] \tag{17}$$

Consider ΔJ_1 from (17) and substituting (12) to get

$$\begin{aligned}\Delta J_1 &= e_{r,i}^T(l+1)e_{r,i}(l+1) - e_{r,i}^T(l)e_{r,i}(l) \\ &= \left(k_r e_{r,i}(l) + \tilde{\Phi}_i{}^T(l)\psi_i(l)\right)^T\left(k_r e_{r,i}(l) + \tilde{\Phi}_i{}^T(l)\psi_i(l)\right) - e_{r,i}^T(l)e_{r,i}(l)\end{aligned} \tag{18}$$

Taking the second term of the first difference from (18) and substituting (14) yields

$$\begin{aligned}\Delta J_2 &= \frac{1}{\sigma}\kappa\left[\tilde{\Phi}_i{}^T(l+1)\tilde{\Phi}_i(l+1) - \tilde{\Phi}_i{}^T(l)\tilde{\Phi}_i(l)\right] \\ &= -2\left[k_r e_{r,i}(l)\right]^T\tilde{\Phi}_i{}^T(l)\psi_i(l) - 2\left[\tilde{\Phi}_i{}^T(l)\psi_i(l)\right]^T\left[\tilde{\Phi}_i{}^T(l)\psi_i(l)\right] + \\ &\sigma\psi_i{}^T(l)\psi_i(l)\left[k_r e_{r,i}(l) + \tilde{\Phi}_i{}^T(l)\psi_i(l)\right]^T\left[k_r e_{r,i}(l) + \tilde{\Phi}_i{}^T(l)\psi_i(l)\right]\end{aligned} \tag{19}$$

Combining (17) and (19) to get

$$\begin{aligned}\Delta J &= -e_{r,i}^T(l)\left[I - \left(1+\sigma\psi_i{}^T(l)\psi_i(l)k_r{}^T k_r\right)\right]e_{r,i}(l) \\ &+ 2\sigma\psi_i{}^T(l)\psi_i(l)\left[k_r e_{r,i}(l)\right]^T\left[\tilde{\Phi}_i{}^T(l)\psi_i(l)\right] - \left(1-\sigma\psi_i{}^T(l)\psi_i(l)\right)\left[\tilde{\Phi}_i{}^T(l)\psi_i(l)\right]^T\left[\tilde{\Phi}_i{}^T(l)\psi_i(l)\right] \\ &\le -\left(1-\delta k_{v\max}^2\right)\left\|e_{r,i}(l)\right\|^2 - \left(1-\sigma\left\|\psi_i(l)\right\|^2\right)\left\|\tilde{\Phi}_i{}^T(l)\psi_i(l) - \frac{\sigma\left\|\psi_i(l)\right\|^2}{1-\sigma\left\|\psi_i(l)\right\|^2}k_r e_{r,i}(l)\right\|^2\end{aligned} \tag{20}$$

where δ is given after (15). Taking now expectations on both sides yields

$$E(\Delta J) \le -E\left(\left(1-\delta k_{v\max}^2\right)\left\|e_{r,i}(l)\right\|^2 - \left(1-\sigma\left\|\psi_i(l)\right\|^2\right)\left\|\tilde{\Phi}_i{}^T(l)\psi_i(l) + \frac{\sigma\left\|\psi_i(l)\right\|^2}{1-\sigma\left\|\psi_i(l)\right\|^2}k_r e_{r,i}(l)\right\|^2\right) \tag{21}$$

Since $E(J) > 0$ and $E(\Delta J) \le 0$, this result demonstrates the stability in the mean via sense of Lyapunov provided the conditions (14) and (15) hold, so $E\left[e_{r,i}(l)\right]$ and $E\left[\tilde{\Phi}_i(l)\right]$ (and hence $E\left[\hat{\Phi}_i(l)\right]$) are bounded in the mean if $E\left[e_{r,i}(l_0)\right]$ and $E\left[\tilde{\Phi}_i(l_0)\right]$ are bounded in a mean. Summing both sides of (21) and taking the limit $\lim_{l\to\infty} E(\Delta J)$, the rate error $E\left[\left\|e_{r,i}(l)\right\|\right] \to 0$.

From (8), clearly the closed-loop error system is the network congestion estimation error. If the congestion levels are properly estimated, then estimation error tends to be zero. In this case, equation (8) becomes (10). In the presence of error in estimation, only

boundedness of error can be shown similar to the case of adaptive power control (see next section where similar results are shown for joint power and rate control). Provided the system uncertainties are properly estimated the rates can be shown to approach the target values. Additionally, in this case the parameter update will only converge close to the actual values under the assumption of persistently excitation (PE) of the inputs. Since the PE condition is difficult to guarantee over all time it will be relaxed in the joint controller framework presented later through modification of the parameter update law.

E. SIR Dynamics and Control

Control of the transmission power impacts the SIR at all other nodes. A distributed controller providing estimates of channel conditions through adaptive methods is introduced in this section. The goal of transmitter power control is to maintain a target SIR threshold for each wireless link through adjustment of transmitter power to meet capacity requirements and preserve energy while countering channel uncertainties. In this section the dynamics of the SIR system are presented.

Suppose there are $N \in Z_+$ links in the network. Let g_{ij} be the power loss (gain) from the transmitter of the j^{th} link to the receiver of the i^{th} link. The power attenuation is considered to follow the relationship given in equation (2).

Calculation of SIR, $R_i(t)$, at the receiver of i^{th} link at the time instant t, is given by

$$R_i(t) = \frac{g_{ii}(t)P_i(t)}{I_i(t)} = g_{ii}(t)P_i(t) \Big/ \sum_{j \neq i} g_{ij}(t)P_j(t) + \eta_i(t) \tag{22}$$

where $i, j \in \{1,2,3,\ldots,n\}$, $I_i(t)$ is the interference, $P_i(t)$ is the link's transmitter power, $P_j(t)$ are the transmitter powers of all other nodes, and $\eta_i(t)>0$ is the variance of the noise at its

receiver node. Now the dynamic system is presented in (23), for a full treatment of the SIR dynamics see [12].

The dynamics of the SIR are given in (23) where $\omega(l)$ is the zero mean stationary stochastic channel noise with $n_i(l)$ as its coefficient. The SIR of each link at time instant l is obtained as

$$y_i(l+1) = \alpha_i(l) y_i(l) + \beta_i(l) v_i(l) + n_i(l)\omega_i(l) \tag{23}$$

Carefully observing (23), one sees that the SIR at the time instant $l+1$ is a function of channel variations, $\alpha_i(l)$, from time instant l to $l+1$. The channel variation is not known a priori, making the DPC scheme development challenging. Since $\alpha_i(l)$ is not known, is must be estimated for the development of DPC methods. Note that available DPC schemes [3-4, 24] ignore the channel variations and therefore render unsatisfactory performance. Equation (23) can be rewritten as

$$y_i(l+1) = \begin{bmatrix} \alpha_i(l) & n_i(l) \end{bmatrix} \begin{bmatrix} y_i(l) \\ \omega_i(l) \end{bmatrix} + \beta_i(l) v_i(l) = \tau_i^T(l)\gamma_i(l) + \beta_i(l) v_i(l) \tag{24}$$

where the regression vector $\gamma_i(l) = \begin{bmatrix} y_i(l) \\ \omega_i(l) \end{bmatrix}$ and the unknowns are $\tau_i^T(l) = [\alpha_i(l) \quad n_i(l)]$ and

$$\alpha_i(l) = [\Delta g_{ii}(l)/g_{ii}(l)] - \left[\left(\sum_{j \neq i} \Delta g_{ij}(l) P_j(l) + \Delta P_j(l) g_{ij}(l) \right) \Big/ (I_i(l)) \right] \tag{25}$$

and $\beta_i(l) = g_{ii}(l)$ (26)

The control input can be selected as (21) and

$$v_i(l) = P_i(l+1)/I_i(l) \tag{27}$$

Remark: The control of the SIR is similar to the case of previously presented rate control. Therefore the SIR control can similarly be shown to be asymptotically stable and bounded with removal of the PE condition. Due to PE condition being difficult to guarantee the PE condition will be relaxed in the next section when the joint controller framework is proposed and proofs are given.

F. Joint Controller

Using the system dynamics described above a joint control system is now proposed. The two control systems are concatenated to form a single state-space system and then control methods are applied. A rigorous treatment of the stability and performance is shown. The proofs reflect the performance of the control methods in an ideal environment and in a non-ideal setting with estimation errors. When estimation errors are present, the control method can be shown to have a bounded error. Rewrite (11) and (24) as

$$x_i(l+1)=\begin{bmatrix}\Phi_i(l)r_i(l)+\mu_i d_i(l)\\ \psi_i(l)y_i(l)+\beta_i v_i(l)\end{bmatrix}=\begin{bmatrix}\Phi_i(l) & 0\\ 0 & \tau_i(l)\end{bmatrix}\begin{bmatrix}r_i(l)\\ \gamma_i(l)\end{bmatrix}+\begin{bmatrix}\mu_i & 0\\ 0 & \beta_i\end{bmatrix}\begin{bmatrix}d_i(l)\\ v_i(l)\end{bmatrix} \quad (28)$$
$$=\theta_i(l)x_i(l)+B_i(l)u_i(l)$$

where $x_i(l+1)=\begin{bmatrix}r_i(l+1)\\ y_i(l+1)\end{bmatrix}$, $\theta_i(l)=\begin{bmatrix}\Phi_i(l) & 0\\ 0 & \tau_i(l)\end{bmatrix}$, and $x_i(l)=\begin{bmatrix}r_i(l)\\ y_i(l)\end{bmatrix}$.

The control input is given as

$$u_i(l+1)=B_i^{-1}\left[K_{v,i}e(l)+x_{d,i}(l+1)-\hat{\theta}_i(l)x_i(l)\right] \quad (29)$$

which yields the following joint power and bit-error dynamics as

$$e_i(l+1)=\begin{bmatrix}K_v & 0\\ 0 & K_{SNR}\end{bmatrix}e_i(l)+\begin{bmatrix}\tilde{\Phi}_i(l) & 0\\ 0 & \tilde{\gamma}_i(l)\end{bmatrix}\begin{bmatrix}r_i(l)\\ y_i(l)\end{bmatrix}=K_{v,i}e(l)+\tilde{\theta}_i^T(l)x_i(l)+\varepsilon(l) \quad (30)$$

The ADRPC method is now shown for the case when both $\theta_i(l)$ and $B_i(l)$ are unknown. The control method is shown to have mean estimation error along with the mean state error that is bounded. The cases where both $\theta_i(l)$ and $B_i(l)$ are known are similar to proofs presented earlier in the rate control methods. In this section, an update method for the adaptive estimation is presented that allows for relaxation of the PE condition providing stable bounded estimations and control in the mean.

When $\theta_i(l)$ and $B_i(l)$ are unknown, equation (28) can be expressed as an affine nonlinear discrete-time system

$$x_i(l+1) = \theta_i(l)x_i(l) + B_i(l)u_i(l) \tag{31}$$

Now selecting feedback control as

$$u_i(l+1) = B_i^{-1}\left[K_{v,i}e(l) + x_{d,i}(l+1) - \hat{\theta}_i(l)x_i(l)\right] \tag{32}$$

where $\hat{\theta}_i(l)$ is the estimate of $\theta_i(l)$. Then the state error system is expressed as

$$e_i(l+1) = K_{v,i}e(l) + \tilde{\theta}_i^T(l)x_i(l) + \varepsilon(l) \tag{33}$$

where $\tilde{\theta}_i(l) = \theta_i(l) - \hat{\theta}_i(l)$ is the parameter estimation error . From (33), clearly the closed-loop error dynamics are driven by channel estimation and the network congestion estimation errors. If the channel uncertainties and congestion levels are properly estimated, then estimation error tends to be zero. In the presence of error in estimation, only boundedness of error in can be shown. The actual system states can be shown to approach the target provided the system uncertainties are properly estimated. To proceed further, the Assumption 1 is used as in [6].

Assumption 1: The channel and congestion levels change slowly in comparison to parameter updates.

Remark: The above assumption is standard in the adaptive control literature. A safety factor is included to compensate for both sudden changes in the SIR and the congestion of the network.

Consider now the closed-loop error dynamics with parameter estimation error, *ε(l)*, as

$$e_i(l+1) = K_{v,i} e_i(l) + \tilde{\theta}_i^T(l) \cdot \psi_i(l) + \varepsilon(l) \tag{34}$$

where *ε(l)* is considered bounded above $\|\varepsilon(l)\| \le \varepsilon_N$, with ε_N a known constant. It is typical in the standard adaptive control [25] to assume the estimation errors to be zero and demonstrating stability by applying certainty equivalence principle. By contrast, in this work, the estimation errors are not considered as zero and certainty equivalence principle is relaxed.

Theorem 2: Given the adaptive scheme above with channel and congestions uncertainties for a wireless network, if the feedback from the scheme is selected as (33), then the mean estimation and state errors are bounded, if the parameter updates are taken as

$$\hat{\theta}_i(l+1) = \hat{\theta}_i(l) + \sigma \psi_i(l) e_i^T(l+1) - \Gamma \left\| I - \sigma \psi_i^T(l) \psi_i(l) \right\| \hat{\theta}_i(l) \tag{35}$$

where $\Gamma > 0$ is a design parameter, and $\psi_i(l) = x_i(l)$, where $\psi_i(l)$ is formally known as the regression matrix and is simply the states for our purposes. Then the mean error in states $e_i(l)$ and the mean estimated channel and congestion parameters, $\hat{\theta}_i(l)$,are bounded without the need for the PE condition, with the bounds specifically given by (44) and (46) provided

$$\sigma\left\|\psi_i(l)\right\|^2 < 1 \tag{36}$$

$$0 < \Gamma < 1 \tag{37}$$

$$K_{v,i:\max} < 1/\sqrt{\delta} \tag{38}$$

where

$$\delta = \eta + 1/\left(1-\sigma\left\|\Psi_i(l)\right\|^2\right)\left[\Gamma^2\left(1-\sigma\left\|\Psi_i(l)\right\|^2\right)^2 + 2\sigma\Gamma\left\|\Psi_i(l)\right\|^2\left(1-\sigma\left\|\Psi_i(l)\right\|^2\right)\right] \tag{39}$$

and σ is the adaptation gain.

Note: The parameters σ, η, δ are dependent upon the desired SIR and bit-rate values with time.

Proof: Select a Lyapunov function candidate as

$$J_i = e_i^T(l)e_i(l) + \frac{1}{\sigma}\kappa\left[\tilde{\theta}_i^T(l)\tilde{\theta}_i(l)\right] \tag{40}$$

Using the estimation error shown in (34) and parameter tuning mechanism (35) to obtain

$$\begin{aligned}\Delta J \le &-\left[1-\sigma K_{v,i:\max}^2\right]\left\|e_i(l)\right\|^2 - \left[1-\sigma\Psi_i^T(l)\Psi_i(l)\right]\cdot\left\|\tilde{\theta}_i^T(l)\Psi_i(l) - \frac{1}{\left(1-\sigma\Psi_i^T(l)\Psi_i(l)\right)}\left(\sigma\Psi_i^T(l)\Psi_i(l) + 2\Gamma\left\|I-\sigma\Psi_i(l)\Psi_i^T(l)\right\|\right)\right. \\ &\left.\cdot\left(K_{v,i}e_i(l)+\varepsilon(l)+d(l)\right)\right\|^2 + 2\gamma K_{v,i:\max}\left\|e_i(l)\right\| + \rho - \frac{1}{\sigma}\left\|I-\sigma\Psi_i(l)\Psi_i^T(l)\right\|^2\left[\Gamma(2-\Gamma)\left\|\hat{\theta}_i(k)\right\|\theta_{max} - \Gamma^2\theta_{max}^2\right]\end{aligned} \tag{41}$$

where

$$\gamma = \left[\eta\left(\varepsilon_N + d_M\right) + \Gamma\left(1-\sigma\left\|\Psi_i(l)\right\|^2\right)\left\|\Psi_i(l)\right\|\theta_{max}\right] \tag{42}$$

and

$$\rho = \left[\ \eta\left(\varepsilon_N + d_M\right)^2 + 2\Gamma\left(1-\sigma\left\|\Psi_i(l)\right\|^2\right)\left\|\Psi_i(l)\right\|\theta_{max}\left(\varepsilon_N + d_M\right)\ \right] \tag{43}$$

Completing the squares for $\widetilde{\theta}_i(l)$ in (41) and taking the expectation of both sides results in $E(J)>0$ and $E(\Delta J)\le 0$, this result shows the stability in the mean via sense of Lyapunov provided the conditions (36) and (38) hold. The results demonstrate that $E(\Delta J)$ is negative outside a compact set U. According to a standard Lyapunov extension, the state error $E[e_i(l)]$ is bounded for all $l \ge 0$ and the upper bound on the mean state error is given by

$$E(\|e_i(l)\|) > 1/\left(1-\sigma K^2_{v,i:\max}\right)\left[\gamma K_{v,i:\max} + \sqrt{\rho_1\left(l-\sigma K^2_{v,i:\max}\right)}\right] \tag{44}$$

where

$$\rho_1 = \rho + \frac{1}{\sigma}\frac{\Gamma}{2-\Gamma}\left(1-\sigma\|\theta(l)\|^2\right)^2\theta^2_{max} \tag{45}$$

On the other hand, completing the squares for $\|e_i(l)\|$ in (41) results in $E(\Delta J)\le 0$ as long as the conditions (36)-(38) are satisfied and

$$E(\|\widetilde{\theta}_i(l)\|) > \left(\Gamma(1-\Gamma)\theta_{max} + \sqrt{\Gamma^2(1-\Gamma)^2\theta^2_{max} + \Gamma(2-\Gamma)\Theta}\right)\Big/\left(\Gamma(2-\Gamma)\right) \tag{46}$$

where

$$\Theta = \left[\Gamma^2\theta^2_{max} + \sigma\rho_1\Big/\left(1-\sigma\|\Psi_i(l)\|^2\right)^2\right] \tag{47}$$

and

$$\rho_1 = \rho + \frac{\gamma^2 k^2_{vmax}}{\left(1-\delta k^2_{vmax}\right)} \tag{48}$$

In general $E(\Delta J) \leq 0$ in a compact set as long as (36) and (39) are satisfied and either (44) or (46) hold. According to the standard Lyapunov extension theorem [25] demonstrates that the tracking error and the error in parameter estimates are bounded.

Remarks:

a) For practical purposes, (44) and (46) can be considered as bounds for $\|e_i(l)\|$ and $\|\tilde{\theta}_i(l)\|$.

b) Parameter reconstruction error bound ε_N and the bounded channel disturbances d_M increase the bounds on $\|e_i(l)\|$ and $\|\tilde{\theta}_i(l)\|$ in a very interesting way.

c) Lyapunov proofs for nonlinear discrete-time systems are relatively difficult to pursue when compared to continuous-time since the first difference of the Lyapunov function is quadratic with respect to the states whereas in the continuous-time case it is linear. The discrete-time nature makes the proof quite complicated and this result is provided here.

III. GHJB Optimal Control Methodology

The transmission power and bit-rate system described above is represented as an affine non-linear discrete-time form as given by

$$x(l+1) = f(x(l)) + g(x(l))u(l) \tag{49}$$

where $x(l) \in \Omega \subset \Re^n$, $u: \Re^n \to \Re^m$, $f: \Re^n \to \Re^n$ and $g: \Re^n \to \Re^{n \times m}$. For the system presented earlier in Section II, define $f(x(l)) = \theta_i(l)x_i(l)$ and $g(x(l)) = B_i(l)$ to represent the dynamics in the form of (49). Also, for the application of the GHJB optimal methods, the assumption that .the nonlinear system terms are explicitly known initially is made.

The system shown in (49) is assumed controllable such that there exists a control signal on Ω that will stabilize the system. A desirable control function $u: \Re^n \to \Re^m$, is one that minimizes the generalized quadratic cost function

$$J(x(0);u)=\sum_{l=0}^{\infty}\left(x(l)^T Qx(l)+u(x(l))^T Ru(x(l))\right)+\phi(x(\infty)) \tag{50}$$

where Q is a positive definite matrix, R is a symmetric positive definite matrix, and $\phi:\Re^n\to\Re$ is the final positive definite state punishment function.

Admissible controls are denoted by a set Ω_u. A control function $u:R^n\to R^m$ is defined to be admissible with respect to the state penalty function $(x(l))^T Q(x(l))$ and control energy penalty function $(u(x(l)))^T R(u(x(l)))$ on Ω, denoted as $u\in\Omega_u$, if

- u is continuous on Ω,
- $u(x)|_{x=0}=0$,
- $u(x)$ stabilizes system (2-1) on Ω
- $J(x(0);u)=\sum_{k=0}^{\infty}\left(\left(x(l)^T Qx(l)\right)+u(x(l))^T Ru(x(l))\right)+\phi(x(\infty))<\infty,\ \forall x(0)\in\Omega$

Theorem 3 [13]: Assume $u(x)\in\Omega_u$ is an admissible control law arbitrarily selected for the nonlinear discrete time system. If there exists a positive definite, uniformly convex, and continuously differentiable value function $V(x)$ on Ω satisfying the following

$$\begin{aligned}&\frac{1}{2}\left(f(x)+g(x)u(x)-x\right)^T\cdot\nabla^2V(x)\cdot\left(f(x)+g(x)u(x)-x\right)\\&+\nabla V(x)^T\left(f(x)+g(x)u(x)-x\right)+x^TQx+u(x)^TRu(x)=0\end{aligned} \tag{51}$$

$$V(x(\infty))=\phi(x(\infty)) \tag{52}$$

where $\nabla V(x)$ and $\nabla^2V(x)$ are the gradient vector and Hessian matrix of $V(x)$ respectively. Then $V(x(j))$ is the value function of the system defined in (49) for all $j=0,....\infty$ applying the feedback control $u(x)$ and $V(x(0))=J(x(0);u)$

To extend the capabilities of the controller an optimal control input for the bit-rate and transmission power system is introduced. Optimal solutions are often formulated in the linear regime and lack the abilities required for compensation of nonlinear dynamics. The proposed controller provides near-optimal control effort to the system via the nonlinear discrete-time GHJB equation given by

$$\frac{1}{2}\Delta x^T \nabla^2 V(x)\Delta x + \nabla V(x)^T \Delta x + x^T Q x + u(x)^T R u(x) = 0 \tag{53}$$

$$V(x)\big|_{x=0} = 0 \tag{54}$$

where $\Delta x = f(x) + g(x)u(x) - x$. The optimal control input is obtained as

$$u^{(i+1)}(x) = -\left[g(x)^T \nabla^2 V^{(i)}(x) g(x) + 2R\right]^{-1} g(x)^T \left(\nabla V^{(i)}(x) + \nabla^2 V^{(i)}(x)(f(x) - x)\right) \tag{55}$$

Next, a discussion of how the GHJB system can be used as an optimal control will be given for a general system, and then formulation of an optimal tracking control follows.

Theorem 4 (Improved Control) [13]: If $u^{(i)}(x) \in \Omega_u$ and $x(0) \in \Omega$, and the positive definite and convex function $V^{(i)}$ satisfies $GHJB\left(V^{(i)}, u^{(i)}\right) = 0$ with the boundary condition $V^{(i)}(0) = 0$, then the updated control function derived in (55) results in an admissible control for the system (49) on Ω. Moreover, if $V^{(i+1)}$ is the unique positive definite function satisfying $GHJB\left(V^{(i+1)}, u^{(i+1)}\right) = 0$, then

$$V^{(i+1)}(x(0)) \le V^{(i)}(x(0))$$

Theorem 5 (Convergence of Successive Approximations): Given an initial admissible control, $u^0(x) \in \Omega_u$ for the joint power and bit rate adaptation, by iteratively solving the GHJB equation (53) and updating the control function using (55), the sequence of solutions $V^i(x)$ will converge to the optimal HJB solution $V^*(x)$.

Proof [13]: It can be shown that after iteratively solving the GHJB equation and updating the control, the sequence of solutions $V^{(i)}$ is a decreasing sequence with a lower bound. Since $V^{(i)}$ is a positive definite function, $V^{(i)} > 0$ and $V^{(i+1)} - V^{(i)} \le 0$, the sequence of solutions $V^{(i)}$ will converge to a positive definite function $V^{(i)} = V^{(i+1)} = V^d$ when $i \to \infty$. Due to the uniqueness of solutions of the HJB equation [26], now it is shown that $V^d = V^*$. When $V^{(i)} = V^{(i+1)} = V^d$, from

$$\begin{aligned} &V^{(i+1)}(x(0)) - V^{(i)}(x(0)) = \\ &-\frac{1}{2}\sum_{k=0}^{\infty}\left(u_k{}^{(i+1)} - u_k{}^{(i)}\right)^T\left(g_k{}^T\nabla^2 V_k{}^i g_k + 2R\right)\left(u_k{}^{(i+1)} - u_k{}^{(i)}\right) \le 0 \end{aligned} \tag{56}$$

shows that $u^{(i)}(x) = u^{(i+1)}(x)$. Using (55) and taking $u^{(i)}(x) = u^{(i+1)}(x)$ f to get

$$u^{(i)}(x) = u^{(i+1)}(x) = -\left[g(x)^T\nabla^2 V^*(x)g(x) + 2R\right]^{-1} g(x)^T\left(\nabla V^*(x) + \nabla^2 V^*(x)(f(x) - x)\right) \tag{57}$$

The GHJB equation for $u^{(i)}$ can now be expressed as

$$\begin{aligned} &\frac{1}{2}\left(f(x) + g(x)u^{(i)}(x) - x\right)^T\nabla^2 V^{(i)}(x)\left(f(x) + g(x)u^{(i)}(x) - x\right) + \\ &\nabla V^{(i)}(x)^T\left(f(x) + g(x)u^{(i)}(x) - x\right) + x^T Qx + u^{(i)}(x)^T Ru^{(i)}(x) = 0 \end{aligned} \tag{58}$$

$$V^{(i)}(x)\big|_{x=0} = 0 \tag{59}$$

where (57) and (58) are nothing but the well-known HJB equation. This result implies that $V^{(i)} \to V^*$ and $u^{(i)} \to u^{(*)}$ as $i \to \infty$.

Remark: Consistency between GHJB method and Dynamic programming method: From the Dynamic programming principle [27], the optimal controller can be given as

$$u^{(*)}(x) = -\frac{1}{2}R^{-1}g(x)^T\frac{dV^*(x(l+1))}{dx(l+1)} \tag{60}$$

Using an optimal control input the solution for $V(x)$ is approximated by a neural network (NN). The NN is used to provide an approximation to $V(x)$ since no closed form

solution is available [13]. Additionally, other recursive approximation methods of the GHJB require a large amount of effort to solve. Therefore, a NN of the form

$$V_L(x) = \sum_{j=1}^{L} w_j \sigma_j(x) = W_L^T \bar{\sigma}_L(x) \tag{61}$$

is used to provide a solution where the activation function vector $\sigma_j(x): \Omega \to \Re$, is continuous, $\sigma_j(x)|_{x=0} = 0$ and the neural network weights are w_j and L is the number of hidden layer neurons. The vectors $\bar{\sigma}_L(x) \equiv [\sigma_1(x), \sigma_2(x), \cdots, \sigma_L(x)]^T$ and $W_L \equiv [w_1, w_2, \cdots, w_L]^T$ are the vector of activation function and NN weight matrix respectively. The neural network weights will be tuned to minimize the residual error in a least-squares sense within the stability region of the initial stabilizing control. A least squares solution [28] attains the lowest possible residual error with respect to the neural network weights.

For the $GHJB(V,u) = 0$, V is replaced by V_L having a residual error as

$$GHJB\left(V_L = \sum_{j=1}^{L} w_j \sigma_j, u\right) = e_L(x) \tag{62}$$

To find the least-squares solution, the method of weighted residuals is used [28]. The weights w_j are determined by projecting the residual error onto $\frac{\partial(e_L(x))}{\partial W_L}$ and setting the result to zero $\forall x \in \Omega$, i.e.

$$\left\langle \frac{\partial(e_L(x))}{\partial W_L}, e_L(x) \right\rangle = 0 \tag{63}$$

When expanded, the above equation becomes

$$\left\langle \nabla\bar{\sigma}_L \Delta x + \frac{1}{2}\Delta x^T \nabla^2 \bar{\sigma}_L \Delta x, \nabla\bar{\sigma}_L \Delta x + \frac{1}{2}\Delta x^T \nabla^2 \bar{\sigma}_L \Delta x \right\rangle W_L + \left\langle x^T Q x + u^T R u, \nabla\bar{\sigma}_L \Delta x + \frac{1}{2}\Delta x^T \nabla^2 \bar{\sigma}_L \Delta x \right\rangle = 0 \tag{64}$$

where $\nabla\bar{\sigma}_L = \left[\frac{\partial\sigma_1(x)}{\partial x}, \frac{\partial\sigma_2(x)}{\partial x}, \ldots, \frac{\partial\sigma_L(x)}{\partial x}\right]^T$ $\nabla^2\bar{\sigma}_L = \left[\frac{\partial^2\sigma_1(x)}{\partial x^2}, \frac{\partial^2\sigma_2(x)}{\partial x^2}, \ldots, \frac{\partial^2\sigma_L(x)}{\partial x^2}\right]^T$ and $\Delta x = f + gu - x$.

Remark: The assumption that control coefficients and the nonlinear $f(x)$ are known is a stringent assumption. For the case of rate control the control coefficients is explicitly known, in contrast to the power control where the interference must be estimated. Future work will concentrate on relaxing these assumptions.

Theorem 6(GHJB Tracking Control): Assume $u(e) \in \Omega_u$ is an admissible control law arbitrarily selected for the nonlinear discrete time system. If there exists a positive definite, uniformly convex, and continuously differentiable, value function $V(e)$ on Ω satisfying (65). Consider the modified GHJB equation for the rate and power dynamics using tracking error as

$$\frac{1}{2}\Delta e^T \nabla^2 V(e) \Delta e + \nabla V(e)^T \Delta e + e^T Q e + u(e)^T R u(e) = 0 \tag{65}$$

where $\Delta e = e(l+1) - e(l) = f_l + g_l u_l - x_d(l+1) - e(l)$. Define u_l* be the new optimal control law defined as

$$u_l^* = -\left[g_l^T \nabla^2 V_l^* g_l + 2R\right]^{-1} g_k^T \left(\nabla V_l^* + \nabla^2 V_l^* \left(f_l - x_d(l+1) - e(l)\right)\right). \tag{66}$$

where $f(x(l)) = \theta_i(l) x_i(l)$ and $g_l = B_i(l)$ as shown in previous section. Given an initial admissible control, $u^0(x) \in \Omega_u$ for the joint power and bit rate adaptation, by iteratively solving the GHJB equation (65) and updating the control function using (66), the sequence of solutions $V^i(x)$ will converge to the optimal HJB solution $V^*(x)$.

Proof: Using (53) and substitution of Δe for Δx a form of the GHJB that uses the error for derivation of control as shown in (66) is produced.

Using the form of the GHJB equation in (66), a new tracking control is realized and can be applied to the rate and power control dynamics. Results from both controllers are shown later via simulation.

IV. Results and Discussion

Simulations were carried out for random topologies with between 1 and 250 links, or 2 to 500 nodes, over a 50 by 50 m^2 area in order to evaluate the performance of the controller as placement, density, and congestion vary. Each number of links was simulated 5 times with random node placements and averaged results are presented.

First, to highlight the ability of the adaptive method to appropriately estimate congestion levels initial simulations were performed. The adaptive controller was tested with a varying congestion level to show tracking of the unknown parameters. A triangular pulse with a mean random noise was input as a congestion level to provide an example. The congestion levels and the estimated values are shown in Fig. 1. Results from Fig. 1 demonstrate that the control method is able to estimate and therefore compensate for varying congestion levels. This result is important since congestion levels can impact the QoS of the network.

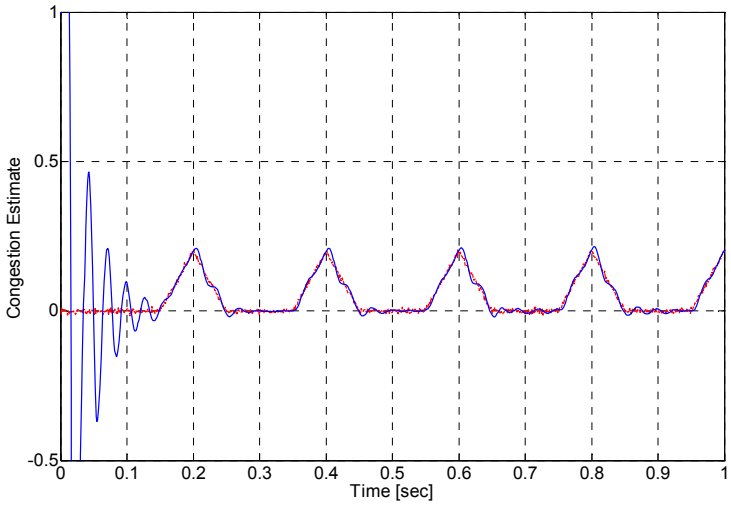

Fig. 1 Results of congestion estimation

For these simulations, a dynamic channel described in (2) was used. The channel is assumed to vary slower than the update of the controller system as assumed in the controller formulation. For each simulation a random channel state over time is generated so each run must overcome a unique channel sequence. A stochastic channel gain allows the performance of the controller to be evaluated in a more realistic manner. An example of a typical channel gain pattern is shown in Fig. 2.

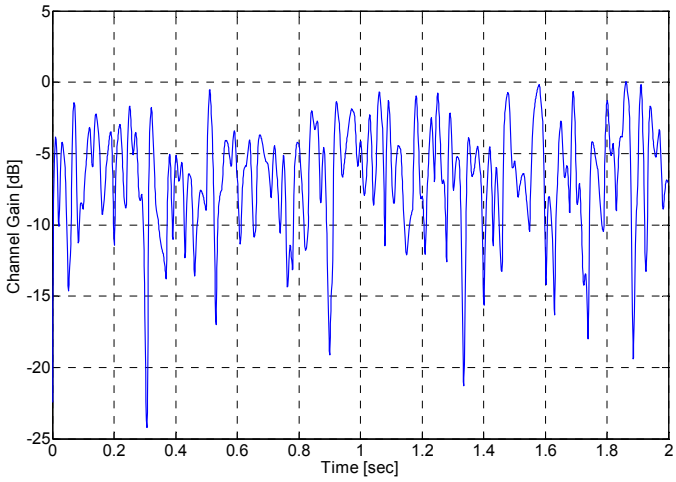

Fig. 2 Channel gain .vs. time

Simulations included several aspects to provide realism for testing of the control methods performance. These aspects included comparisons of continuous and discrete power levels, dropped bits, random channel noise, and varying congestion levels. Power levels for the network were simulated with continuous power levels to show the ability of the controller to provide efficiency. Next, simulations with discrete power levels were used to highlight energy consumption in a more realistic manner for embedded systems. The level of discretization of the power was varied to show performance changes as a continuous power method is approached. Additionally, a probabilistic dropping of bits was included to provide data loss due to channel conditions.

Results show that the joint adaptive controller and the GHJB method are capable of meeting QoS requirements as proposed in previous sections. The ADRPC and GHJB were simulated in MATLAB® with a dynamic channel, a probabilistic bit drop-rate, and with varying numbers of nodes in random topologies.

Results for the ADRPC and GHJB methods show that both are capable of controlling the bit-rate and SIR for each link. Both controllers provide suitable QoS by providing a desired bit-rate based on user-defined time of arrival. In Figs. 3 and 4, the controlled SIR levels can be seen for the ADRPC method and GHJB method respectively. The SIR is held higher to accommodate traffic flow during transfer and next falls to conserve power.

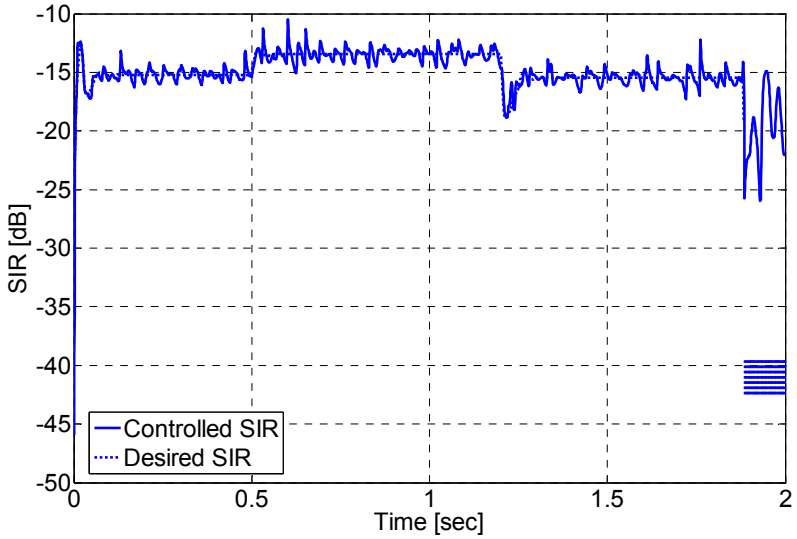

Fig. 3 ADRPC SIR levels

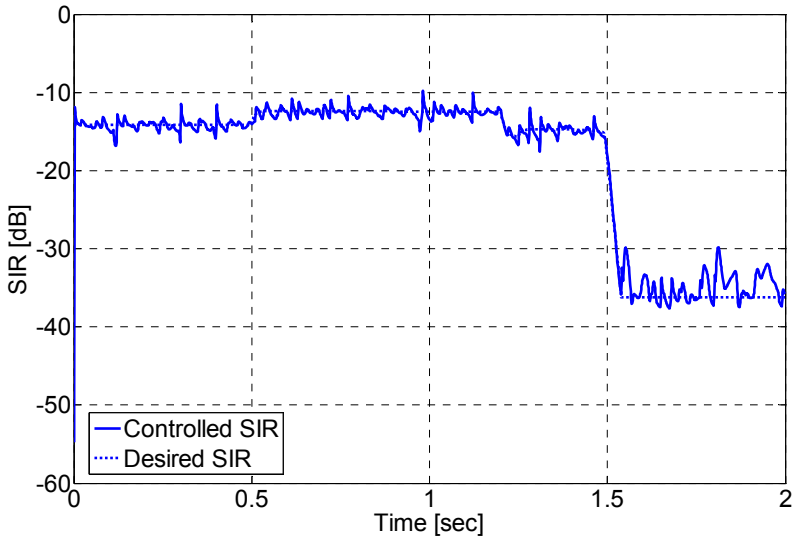

Fig. 4 GHJB SIR levels.

Figure 5 depicts the controlled and desired bit-rates for the ADRPC. The ADRPC method provides suitable control and delivers the desired payload in the time period set by the QoS. However, for the rate dynamics the GHJB method performs with slightly higher state error as shown in Fig. 6. The ability of the GHJB controller to provide an optimal control input allows for less overshoot and less control effort; however it causes a slight

increase in error. The increase in error can be tolerated however since the controller is still able to maintain the QoS levels while providing energy savings.

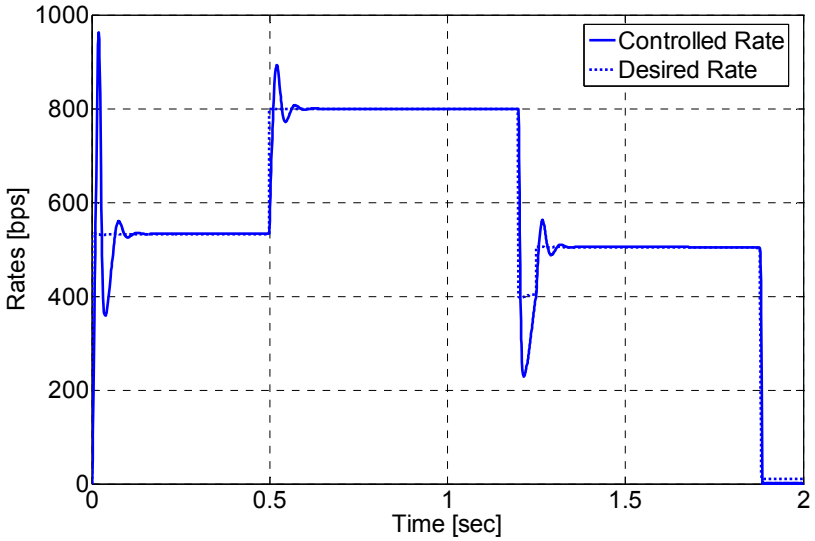

Fig. 5 ADRPC rates.

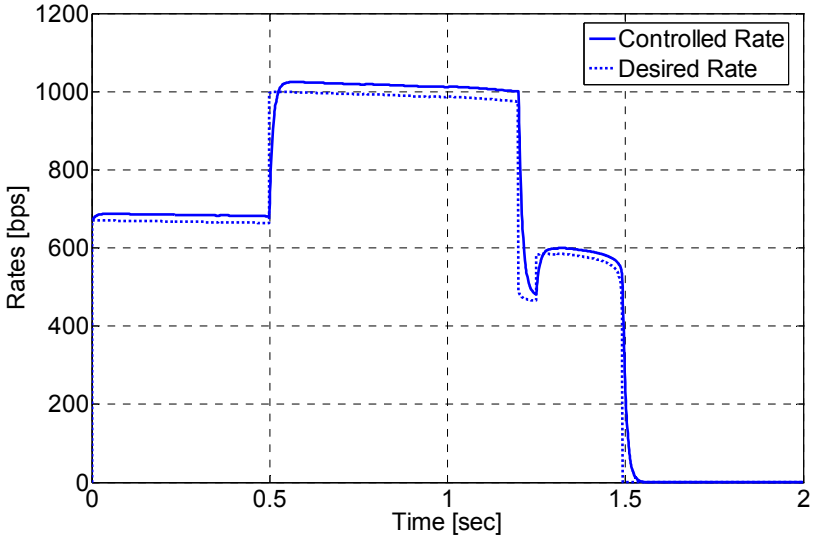

Fig. 6 GHJB rates

It is also observed that the power levels change to compensate for the channel gain; however, they do not stay at a higher level for long periods of time. Thus, the total energy expended to transmit data is minimized allowing for management of the channel capacity

while simultaneously conserving energy over the lifetime of the wireless device. Fig. 7 displays a representative example of the power levels; similar levels were found for each controller.

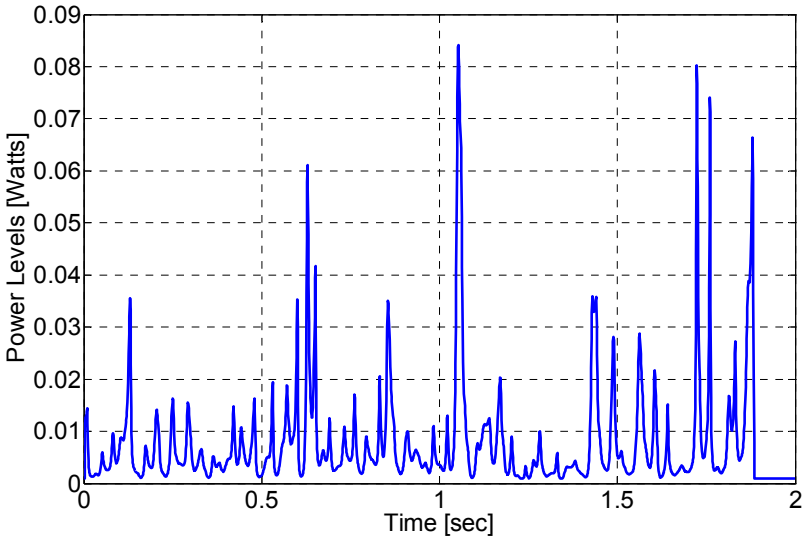

Fig. 7 Power levels

The impact of the performance of the network was next considered for discrete power levels. When the power levels are discretized, the performance of the control method suffers since one must choose the next most powerful setting to achieve the desired channel capacity. Therefore, the power settings were separated into levels of 5, 10, 100 and 1000 over a one Watt range. Table I shows how the SIR error is affected by the discretization of the power levels. If there are too few levels a node expends more energy than needed to meet QoS levels since it selects a level higher than needed.

TABLE I AVERAGE SIR ERROR WITH VARYING DISCRETIZED POWER LEVELS

	5 Levels	10 Levels	100 Levels	1000 Levels
ADPRC	0.4307	0.5889	0.0040	0.0008
GHJB	8.3515	1.8928	0.0054	0.0004

Next, a study of how network density affects both the controller error and the QoS of the data transfer was undertaken. In Tables II and III, these results are shown. Table II shows that the errors in the controller are statistically similar for all node densities considered. The adaptive controller provides tighter control over the bit-rate; however it does not provide substantial gains in the SIR control. Also noticed is a consistent error for the GHJB rate control; this error is likely due to the reduction of controller effort as contrasted to the adaptive method.

TABLE II ERRORS LEVELS WITH VARYING NUMBERS OF LINKS

	Error	1 Link	10 Links	25 Links	50 Links	100 Links	150 Links	200 Links	250 Links
ADPRC	Rate [bps]	-7.7456	-1.5701	-3.4432	-1.3666	-2.3766	-2.4001	-1.5095	-2.4538
	SIR [unitless]	0.3977	0.0005	-0.0109	-0.0274	-0.0279	-0.0306	-0.0310	-0.0318
GHJB	Rate [bps]	22.1870	21.2750	20.6330	20.6520	21.2270	20.7750	21.1590	21.1800
	SIR [unitless]	-0.0004	-0.0245	-0.0181	-0.0198	-0.0281	-0.0305	-0.0324	-0.0317

Similarly, the results in Table III show that both bit-rate controllers are capable of meeting the QoS metric based on the arrival times of the data. There are more consistent results using the GHJB optimal control. This increase in performance is due to the control effort of the adaptive method having overshoot and not providing the needed bit-rate as fast as the GHJB based optimal method. This result is seen by comparison of Figs. 5 and 6, which shows the controlled and desired bit-rates. The GHJB provides arrival times within 0.7 milli-seconds of the desired times as compared to the adaptive method.

TABLE III QOS ERROR WITH VARYING NUMBERS OF LINKS (UNITS IN SECONDS)

	1 Link	10 Links	25 Links	50 Links	100 Links	150 Links	200 Links	250 Links
ADPRC	6.78E-04	2.65E-04	6.42E-03	1.79E-03	3.55E-03	3.21E-03	2.20E-03	2.35E-03
GHJB	-1.78E-04	9.73E-04	7.38E-04	8.56E-04	7.97E-04	8.85E-04	9.12E-04	9.34E-04

The tighter control on the arrival times shown by the GHJB methods are attributed to the lower amount of overshoot and shorter settling time in the state trajectories as compared to the desired values. Therefore, the GHJB provides a tighter control on the QoS while still decreasing the overall effort of the controller.

Finally, a comparison of control effort was performed. The comparison was performed to investigate the advantage the GHJB tracking method provides in terms of power savings. Since energy-efficiency is of prime concern, conservation of control effort is also desirable. In Fig. 8 a comparison of the control effort of the adaptive and optimal methods is shown. The optimal method provides less control input however it delivers excellent performance when graded on the control of SIR and on the QoS criteria presented. Additionally, in Table IV a comparison of the normalized control energy over the simulation time is shown. The GHJB method exerts less control energy over time; thus increasing available energy and network lifetime.

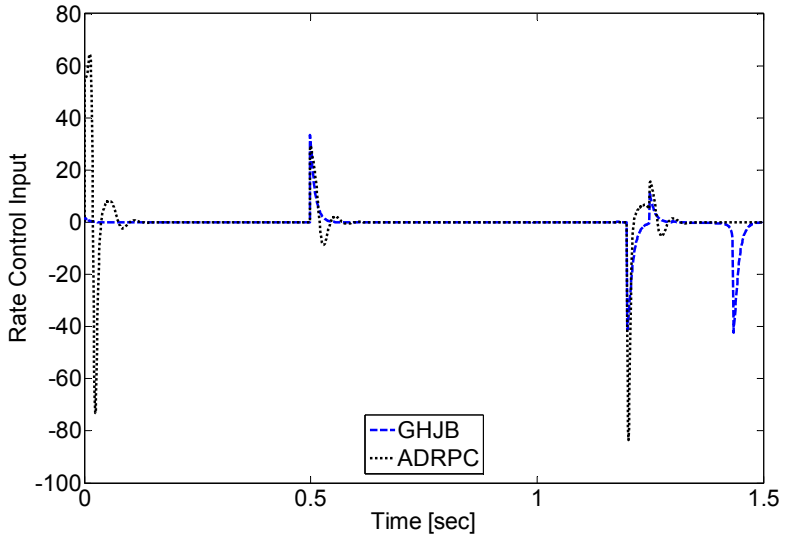

Fig. 8 Comparison of ADRPC and GHJB control inputs

TABLE IV CONTROL ENERGY COMPARISON

	Normalized Energy
ADPRC	100 %
GHJB	51.5 %

V. CONCLUSIONS

In this paper a novel adaptive controller and an optimal tracking controller for the joint control of bit-rate and transmission power for wireless networks are introduced. The adaptive controller approach uses Lyapunov-based stability analysis to provide analytically guaranteed convergence of the estimated parameters and asymptotic stability in terms of the error. Development of a GHJB tracking controller was also presented. This controller provides near-optimal control effort for the system dynamics and increases energy-efficiency. The combined bit-rate and power control provide a novel method for controlling the actual capacity of the channel for the current bit-rate being produced. By mathematically guaranteeing the channel capacity and estimating both the channel state and the congestion level in a distributed manner provide a robust protocol with little overhead.

Results from simulation show that the control methods presented provide adequate control with a minimal overhead. Analysis of the simulations reveals that the control methods perform well under dynamic channel conditions and with network congestion. Applying these controllers allows for power savings over other methods. Additionally, use of the proposed controllers overpowering of the channel is avoided to increase energy-efficiency and to decrease interference among nodes in the network. Finally, QoS is enhanced since the bit-rates are set by user perceived metrics and these are met by the controller through tracking a desired rate trajectory allowing for greater network flexibility and performance.

REFERENCES

[1] K.T. Kornegay, Q. Gang, and M. Potkonjak, "Quality of service and system design," *Proc. of the IEEE Computer Society Workshop on VLSI*, Orlando, FL, pp. 112-117, 1999.

[2] E. Babulak, "User perception of quality of service provision," *Proc. of the IEEE Int. Conf. on Industrial Technology*, vol. 2, pp. 1022-1025, 2003.

[3] N. Bambos, S. Chen and G. J. Pottie, "Channel access algorithms with active link protection for wireless communication networks with power control," *IEEE/ACM Trans. on Networking*, vol. 8, no. 5, pp.583-597, October 2000.

[4] S. Dontula and S. Jagannathan, "Active link protection for wireless peer-to-peer and cellular networks with power control," *Proc. of the World Wireless Congress*, pp. 612-617, May 2004.

[5] S. Jagannathan, A. T. Chronopoulos and S. Ponipireddy, "Distributed power control in wireless communication systems," *Proc. of the IEEE Int. Conf. on Computer Communications and Networks*, pp. 493-496, Nov. 2002.

[6] S. Jagannathan, M. Zawodniok and Q. Shang, "Distributed power control of cellular networks in the presence of channel uncertainties," *Proc. of the IEEE Conf. on Computer Communications (INFOCOM)*, vol.2, pp. 1055- 1066, March 2004.

[7] S-J. Park and R. Sivakumar, "Quantitative analysis of transmission power control in wireless ad-hoc networks," *Proc. of the IEEE Int. Conf. on Parallel Processing Workshops (ICPPW)*, pp. 56-63, 2002.

[8] E-S. Jung and N. H. Vaidya "A power control MAC protocol for ad hoc networks," *Proc. of the ACM/IEEE Int. Conf. on Mobile Computing and Networking (MobiCom)*, pp. 36-47, 2002.

[9] J. Gomez, A.T.Campbell, M.Naghshineh, and C.Bisdikian, "Conserving transmission power in wireless ad hoc networks," *Proc. of the IEEE Int. Conf. on Network Protocols (ICNP)*, pp. 24-34, 2001.

[10] P. Karn, "MACA – A new channel access method for packet radio," *Proc. ARRL Computer Networking Conf.*, pp. 134-140, 1990.

[11] M.B.Pursley, H.B. Russell, and J.S.Wysocarski, "Energy efficient transmission and routing protocols for wireless multiple-hop networks and spread-spectrum radios," *Proc. of the IEEE/AFCEA Conf. on Information Systems for Enhanced Public Safety and Security (EUROCOMM)*, Munich, Germany, pp. 1-5, 2000.

[12] M. Zawodniok and S. Jagannathan, "A distributed power control MAC protocol for wireless ad-hoc networks," *Proc. of the IEEE Conf. on Wireless Communications and Network Conference (WCNC),* vol. 3 pp. 1915-1920, 2004.

[13] Z. Chen and S. Jagannathan, "Generalized Hamilton-Jacobi-Bellman formulation-based neural network control of affine nonlinear discrete-time systems," *IEEE Tran. on Neural Networks,* vol 10, no. 1, pp. 90-106, 2008.

[14] L. T. Ong, A. Olfat, and M. Shikh-Habaei, "Joint SNR target and rate adaptation based on bit error rate estimation," *Proc. of Int. Symp. on Wireless Communication Systems*, pp. 470-474, 2005.

[15] Y.E. Sagduyu, and A. Ephriemides, "Power and rate adaptation as stochastic games for random access," *Proc. of the IEEE Conf. on Decision and Control (CDC)*, vol. 4, pp. 4202-4207, 2003.

[16] A. K. Karmokar, D. V Djonin, and V. K. Bhargava, "Delay constrained rate and power adaptation over correlated fading channels," *Proc. of the IEEE Global Communications Conf. (GLOBECOM)*, vol. 6, pp. 3448-3453, 2004.

[17] C-H. Jiang, and J-K. Lain, "Adaptive neuro-fuzzy power control and rate adaptation for multirate CDMA radio systems," *Proc. of the IEEE Int. Conf. on Networking, Sensing, and Control*, pp. 1307-1312, 2004.

[18] M. Shikh-Bahaei, M. Mouna-Kingue, and G. Charbit, "Joint optimisation of outer-loop power control and rate adaptation over fading channels," *Proc. of the IEEE Vehicular Technology Conf. (VTC-Fall)*, vol. 3, pp. 2205-2209, 2004.

[19] S. A. Jafar, and A. Goldsmith, "Optimal rate and power adaptation for multirate CDMA," *Proc. of the IEEE Vehicular Technology Conf. (VTS-Fall)*, Boston, MA, vol. 3, pp. 994-1000, 2000.

[20] A. Subramanian and A. H. Sayed, "Joint rate and power control algorithms for wireless networks," *IEEE Trans. on Signal Processing*, vol. 53, no. 11, pp. 4204-4214, 2005.

[21] T. S. Rappaport, Wireless communications, principles and practices, Prentice Hall, New Jersey, 1999.

[22] R. Canchi and Y. Akaiwa, "Performance of adaptive transmit power control in $\pi/4$ DQPSK mobile radio systems in flat Rayleigh fading channels," *Proc. of the IEEE Vehicular Technology Conf. (VTC)*, Houston, TX, vol.2, pp.1261-1265, 1999.

[23] W. Stallings, *Wireless Communication and Networks*, Prentice Hall, New Jersey, p. 144, 2002.

[24] R. Jantti and S. L. Kim, "Second-order power control with asymptotically fast convergence," *IEEE Journal on Selected Areas in Communications*, vol.18, no. 3, 2000.

[25] S. Jagannathan, Discrete-time Neural Network Control of Nonlinear Systems, Taylor & Francis, London, UK, 2006.

[26] G.N. Saridis,and C. S. Lee, "An approximation theory of optimal control for trainable manipulators," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 9, no. 3, pp. 152-159, 1979.

[27] D.P Bertsekas, and J. N. Tsitsiklis, *Neuro-Dynamic Programming,* Athena Scientific, Belmont, MA, 1996.

[28] B.A. Finlayson, The Method of Weighted Residuals and Variational Principles, Academic Press, New York, NY, 1972.

PAPER 4

Fault Detection and Prognostics Using NN/RISE Observer with Applications to Civil Infrastructure Monitoring

James W. Fonda[1,2,*], S. Jagannathan[1], and Steve E. Watkins2
Embedded Systems and Networking Laboratory[1]
Applied Optics Laboratory[2]
Department of Electrical and Computer Engineering
Missouri University of Science and Technology, Rolla, Missouri 65409-0040
jfonda@ieee.org, sarangap@mst.edu, and steve.e.watkins@ieee.org

Abstract—A framework for fault detection and prognostics in continuous-time is introduced. The proposed method uses a novel neural network (NN) observer formulated with the robust integral sign of the error (RISE) method allowing for semi-global asymptotic stability in the presence of approximation errors, disturbances and unmodeled dynamics. Additionally, the combination of NN and the RISE method provides a robust observer, which renders better accuracy in the estimated states, over other NN observers in the presence of noise. The observer NN weights are tuned online without any offline learning phase. Next, by using linearity in the unknown parameter assumption on certain uncertain nonlinear terms, which is standard in adaptive control, parameter tuning law for the observer is utilized to provide the time-to-failure (TTF) prediction using known thresholds on key parameters. Use of adaptive methods for estimation of physical parameters aids in the scheduling of maintenance, provides intuitive inspection, and detects alarm levels.

Application examples for both of the proposed methods are shown through simulation. First a simple nonlinear system for performance testing of the NN/RISE observer system is considered where noise levels are added to the sensor measurements and performance is compared to a traditional NN observer. Civil infrastructure examples are shown using both a simply supported beam under cyclic loading and a levee wall subject to foundation failure. Both examples are formulated using FEA methods and used to simulate failures to provide example data for the observer and prognostics methods. Results show that the proposed NN/RISE method provides as much as 17% increased accuracy in the state estimation compared to a NN observer. Consequently, TTF prediction can be obtained accurately due to robust observer accuracy.

Index Terms- Neural Networks, Fault Detection, Prognostics, Robust Observer, Wireless Sensor Networks

---

Research supported in part by Dept. of Education GAANN Fellowship, NSF I/UCRC on Intelligent Maintenance Systems, Air Force Research Laboratory Grant (FA8650-04-C-704) and Intelligent Systems Center.

I. Introduction

Maintenance is critical to the long term usability and stability of systems. In order to concentrate effort on failing components and limit expenditure, the use of condition-based and predictive maintenance methods is desirable. In large-scale systems, such as civil infrastructure, methods that provide distributed diagnostics and prognostics capabilities also enhance maintenance needs. Components such as bridges, levees, and buildings are susceptible to faults that are not easily inspected, or readily identified by traditional methods such as visual inspection, and can go unnoticed. Due to reliability concerns and the desire for early detection of incipient failures practical methods that can be used with wireless sensor networks (WSN) for deployment of distributed prognostics are required.

Monitoring of components is desirable when employed over long periods of time; however, over time sensor noise levels change and the accuracy of state estimation methods can suffer. Therefore, robust observer methods that resist noise levels are desirable. Previous methods have addressed these needs for observability through adaptive estimation [1] and robust observers [2-5]. Observers act as virtual sensors when the states are difficult to measure. Since many observers are for control purposes accurate state estimation in the face of uncertainties is required to provide adequate basis for calculation of control inputs. In the case of fault detection, the observer, or on-line approximator (OLA), is used to produce system states for the purpose of detection. Therefore, just as in a controls application the estimation of these states must be as accurate as possible to determine changes in the health of a system. Moreover, if these states have to be utilized for prediction, then accurate estimation of the states is a key requirement. The OLA methods combine observer formulations with a learning scheme

such as adaptive [6] methods in order to detect changes in system behavior beyond a particular threshold. The OLA's are typically used for fault detection and characterization during the progression of fault dynamics in a nonlinear system.

Prognostics are used to provide predictive services for maintenance applications. Often, understanding how long a component will survive before replacement is desirable to performance maintenance. In large-scale systems however, the interaction of failing components can also lead to catastrophic failures. Existing prognostic methods rely on either rule-based or statistical methods [7-10]. While these methods are adequate for many applications some require more certainty, or ability, to provide prediction and fault detection. Therefore, physical parameters and their impact on either the stability, or the performance, of a system are required to be estimated and a prediction method utilized to determine when a particular parameter will meet a threshold value. Many systems have critical values for parameters, such as the stiffness for a structural system, which can be exploited to determine fault occurrence.

Additionally, the ubiquitous use of embedded systems in predictive maintenance applications provides a growing shift to the network enabled manufacturing (NEM) paradigm. NEM methods provide visibility and advanced application spaces to large-scale systems where incipient failures at the component level can impact a large number components. Mitigation of failures and prediction of time-to-failure (TTF) at the component level can result in cost savings and focused maintenance. These methods can also be applied to provide visibility to aging and failing infrastructure. Additionally, NEM methods allow distributed data collection services that employ small, low-power, low-cost networked processing platforms [11]. Commonly referred to as motes, these

devices provide networked and observable systems. As these devices become ubiquitous, the techniques for maintenance must evolve to meet the need for the increased visibility and processing capabilities they provide. Therefore, algorithms for robust monitoring and prognostics that can be used with embedded devices such as WSN platforms, or motes, are desired.

First a method for robust observers as an OLA using a neural network (NN) with RISE feedback is proposed. This method provides increased robustness for observers and allows for increased performance in practical deployments where uncertainties are introduced. The observer is designed to decrease sensitivity to noise levels thus providing robustness. The observer is presented in the context of an OLA for fault detection as developed in [6]; however, formulation is presented in a general manner to support a variety of applications. Additionally, NN's are often applied as the OLA providing little insight to physical parameters and their impact on system health. In this work, an adaptive observer for use in conjunction with the NN/RISE methods to address this issue is also formulated. Additionally, a method for parameter estimation and prediction for prognostics is developed to provide estimations of TTF of components while making the linearity in the unknown parameter assumption on certain system nonlinearities.

This method extends the parameter estimation properties of the OLA and adaptive observer to predict when parameters will reach critical values. These critical parameter values are set by physically meaningful thresholds in contrast with use of states since they may not represent physical parameters. Next, a simulation study of the performance gains the NN/RISE method is explored using a simple nonlinear system. Finally, finite element analysis (FEA) methods are applied to the problems of cyclic failure of a simply

supported beam and erosion failure of a levee wall to evaluate the performance of the prognostics method.

II. Observer Formulation

Detection of fault relies on the accurate estimation of system states based on directly measurable state information. To provide an observer capable of detection of faults and characterization of fault dynamics, and to enable prognostics, robust observers are desirable. In this section, a discussion of the requirements, the formulation, and the performance of such an observer is discussed. The proposed observer employs a NN for learning and provides robustness via RISE feedback [12].

Observer performance in the presence of nonidealities, or its robustness, is of concern related to its impact on the error levels. To maintain accuracy while noise is present allows for the detection of failures even when sensor measurements are non-ideal. In many applications there is no guarantee of the magnitude or statistical variation of measurement noise. Therefore, robust methods for estimation of states in the presence of sensor noise and unmodeled dynamics are desirable.

Many observer systems provide stability in the mean, and this level of performance is acceptable on many systems; however, systems require enhanced performance where sensor noise is unavoidable and degradation of the performance introduces risk require enhanced performance. In performance-oriented applications, robust observer methods provide more confidence in state estimates allowing for reliability when those estimates are utilized for decision and/or control.

In applications of robust methods, such as in controls, added “effort” of the control input is an issue and can lead to actuator saturation. Saturation in digital and analog

systems is a well-known phenomenon and can cause limitation in performance. However, use of embedded digital platforms for these applications relaxes many of these issues if they are of sufficient resolution. These issues must be considered at the time of realization due to the physical demand of the system on the observer.

In this section a state observer that satisfies the above requirements is developed. The method extends the fault detection work presented in Demetriou et. al [6], and by Xian et. al [12] by using an innovative NN/RISE observer structure which can support prognostics. While the RISE method [12, 13] has been considered for improvement in control applications, to date, no observer formulations have been shown in the literature that utilizes RISE feedback. Work presented by Demetriou et. al [6], and by Xian et. al [12] encompassed methods for fault detection and detection using NN's respectively. These methods need to be extended for fault detection and prediction to provide robustness in the field although a robust control system can hide the occurrence of faults, especially a small and slowly progressing incipient failure. However, the use of RISE method along with OLA provides robustness which makes the estimation more robust in the presence of approximation errors, noise and disturbances thus rendering more accurate prediction. As a consequence, the issue of missing incipient faults by a robust control system is mitigated through the proposed work.

While work presented in Demetriou et al. [6] provides a suitable framework for detection of fault dynamics through the use of an adaptive agent (i.e. adaptive parameter estimation, NN's, and others) it does not extend into the prognosis of critical dynamic behavior. In contrast, authors have demonstrated how the OLA based framework can be extended to deliver prognostic capabilities.

A. System Dynamics

Consider a class of nonlinear systems that can be described as

$$\begin{aligned} \dot{x} &= \xi(x,u) + B(t-T)f(x,u) \\ y &= h(x) \end{aligned} \tag{1}$$

where

$$B(t-T) = diag\{\beta_1(t-T), \beta_2(t-T), ..., \beta_n(t-T)\} \tag{2}$$

$$\beta_i(\tau) = \begin{cases} 0, & if\ \tau < 0 \\ 1 - e^{-\rho_i \tau}, & if\ \tau \geq 0 \end{cases} \quad i = 1,2,...,n \tag{3}$$

and the output, y , is a selection of measurable states. The system description (1) includes the nominal system dynamics, $\xi(x,u)$ and the unknown failure dynamics $B(t-T)f(x,u)$. The failure term is assumed to be an unknown smooth function where the failure begins at a time T . The failure dynamics are assumed to have the form shown in the second term of (2), where the $f(x,u)$ term comprises the actual fault dynamics to be isolated, and the $B(t-T)$ term is a smooth function providing continuity between nominal and faulty dynamic behavior. The exponential terms in β_i , as defined in (3), that includes unknown rates of onset, ρ_i , for the faults. The nominal system dynamics are assumed to be sufficiently known and to include unmodeled dynamics as presented in assumption ***A1***.

A1) The nominal system dynamics may be expressed as a combination of known and unknown dynamics such that

$$\dot{x}_n = \xi(x_n(t), u(t)) \equiv \xi^*(x_n(t), u(t)) + \tilde{\xi}(x_n(t), u(t)) \tag{4}$$

where ξ^* represents the known nominal dynamics of the healthy system and $\tilde{\xi}(\cdot)$ represents unmodeled dynamics [6].

A2) There exists compact sets $\chi \subset \Re^n, U \subset \Re^m$ such that $x(t) \in \chi$ and $u(t) \in U$ for all $t \geq 0$. Additionally, the inputs and states are bounded prior to and after the occurrence of the fault.

A standard assumption is made as to the boundedness of both the states and the inputs to the system in Assumption ***A2***.

Remark 1: As $t \to \infty$ then $\beta_i \to 1$ and $\dot{\beta}_i \to 0$ and, therefore the dynamics will converge to the form $\dot{x} = \xi(x,u) + f(x,u)$ in a finite amount of time after the onset of a failure. The case given in (1) is considered for the development of the observer, but this important feature should be noted. Also, on exploration of the failure dynamics, note that the derivative of $B(t-T)$ with respect to time yields $\dot{B}(t-T) = QB(t-T)$, where Q is a diagonal matrix consisting of the rate of failure ρ_i. This matrix is upper-bounded by the physics of the particular system being observed given as the largest singular value of $Q_{\max}$.

B. Nonlinear Observer with RISE

An observer for a class of systems described in (1) can be constructed as

$$\dot{\hat{x}} = \xi^*(x,u) + \hat{g}(x,u) + \mu(t) \tag{5}$$

where $\hat{x} \in \Re^n$ is the estimated state vector, $\hat{g}$ is the output of an OLA used to characterize the unknown fault dynamics and $\mu(t)$ is the RISE term added to provide robustness. It should be noted that the normal proportional term in (5) is missing as compared to [6], since it is included in the RISE feedback as

$$\mu(t) \equiv (k_s + 1)e_1(t) - (k_s + 1)e_1(0) + \int_0^t \left[(k_s + 1)\alpha(\tau)e_1(\tau) + (\beta(\tau) + 1)\text{sgn}(e_1(\tau))\right]d\tau \tag{6}$$

where $k_s, \alpha(t), \beta(t)$ are positive control gains, and $\text{sgn}(\cdot)$ is the standard signum function. Also, the derivative of $\mu(t)$ is shown to be

$$\dot{\mu}(t) = (k_s + 1)r + \left(\beta(t) + 1\right)\text{sgn}(e_1(\tau)) \tag{7}$$

The RISE input introduces robustness to noise and unmodeled dynamics; both of which normally restrict performance to boundedness in the stability analysis [14]. With the RISE feedback, reaching a semi-global asymptotically stable (AS) over a defined region of attraction is possible. This result is shown in the following sections through the stability and performance analysis using Lyapunov methods. Also, the estimation of the fault dynamics is performed with an online approximator such as a NN in this work, similar to the online approximation [6] based scheme. For the scope of this work, the construction of the observer includes a NN or any online approximator as the approximation agent, and the RISE as the robustifying term.

The *universal approximation property* (UAP) for NN [14] states that for any smooth function $F(x,u)$ there exists a NN with ideal weights θ and V such that $F(x,u) = \theta^T\left(\sigma\left(V^T z\right)\right) + \varepsilon$ where ε is the NN functional approximation error and $\sigma(\circ): \Re_a \to \Re_b$ is the activation function in the hidden layers. Also, the approximation error is considered bounded in a compact set such that $\left\|\varepsilon(t)\right\| \le \varepsilon_m$ and the hidden layer to the output layer NN weights are tuned whereas the input to the hidden-layer weights are selected initially at random and held constant then according to [15, 16], the activation function vector forms a basis. Consequently, the two-layer NN possesses the UAP and is considered as a nonlinear approximator. For a full treatment of the properties of NN's please refer to [14].

C. Stability and Performance Analysis

The observer presented in the previous section provides suitable characteristics in terms of performance and fault detection and is now analyzed for stability and performance. Lyapunov techniques are applied to demonstrate that non-ideal systems with disturbances and estimation errors can be shown to be semi-globally asymptotically stable (AS) over a region by inclusion of the RISE feedback. In contrast, existing framework for incipient fault detection [6] to be AS only under ideal conditions. The use of RISE feedback relaxes this condition and provides greater flexibility to applications where measurement noise or unmodeled dynamics exist. The capability to provide accurate state estimates under non-ideal conditions provides a practical method that can be applied to a wide range of applications, including fault detection. In this work, contrasts of performance to traditional NN observers are discussed as well as providing application examples. Now the stability analysis of the proposed observer is shown.

To ease analysis a general function encompassing the unknown fault dynamics is defined as.

$$g(x,u) \equiv B(t-T)f(x,u) \tag{8}$$

The observer state error is defined as

$$e_1 = x - \hat{x} \tag{9}$$

Now, differentiating the error dynamics (9) yields

$$\dot{e}_1 = \tilde{\xi}(x,u) - \mu(t) + B(t-T)f(x,u) - \hat{g}(x,u) = g(x,u) - \hat{g}(x,u) - \mu(t) + \tilde{\xi}(x,u) \tag{10}$$

Next, the filtered tracking error (FTE) is defined as

$$r = \dot{e}_1 + \alpha(t)e_1 \rightarrow \dot{e}_1 = r - \alpha(t)e_1 \tag{11}$$

where

$$\alpha(t) = \alpha_0 + \alpha_1(t) \tag{12}$$

with $\alpha(t)$ is defined in (6). It should be noted at this time that $\alpha(t)$ consists of a constant and time variant term as shown in (12). To bring in the dynamics, the derivative of the FTE is taken as

$$\dot{r} = \ddot{e}_1 + \alpha(t)\dot{e}_1 = \ddot{e}_1 + \alpha(t)\dot{e}_1 + \dot{\alpha}(t)e_1 = \ddot{e}_1 + \tilde{N} \tag{13}$$

where $\widetilde{N}$ is defined as the error in the immeasurable nonlinear terms in reference to a desired set of dynamics given by

$$\widetilde{N} \equiv \alpha(t)\dot{e}_1 + \dot{\alpha}(t)e_1 \tag{14}$$

with the function N includes immeasurable nonlinear terms in the system dynamics.

***Remark 2*:** Since $\widetilde{N}$ is defined as a continuously differentiable function and it can be shown that $\widetilde{N}$ can be upper bounded as

$$\left\|\widetilde{N}\right\| \le \Omega\left(\|b\|\right)\|b\| \tag{15}$$

where $\|\circ\|$ denotes the Euclidian norm and $b(t)$ is defined as $b = \begin{bmatrix} e_1 & r \end{bmatrix}$ for some $\Omega : \Re_{\ge 0} \to \Re_{\ge 0}$ that is globally invertible and a nondecreasing function.

The NN UAP is now used to estimate the $\dot{g}(x,u)$ as

$$\dot{g}(x,u) = \theta^T\left(\sigma\left(V^T z\right)\right) + \varepsilon \tag{16}$$

Additionally, the NN is placed in the form of

$$\hat{\dot{g}}(x,u) \equiv \hat{\theta}^T\left(\sigma\left(V^T z\right)\right) \tag{17}$$

where the z is a vector of inputs to the hidden layer $z = \begin{bmatrix} \hat{x} & y & e_1 \end{bmatrix}$ including the observer states, $\hat{x}$, the measured outputs of the system, y, and the state error, e_1. A common assumption in the literature is to approximate the nonlinear function whereas in this paper, the derivative of the nonlinear function is approximated.

Next, the derivative of the FTE is used by incorporating the system and fault dynamics along with $\dot{\mu}(t)$ as

$$\begin{aligned} &\dot{r} = \ddot{e}_1 + N = \dot{g}(x,u) - \dot{\hat{g}}(x,u) + \dot{\tilde{\xi}}(x,u) - \dot{\mu}(t) + N = \\ &\dot{g}(x,u) - \dot{\hat{g}}(x,u) + \dot{\tilde{\xi}}(x,u) - \dot{\mu}(t) + \widetilde{N} + N_d = \\ &\theta^T\left(\sigma\left(V^T z\right)\right) + \varepsilon - \hat{\theta}^T\left(\sigma\left(V^T z\right)\right) + \varepsilon + \dot{\tilde{\xi}}(x,u) - \dot{\mu}(t) + \widetilde{N} = \\ &\widetilde{\theta}^T\left(\sigma\left(V^T z\right)\right) + \varepsilon + \dot{\tilde{\xi}}(x,u) - \dot{\mu}(t) + \widetilde{N} = \\ &\widetilde{\theta}^T\left(\sigma\left(V^T z\right)\right) + \varepsilon_2 - \dot{\mu}(t) + \widetilde{N} \end{aligned} \tag{19}$$

where the unmodeled dynamics include

$$\varepsilon_2 = \varepsilon + \dot{\tilde{\xi}}(x,u) \tag{20}$$

Before proceeding Lemma 1 is used to aid in the forthcoming stability analysis and to be invoked later in the stability analysis.

Lemma 1: Let an auxiliary function $P(t) \in \Re$ be defined as

$$P(t) \equiv \zeta_b - \int_0^t L(t)\,d\tau \tag{21}$$

where $L(t) \in \Re$ is defined as

$$L \equiv r\left(N_1 - \beta(t)\,\mathrm{sgn}(e_1)\right) - \dot{\beta}(t)\|e_1\| - e_1\widetilde{\theta}^T\sigma\left(V^T z\right) \tag{22}$$

with $N_1 = \varepsilon_2 + \tilde{\theta}^T \sigma\left(V^T z\right)$ and ζ_b a positive constant. Note that $\dot{P}(t) = -L(t)$. The auxiliary function requires that $\beta(t)$ be selected such that

$$\beta(t) \geq \left\|N_1\right\|_{L_\infty} + \frac{1}{\alpha(t)}\left\|\dot{N}_1\right\|_{L_\infty} \tag{23}$$

where $\left\|\cdot\right\|_{L_\infty}$ is the L_∞ norm, then the auxiliary function can be shown to be bounded as shown in (24) where $\alpha(t) > 1$. Through selection of $\alpha(t) > 1$, it can be shown that $L(t)$ is bounded such that

$$\int_0^t L(t)d\tau \leq \zeta_b \tag{24}$$

where the positive constant is given by

$$\zeta_b = \beta(0)\left\|e_1(0)\right\| - e_1(0)N_1(0) \tag{25}$$

Proof: See Appendix A

After showing $P(t)$ is positive-definite, the stability analysis of the fault error dynamics can be shown in Theorem 1.

Theorem 1 (Observer Asymptotic Performance Guarantee): Given the system dynamics of (10) let the NN/RISE observer be applied to estimate the system states. If the $\beta(t)$ and $\alpha(t)$ functions are selected from Lemma 1, with k_s defined as

$$k_s > \frac{\Omega^2\left(\left\|b\right\|\right)}{4\lambda} \tag{26}$$

and by taking the hidden-to the-output-layer weight update law as

$$\dot{\hat{\theta}} = \Gamma\sigma\left(V^T z\right)e_1 \tag{27}$$

then the NN/RISE observer provides semi-global asymptotically stable (AS) performance over a domain of

$$D \equiv \left\{ q \in \Re^n \times \Re_{\geq 0} \mid \|q\| < \Omega^{-1}\left(2\sqrt{\lambda k_s}\right) \right\} \tag{28}$$

with the NN weights remaining bounded.

Proof: A Lyapunov candidate can be defined as

$$v = \frac{1}{2}e_1^2 + \frac{1}{2}r^2 + P + \frac{1}{2}tr\left\{\tilde{\theta}^T \Gamma^{-1} \tilde{\theta}\right\} \tag{29}$$

Taking the derivative of (29), collecting terms, substitution of $\dot{r}$ and with substitution for $\dot{e}_1$ from (11) to get

$$\begin{aligned} \dot{v} = e_1\dot{e}_1 + r\dot{r} + \dot{P} + tr\left\{\tilde{\theta}^T \Gamma^{-1} \dot{\tilde{\theta}}\right\} = e_1\left(r - \alpha(t)e_1\right) + r\dot{r} - L + tr\left\{\tilde{\theta}^T \Gamma^{-1} \dot{\tilde{\theta}}\right\} = \\ e_1 r - \alpha(t)e_1^2 + r\dot{r} - L + tr\left\{\tilde{\theta}^T \Gamma^{-1} \dot{\tilde{\theta}}\right\} \end{aligned} \tag{30}$$

Now substitute the FTE dynamics into the Lyapunov system to arrive at

$$\dot{v} = e_1 r - \alpha(t)e_1^2 + r\left(\tilde{\theta}^T\left(\sigma\left(V^T z\right)\right) + \varepsilon_2 - \dot{\mu}(t) + \tilde{N}\right) - L + tr\left\{\tilde{\theta}^T \Gamma^{-1} \dot{\tilde{\theta}}\right\} \tag{31}$$

Substitution of $L(t)$ into (31) and simplification results in

$$\begin{aligned} \dot{v} = e_1 r - \alpha(t)e_1^2 + r\left(\tilde{\theta}^T\left(\sigma\left(V^T z\right)\right) - (k_s + 1)r - \left(\beta(t) + 1\right)\mathrm{sgn}(e_1) + \varepsilon_2 + \tilde{N}\right) - \\ r\left(\varepsilon_2 + \tilde{\theta}^T\sigma\left(V^T z\right) - \beta(t)\,\mathrm{sgn}(e_1)\right) + \dot{\beta}(t)\|e_1\| + e_1\tilde{\theta}^T\sigma\left(V^T z\right) + tr\left\{\tilde{\theta}^T \Gamma^{-1} \dot{\tilde{\theta}}\right\} \end{aligned} \tag{32}$$

or

$$\dot{v} = e_1 r - \alpha(t)e_1^2 - (k_s + 1)r^2 - r\tilde{N} + \dot{\beta}(t)\|e_1\| + r\,\mathrm{sgn}(e_1) + e_1\tilde{\theta}^T\sigma\left(V^T z\right) + tr\left\{\tilde{\theta}^T \Gamma^{-1} \dot{\tilde{\theta}}\right\} \tag{33}$$

Application of $ab \leq \frac{1}{2}\left(a^2 + b^2\right)$ in (33) to simplify the $e_1 r$ term yields

$$\dot{v} \le \frac{1}{2}e_1^2 + \frac{1}{2}r^2 - \alpha(t)e_1^2 - r^2 - k_s r^2 - r\widetilde{N} + \dot{\beta}(t)\|e_1\| + r\,\mathrm{sgn}(e_1) + e_1\widetilde{\theta}^T\sigma\left(V^T z\right) + tr\left\{\widetilde{\theta}^T\Gamma^{-1}\dot{\widetilde{\theta}}\right\} \tag{34}$$

Collecting terms to get

$$\dot{v} \le \left(\frac{1}{2} - \alpha(t)\right)e_1^2 - \frac{1}{2}r^2 - k_s r^2 - r\widetilde{N} + \dot{\beta}(t)\|e_1\| + r\,\mathrm{sgn}(e_1) + e_1\widetilde{\theta}^T\sigma\left(V^T z\right) + tr\left\{\widetilde{\theta}^T\Gamma^{-1}\dot{\widetilde{\theta}}\right\} = \left(\frac{1}{2} - \alpha(t)\right)e_1^2 - \frac{1}{2}r^2 - k_s r^2 - r\widetilde{N} + \dot{\beta}(t)\|e_1\| + r\,\mathrm{sgn}(e_1) + tr\left\{\widetilde{\theta}^T\left(\Gamma^{-1}\dot{\widetilde{\theta}} - \sigma\left(V^T z\right)e_1\right)\right\} \tag{35}$$

From this result $\alpha(t) > 1/2$ to provide stability; however, from Lemma 1 $\alpha(t) > 1$. Therefore selecting the condition from Lemma 1 will satisfy both conditions. The first derivative of the Lyapunov function can be written as

$$\dot{v} \le \left(\frac{1}{2} - \alpha(t)\right)e_1^2 - \frac{1}{2}r^2 - k_s r^2 - r\widetilde{N} + \dot{\beta}(t)\|e_1\| + r\,\mathrm{sgn}(e_1) + tr\left\{\widetilde{\theta}^T\left(-\sigma\left(V^T z\right)e_1\sigma\left(V^T z\right)e_1\right)\right\} = \left(\frac{1}{2} - \alpha(t)\right)e_1^2 - \frac{1}{2}r^2 - k_s r^2 - r\widetilde{N} + \dot{\beta}(t)\|e_1\| + r\,\mathrm{sgn}(e_1) \tag{36}$$

Expansion of $r\,\mathrm{sgn}(e_1)$ yields

$$\dot{v} \le \left(\frac{1}{2} - \alpha(t)\right)e_1^2 - \frac{1}{2}r^2 - k_s r^2 - r\widetilde{N} + \dot{\beta}(t)\|e_1\| + \dot{e}_1\,\mathrm{sgn}(e_1) + e_1\,\mathrm{sgn}(e_1) = \left(\frac{1}{2} - \alpha(t)\right)e_1^2 - \frac{1}{2}r^2 - k_s r^2 - r\widetilde{N} + \dot{\beta}(t)\|e_1\| + \dot{e}_1\,\mathrm{sgn}(e_1) - \alpha(t)\|e_1\| \tag{37}$$

Expansion of the $\alpha(t)$ term into its constant and time varying components such as in (12) further yields

$$\dot{v} \le \left(\frac{1}{2} - \alpha(t)\right)e_1^2 - \frac{1}{2}r^2 - k_s r^2 - r\widetilde{N} + \dot{\beta}(t)\|e_1\| + \dot{e}_1\,\mathrm{sgn}(e_1) - \alpha_1(t)\|e_1\| - \alpha_0\|e_1\| = \le \left(\frac{1}{2} - \alpha(t)\right)e_1^2 - \frac{1}{2}r^2 - k_s r^2 - r\widetilde{N} + \left(\dot{\beta}(t) - \alpha_1(t)\right)\|e_1\| + \dot{e}_1\,\mathrm{sgn}(e_1) - \alpha_0\|e_1\| \tag{38}$$

By selection of $\alpha_1(t) = \left\|\dot{\beta}(t)\right\|$ the fifth term of the Lyapunov is eliminated. Additionally, from Remark 2 $\widetilde{N}$ is bounded from (15); therefore, the last two terms of (38) are also bounded in a similar manner. By noticing that the last two terms are similar to the FTE as defined earlier it can be shown they are bounded by $\|r\|$. Based upon (15) it is practical to assume that a second bounding function can be defined as

$$\left\|\widetilde{N}\right\| + 1 \le \Omega'\left(\|b\|\right)\|b\| \tag{39}$$

By utilizing (39) and taking the norm of the entries

$$\dot{v} \le -\left(\alpha(t) - \frac{1}{2}\right)\|e_1\|^2 - \frac{1}{2}\|r\|^2 - k_s\|r\|^2 + \|r\|\Omega\left(\|b\|\right)\|b\| = -\left[\alpha(t) - \frac{1}{2} \quad \frac{1}{2}\right]\|b\|^2 - k_s\|r\|^2 + \|r\|\Omega\left(\|b\|\right)\|b\| =$$
$$-\lambda\|b\|^2 - k_s\|r\|^2 + \|r\|\Omega\left(\|b\|\right)\|b\| \tag{40}$$

where $\lambda = \frac{1}{2}$.

Completing the squares with respect to $\|r\|$, yeilds the inequality

$$\dot{v} \le -\left(\lambda - \frac{\Omega^2\left(\|b\|\right)}{4k_s}\right)\|b\|^2 \le -c_1\|b\|^2 \tag{41}$$

provided

$$k_s > \frac{\Omega^2\left(\|b\|\right)}{4\lambda} \text{ or } \|b\| < \Omega^{-1}\left(2\sqrt{k_s\lambda}\right) \tag{42}$$

where c_1 is a small constant. Now define $U(s) = c\|b\|^2$ as a continuous positive-semi-definite function for some real positive constant *c* defined on the domain *D* such that

$$D \equiv \left\{q \in \Re^n \times \Re_{\ge 0} \mid \|q\| < \Omega^{-1}\left(2\sqrt{\lambda k_s}\right)\right\} \tag{43}$$

where $q \equiv \begin{bmatrix} b^T & \sqrt{P} \end{bmatrix}^T$

This result follows similar methods shown in Theorem 1 of [12] where $m(x) = 1$, $\lambda_1 = 1/2$, and $\lambda_2 = 1$ in the notation of that work. Using this result yeilds (43). Also, following similar arguments as in [12] a region of attraction for the semi-global AS condition is defined S to be (44) [17].

$$S \equiv \left\{ q \in D \mid \|q\| < \left(\Omega^{-1} \left(2\sqrt{\lambda k_s} \right) \right)^2 \right\} \tag{44}$$

Remark 3: It should be noted that to accommodate both desired regions of attraction and initial conditions the gain k_s must be selected sufficiently large so the region of attraction can be arbitrarily large [12]. The proposed methods and the ones in [6] are not capable of distinguishing between instability and fault progression due to the constraints on initial conditions. Implementation of a sliding-mode type k_s can compensate for unknown initial conditions. The gain, or control effort, can be set to a large value to encompass initial conditions and then be slid into a smaller value to provide performance without large control efforts.

The methods proposed in this section provide a practical observer that can be applied to many systems. Not only does it provide state observations but these are obtained under nonideal conditions where many methods fail to reach an AS state. While the proposed observer can only attain a semi-global AS state, the region of stability can be made arbitrarily large via gain adjustment allowing for inclusion of initial conditions. The AS stability is an advantage over traditional observer methods where only the mean of the error can be reduced while providing stability. The performance of this observer and its ability to support practical applications will be discussed later in the work.

Remark 4: In [6], boundedness of the state estimation error is demonstrated in the presence of bounded uncertainty. The threshold value is selected as a function of bounded uncertainty of the nonlinear system. There is difficulty in identification of the bound on the system uncertainties. By contrast, in this work, despite the presence of unmodeled dynamics, approximation errors and bounded disturbances, asymptotic stability is demonstrated which is more powerful than boundedness. As a consequence, the deadzone operator normally used for fault detection in [6] is not required. Additionally, the threshold can be selected to be zero or small.

Remark 5: In [6], projection scheme is employed for the parameter update of the OLA scheme whereas in this work, no modification of the parameter or NN update law is utilized. In fact, a two layer NN, which is considered as a nonlinear approximator, is utilized in this work. However, one can simply employ any linear approximator such as the ones in [4, 5, 18].

Remark 6: For control design using NNs, boundedness is commonly shown through a bounded tracking error [17] whereas using the RISE approach, one can show asymptotic convergence of the tracking error. However, control architecture is outside the scope of this effort and has been previously shown. Similarly, the proposed scheme can be utilized to show asymptotic stability of fault isolation and accommodation.

Remark 7: Asymptotic convergence of the state estimation error can be utilized to predict time to failure by projecting the estimated states to their limit. Unfortunately, states in a state space representation may not represent physical parameters of a nonlinear system. Additionally, parameter limits are normally set according to physical quantities, for instance, stiffness in the case of bridges etc. Therefore, the parameter update law is

utilized to obtain a TTF under the assumption that the considered system can be expressed as linear in the unknown parameters (LIP) as discussed next.

III. Adaptive Parameter Estimation

Prognostics require estimates of system health through parametric representations systems to isolate the root cause of failing components. Parametric models may be coupled with parameter estimation for model matching to the physical process. To provide both parametric modeling and health estimation, and suitable performance, for fault detection and robustness, a combination of adaptive control methods as well as the proposed NN/RISE methods is desirable.

In Fig. 1 the topology of the proposed method for prognostics is shown. An adaptive observer is now proposed to compensate for the lack of physical parameter estimations so that fault detection can be performed. The use of the adaptive estimation loop (AEL) allows for physically meaningful data to be extracted while maintaining the performance of the NN/RISE observer for state estimations. Use of the robust observer states as proposed above will also aid in the robustness of the AEL and is also subject to UUB error dynamics in the presence of noise.

The AEL block consists of an adaptive based model where the system model can be shown to be LIP. Many systems do not satisfy the LIP condition and therefore other methods such as NN are employed. However, in many systems estimation of certain parameters is desirable to determine the component health. In many cases when considering civil infrastructure components, adaptive methods can be applied to a LIP model based on FEA methods while the structures are operating in a normal manner [1].

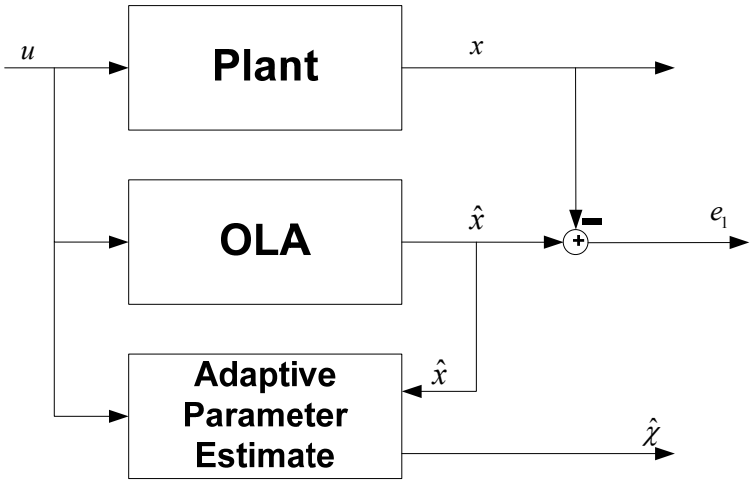

Fig. 1 Block diagram of proposed OLA and parameter estimation

Before proceeding, two assumptions for the AEL in this topology are stated.

A3) A suitable LIP model is available for the system to be monitored and nominal model parameters exist for normal operation. This assumption is critical to relate the nonlinear system to the physical parameters unlike in the general case where such a LIP assumption can restrict the type of systems under consideration.

A4) Based on the physical process there exist levels of the estimated parameters that are known to be critical values such that they are indicators of failure progression as they change.

A. Adaptive Observer

Using the above assumptions parameter estimates via adaptive methods are provided. The basic adaptive observer methods will be shown with the appropriate parameter update law. Then, extension of the update method will be shown in order to provide estimation of the time a parameter will reach a critical value, or the TTF. The thresholds for the parameters can be set by maintenance personnel in order to provide notification of impending failure conditions.

Remark 7: The development of the standard adaptive observer architecture is shown to contrast with the proposed NN/RISE methods. For brevity only the augmented tuning law is shown that is used to relax requirements for the persistency of excitation [14]. Other forms of the proof can found in [18] where ideal conditions are considered that will result in AS estimations. Since improvements in the performance are found by relaxing the impact of noise level with NN/RISE observer the AS state is approached, however for completeness and since noise levels can not be guaranteed the nonideal case is shown for the AEL proof.

The AEL system is constructed as in (46) where $\hat{\theta}$ is a vector of parameter estimates and ψ is the regression vector that is constructed of known system model terms.

$$\dot{\hat{x}}_\theta = f_\theta(x,u,\hat{\theta}) = \psi^T\hat{\theta} + K_v e_{id} \tag{46}$$

It is assumed similarly to the UAP given for NN that the system can be represented in the form

$$f(x,u,\theta^*) = \psi^T\theta^* + \upsilon \tag{47}$$

where θ^* are ideal parameter values and υ is the unmodeled dynamics and are considered bounded in a compact set such that $\|\upsilon(t)\| \le \upsilon_m$. The error in the estimated states is given as

$$e_{id} = \hat{x} - \hat{x}_\theta \tag{49}$$

The error dynamics then can be given by

$$\dot{e}_{id} = \dot{\hat{x}} - \dot{\hat{x}}_\theta = f(x,u,\theta^*) - f_\theta(x,u,\hat{\theta}) = \psi^T\theta^* - \psi^T\hat{\theta} - K_v e_{id} = \psi^T\tilde{\theta} - K_v e_{id} + \upsilon \tag{50}$$

where $\hat{\theta} = \theta^* - \hat{\theta}$ and $\theta, \hat{\theta}, \tilde{\theta} \in \Re^n$.

For the error dynamics in (50), let the parameter update law be similar to the NN weight update as

$$\dot{\hat{\theta}} = F\psi e_{id}^T - \kappa F \|e_{id}\| \hat{\theta} \tag{51}$$

where $F > 0$ and κ is a small positive constant. If K_v is selected to satisfy a minimum value as $K_v > \upsilon + \frac{\kappa \theta_{\max}^2}{4}$

then e_{id} and the parameter estimates are bounded. The proof can be presented as follows. Taking the Lyapunov candidate as

$$L_\theta = \frac{1}{2} e_{id}^2 + \frac{1}{2} tr\left\{\tilde{\theta}^T F^{-1} \tilde{\theta}\right\} \tag{52}$$

Taking the derivative of the Lyapunov candidate renders

$$\begin{aligned} &\dot{L}_\theta = e_{id}\dot{e}_{id} + tr\left\{\tilde{\theta}^T F^{-1}\dot{\tilde{\theta}}\right\} = e_{id}\psi^T\tilde{\theta} + e_{id}\upsilon - K_v e_{id}^2 + tr\left\{\tilde{\theta}^T F^{-1}\dot{\tilde{\theta}}\right\} = \\ &tr\left\{\tilde{\theta}^T\left(F^{-1}\dot{\tilde{\theta}} + \psi e_{id}^T\right)\right\} - K_v e_{id}^2 + e_{id}\upsilon \end{aligned} \tag{53}$$

Next, substitution of the tuning law (51) yields

$$\begin{aligned} &\dot{L}_\theta = tr\left\{\tilde{\theta}\left(-\psi e_{id}^T + \kappa\|e_{id}\|\hat{\theta} + \psi e_{id}^T\right)\right\} - K_v e_{id}^2 + e_{id}\upsilon = tr\left\{\tilde{\theta}^T\left(\kappa\|e_{id}\|\hat{\theta}\right)\right\} - K_v e_{id}^2 + e_{id}\upsilon \\ &= \kappa\|e_{id}\| tr\left\{\tilde{\theta}^T\hat{\theta}\right\} - K_v e_{id}^2 + e_{id}\upsilon = \kappa\|e_{id}\| tr\left\{\left(\theta - \hat{\theta}\right)^T\hat{\theta}\right\} - K_v e_{id}^2 + e_{id}\upsilon \end{aligned} \tag{54}$$

$$\dot{L}_\theta \le \kappa\|e_{id}\| tr\left\{\left(\theta - \hat{\theta}\right)^T\hat{\theta}\right\} - K_v e_{id}^2 + e_{id}\upsilon \tag{55}$$

Taking the norms yields

$$\dot{L}_\theta \le -K_v\|e_{id}\|^2 + \|e_{id}\|\upsilon_m + \kappa\|e_{id}\|\|\theta\|\|\hat{\theta}\| + \kappa\|e_{id}\|\|\hat{\theta}\|^2 \tag{56}$$

Next, completing the squares with respect to $\left\|\hat{\theta}\right\|$ and omitting the negative-definite terms will result

$$\dot{L}_\theta \le -\left\|e_{id}\right\| \cdot \left[K_v \left\|e_{id}\right\| + \upsilon_m + \frac{\kappa \theta_{\max}^2}{4} \right] \tag{57}$$

where $\theta_{\max} \le \left\|\theta\right\|$ is the bound on the ideal parameter values.

In order to maintain a positive definite function a $K_{v,\min}$ is found to be

$$K_{v,\min} > \upsilon_m + \frac{\kappa \theta_{\max}^2}{4} \tag{58}$$

This parameter constraint results in $\dot{L}_\theta$ being negative outside a compact set, b_r, which is guaranteed as long as

$$\left\|e_{id}\right\| > \frac{\upsilon_m + \kappa \theta_{\max}^2 \big/ 4}{K_{v,\min}} \equiv b_r \tag{59}$$

Similarly, completing the square for the parameter using (56) to get

$$\left\|\tilde{\theta}\right\| < \frac{\theta_{\max}}{2} - \sqrt{\frac{\theta_{\max}^2}{4} + \frac{\upsilon_m}{\kappa}} \equiv b_\theta \tag{60}$$

This result demonstrates the boundedness of both the error and parameter estimates.

The work for the NN/RISE observer can also be applied to adaptive observer methods. Next a corollary to Theorem 1 is introduced to provide continuity to the adaptive methods for parameter estimation.

Corollary 1: Consider the hypothesis presented in Theorem 1 except an adaptive observer is considered through the replacement of $\sigma\left(V^T z\right)$ with an appropriate regression

matrix, ψ under the LIP assumption on system nonlinearities as discussed in Assumption 3. Now assume that the RISE method is used in conjunction with the adaptive observer. With the parameter update selected as (27), the estimation error and the parameter estimation errors converge asymptotically to zero under the persistence of excitation condition and in the absence of unmodeled dynamics and disturbances.

Proof: Follow steps similar to Theorem 1.

Remark 8: In theorem 1 $e_1(t)$ is shown to be semi-globally asymptotically stable using the using the RISE feedback and the parameter update, $\dot{\hat{\theta}} = F\psi e_1^{\,T}$. Additionally, one can use adaptive observer framework as described in theorem 1 and Corollary 1 and the resulting system developed for a nonideal case where $e_1(0)$ and the modeling uncertainties, $\tilde{\xi}(x,u)$, are non-zero results in asymptotically stable states and a bounded value for the parameter estimates.

Remark 9: For an ideal case, i.e. $e_1(0) = 0$ and $\tilde{\xi}(x,u) = 0$, where the adaptive framework in Corollary 1 is used with the RISE feedback and if the regression matrix ψ is found to be persistently exciting (PE) such that $\lambda_1 I \le \int_{t_0}^{t_0+T} \psi^T \psi \, d\tau \le \lambda_2 I$ for some positive constants λ_1, λ_2, and T then $\theta^* - \hat{\theta} = \tilde{\theta} \to 0$ as $t \to \infty$. In other words, for an ideal case given a regression matrix that spans the system the parameter estimation error is also asymptotically stable about the origin and will converge and track the true parameter value. For detailed treatment of the PE condition and its effect on AS parameter estimates please refer to [18].

Remark 10: In many applications development of a suitable LIP model is difficult via construction of a regression vector of known system model terms; in this case a single-layer NN may be employed and the vector ψ will then simply become the activation function, $\sigma(\bar{x})$, of that NN where $\bar{x}$ is a vector of NN inputs. This modification allows for application of this method to systems when a suitable LIP model is not practical.

***Remark 11*:** While the parameter estimates are shown to be bounded in Theorem 1 through 3, when fault dynamics are growing the physical parameters will continue to grow with them and the bounds may not hold under these conditions.

While use of a second observer provides estimates of physical parameters that the NN methods do not, a primary observer is used to remove LIP assumptions. Since LIP conditions will only hold in some applications, and even then only for particular behaviors it requires both methods to complete the comprehensive prognostics framework that is sought. Therefore, the proposed NN/RISE observer system can be employed when a system is failing in a highly non-linear manner and the LIP condition will not hold, allowing for characterization of these types of behaviors. However, when systems fail gradually they can exhibit dominant linear behavior, such as structural systems, then the prognostics capabilities can be provided by the AEL and TTF estimates be provided to aid in maintenance needs. Next the prognostics methods are developed.

B. Prognostics via Adaptive Parameter Projection

Through inspection of the parameter update law used in the AEL it is seen that it takes the form of a first order ordinary differential equation with time-varying coefficients. The solution to this equation is well known and can be found in [19].

Through use of the solution and solving for $(t - t_o)$ for the time to a particular threshold value, prognostics capabilities are provided.

To estimate the TTF the current values of the error and the parameters are used since future values are not known. Threshold values set by the application's physics allow for estimation of the TTF. The parameter update law for the adaptive estimator is shown in (51). To predict the failure point (51) must be solved for the time, *t*, when $\hat{\theta}$ reaches $\hat{\theta}_{i,thresh}$. The threshold value $\hat{\theta}_{i,thresh}$ is a design parameter that must be chosen based on the physical system and the maintenance needs. Selection of $\hat{\theta}_{i,thresh}$ should be performed using input from maintenance personnel and designers in order to provide sufficient warning of failures and margins of safety.

Note: The estimated TTF may or may not be the actual time a system physically fails, rather is based on a failure condition set by design parameters for components. In other words, the TTF is the time remaining till the parameter reaches a predetermined threshold that, $\hat{\theta}_{i,thresh}$, is considered to be a failing component, or system.

Theorem 2: Given the parameter update law (51) and a known threshold value, $\hat{\theta}_{i,thresh}$, for the given parameter then the projected TTF can be shown to be

$$TTF_i = \frac{1}{\kappa F \left\| e_{id} \right\|} \left[\frac{\frac{\left[\psi e_{id}^T \right]_i}{\kappa \left\| e_{id} \right\|} - \hat{\theta}_{i,thresh}}{\theta_o - \frac{\left[\psi e_{id}^T \right]_i}{\kappa \left\| e_{id} \right\|}} \right] \tag{61}$$

Proof: The solution of (51) is an ordinary differential equation is given as

$$\hat{\theta}_i(t) = e^{-\int_{t_o}^{t}(\kappa F\|e_{id}\|)d\tau}\hat{\theta}(t_o) + e^{-\int_{t_o}^{t}(\kappa F\|e_{id}\|)d\tau}\cdot\int_{t_o}^{t} e^{\int_{t_o}^{s}(\kappa F\|e_{id}\|)d\tau} F\left[\psi e_{id}^T\right]_i ds \tag{62}$$

where $\left[\psi e_{id}^T\right]_i$ is the i^{th} entry of the ψe_{id}^T product used in the parameter estimation update law. By using the current values of e_{id}^T, ψ, and $\hat{\theta}(t_o)$, which is the current parameter value, an explicit solution for $(t-t_o)$ is found, which is the TTF. Evaluating the integral present in the exponential provides the result in (63)

$$\int_{t_o}^{t}\kappa F\|e_{id}\|d\tau = \kappa F\|e_{id}\|(t-t_o) = \kappa F\|e_{id}\|TTF_i \tag{63}$$

Substitution of (63) into (62) yields the result in (64).

$$\hat{\theta}_i(t) = e^{-\kappa F\|e_{id}\|TTF}\hat{\theta}(t_o) + F\left[\psi e_{id}^T\right]_i e^{-\kappa F\|e_{id}\|TTF}\cdot\int_{t_o}^{t} e^{\kappa F\|e_{id}\|(s-t_o)} ds \tag{64}$$

Next, evaluation of the final integral yields (65), a result containing the needed parameters.

$$\hat{\theta}_i(t) = e^{-\kappa F\|e_{id}\|TTF_i}\hat{\theta}(t_o) + \frac{\left[\psi e_{id}^T\right]_i}{\kappa\|e_{id}\|}\cdot\left[1 - e^{\kappa F\|e_{id}\|TTF_i}\right] \tag{65}$$

Finally, solving TTF_i and renaming $\hat{\theta}_i(t)$ to $\hat{\theta}_{i,thresh}$ yields

$$TTF_i = \frac{1}{\kappa F\|e_{id}\|}\left[\frac{\frac{\left[\psi e_{id}^T\right]_i}{\kappa\|e_{id}\|} - \hat{\theta}_{i,thresh}}{\theta_o - \frac{\left[\psi e_{id}^T\right]_i}{\kappa\|e_{id}\|}}\right] \tag{66}$$

where θ_{thresh} is the maximum parameter value allowed by maintenance criteria, and θ_o is the current value of the parameter. In the explicit solution for TTF_i allows for prediction of the failure time based on particular parameter values. Performance of this method and implementation issues will be discussed in simulation studies later in this work.

Use of the TTF calculation gives predictive qualities to the adaptive estimation method shown above. The use of the TTF analysis allows for predictive maintenance, safety actions to be taken, or for shutdown of equipment to be performed before failure occurs. In the next sections example applications will be covered before simulation studies that use both the TTF estimation and the NN/RISE observer system to illustrate performance of both methods.

Remark 12: The TTF must be calculated iteratively given the state estimation error. Then the parameters are updated after projection and are utilized as the initial values for the next iteration to obtain state estimation error.

Remark 13: In the above theorem, parameter update law (51) is utilized to show the TTF. Instead, one can use (27) without the e-modification term to obtain TTF by following the same steps as given in the theorem. Without the extra term, the proof will be much simpler than the case with the term.

IV. Simulation Examples

The maintenance of infrastructure components is of prominent concern. In particular, bridges have recently come under increased scrutiny due to catastrophic failures [20-22]. Additionally, during storm conditions, such as those in the hurricane Katrina, the failure of levee systems illustrates the need for early warning systems to predict eminent failures for these systems. Failures of infrastructure systems incur significant capitol losses and

risk to users. Methods for the prediction of failures and detection of fault dynamics therefore provide incentives to maintainers of these large systems who must provide reliable operation.

In this section simulation studies for the proposed observer and prognostics methods are shown. Examples include models of a nonlinear mass-spring-damper system a simply supported beam, and a levee wall. The beam example is formulated as a composite beam similar to the Missouri University of Science and Technology (MST) Smart Composite Bridge (SCB) [23, 24] construction and is subject to fatigue. For the levee example, foundation failure due to under-seepage of water is considered. Simulation results for each system showcase the ability of the proposed methods to provide performance for their given purpose. It should be noted that all simulations present accelerated failure results to show the behavior of the systems while experiencing failure dynamics.

A. Non-linear Mass-Spring-Damper System

Evaluation of performance for the NN/RISE observer was performed using a nonlinear mass-spring-damper (MSD) system. This system provides a suitable system for evaluation of the performance of the observer in the presence of measurement noise. The MSD system is described in (67), where m is the mass, c is the damping coefficient, k is the stiffness, and $f_1(t)$ is the forcing function. The failure dynamics for the MSD system are considered to be in the form of a hardening non-linear spring, in other words as the spring ages it becomes stiffer. In order to introduce a failure a non-linear term is introduced into the system to provide additional dynamics as defined in (68) where $f_1(t) = 5\sin(2\pi f_o t)$ with $f_o = 1/2$ and ρ is the unknown onset rate of failure. Constants for the simulation are shown in Table I. The nonlinear MSD system provides a

baseline for both the performance of the NN/RISE observer subject to noise as well as an adaptive observer for prognostics.

$$m\ddot{x} + c\dot{x} + kx + H(t) = f_1(t) \tag{67}$$

$$H(t) = \left(1 - e^{-\rho(t-t_f)}\right) k a_1^2 x \tag{68}$$

TABLE I SIMULATION CONSTANTS FOR NON-LINEAR MSD SYSTEM

Variable	m [kg]	c [kg/sec]	k [N/m]	ρ [unitless]	a_1 [unitless]	t_f [sec]
Value	1.0	0.5	0.5	0.05	1.1	30

B. Cyclic Failure of a Composite Beam

Cyclic failure testing is often used on structural specimens to determine the reliability and failure modes of components. Often the failure mode is noticed only during inspection of the specimen on a periodic inspection cycle, and the dynamics of the failure may or may not be recorded. To demonstrate the proposed prognostic/diagnostic method presented above a simulation of a composite member used in the construction of the MST SCB is presented. The simulation is an FEA simulation in MATLAB® where the beam is subjected to fatigue-based failure.

1) Description of Beam

The beam system is a simply supported beam and its dimensions and loading is shown in Fig. 2, while the physical dimensions and properties are shown in Table III. The beam is constructed of composite and the simulation was constructed from actual failure test data where the ultimate strength of the member was discovered [24]. By merging experimental data and FEA methods a simulation that provides a suitable testing environment for the observer and prognostics methods previously proposed is produced. The FEA model was formulated using a displacement model of an Euler-Bernoulli beam

system. The model was constructed of simple elements that are considered linear with normal configurations of tensile and shear forces.

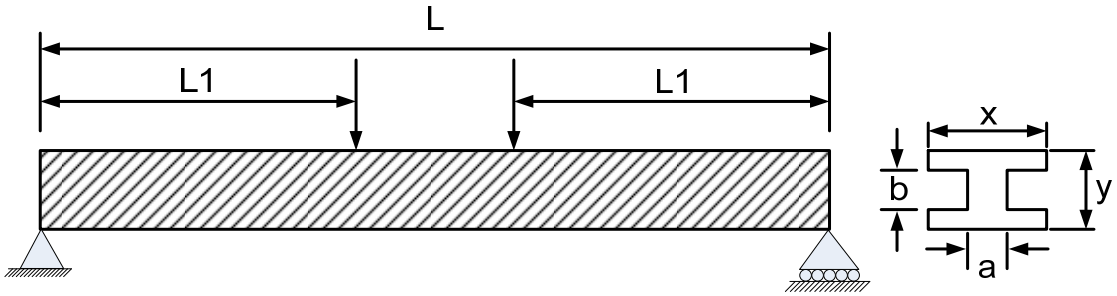

Fig. 2 Schematic of a composite beam for FEA simulations

TABLE II COMPOSITE BEAM DIMENSIONS AND PROPERTIES

Quantity	Length [m]	Loading Location [m]	Beam Width [m]	Beam Height [m]	Web Width [m]	Web Height [m]	Young's Modulus [GPa]	Moment of Inertia [m^3]
Symbol	L	L1	x	y	a	b	E	I
Value	8.534	3.658	0.610	0.610	0.305	0.229	34.255	0.01826

2) Description of Simulation

The beam was subjected to cyclic loading with a frequency of ½ Hz and a total force of 70 kN (15,735 lbs) was applied. In the FEA simulation the beam experiences fatigue in the form of a reduction in the stiffness of the beam. As the simulation progresses the beam begins to weaken at an unknown time to the observer system. The beam displacement at mid-span was used to supply the observer with a measured state value. Selection of the failure rate was performed by using both linearly and exponentially aging stiffness for the beam. The results from the simulation study allow for greater evaluation of the prognostics method to calculate the TTF as proposed in (66).

C. Levee System Simulation

The use of FEA methods was also applied to the simulation of the levee system. The levee was subjected to a foundation under-seepage failure mode in which erosion of the

foundation causes the levee wall to fail. After noting some limitations and reasons for the model used in simulation the development of the model and description of simulations are shown.

***Remark 14*:** Simulation of a levee system is a complex and meticulous task, the authors have created a model that will help to showcase the capabilities of both the NN/RISE and prognostics framework presented in this work. The goal of the model is to provide a reasonable approximation to the behavior of a levee system under tidal and storm surge loadings to emulate the behavior for under-seepage foundation failures. More thorough modeling would be required to provide the behavior of a levee system located in a specified environment.

1) Levee Model

The levee FEA model is a modified beam model that has a non-uniform cross-section to approximate the structure shown in Fig. 3. The loading on the wall due to the water is a distributed load based on the average water level, $\overline{h}$ and a tide height component. Additionally, an example of how the levee will start to lean as the foundation is eroded due to under-seepage is shown with respect to the angle θ in Fig. 3. The observer was supplied with the angle of the levee by assumption that a tilt-meter was used to instrument the levee at the top of the wall.

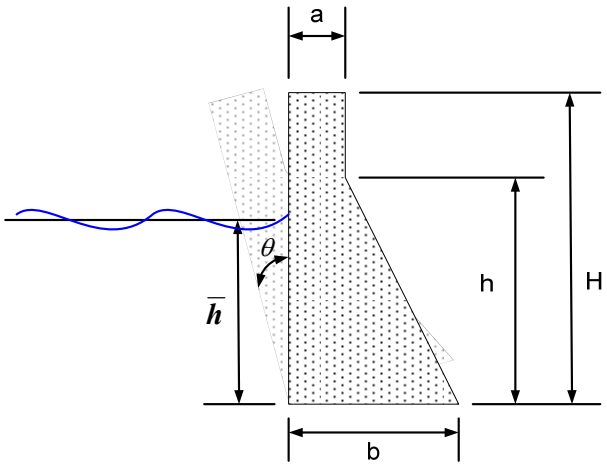

Fig. 3 Diagram of levee wall cross-section for FEA simulations

The model for the tide level is shown in (70) where ω_0 should be a slow varying frequency with a period on the order of 12 hours. The amplitudes of the water level model are set based on the average water level, $\overline{A}$, and the fluctuation in that average, A_0. The surface wave force, f_{wave}, is assumed to a be a sinusoidal force that occurs along the surface of the water and is modeled as in (71) where ω_1 is a frequency of the crests. The surface wave frequency, and its associated amplitude, should be set to a reasonable quantity based on the assumed location of the levee. This modeling of the forces allows for the model to provide realism as well as remain practical enough for use in the analysis of the prognostics methods proposed in this work.

Levee wall loading from the force of the water is given by

$$f(\overline{h}) = P_o + \rho_w g \overline{h} + f_{wave} \tag{69}$$

where

$$\overline{h} = \overline{A} + A_0 \sin(\omega_0 t) \tag{70}$$

and

$$f_{wave} = A_1 \sin(\omega_1 t) \tag{71}$$

where the effect of the force of the water due to water level $\overline{h}$, water density ρ_w, gravity g, and the ambient pressure P_o, is shown. The loading of the levee is based on the water level and the wave forces. The water level is varied and can be considered to change with the tide level; or for detailed simulations with storm surges. Table III and Table IV describe the parameters used for simulation results presented in the next section.

TABLE III LEVEE WALL PARAMETERS

Variable	a [m]	b [m]	h [m]	H [m]
Value	1	10	26	28

TABLE IV LEVEE SIMULATION PARAMETERS

Variable	$\overline{A}$ [m]	A_0 [m]	A_1 [N]	ω_0 [Hz]	ω_1 [Hz]	P_o [Pa]	ρ_w [kg/m^3]
Value	10	15	50	2.315E-5	1	101,325	1000

2) Foundation Under-seepage Model

In order to provide a realistic simulation of levee foundation a failure mode due to under-seepage of the levee wall was used [25]. It should be noted that while overtopping is also of great concern methods for detection of under-seepage when a tilt-meters are used for sensing are similar to overtopping; differences are found in the tilt of the wall as failure increases is to the opposite direction to that of under-seepage. Therefore, the detection, monitoring, and prognostics methods will differ only in the alarm levels for the angle of inclination, θ, of the wall as the failure proceeds.

For under-seepage failures a crack in the floor of the water galley, or the base of the foundation, provides a path for water to seep under the levee wall [25]. As the failure proceeds, the foundation of the levee wall begins to loose its footing and will eventually topple towards the water. In Fig. 3, the tilt is illustrated where wall leans towards the

water since there will be a greater distribution of force at the bottom of the wall as compared to the top due to the water pressure gradient. To simulate this failure mode the boundary conditions at the base of the levee are aged over time as the volume of water seepage increases. The seepage will eventually "unclamp" the levee from its foundation and lead to catastrophic failure. The assumption is made that during the simulation time the failure goes unnoticed and that no repairs are made to the levee foundation.

V. RESULTS

Results of the simulation studies are now presented. An analysis of the NN/RISE method robustness to noise as compared to a NN observer is shown. The observer was simulated with and without the RISE technique to illustrate the performance of the observer. For each case the noise level was increased to demonstrate the ability of the RISE term to provide suitable stability in the presence of noise. Additionally, simulations were performed using the NN/RISE method with the addition of the AEL and the TTF prediction. The performance of the TTF method is also presented using the nonlinear spring failure. For the infrastructure examples the prognostics results are shown since the performance of the NN/RISE is characterized in the non-linear spring result. The NN weights are initialized to zero and no offline training was used for simulation. Additionally, the following general parameters were used for all simulations presented in this section:

- Systems gains were implemented as dynamic terms as shown in the development of the NN/RISE methods, with the exception of k_s which was set to 0.9995
- A two-layer NN with tuning of only the outer layer

- Static random inner layer of weights
- Sigmoid activation functions, and
- Eleven neurons used for the NN

Use of the proposed methods requires that the user be aware of some limitations. Since the TTF estimate is error driven, the estimate can change rapidly with the error as it fluctuates with the observer system. For a steadily growing error the TTF estimates function in a more predictable manner and makes this method suitable for slowly progressing failures. Additionally, the methods developed are not well suited for sudden failures, but rather those that are incipient and can be tracked over a time period that is not infinitesimally small as compared to the system dynamics. For slow failures as the TTF roughly tracks the degradation of the system. For sudden failures the TTF is still an indication of degradation; however, the changes do not necessarily track the exact progression. A sudden transition in the TTF does not always indicate a sudden failure, but instead merely indicates that a change is happening. This behavior can be seen in the beam and levee examples where the TTF estimate changes after the onset of the failure and at some points has dramatic swings due to the dynamics of the estimate. Finally, some systems are better observed through the absolute value of the change in the estimated parameters. This change in observation is due to the relative change in a parameter being easier to observe due to scaling issues. The beam and levee systems are presented in this manner later in this section.

Application of the proposed methods to a structural system, for example, would require some prior data. The required data would be needed to form safety guidelines for the estimated parameters, and/or changes in those parameters, as well as providing the

opportunity to test differing model structures for the adaptive observer. Knowledge of the $\hat{\theta}_{i,thresh}$ values is important to deployment. Pre-recorded data is needed to establish failure characteristics for online system, otherwise the threshold process must be done during service and some failures may not be predicted. The model structure is important since adaptive methods are more susceptible to modeling uncertainty issues. The user of the system would then be able to install the proposed methods and use them to provide interpretation of the failure progression.

A. Performance Analysis of Fault Detection

Simulations were performed on the system shown in (67). During the simulation the noise variance was changed to investigate the performance of the NN/RISE observer under nonideal conditions. For comparison of the NN/RISE method to a normal NN observer the change is minor. Here the NN observer is generated without the RISE term by keeping the number of layers and nodes same as that of the case with NN/RISE design. Revisiting (6) the proportional gain component, (k_s+1), and the two extra terms based on the initial conditions and the integral term form the RISE input. It is the last two terms that comprise the differences between the NN and NN/RISE observer systems. These terms provide the additional performance gains that are seen in the simulation results. The proportional gain term, k_s, must be help constant between the NN and NN/RISE methods for fair comparison of the two methods. The simulation results were obtained through modification of (6) in this manner.

In Fig. 4 the mean squared error (MSE) of each method is shown. As the variance of the noise is increased the ability for the NN/RISE method to decrease the error in the

state estimates is shown. The performance was increased up to 25% at the highest noise variance considered.

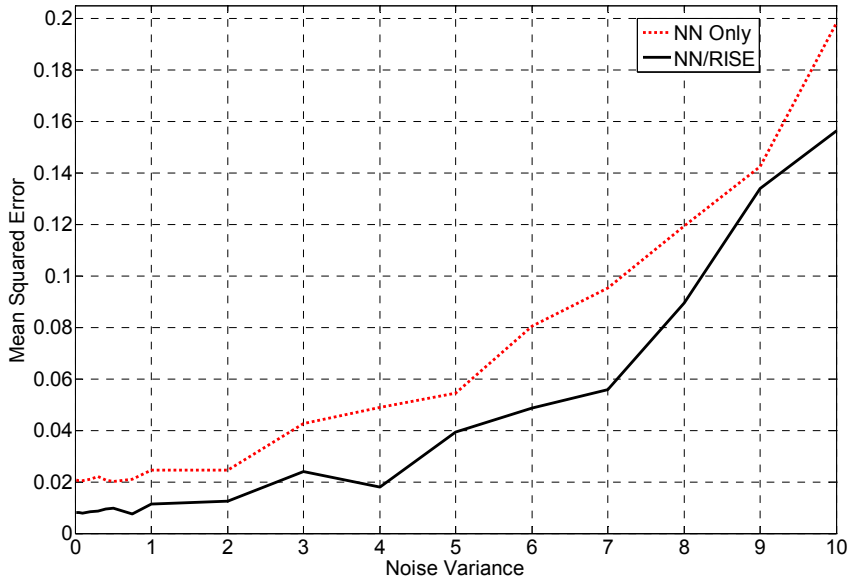

Fig. 4 Comparison of the mean squared error

Additionally, performance differences can also be seen in plots of the states. In Fig. 5 an example of the state estimates are shown. The figure shows that the NN method has more variance in the state estimates as compared to the NN/RISE methods. The NN method follows the noise more than the NN/RISE method showing the performance gains the NN/RISE method offers.

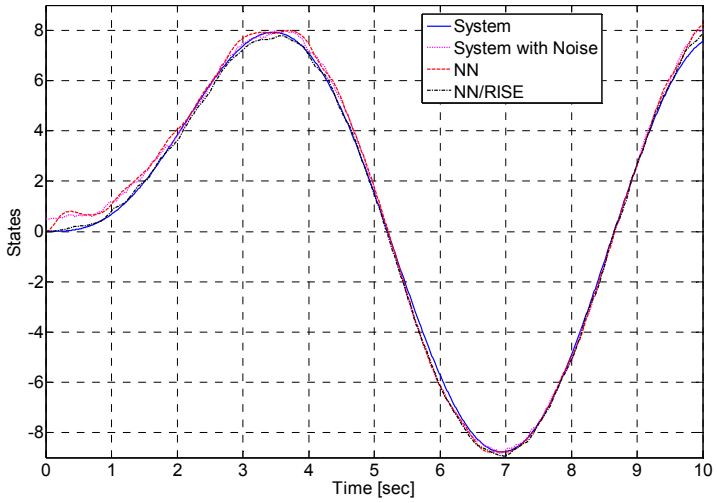

Fig. 5 Actual and observed states of the system

In Figs. 6 and 7 zoomed views are presented to provide a closer inspection of the states. Fig. 6 shows that the NN/RISE method converges sooner than the NN method as the simulation time begins and also follows the actual state with more accuracy. Finally, in Fig. 7 it can be seen that the NN method actually follows the fluctuations introduced by the noise levels while the NN/RISE method provides a more robust approximation of the actual state value.

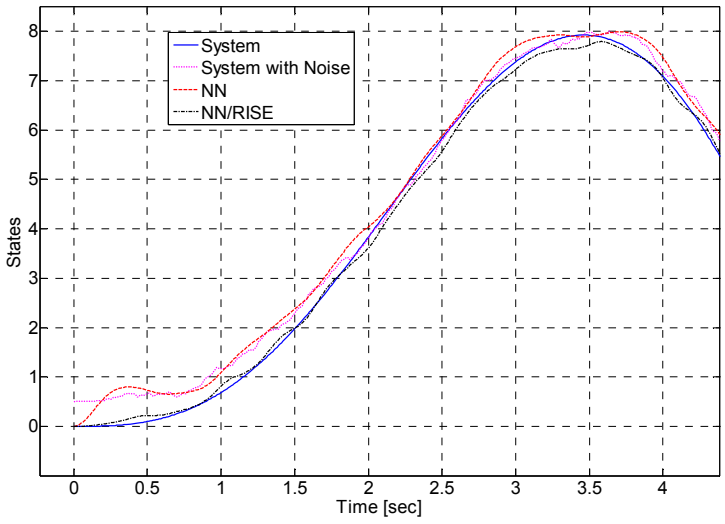

Fig. 6 Actual and observed states of the system

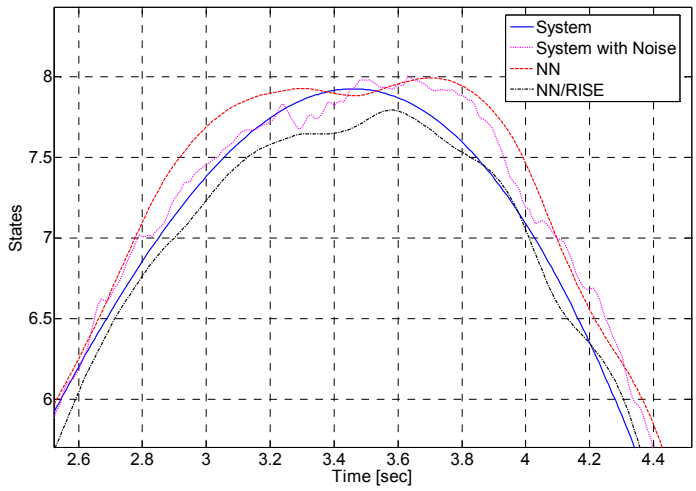

Fig. 7 Actual and observed states of the system

Simulation results for the prognostics method for TTF prediction are shown in Fig. 8 and Fig. 9. In Fig. 8 the estimation of the spring-stiffness is shown along with the maximum safe level. The prognostics method predicts the TTF before the estimation reaches the threshold as shown in Fig. 9. Note that the spring-stiffness converges to 0.75 and then begins to harden at 30 seconds. The TTF result not only shows that the result converges to near the failure point; roughly 142 seconds into the simulation, but also the actual TTF will improve as the failure point approaches. The TTF estimate also starts to change rapidly after 40 seconds as the parameter value begins to change more rapidly. This behavior is typical of the results obtained and is due to the progression of the fault as well as the growth in the error values. Also, it should be considered as an implementation issue with the method that results when the errors of the states are at their peak provide the best estimates. The implementation of the TTF prediction employed a peak detector in order to provide the best performance.

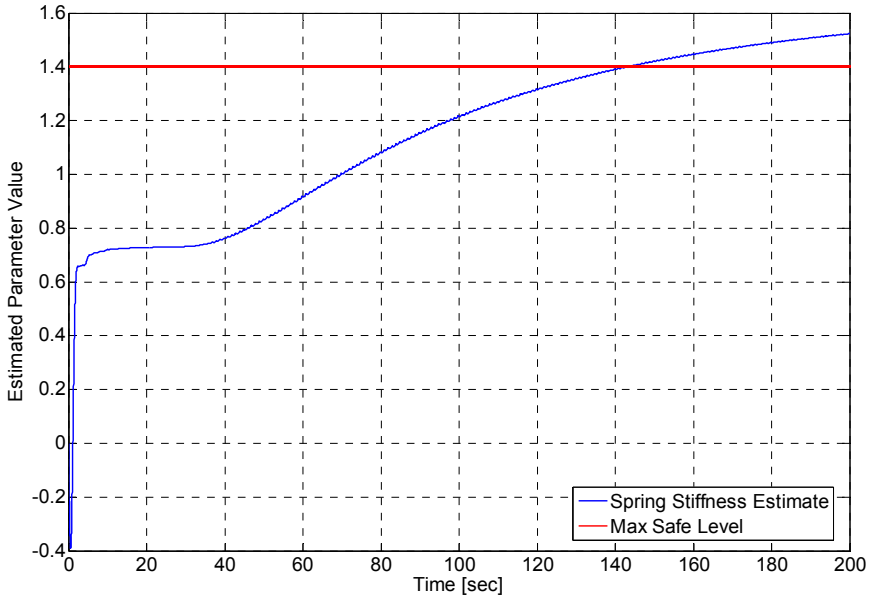

Fig. 8 Spring stiffness estimate

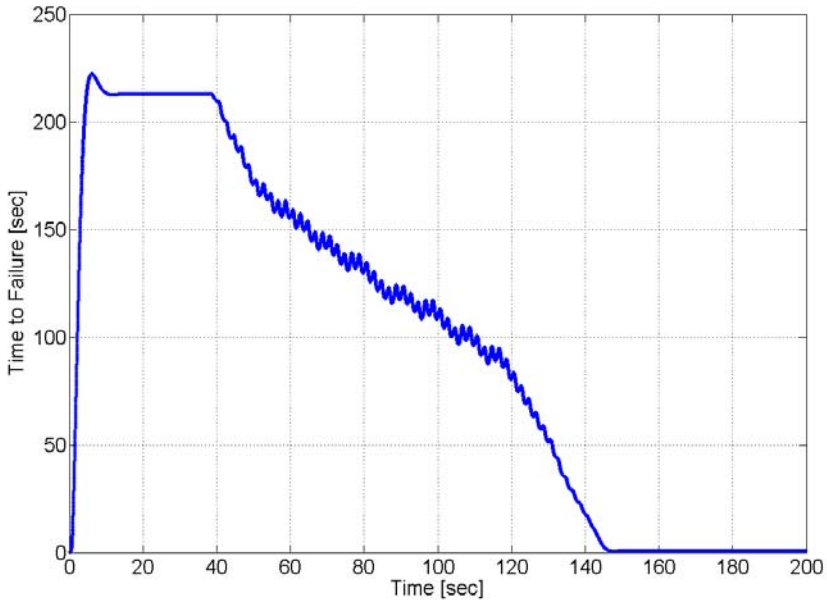

Fig. 9 TTF Prediction for nonlinear spring example

B. Beam FEA Simulation Results

Results from the beam simulations show that the proposed method tracks changes in the overall behavior of the beam. Additionally, the AEL component of the prognostics

framework allows oversight of the stiffness coefficient of the structural element as it ages. Figure 10 shows the estimation of the beam stiffness. As the failure begins, at 50 seconds, there is a sudden change in the value of the parameter and it then settles to the proper value again. Again the progression of the fault and growing error values introduce this behavior. This behavior was also seen in Fig. 11 where the TTF become small and then the estimate grows again to compensate as the failure proceeds the TTF estimate becomes more accurate as seen before in the previous results.

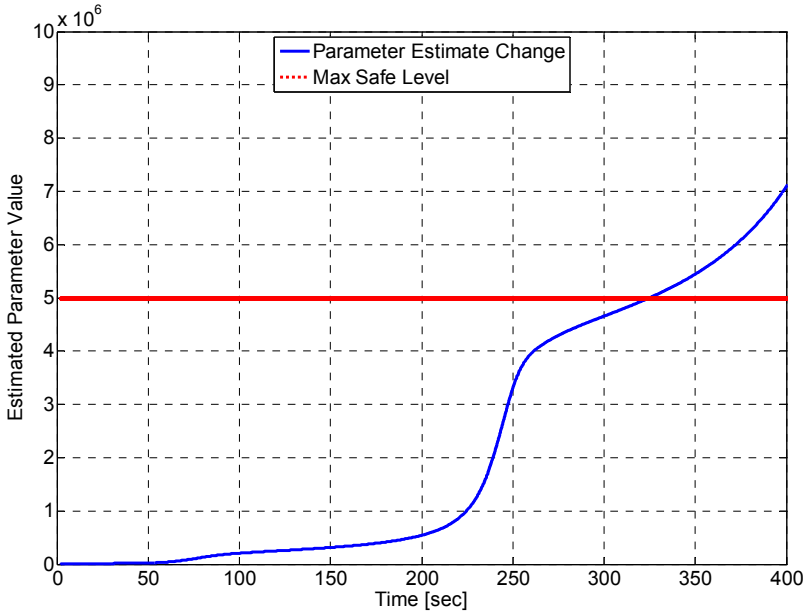

Fig. 10 Beam stiffness estimate

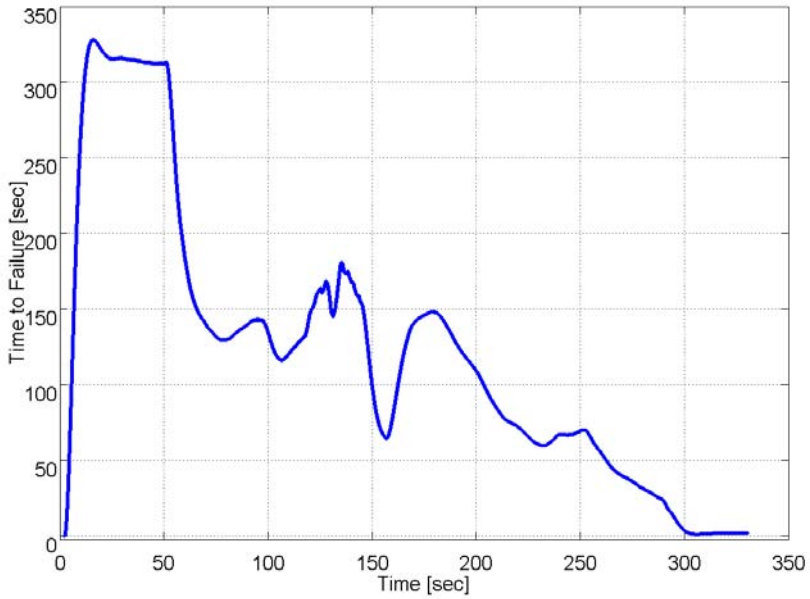

Fig. 11 TTF prediction for beam example

C. Levee Simulation Result

Simulation of the levee system provided insight to the failure modes seen during erosion leading to foundation failures. Results show that when erosion of the foundation is sufficient the boundary conditions of the levee footing are compromised and the levee wall will topple. During simulation the primary indicator used for levee failure was the angle of inclination. A tilt-meter was used with the TTF prediction and the adaptive observer. As the boundary condition of the levee fails the magnitude of the angle changes, and this translates into a change in the estimated stiffness of foundation. Figure 12 shows the estimate of the stiffness is shown, the estimate converges to the "healthy" state and then at 50 seconds the failure begins and the estimate slowly changes with the health of the structure. In Fig. 13 the estimate of the TTF shows that the predictive method provides adequate estimations and as before in the other examples it become more accurate as the actual failure-time gets closer. Additionally, the onset of the failure

and increasing errors continue to introduce rapid changes into the TTF estimate that are nonlinear.

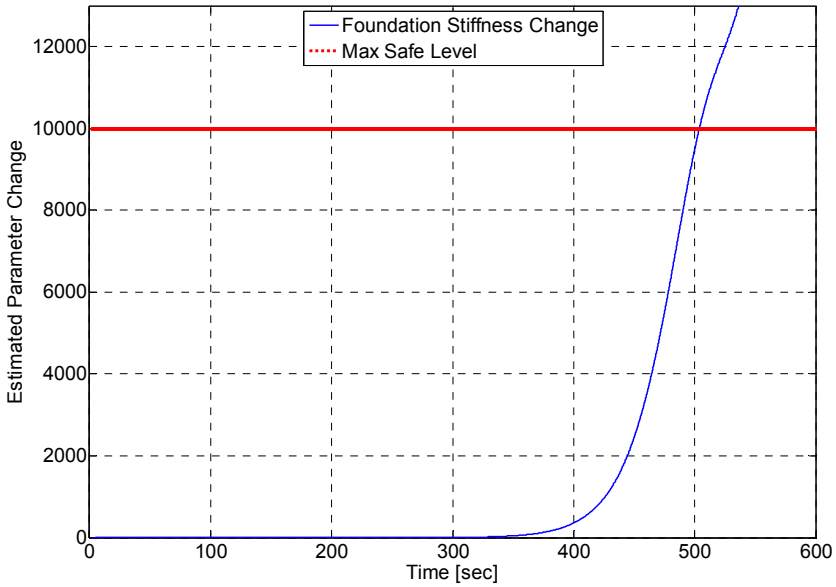

Fig. 12 Foundation stiffness estimate

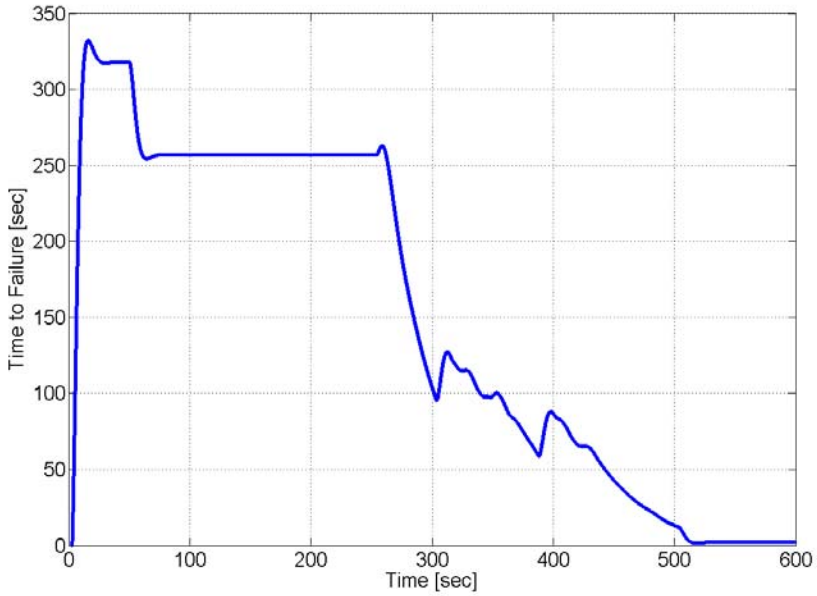

Fig. 13 TTF prediction for levee system

The preceding results show that the parameter estimation methods coupled with the TTF prediction provide prognostics capabilities. Estimation of TTF allows for inspection cycle planning and increased effectiveness of maintenance methods. As estimations proceed the accuracy increases. However, some of this behavior is due to the accelerated failure examples. In many systems the failure will be more progressive and the method will have more opportunity to learn the change and provide accurate prediction.

VI. Conclusions

In this work a novel method for fault detection and prognostics is applied to the problem of large-scale systems through examples of civil infrastructure. Due to the nations failing infrastructure components, a distributed and computationally viable method for detecting faults in-situ is desirable. The use of wireless sensor networks and intelligent data processing with analytically guaranteed methods can provide the needed performance that infrastructure component monitoring demands.

The proposed NN/RISE OLA provides suitable performance of fault dynamics in the presence of modeling uncertainties. The inclusion of the RISE term provides robustness in the detection of fault dynamics while the NN allows for adaptation to changing system dynamics as faults develop. Simulation results show increases of performance in the mean-squared-error of 25% over standard NN observer techniques. This result illustrates the NN/RISE's ability to deliver performance in the presence of uncertainties and to deliver appropriate state estimates. The addition of an adaptive parameter estimation that is used to observe internal states of the structure, or other applicable system, as it ages allows for health monitoring and extensions into predictive maintenance and prognostics. Additionally, extension of adaptive parameter estimates to estimate the time when alarm

levels will occur allows for application based prognostics that are meaningful to a maintenance program. For instance, the use of a change in structural stiffness of a beam could be used as an input to a crack model allowing for physically meaningful information to be provided from the prognostics methods.

The contribution of this paper is to propose not only analytical methods to provide robust estimation of states and associated physical parameters; but to present them in a context that demonstrates how the components of an effective maintenance program can be assembled. The use of these components can enhance the ability of maintenance personnel by allowing accurate estimation of the most useful variables as well as give them the predictive nature that is sought in practice. Future work should encompass the abilities of the NN/RISE observer to be extended to other classes of nonlinear system and applications. Also, work should be put forth to extend the prognostic prediction methods for the physical parameters. Finally, relation of the NN weights to prognostics conditions for a particular application would also be of interest.

REFERENCES

[1] J. W. Fonda and Watkins, S. E., "Health monitoring of a truss bridge using adaptive identification," Proc of the IEEE Conf. on Intelligent Transportation Systems Conf. (ITSC), pp.944-949, 2007.

[2] R. Marine, G. L. Santosuosso, and P. Tomei, "Robust adaptive observers for nonlinear systems with bounded disturbances," IEEE Trans. on Automatic Control, vol.46, no.6, pp.967-972, 2001.

[3] V. Stepanyan and N. Hovakimyan, "Robust adaptive observer design for uncertain systems with bounded disturbances," IEEE Trans. on Neural Networks, vol.18, no.5, pp.1392-1403, 2007.

[4] H. Wang and S. Daley, "Actuator fault diagnosis: an adaptive observer-based technique," IEEE Trans. on Automatic Control, vol.41, no.7, pp.1073-1078, Jul 1996

[5] Da-Wei Gu and Fu Wah Poon, "A robust state observer scheme," IEEE Trans. on Automatic Control, vol.46, no.12, pp.1958-1963, 2001.

[6] M.A Demetriou and M.M. Polycarpou, "Incipient fault diagnosis of dynamical systems using online approximators," IEEE Trans. on Automatic Control, vol.43, no.11, pp.1612-1617, 1998.

[7] N. Lybeck, S. Marble, B. Morton, "Validating prognostic algorithms: A case study using comprehensive bearing fault data," Proc. of the IEEE Aerospace Conf, pp.1-9, 2007.

[8] L. Jianhui, M. Namburu, K. Pattipati, Q. Liu, M. Kawamoto, and S. Chigusa, "Model-based prognostic techniques [Maintenance applications]," Proc. of the IEEE Systems Readiness Technology Conf. (AUTOTESTCON), pp. 330-340, 2003.
[9] N. M. Vichare and M. G. Pecht, "Prognostics and health management of electronics," IEEE Trans. on Packaging Technologies, vol.29, no.1, pp. 222-229, 2006.
[10] L.P. Su, M. Nolan, G. DeMare, and D. R. Carey, "Prognostics framework [for weapon systems health monitoring]," Proc. of the IEEE Systems Readiness Technology Conf. (AUTOTESTCON), pp.661-672, 1999.
[11] K. Mitchell, S. E. Watkins, J. W. Fonda, and S. Jagannathan, "Embeddable Modular Hardware For Multi-Functional Sensor Networks," Smart Mater. Struct., vol. 16, no. 5, pp. N27-N34, 2007.
[12] B. Xian, D. M. Dawson, M.S. de Queiroz, and J. Chen, "A continuous asymptotic tracking control strategy for uncertain nonlinear systems," IEEE Trans. on Automatic Control, vol.49, no.7, pp. 1206-1211, 2004.
[13] P. M. Patre, W. MacKunis, K. Kaiser, and W. E. Dixon, "Asymptotic tracking for uncertain dynamic systems via a multilayer NN feedforward and RISE feedback control structure," American Control Conference (ACC), pp.5989-5994, 2007.
[14] F.L. Lewis, S. Jagannathan, and A. Yesilderek, Neural network control of robot manipulators and nonlinear systems, Taylor and Francis, London, UK, 1999.
[15] B. Igelnik and Y.-H. Pao, "Stochastic choice of basis functions in adaptive function approximation and the functional-link net," IEEE Trans. on Neural Networks, vol. 6, no. 6, pp. 1320-1329, 1995.
[16] Y.-H. Kim, "Intelligent closed-loop control using dynamic recurrent neural network and real-time adaptive critic" Ph.D dissertation, Dept. Elect. Eng., The University of Texas at Arlington, Arlington, TX, 1996.
[17] H. K. Khalil, Nonlinear Systems, 2nd ed. Upper Saddle River, NJ, Prentice-Hall, 1996.
[18] K. S. Narendra, and A. M. Annaswamy, Stable Adaptive Systems, Dover Publications, NY, 2005.
[19] W. Kelly, and A. Peterson, The theory of differential equations classical and qualitative, Pearson Education, Upper Saddle River, NJ, 2004.
[20] P. Petchprayoon, "The prediction of flash flood caused by dam failure: A Case Study at the Tha Dan Dam, Thailand," Proc. of the Asian Conf. on Remote Sensing, pp. 337-342, 2001.
[21] T. Tingsanchali and C. Chinnarasi, "Numerical modeling of dam failure due to flow overtopping," Journal-des Sciences Hydrologiques, vol. 46, no. 1, Feb. 2001.
[22] National Public Radio, NPR:Minneapolis Bridge Collapses, available online: http://www.npr.org/templates/story/story.php?storyId=12463292, 2007.
[23] S. E. Watkins, "Smart bridges with fiber optic sensors," IEEE Instrument and Measurements Magazine, vol. 6, no. 2, pp. 25-30, 2003.
[24] K. Chandrashekhara, S.E. Watkins, A. Nanni, and P. Kumar, "Design and technologies for a smart composite bridge," Proc. of the IEEE Conf. on Intelligent Transportation System (ITSC), pp. 954-959, 2004.
[25] Engineer Manual No. EM 1110-2-1913, (2000). Engineering and Design. Design and Construction of Levees. Department of the Army. U.S. Army Corps of Engineers, Washington DC, 2000.

APPENDIX A

PROOF OF LEMMA 1

Integration of both sides of (22) yields (A.1).

$$\int_0^t L(t)d\tau = r\left(\varepsilon_2 + \widetilde{W}^T\sigma\left(V^T z\right) - \beta(t)\,\mathrm{sgn}(e_1)\right) - \dot{\beta}(t)\|e_1\| - e_1\widetilde{W}^T\sigma\left(V^T z\right) + \kappa\|e_1\|tr\left\{\widetilde{W}^T\hat{W}\right\} \quad \text{(A.1)}$$

Defining $N_1 = \varepsilon_2 + \widetilde{W}^T\sigma\left(V^T z\right)$ to get

$$\begin{aligned}\int_0^t L(t)d\tau &= \int_0^t r\left(N_1 + \varepsilon_2 + \widetilde{W}^T\sigma\left(V^T z\right) - \beta(t)\,\mathrm{sgn}(e_1)\right)d\tau \\ &- \int_0^t \dot{\beta}(t)\|e_1\|d\tau - \int_0^t e_1\widetilde{W}^T\sigma\left(V^T z\right)d\tau + \int_0^t \kappa\|e_1\|tr\left\{\widetilde{W}^T\hat{W}\right\}d\tau\end{aligned} \quad \text{(A.2)}$$

$$\int_0^t L(t)d\tau = \int_0^t r\left(N_1 - \beta(t)\,\mathrm{sgn}(e_1)\right)d\tau - \int_0^t \dot{\beta}(t)\|e_1\|d\tau - \int_0^t e_1\widetilde{W}^T\sigma\left(V^T z\right)d\tau + \int_0^t \kappa\|e_1\|tr\left\{\widetilde{W}^T\hat{W}\right\}d\tau \quad \text{(A.3)}$$

Substitution of the expanded form of the FTE and expansion of the equation yields

$$\begin{aligned}\int_0^t L(t)d\tau &= \int_0^t \dot{e}_1 N_1 d\tau + \int_0^t \alpha(t)\left(e_1 N_1 - \beta(\tau)\|e_1\|\right)d\tau - \int_0^t \dot{e}_1\beta(t)\,\mathrm{sgn}(e_1)d\tau \\ &- \int_0^t \dot{\beta}(\tau)\|e_1\|d\tau - \int_0^t e_1\widetilde{W}^T\sigma\left(V^T z\right)d\tau + \int_0^t \kappa\|e_1\|tr\left\{\widetilde{W}^T\hat{W}\right\}d\tau\end{aligned} \quad \text{(A.4)}$$

Integration by parts on the first and third terms render

$$\int_0^t \dot{e}_1 N_1 d\tau = \left[e_1 N_1\right]_0^t + \int_0^t e_1\dot{N}_1 d\tau = e_1(t)N_1(t) - e_1(0)N_1(0) + \int_0^t e_1\dot{N}_1 d\tau \quad \text{(A.5)}$$

$$\begin{aligned}&\int_0^t \dot{e}_1\beta\,\mathrm{sgn}(e_1)d\tau = \left[e_1\beta(\tau)\,\mathrm{sgn}(e_1)\right]_0^t - \int_0^t e_1\dot{\beta}(\tau)\,\mathrm{sgn}(e_1)d\tau = \\ &\beta(t)\|e_1(t)\| - \beta(0)\|e_1(0)\| - \int_0^t \dot{\beta}(\tau)\|e_1\|d\tau\end{aligned} \quad \text{(A.6)}$$

Substitution of the IBP terms into (A.4) yields

$$\int_0^t L(t)d\tau = e_1(t)N_1(t) - e_1(0)N_1(0) - \beta(t)\|e_1(t)\| + \beta(0)\|e_1(0)\| + \int_0^t e_1\dot{N}_1 d\tau + \int_0^t \alpha(t)\left(e_1 N_1 - \beta(\tau)\|e_1\|\right)d\tau - \int_0^t e_1\tilde{W}^T\sigma\left(V^T z\right)d\tau + \int_0^t \kappa\|e_1\|tr\left\{\tilde{W}^T\hat{W}\right\}d\tau \tag{A.7}$$

Application of the bounds on $L(t)$ shows that

$$\int_0^t L(t)d\tau \le \beta(0)\|e_1(0)\| - e_1(0)N_1(0) + - \beta(t)\|e_1(t)\| + \|e_1(t)\|\|N_1(t)\| + \int_0^t \left[\|e_1\|\|\dot{N}_1\| + \alpha(\tau)\left(\|e_1\|\|N_1\| - \beta(\tau)\|e_1\|\right) - \|e_1\|\|\tilde{W}\| + \kappa\|e_1\|\left\|tr\left\{\tilde{W}^T\hat{W}\right\}\right\|\right] d\tau \tag{A.8}$$

By evaluation of the integrand of the last term and factoring out $\|e_1\|$ yields a bound on both $\alpha(t)$ and $\beta(t)$ that also satisfies the condition on $\alpha(t)$ from the Lyapunov system. Application of the bounds on the weights and estimated weights and taking the portion of interest from (A.8) shows that $\beta(t)$ is lower bounded as:

$$\|\dot{N}_1\| + \alpha(t)\left(\|N_1\| - \beta(t)\right) \rightarrow \beta(t) \ge \|N_1\| + \frac{1}{\alpha(t)}\left(\|\dot{N}_1\| + \|W\| + \|\hat{W}\| + \kappa\|\tilde{W}\|\|\hat{W}\|\right) \tag{A.9}$$

where $\alpha(t) > 1$.

Finally, in order to properly set the bounds on $\beta(t)$ the bounds on $\|N_1\|$ and $\|\dot{N}_1\|$ are evaluated. These are shown next as

$$\|N_1\| \le \|\varepsilon\| + \left\|\dot{\tilde{\xi}}\right\| + \|\tilde{W}\| = \|\varepsilon\| + \left\|\dot{\tilde{\xi}}\right\| + \|W\| + \|\hat{W}\| \tag{A.10}$$

$$\|\dot{N}_1\| \le \|\dot{\varepsilon}\| + \left\|\ddot{\tilde{\xi}}\right\| + \|\dot{\tilde{W}}\| = \|\dot{\varepsilon}\| + \left\|\ddot{\tilde{\xi}}\right\| + \Gamma\|e_1\| + \kappa\Gamma\|e_1\|\|\hat{W}\| \tag{A.11}$$

Use of the shown bound on $\beta(t)$ will guarantee that $L(t)$ is a positive-definite term.

APPENDIX

Embeddable Sensor Mote for Structural Monitoring

James W. Fonda[1,2,*], Steve E. Watkins[2,] S. Jagannathan[1], and Maciej Zawodniok[1]
Embedded Systems and Networking Laboratory[1]
Applied Optics Laboratory[2]
Department of Electrical and Computer Engineering
Missouri University of Science and Technology, Rolla, Missouri 65409-0040
jfonda@ieee.org, steve.e.watkins@ieee.org, sarangap@mst.edu, and mjzx9c@mst.edu

Abstract—These An embeddable sensor mote for structural monitoring is described. The Missouri University of Science and Technology (MST) F1 mote is designed to provide a general platform for sensing, networking, and data processing. The platform consists of an 8051 variant processor, two 802.15.4 variant radio platform options, micro Smart Digital (SD™) flash storage, USB connectivity, RS-232 connectivity, and various sensor capabilities. Sensor capabilities include, but are not limited to, strain gauges, a three-axis multi-range accelerometer, thermocouples, and interface options for other digital and analog sensors via a screw terminal block. In its default configuration the strain conditioning channel is appropriate for structural monitoring, but through reconfiguration it can be used with other resistive bridge transducers for pressure, force, displacement, etc. The F1 mote provides capabilities for strain, temperature, and vibration sensing in a small package. The mote is used at MST for networked monitoring of structures and networked robotic vehicles. In this paper an overview of the F1 mote will be given that emphasizes its operating architecture and potential applications. Applications include infrastructure monitoring for structures such as bridges, levees, and buildings as well as robotics, machine monitoring, and sensor networks. The described platform provides novelty in that it has the ability to be a dedicated structural monitoring system, however can be also used in development of other systems. The F1 platform was designed to combine features of available dedicated platforms and available development kits. The F1 provides a novel combination of sensing, processing, and application possibilities for the targeted application areas.

***Index Terms*—Infrastructure Monitoring, Embedded Sensors, mote, sensor networks, wireless**

I. INTRODUCTION

Embedded systems for monitoring and networking are designed for hostile, hard-to-reach, or remote environments. These systems must integrate sensing, computation,

Research supported in part by Dept. of Education GAANN Fellowship, NSF I/UCRC on Intelligent Maintenance Systems, Air Force Research Laboratory Grant (FA8650-04-C-704) and Intelligent Systems Center.

networking, and low-power capabilities into a constrained platform. Civil engineering infrastructure monitoring and prognostics, robotics, machine prognostics, and other military, industrial, and security applications can benefit from such features. For instance, the management and maintenance of civil engineering infrastructure depend on information about the environment, loading, structural health, etc. With many infrastructure components in poor condition [1] and recent prominent structural failures in both bridges and mining applications, the ability to sense impending failures and to alert pertinent personnel is desirable. Embedded systems are often highly tailored to the intended application, but flexible systems address the need for rapid deployment and the desire to reduce development cost. However, the use of individual development kits for prototyping of embedded systems often lacks the overall needs for temporary deployment. An embedded system that can be rapidly configured for a specific implementation has value for both applications and development. Therefore, it is also desirable to maintain an ability to use the F1 in development applications. Applications of the hardware as configured in the base state include diagnostics, prognostics, state observers, networking [2-7] and controls applications.

In this work, the Missouri University of Science and Technology (MST), formerly the University of Missouri-Rolla (UMR), F1 mote is described which provides a general platform for sensing, networking, and data processing. The platform consists of an 8051 variant processor, two 802.15.4 variant radio platform options, micro Smart Digital (SD™) flash storage, USB connectivity, RS-232 connectivity, and various sensor capabilities. The design, construction and an application example are shown for the F1 mote. The mote features:

- A small footprint to aid in embedding the device into structures and machinery,
- Low-power consumption for durability (generally less than 0.15 Watts),
- Integrated sensing support, and
- Computational abilities for distributed processing and intelligent networking

The F1 mote platform integrates the needed functionalities for several applications into a single deployable package. Also, the base design has suitable resources to directly support monitoring applications and related prototyping. While other platforms have been proposed that support the listed types of applications [8-10], many are only networking platforms when in their basic form; requiring daughter boards for any sensor usage. The F1 is a comprehensive attempt to transition from motes to sensor nodes by including an assortment of sensors in the basic configuration. As sensors become more integrated and systems-on-a-chip (SoC) becomes more commonplace, the abilities of wireless sensor networks (WSN) will gravitate towards a truly integrated system with a suite of sensing capabilities. Currently few systems exist that satisfy the needs of a development environment and already include basic sensing and networking capabilities to support direct deployment. The F1 attempts to fill this need with a novel combination of deployment, development, and research features to provide a multi-role mote.

II. System Architecture

The architecture for the F1 mote provides an adaptable configuration to facilitate use in several applications and as an expandable yet usable multi-role system. Many mote systems require expansion boards for additional capabilities and while elimination of such modularity is not desirable, a balance is needed to provide more functionality as a base design. Additional boards increase the size and the power consumption. For

instance, a network application normally requires a specialized board for interfacing to the PC or for a wireless link. The F1 platform can be used as a network interface via USB or RS-232 and has a wireless base capability providing greater flexibility in its roles on the network.

Two daughter boards for a thermocouple and expansion connector were designed to supplement the main board. . They are part of the overall mote when applications warrant; however, these boards can be removed when not needed. In Fig. 1 the printed circuit board (PCB) of the F1 mote is shown. Perforations can be seen where the main board and the thermocouple and expansion connector boards may be separated when needed. When not separated all the boards are electrically connected and function as a single unit allowing for adaptability when being considered for deployment. When removed, both daughter boards can be used via cabled connections as needed.

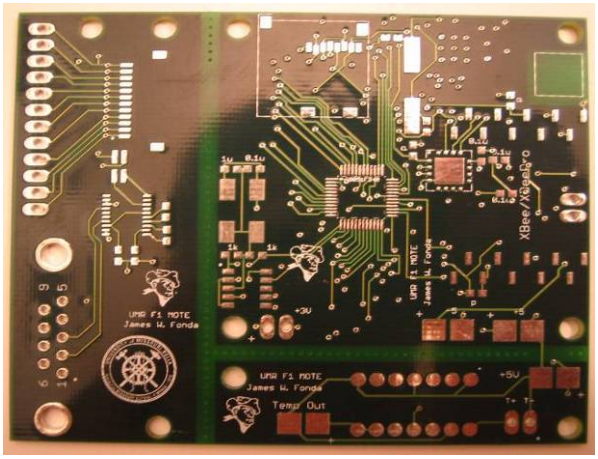

Fig. 1. F1 Mote PCB showing the main-board (top-right), the thermocouple board (bottom-right), and the expansion connector (left)

Fig. 2 is a block diagram of the system architecture for the F1 mote. The different interfaces and interconnections among devices are shown. Interconnections among different subsystems present on the processor and to the rest of the mote are also shown. The serial peripheral interface (SPI) bus interconnects the processor with the Chipcon

CC2420, the SD™ card for non-volatile data logging, and a multiplexer. The C8051F340 was selected due to the inclusion of an external memory interface (EMI) which facilitates the interface to the external ram (XRAM). The XRAM footprint was chosen based on popular 128 kB models and several variants can be populated on the board. UART0 is used to interface the other 802.15.4 radio that is the XBee/XBeePro variant. Finally, there are two interface types for PC interfacing. The C8051F340 is capable of being a USB device and communicating with a PC while a traditional RS-232 interface is also included to communicate with other devices as well as a PC.

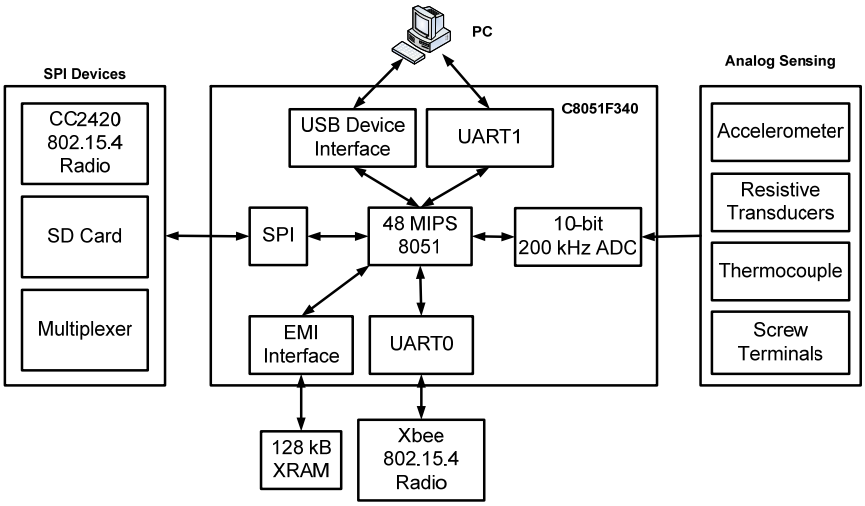

Fig. 2. F1 Mote Architecture

The F1 mote is pictured in Fig. 3 showing the components of the populated PCB. The mote is constructed on a 4-layer PCB in order to provide extended trace routing capabilities, to improve noise performance, and to allow for a more compact design. The F1 mote main-board is a 2” by 2” board that includes two radios, a processor, resistive transducer conditioning, DC-DC converter, RAM, and micro-SD card. Also, a multiplexer creates a switch-bank for use in enabling and disabling devices to conserve general purpose input-output (GPIO) pins on the microcontroller. The multiplexer is used

to interface with reset and enable lines that do not have to be switched at a fast rate. This feature conserves port-pins for analog interfacing and other communication busses such as the SPI and the I2C buss.

While the F1 does not provide the modular methods as seen in Mitchell et. al. [8], it does rely on modular layout of the PCB to accomplish the design criteria. During layout of the PCB the components are grouped according to functionality. Each subsection is then completed independent of the others, for instance the XRAM and CC2420 sections were completed independently of the rest of the PCB. The independent segments can then be laid out with the specified considerations, such as constraints introduced with RF devices. Specifically, the balun circuit and antenna layout on the CC2420 must assure good power transfer as well as reduce electromagnetic interference (EMI). Finally, interconnection of the subsystems can be made by through the leftover PCB area. Additionally, the 4-layer PCB design allows for easier layout and more compact design while also helping to reduce EMI issues such as cross-talk and grounding issues.

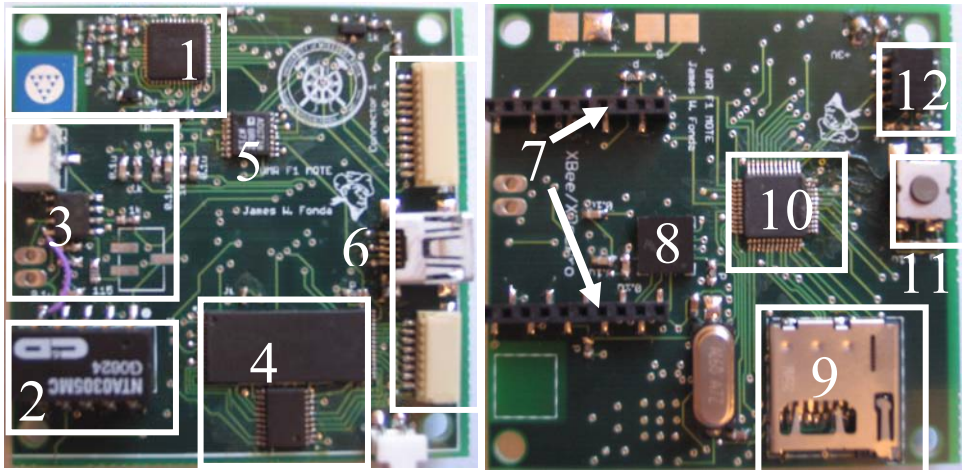

Fig. 3. F1 Mote. (a) Bottom side of the mote showing: 1-Chipcon CC2420, 2-DC-DC Converter, 3-Resistive Transducer Bridge, 4- XRAM, 5-multiplexer, and 6- USB and ribbon-cable connectors. (b) Top side of the mote showing 7-XBee Connectors, 8-accelerometers, 9-micro-SD card slot, 10-C8051F340, 11-reset button, and 12-JTAG debugger connector

III. Component Selection

Selection of components was carried out with consideration to both cost-effectiveness and performance criteria. Through integration of processing, memory, communications, and peripheral support the F1 mote provides a complete solution for embedded monitoring and networking.

A. Microcontroller

The microcontroller, or processor, was selected from the Silicon Laboratories® 8051 family. A USB enabled microcontroller was selected to provide connectivity options other than standard RS-232. The C8051F340 variant was chosen and is shown in Fig. 3(10). The processor includes builtin analog-to-digital converter with up to 29 inputs, an on-board voltage reference, a temperature sensor, a 16-bit programmable counter array (PCA), and two comparators. Also included are on-board voltage regulator provides 3.3 V_{DC} power from a $5V_{DC}$ input via battery or the USB connection. Additionally, an external memory interface (EMI) to provide support for parallel memory chips is included. Finally, an adjustable clock source is included on the processor with high and low clock-speed ability. The combination of features provides a single chip solution for many of the functions needed for support for sensors and peripheral devices.

B. Processor EMI Accessible RAM

Expansion of memory space was necessary to encompass various deployment needs such as the ability to queue packets from multiple sources during networking. Parallel RAM was added to the F1 to increase the total memory available from 5kB to around 260kB. An address latch, to conserve GPIO pins, and a TSOP package RAM footprint was added as shown in Fig. 3(4). The parallel RAM interface is supported by the microcontroller EMI and was physically designed to support both TSOP and short-TSOP

packages to increase the number of memory variants that are usable. Notably, special attention was given for the inclusion of ferrous RAM (FRAM) to provide non-volatile memory access on the EMI. While FRAM is non-volatile, it has a lower access speed and can be ill-suited for some applications and is withheld from normal deployments. For applications where FRAM is not suitable but non-volatile memory is needed, a micro-SD card was also added.

C. SD Card

A micro-SD card, shown in Fig. 3(9), was also added to the F1 mote to increase non-volatile storage capacity. Additionally, the use of the SD memory was accomplished with the SPI bus. In addition to data-logging, other aspects of the SD card can be used for storage of large amounts of information due to their low-cost and high memory-density. One desired capability was to provide a code bank for application specific code that cannot fit on the base code-space of the processor.

D. Radio Platforms

Two radio platforms where chosen for the F1 mote. A Maxstream XBee/XBeePro 802.15.4 platform, Fig. 3(7), was chosen for its ease of use and for its FCC certification. The XBee series provides a practical solution for many applications. However, for purposes of networking research, there are times when more access to the physical layer is desired, i.e. direct interface over the SPI bus. For this reason, the Chipcon CC240 chipset, an 802.15.4 variant, was also included as shown in Fig. 3(1).

E. Connectors

Interconnections on the F1 support daughter boards as well as external sensors. The main interconnects are via two 1mm pitch ribbon cable connectors providing access to all main communication busses, ADCs and GPIO. In Fig. 4, the expansion connector board

is shown with screw terminals and a DB9 RS-232 connector. The screw terminals allow easy connection to new sensors and systems. Additionally, support for RS-232 devices is provided by a Maxim IC MAX3380 transceiver through the DB9. Additionally, TTL level signals can be routed through the screw terminals. A set of jumpers allows for routing of the output of the RS-232 signals through either the screw terminals or the transceiver.

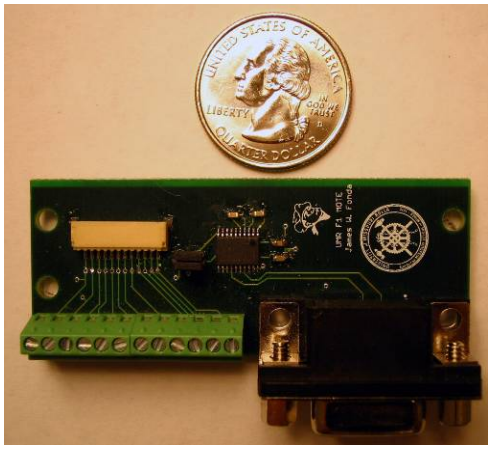

Fig. 4. Screw Terminal Breakout Board featuring DB9 RS-232 (bottom-right), screw-terminals (bottom-left), ribbon cable connector (top-left), and RS-232 transceiver (middle).

F. Power Supply

Main power can be supplied from either USB or battery connections. USB connections can supply $5V_{DC}$ to the microcontroller to power the on-board regulator. The regulator can then be used to power the main board; however, the regulator has a limitation of only 100 mA. Therefore, for full utilization of all components, battery power is required. Battery selection for the F1 mote is application specific since differing applications will require different mote lifetimes; however, any type of $3.3V_{DC}$ nominal battery will allow the mote to function.

In addition to battery connectors and the on-board regulator of the microcontroller, a DC-DC converter was added to provide power for sensors that require $\pm 5V_{DC}$ as well as

the resistive transducer bridge. A C&D Technologies NTA0305MC, shown as Fig. 3(2), was selected to satisfy power requirements. The DC-DC converter provides $\pm 5V_{DC}$ from a $3V_{DC}$ input with a maximum of 100 mA on each leg for a total output of 1 Watt. With an efficiency of 78% the DC-DC converter provides the ability to power many embedded sensors that can be used with the F1 mote while not taxing the energy reserve. The outputs of the DC-DC converter are also routed to pads to allow for external supplies to be used when the DC-DC converter is not present, or to route power out to other devices easily.

G. Sensors

Sensor selection for the F1 mote includes a three-axis accelerometer, resistive transducer bridge, and thermocouples. This selection of sensors supports several applications such as civil infrastructure monitoring, machine monitoring, and robotics. While all sensor types were not included in the base design, the F1 mote provides several useful capabilities in one package. The ability for the F1 to be modular through daughter boards and retain the advantages of a more dedicated design allows it to be fielded or be employed for development work.

1) Accelerometer

An accelerometer was included on the main PCB. The selected part was the Freescale Semiconductor MMA7261QT unit as shown in Fig. 3(8). The unit is a three-axis micromachined accelerometer with selectable ±2.5, ±3.3, ±6.7, and ±10g ranges in a 6mm by 6mm footprint. Through the use of a simple two-pin interface, the accelerometer range can be changed dynamically to suite the needs of the application. Additionally, the ranges can be changed when a particular level is reached to maintain dynamic range. Also, the accelerometer includes sleep mode and low active mode power consumption;

however, even in active mode, the unit only consumes 1.8 mW. By comparison, the processor and radio nominally consume a combined 160 mW.

2) Resistive Transducer Bridge

A resistive transducer bridge channel was also included on F1, shown in Fig. 3(3), due to the wide array of sensors that can be used with this type of circuitry. Strain, temperature, position, light level, etc. are common measurands for industrial sensors that are based on resistance change resistance. In Fig. 5 the circuit diagram is shown where V_{ex} is the excitation voltage, R_{gauge} is the active sensor branch, and R_{gain} is the gain adjustment potentiometer. Additionally, there is a 10Ω potentiometer placed in the lower branch to provide offset compensation. The resistive bridge circuitry as shown is configured for electrical resistance strain sensors (ESG), and ESG sensors provide suitable measurements for load testing of structures and devices as well as being embeddable. Additionally, the resistive bridge circuitry is reconfigurable by changing the resistances in order to support other available sensors for measurement of force, pressure, and displacement.

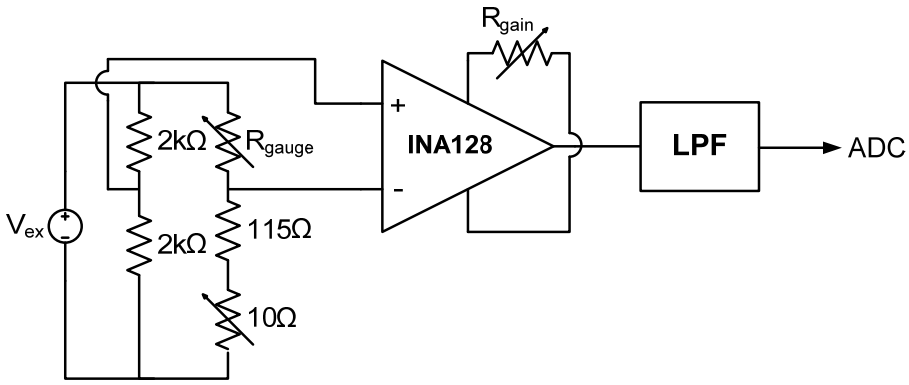

Fig. 5. Resistive Transducer Bridge Circuitry

3) Thermocouple Amplifier

Measurement of temperature at point locations is required for many applications. Thermocouples are an alternative to the temperature sensor on the microcontroller. An AD595 thermocouple amplifier was selected from Analog Devices and is shown in Fig. 6. A components-off-the-shelf (COTS) solution for the amplifier was an advantage due to the calibration needs of this chipset. The amplifier is laser-trimmed for calibration at the factory to provide off-the-shelf 1°C accuracy. The thermocouple amplifier was placed on a daughter board due to the size of the amplifier package. Like the connector expansion card, the thermocouple is also attached to the main PCB unless removed during the population of parts giving it the same flexibility as described for the connector expansion board.

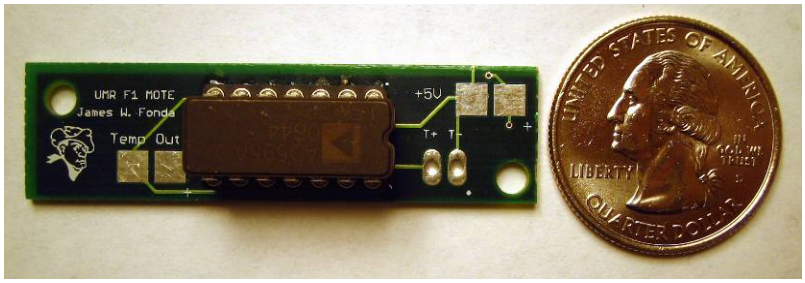

Fig. 6. Thermocouple PCB

IV. Comparison to Existing Platforms

Many systems used in laboratory experimentation are large and require AC power sources which limits them to for laboratory applications or temporary field testing. In order to provide extended capabilities, embedded systems are desired.

To support civil infrastructure testing, a field capable system was developed at MST and is shown in Fig. 7. The MST Orange Box is shown that utilizes National Instruments PCI-X systems and a PC interface. While the Orange Box is used for field testing, it is ill-suited for continuous monitoring due to size and power requirements. Previous work to

shrink the capabilities of larger monitoring systems to an embedded platform can be seen in via the 2” by 2” Generation-4 Smart Sensor Node (G4-SSN) [8] as shown in Fig. 8. The G4-SSN was originally developed at MST and subsequently updated at St. Louis University. The G4-SSN has various abilities for sensing and processing. The former include strain gauges, accelerometers, thermocouples, passive and active filtering, and general A/D sensing. The later include analog filtering, CF memory interfacing, and 8-bit data processing at a maximum of 100 MIPS. Comparison of the G4-SSN to other platforms can be found in Mitchell et. al. [8].

Fig. 7. MST Orange Box. (a) Front of the Orange Box showing the PC interface. (b) Rear panel showing connectors for sensors that include thermocouples, load cells, strain gauges, and LVDT’s.

Fig. 8. MST/SLU G4-SSN Platform.

While both the MST Orange Box and the MST/SLU G4-SSN provide adequate solutions for their application space, a more compact solution is needed for long-term

applications. The solution proposed by the F1 mote encompasses the G4-SSN's ability to provide a development platform and for modular expansion. While the stack configuration is abandoned for the F1 mote, ribbon cable connectors provide a method for connecting to new sensor PCB's or other equipment outputs. Also, for processing power the G4-SSN provides more pure computational resources; however, for many applications, the resources of the C8051F340 are more than adequate. Also, in comparison to other platforms, the F1 mote provides a more integrated research and deployment platform through inclusion of two RF interfaces, selectable range accelerometer, and support for a wide array of sensors via the resistive transducer bridge. These capabilities allow the F1 to provide a wider application space for deployment situations while maintaining a small footprint.

V. Applications fo the F1 Mote

The MST smart composite bridge (SCB) provides an example of how distributed wireless sensor networks with F1 motes can be deployed to infrastructure monitoring. The SCB was installed at MST in 2000 and is used as a testbed for improvements in civil infrastructure monitoring. The SCB is shown in Fig. 9 with a full-size pickup used in load testing and the locations for the F1 WSN nodes and instrumentation panel. The bridge is a fiberglass and carbon-fiber structure comprised of layers of square tubing. The load-bearing bottom and top layers are thirty-three tubes wide and these lay on the top and bottom of a network of four I-beam configuration. For a description of the structure and previous fatigue testing refer to Chandrashekhara and Watkins et. al. [11, 12].

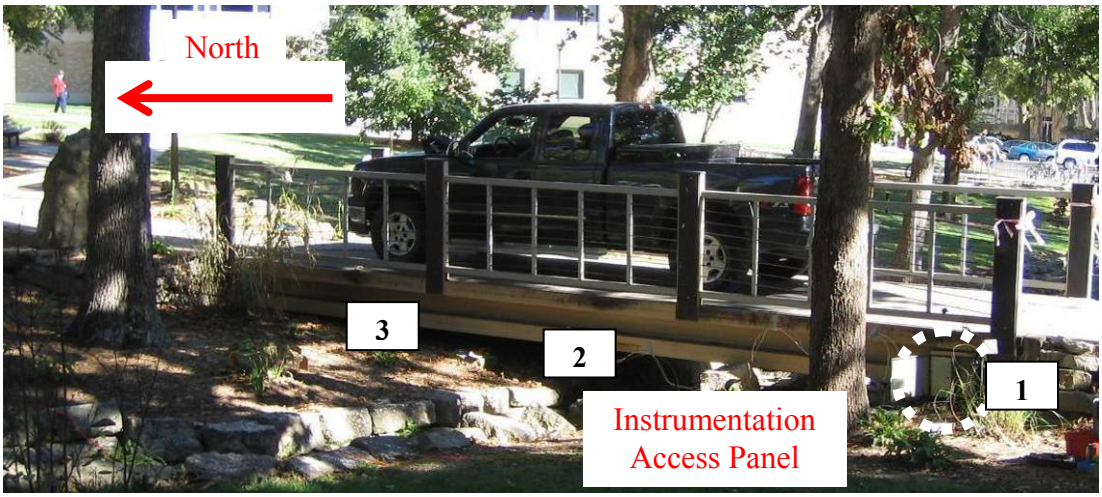

Fig. 9 Photo of the MST SCB

Installation of hardware takes advantage of the small size of the sensors and the motes. By embedding the sensors into the structure using adhesives they are protected for the lifetime of the sensing project. Also, due to the small size of the F1 mote it can be attached to the structure, or machine, using a small enclosure that can protect it from the weather. In general, the use of installation guidelines such as ones presented in Fonda et. al. [13, 14] were used. The motes were attached to the structure using small screws while the ESG sensors were attached using standard acrylic adhesive. The ESG sensors are MicroMeasurements model number EA-06-250BG-120 with a nominal resistance of 120Ω and a gauge factor (GF) of 2.070. The instrumented locations are at the mid-span and quarter-span lengths on the middle tube of the load-bear layer on the bottom of the bridge.

The SCB was subjected to both static and dynamic loading while monitored by the WSN. The loading was performed with a full-size pickup where the heaviest loading is seen over the front axles of the vehicle. Loading was varied by placing weight in the rear of the truck to provide a light and heavy loading. The difference in the loading in the rear of the truck was 400 lbs, which is approximately to 7.5% of the gross vehicle weight

which is approximately 5250 lbs. Since many trucks have 60% of the weight on the front and 40% on the rear this effectively even out the weight of the truck on the axles. The change in weight is used to show how strain profiles change with distribution.

Strain readings for the dynamic loading are shown in Fig. 10. The figure shows the profile of the strain as the load vehicle drives slowly across the structure. The data was filtered with a low-pass Butterworth filter of 2nd order and a cutoff frequency of 10Hz. Even with the filtering it can be seen that the strain profile still contains noise [13, 14]. A portion of the noise is unavoidable, the remainder is due to the low strain values and the high amplification associated with circuitry. The second curve is a polynomial fit of the data to provide a nominal reading for comparison to other tests.

The use of the strain profiles allows for the investigation of the behavior of the structure under dynamic conditions. It can lead to new load testing procedures that are automated and more rapid. This information coupled with the long-term estimation of physical parameters by the previously proposed prognostics method provides the needed visibly and information flow that infrastructure managers need to maintain deployed systems and future deployments.

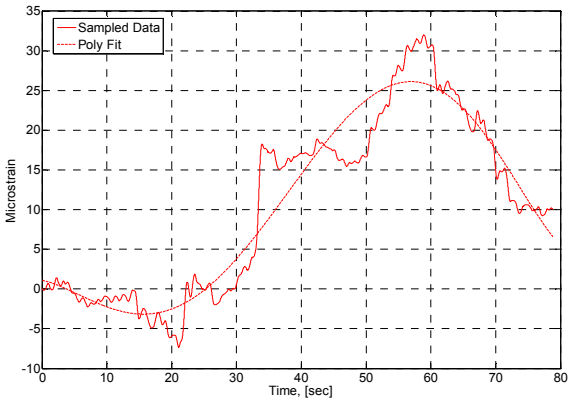

Fig. 10 Example Strain Reading for Dynamic Loading of the MST SCB

The static loading results are shown in Table I. Static results were collected by averaging the strain readings for each gauge over the time that the vehicle is parked in the loading position. While dynamic loading showcases the ability of the WSN to obtain dynamic data for a structure, the actual method used for structural evaluation is typically based in static loading in percentages of the AASHTO scale [15]. The loading positions where selected at intervals of 7.5 feet across the length of characterize the bridge behavior under loading. Results show a difference in strain between the two loadings. The difference between the mid-span and quarter-span for the heavy and light loading is negligible. This behavior is most likely due the same weight being over the front wheels for both tests. However, the loading for quarter-span with a heavy truck at position 2 shows the extra weight increases the strain. The embedded motes with their wireless capabilities provide quantitative information which is superior to qualitative inspection, provide for more rapid testing than temporary instrumentation, and can be a necessary component of an automated load testing system [16]

TABLE I STATIC LOAD TESTING RESULTS

Loading	Strain	Position 1 [με]	Position 2 [με]	Position 3 [με]
Heavy	Quarter-span	17	45	55
	Mid-span	15	30	40
Light	Quarter-span	15	24	22
	Mid-span	9	23	20

VI. CONCLUSIONS

The F1 mote provides a comprehensive solution to applications of civil infrastructure monitoring, robotics, and machine monitoring. The F1 mote features a small footprint and low power usage with sensing and computation capabilities in an embedded package. Although it was initially designed for use in civil infrastructure applications such as sensing parameters on buildings, bridges, and levees, the F1 mote can be adapted for use

in robotics, autonomous aircraft, and health monitoring of machinery. For robotics, the three-axis accelerometer contributes to an inertial guidance system for speed and roll-over control. For machine monitoring, the temperature sensing, the accelerometer, and the data-logging feature all provide capabilities needed for providing visibility to machine component health. In each of these application examples, the need for embedded processing and sensing is required in a platform that is easy to maintain and that provides wireless networking capabilities. The design process for the F1 mote illustrates the technical tradeoffs for such hardware. An increase in capabilities must be balanced with increases in size and in power consumption. Efficient design will use the fewest components for the desired capabilities. Also, miniaturization of components and modular layout provide flexibility.

Applications at MST with similar motes include structural monitoring, robotics, pneumatic system diagnostics, and online prognostics. The F1 provides a novel combination of sensing, processing, and application possibilities for the targeted application areas while retaining the usefulness of a development and research oriented platform. Possible mote enhancements include daughter boards with specialized hardware and other sensor types. The wireless capability of the F1 mote offers the possibilities of coordinated networks and of advanced embedded applications.

ACKNOWLEDGMENTS

This research was supported in part by NSF grants ECCS#0633769, EEC-0639182, the GAANN Fellowship and an Air Force Research Contract (FA8650-04-C-704) through Center for Aerospace Manufacturing Technologies. The University of Missouri-Rolla became Missouri University of Science and Technology as of January 1, 2008.

REFERENCES

[1] American Society of Civil Engineers *The 2005 ReportCard for America's Infrastructure* ASCE, available online: http://www.asce.org/reportcard/, 2005.

[2] B. Warneke, M. Last, B. Liebowitz, and K. S. J. Pister, "Smart Dust: Communicating with a Cubic-Millimeter Computer," *Computer,* vol. 34, no. 1, pp. 44-51, 2001.

[3] J. Polastre, R. Szewczyk, C. Sharp, and D. Culler, "The mote revolution:Low power wireless sensor networks," *Proc. of the Symp. on High Performance Chips (HotChips),* Aug 22-24, 2004.

[4] "Low Cost Wireless Technologies Enable New Sensor Networks," *High Freq. Electronics* vol. 4, no**.** 5, pp 32-34, 2005.

[5] R. M. Kling, "Intel Motes: advanced sensor network platforms and applications," *IEEE Int. Microwave Symp. Digest (MTT-S),* pp. 365-368, 2005.

[6] J. W. Fonda, M. J. Zawodniok, J. Sarangapani, and S. E. Watkins, "Adaptive Distributed Fair Scheduling and Its Implementation in Wireless Sensor Networks," *Proc. of the IEEE Conf. on Systems, Man and Cybernetics*, Taipei, Taiwan, pp. 3382-3387, 2006.

[7] J. W. Fonda, M. J. Zawodniok, S. Jagannathan, S. E. Watkins*,* "Development and Implementation of Optimized Energy-Delay Sub-network Routing Protocol for Wireless Sensor Networks," *Proc. of the IEEE Int. Symp. on Intelligent Control*, Munich, Germany, pp. 119-124, 2006.

[8] K. Mitchell, S. E. Watkins, J. W. Fonda, and J. Sarangapani, "Embeddable Modular Hardware for Multi-Functional Sensor Networks," *Smart Mater. Struct*., vol. 16, no. 5, pp. N27-N34, 2007.

[9] Crossbow Technology, Crossbow Technology –Revolutionary Wireless Sensors and Inertial Systems, available online: http://www.xbow.com, Dec. 2007.

[10] TinyNode, TinyNode, available online: htttp://www.tinynode.com, Dec. 2007.

[11] K. Chandrashekhara, S. E. Watkins, A. Nanni, and P. Kumar, "Design and Technologies for a Smart Composite Bridge," *Proc. of the IEEE Conf. on Intelligent Transportation Systems (ITSC),* pp 954-959, 2004.

[12] S. E. Watkins, "Smart Bridges with Fiber Optic Sensors," *IEEE Instrument and Measurements Magazine,* vol. 6**,** no. 2, pp. 25-30, 2003.

[13] S. E. Watkins, J. W. Fonda, and A. Nanni, "Assessment of an Instrumented Reinforced-Concrete Bridge with Fiber-Reinforced-Polymer Strengthening," *Opt. Eng.,* vol. 46, no. 5, 051010 2007.

[14] J. W. Fonda, Watkins, S. E., "Health Monitoring of a Truss Bridge using Adaptive Identification," *Proc of the IEEE Conf. on Intelligent Transportation Systems Conf. (ITSC)*, pp.944-949, 2007.

[15] American Association of State Highway and Transportation Officials, "Standard Specifications for Highway Bridges," 16th ed., *AASHTO*, Washington D.C., 1996.

[16] S. E. Watkins, T. M. Swift, and J. W. Fonda, "Development of Autonomous Triggering Instrumentation," *Smart Structures and Materials 2008: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems,* SPIE Proc, vol. 6932, San Diego, CA, March 2008, In Press, 2008.

VITA

James William Fonda was born August 11th 1978 in Cape Girardeau, Missouri, to David J. and Donna K. Fonda. James received his High School education and Diploma (1996) from Woodland R-4 High School in Marble Hill, Missouri. He then attended Mineral Area Community College in Park Hills, Missouri from August of 1996 to December of 1999. In January of 2000 he began attending the University of Missouri-Rolla (UMR) where he received a BSEE in December of 2001, MSEE in December 2004, and Ph.D. in Electrical Engineering in May 2008.

James spent the fall of 1999 as an exchange student in Canterbury, England at Canterbury Christ Community College University and also was an intern at Continental AG in Hannover, Deutschland in the summer of 2001. James is a member of IEEE and HKN and served as Vice President of HKN-Gamma Theta Chapter in spring of 2001 as well as President of HKN-Gamma Theta Chapter in fall 2001 and spring 2002. James received the Gamma Theta Chapter's Leadership award in the spring of 2001 and was a recipient of the Skitek scholarship in the UMR ECE department and recognized as the Outstanding Student Member for 2003 by the St. Louis Section of IEEE. James has also been a laboratory instructor and lectured AC Circuits at UMR. During his Ph.D. work he resided in the Applied Optics Laboratory and the Embedded Networking and Systems Laboratory as well as being a GAANN Fellow for Wireless Sensor Networks under Jagannathan Sarangapani.