Analysis of Cross-Layer Routing Protocols for Quality of Service in Real-Time Wireless Ad Hoc Sensor Networks

William S. Hortos

Associates in Communications Engineering Research and Technology (ACERT),
9143 Galleon Court
Orlando, FL 32819
+1 407-876-1026

whortos@embarqmail.com

ABSTRACT

The real-time interactions among the nodes of a wireless sensor network (WSN) to cooperatively process data from multiple sensors are modeled. Quality-of-service (QoS) metrics are associated with the quality of fused information: throughput, delay, packet error rate, etc. Multivariate point process (MVPP) models of discrete random events in WSNs establish stochastic characteristics of optimal cross-layer protocols. In previous work, discrete-event, cross-layer interactions in mobile ad hoc network (MANET) protocols have been modeled using a set of concatenated design parameters and associated resource levels by the MVPPs. Characterization of the "best" cross-layer designs for a MANET is formulated by applying the general theory of martingale representations to controlled MVPPs. Performance is described in terms of concatenated protocol parameters and controlled through conditional rates of the MVPPs. Assumptions on WSN characteristics simplify the dynamic programming conditions to yield mathematically tractable descriptions for the optimal routing protocols. Modeling limitations to determination of closed-form solutions versus explicit iterative solutions for ad hoc WSN controls are examined.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Network topology, wireless topology; C.2.2 [**Network protocols**]: Protocol architecture, routing protocols; G.3 [**Probability and Statistics**]: Distribution functions, multivariate statistics, queueing theory, stochastic processes; G.4 [**Mathematical Software**]: Algorithmic design and analysis.

General Terms

Algorithms, Design, Performance, Theory.

Keywords

Continuous-time stochastic optimization, cross-layer protocol design, dynamic programming, marked Markov process, martingale decomposition, multivariate point processes, multisensor fusion, wireless ad hoc sensor networks

1. INTRODUCTION

Wireless sensor networks (WSNs) are self-organizing systems intended largely for environmental sensing of remote terrains, operations in hostile fields, patient monitoring in hospitals, traffic surveillance, and so on. The primary focus of this paper is a study WSNs comprised of nodes capable of forwarding data to and from neighboring nodes; the constituent nodes are randomly placed in *ad hoc* deployments to fulfill mission objectives.

Despite similarities with mobile *ad hoc* communication networks (MANETs), WSNs instead rely on nodes in unattended operation with extremely limited computational, battery, bandwidth, and storage resources. Sensor nodes are typically densely deployed to compensate for limited resources. Dense deployment not only increases the likelihood of end-to-end connectivity but also saves energy expenditure at each node. Protocols for WSNs are more focused on extending the network lifetime of active nodes than on enabling high throughput, as in MANETs.

A generic WSN is composed of the following elements: the sensor nodes scattered in the field, each equipped with a radio for communications with other nodes; an information processing center for handling the information extracted from sensed data generated by the sensor nodes (and, if needed, issuing commands to sensor nodes); sink nodes for collecting the information extracted from the sensed data and forwarding commands from the processing center; and a conventional, wired or wireless, network to connect sink nodes to the processing center for information delivery as shown in Figure 1. Packets are transferred forward from sensor nodes to the processing center, through intermediate nodes, while packets containing commands and messages are sent at a lower rate in the reverse direction on bidirectional links. Only sink nodes communicate with the processing center, implying this node is more powerful than regular sensor nodes in processing and transmission capabilities.

Competitive interaction among nodes in a multimedia MANET is replaced with distributed and collaborative interaction among nodes of a heterogeneous WSN to process data cooperatively from nodes at network's edge to satisfy shared mission objectives. "Multimedia" in MANETs are comparable to data from multiple sensor types, e.g., audio, seismic, imaging, thermal, chemical, etc., in the WSN used to detect object phenomena. Quality-of-service (QoS) metrics associated with MANETs are supplanted with WSN metrics for the quality of fused information flow, i.e., throughput, delay, packet error rate, information gain, etc. Optimality conditions for cross-layer protocol designs, specifically the routing protocol, are developed. Source-to-sink routes are sought to optimize the cross-layer interdependencies to achieve the "best available" performance, represented one or more of the QoS metrics. Control is realized through conditional rates
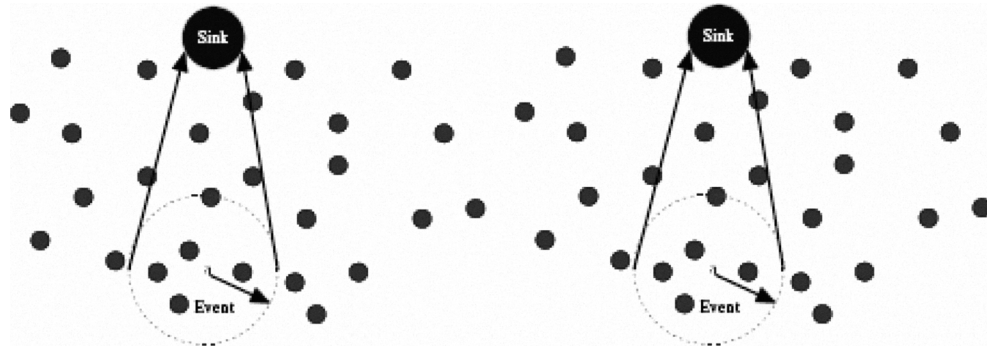
Figure 1. Typical data flows in two clusters of a WSN.

of MVPPs. The optimal cross-layer parameters are characterized by stochastic dynamic programming conditions derived from the packet flow models of the transient WSN over a finite mission horizon.[1], [2] The grouping of sensor nodes into clusters, the feed-forward character of data aggregation (fusion) at intermediate nodes within those clusters, and typically non-empty processor queues of packets at nodes, simplify the stochastic differential equations that describe network dynamics. The simplification leads to explicit iterative solutions to the dynamic programming conditions for optimal controls.

This paper is organized as follows. Section 2 introduces cross-layer interactions among layers of WSN protocols, with parameters that correspond to QoS metrics. The MVPP models for packet flows are formulated to construct random processes of the QoS metrics and network state in Section 3. In Section 4, the theory of MVPP models is used to develop optimality conditions for routing policies that provide best-effort QoS, given complete observations of network history. Section 5 discusses limitations to finding closed-form solutions versus iterative solutions to the explicit optimality conditions for the *ad hoc* WSNs. Section 6 is a summary of results and proposal for further research.

2. QUALITY OF SERVICE
Unlike wired networks, QoS support in WSNs depends not only on *availability of restricted resources* but also on the *transient rate* of such resources. Since WSN nodes are stationary or limited in mobility, the transient resources may cause link failures and disconnected paths, resulting in unwanted network partitioning. Mission applications must adapt to transient and restricted resources. Thus, the QoS allocated to an application depends on the "quality" of the network. "Quality" is a function of the availability and stability of primarily energy-dependent resources. QoS routing in WSNs is viewed as the *provision of a set of resource parameter levels in order to adapt different sensed data, offered at sensor nodes, to the "quality" of the network, while routing information packets through intermediate processing nodes to sink nodes.* From this viewpoint, QoS routing under severe constraints is a mechanism that creates paths in the WSN, then selects and maintains paths satisfying QoS requirements for the packet flows, subject to residual power reserves at nodes and along paths, based on the available network "quality."

2.1 Cross-layer QoS Model
A cross-layer approach is proposed for the QoS model in WSNs. It groups resource parameters and performance metrics associated with protocol layers, i.e., the application, network, link/medium

access control (MAC), and physical (PHY) layers, then maps them to data classes.[3], [4], [5], [6] This approach is suggested by the strict dependence of QoS on network "quality," i.e., interaction and stability of resources on routes and at intermediate nodes.

2.2 Data Classes and Protocol Layers
Streams of sensed data in WSNs can generally be classified as non-real-time, although packets in the streams may be stamped with the time of occurrence. Some data from audio and video sensors can be broadly considered real-time. Real-time data can have distinctive constant bit rates (CBRs), such as 8-kilobits per second (kbps) and 13-kbps audio codecs, or variable bit rates (VBRs), used in interactive video. Excessive delay or delay variation (jitter) noticeably degrades the quality of real-time information. Data aggregation or fusion maintains robustness while decreasing data redundancy, but introduces latency in fused results. Jitter, however, is not a significant source of degradation in most WSN information processing, since stored images and video are used for detection and classification. Non-real-time applications, e.g., file transfers and other delay-insensitive communications, are maintained by available bit rates (ABRs); these applications can be transmitted in high-rate bursts, characterized by "on-off" processes. Packet transmission stops at the end of the data burst, since no information or data is generated during unpredictable "off" intervals. Conversely, transmission of real-time applications is continuous during an active session. Packetized data are transported to intermediate nodes for information extraction at high transmission rates in short-term sessions. Some QoS metrics are expressed in terms of interrelated parameters, e.g., receive signal strength (RSS), signal-to-interference-plus-noise ratio (SINR), bit-error rate (BER), and frame-error rate (FER); other metrics distinguish types of sensed data by bit rate, tolerance to fixed delay, event report delivery ratio, detection accuracy, and energy consumed per packet.[7]

Class 1 sensed events require real-time network connections with very low delay. Wireless video/acoustic sensors deployed in WSNs to monitor and convey events in the form of audio only, video only, or both video and audio. The video and audio streams are examples of Class 1. These streams are different from data-only streams: they consist of a greater number of sequential packets and require stronger guarantees on bandwidth, delay and jitter from the network. Class 1 applications receive higher processing priority than other classes. In energy-conserving protocol design, packet rates may be negotiated between two or

more alternate CBRs based on available bandwidth and other resources.

Class 2 sensed events are non-real-time, delay-sensitive, and connection-oriented with flexible delay requirements. Examples include MPEG-2 video, file transfer protocol (FTP), and similar applications associated with the transport control protocol (TCP). This class receives lower priority than Class 1. Transfer rates may be negotiated as VBRs within an acceptable range, based on mission requirements for QoS delay, subject to resource availability and stability.

Class 3 sensed events are message-oriented and delay-tolerant. Typical services are data file transfers from sensor nodes on the network edge to intermediate nodes, where aggregation, correlation, and other data fusion tasks occur. These data can be conveyed at the earliest possible time, especially if the time of the sensed event is included in data frames. Transfer rates can be adjusted continuously (ABR), based on available bandwidth and other resources, after the QoS requirements for higher-priority classes have first been satisfied.

The requirements for distinct classes of sensed events are mapped into a set of QoS metrics, each defined by application-layer parameters (ALPs) of the OSI model. Classes 1, 2, and 3 and the processing requirements of the classes are mapped to corresponding performance parameters and QoS metrics, as shown in Figure 2. Class 1 corresponds to events with strict *delay* constraints, so it is mapped to the delay metric of the ALPs. Class 2 corresponds to events requiring high *throughput* such as non-iterative video and bulk transactions and is mapped to the throughput metric of the ALPs. Class 3 is not defined by specific constraints and is mapped to *best-effort* processing for ALPs.
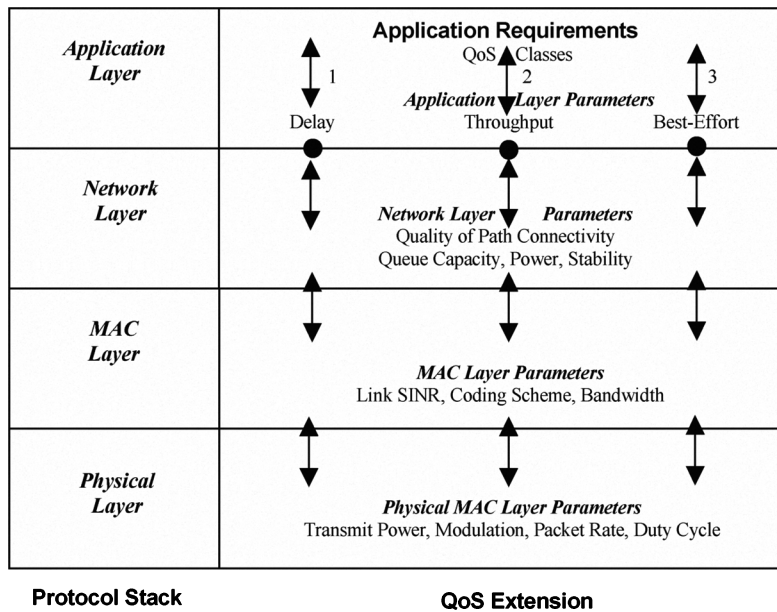

Figure 2. Layer interchange of parameters and metrics in cross-layer QoS model.

Network-layer functions include fast sensor-to-sink path reservation, packet store and forwarding, congestion control such as low-priority packet dropping, and other routing and flow control tasks. The on-off (active/sleep) power state, queue capacity, and route stability characterize network quality. These factors are termed network-layer parameters (NLPs). The power state represents the residual battery capacity on active routes during the mission. Queue capacity is the unallocated buffer space at nodes. Stability refers to link variability as measured by the connectivity variance of a node with respect to its neighboring nodes over time. To determine a path's "quality," concave and additive expressions are used to represent the NLPs of the path, given NLP values for the constituent nodes of the path.

Medium access control (MAC) layer, also termed the link layer, defines the strategy to access wireless channels. An ideal MAC offers high energy efficiency, low access delay, and different access priorities. In a wireless network there are four basic causes of extraneous power loss due to MAC design: (1) collisions, where a receiving node is in the range of two or more transmitting nodes but cannot receive an undistorted signal from either node; (2) overhearing, where nodes sense transmissions addressed to other nodes; (3) overhead, where nodes have to transmit/receive control traffic; and (4) idle listening, where a node senses an idle channel. Idle listening has been shown to consume up to 50% of the energy required for signal reception. Many access schemes must trade off energy efficiency and access delay; the latter is determined by design power-saving features based on active/sleep cycles, medium contention, time synchronization, etc. Access delay and access fairness are considered to be secondary in WSN design to energy efficiency.

Quality in the PHY layer is determined by transmit power, channel code selection, packet rate, modulation type, and antenna mode, e.g., beamforming that supports MIMO operation. The on-off cycle of transmit, receive only (idle), and sleep modes are ascribed to the PHY layer. As PHY-layer parameters (PLPs), these influence battery conservation and network lifetime. Transmit power and on-off cycles directly affect battery reserves at nodes, an NLP. Transmit power also influences channel interference, i.e., SINR, an MLP, as well as hop count and connectivity stability, both NLPs. Channel coding, modulation type, and antenna mode impact packet error rate (PER), link SINR, and hop counts of routes. Bandwidth and packet rate affect

throughput, an ALP. Moreover, packet rate influences queue capacity, an NLP.NLPs, MLPs, and PLPs govern network quality and power usage and thereby characterize energy-efficient paths that support the mission QoS metrics. ALPs determine the path, or paths, among routes with acceptable quality levels, that best satisfy the *application* QoS requirements with the lowest *and* best-distributed energy consumption among nodes. Optimization is a tradeoff among the multiple objectives to *adapt to the available quality of the network.*[8] Protocols can be realized in embedded versions of light-weight distributed multicast routing protocol schema (DMRPS) in [9] to dynamically construct and maintain routing trees between sources and sinks. Table 1 shows one possible mapping from QoS classes/requirements and energy-related metrics to the ALPs, NLPs, MLPs, and PLPs.

Table 1. Mapping QoS classes to protocol layer parameters

QoS Class/Requirement	ALPs	NLPs	MLPs	PLPs
Class 1	Delay	Buffer size, Hop Count, Shaping	SINR, Bandwidth Access Slot	Code, Packet Rate
Class 2	Throughput, Delay Variation	Buffer size, Hop Count, Shaping	SINR, Bandwidth Access Slot	Code, Packet Rate
Class 3	Best-effort	Connection stability, Hop Count	SINR, Access Slot	Antenna Mode, Modulation, Packet Rate
Remaining Battery Capacity at Nodes, on Paths	Throughput, Delay	Hop Count, Shaping	SINR	TX Power, Active/Sleep Duty Cycle, Antenna Mode

3. ANALYTICAL MODELS OF CROSS-LAYER DESIGN

The analytical framework for a wide range of traffic, link distortions, routing control as well as the metrics and constraints on cross-layer protocol design in WSNs is formulated in terms of continuous-time MVPP models. The MVPP models *generalize* Poisson point processes, exponentially distributed message processing, and other renewal processes commonly used to create queueing structures for the evaluation of control schemes for packetized data networks for which an equilibrium state exists. However, the dynamics of many WSNs rarely support an assumption of long-term system equilibrium. With this in mind, the models of transient packetized information flows in a WSN are restricted to a finite mission interval, $[0, T], T < \infty$.

Furthermore, the MVPP approach overcomes the limitations of conventional Markov and Bayesian models of discrete-space events. The latter models incorrectly assume successive observations are independent, so that the probability of a sequence of observations can expressed as the product of probabilities of individual observations. This assumption is not valid due to mobility in the sensed phenomena and competition for constrained resources during the mission. It has been shown that MVPP models also represent self-similar processes with long-range dependence (LRD) that characterize Internet traffic.[2]

Consider the protocol operations that regulate the flows of data packets in a heterogeneous WSN, that is, a network in which each node at the edge has multiple sensor types to detect physical phenomena in the sensed environment. The WSN protocol handles the arrival, aggregation, and information extraction from the sequences of detection events. The source node of a detection event can be any activated sensor node i of the M_{max} nodes comprising the WSN. The network nodes can be grouped or clustered into K neighborhoods (clusters) of M_l nodes, such that,

$$M_{max} \le \sum_{l=1}^{K} M_l,$$ assigned node index l^*. The cluster head may

be connected directly to one or more of the sink nodes or to other intermediate nodes.

The MVPP models are constructed from the counts on underlying network events. Uninterrupted sensor data arrivals occur at random F_t-stopping times, $\tau_0^a, \tau_1^a, ..., \tau_n^a, ...,$ such that the corresponding inter-arrival times, $\tau_1^a - \tau_0^a, \tau_2^a - \tau_1^a,$ $..., \tau_{n+1}^a - \tau_n^a, ...,$ are independent and identically distributed (i.i.d.) random variables with the *common* right-continuous PDF, $F^a(t), \tau_n^a < t \le \tau_{n+1}^a, \forall n \in N_+,$ the set of positive integers. Hence, the corresponding inter-arrival sequence forms a *renewal process*. To represent the statistical dynamics of sensed phenomena, PDFs may change between stopping times, i.e., $F_n^a(t), \tau_n^a < t \le \tau_{n+1}^a, \forall n \in N_+,$ thus violating the renewal assumption. The approach allows slotted and deterministic times of arrivals and processing completions, i.e., $\tau_n = T_n = nT_0,$ for T_0 a known slot time, frame time, or simulation time increment.

To model multi-sensor detections, arriving data from each sensor, from a set of at most S sensor types, may be assumed to require distinct processing. Sensed data from each type may require a different QoS, which, in turn, can be expressed in terms of protocol parameters and metrics. The MVPP models allow any of the S sensor types to be active during the mission and require different resource levels to ensure QoS during the session. Thus, the sequence of processing completion times $\tau_{i,m}^s$, for sensor type $s, s = 1,..., S,$ at node i and the corresponding sequence of inter-completion times, $\tau_{i,1}^s - \tau_{i,0}^s, \tau_{i,2}^s - \tau_{i,1}^s, ..., \tau_{i,m+1}^s - \tau_{i,m}^s, ...,$ may *not* be i.i.d., random variables or share a common PDF with the inter-completion time sequences for packetized data from other sensors. Consequently, superposition of the sequences of the inter-completion times corresponding where the weak inequality indicates that some overlap is allowed among the clusters. In

Figure 1, $K = 2$. Assume that at least one node in each cluster is an intermediate processing node, designated the cluster head to any two or more of the sensor types will *not* form a renewal sequence without the assumption of exponentially distributed inter-service times.[9] The modeling approach only requires the existence of semi-martingale decompositions for the MVPPs that describe the information flows originating from multiple sensor types. The semi-martingale decomposition of each MVPP is the sum of (F_t)-predictable, integrated, non-explosive rate processes and pure-jump martingales, with respect to a probability space $(\Omega; F_t; \wp)$ or controlled probability space $(\Omega; F_t; \wp^U)$.

3.1 Statistical Properties of Packetized Traffic

The packet from a source node at random time τ_n^a is assumed to carry data from one or more of S simultaneously active sensors at that node. The corresponding data arriving at τ_n^a is modeled as an embedded, vector-valued, discrete-space, discrete-time process $B_n = (b_{1,n}, b_{2,n}, \ldots, b_{S,n})$, where $b_{s,n}$ is the data load from sensor type s, $s = 1, 2, \ldots, S$, and can vary from sensed event to sensed event. The condition $b_{s,n} = 0$ indicates sensor s is inactive at τ_n^a. Since the load process $(B_n, n \in \mathbf{N}_+)$ for arrivals and the counting process for the number of new arrivals, denoted $(N_t^a, t \in [0, T])$, have different statistical properties, the possible combinations for the load-arrival process generate a set of hybrid MVPPs, based on the stochastic properties of the constituent processes. For example, if $(B_n, n \in \mathbf{N}_+)$ is a discrete-space, discrete-time Markov process, and $(N_t^a, t \in [0, T])$ is a Poisson process with time-varying rate, $\alpha_t, t \in [0, T]$, the resulting load-arrival process is a non-homogeneous, Markov-modulated Poisson process. Thus, the selection of constituent statistical properties enables creation of random processes used to model wired and wireless telecommunications traffic.[10]

A multi-processor model is constructed to represent the activities at any node i, with the inter-completion events at node processors, for data from sensor type s or information of type s, each characterized by a different PDF, $F_{i,s,t}^d, t \in [0, T]$. The random rate corresponding to type s at node i, conditioned on the event $\{\tau_{i,n}^s \leq t < \tau_{i,n+1}^s\}$, is $\sigma_{i,s,t \wedge \tau_{i,n+1}^s} = -\dfrac{d F_{i,s,t \wedge \tau_{i,n+1}^s}^d / dt}{\left(1 - F_{i,s,t \wedge \tau_{i,n+1}^s}^d\right)}$, where

$\tau_{i,n}^s$ is the time of the n'th processing completion at node i and "$\wedge$" denotes the infimum of two stopping times.[11], [12] As the PDF may also change *after* each time $\tau_{i,n}^s$, the construction allows a marked renewal sequence with a conditional PDF $F_{i,s,n,t}^d$ between the n'th and $n+1$'th processor completions. Martingale representation theory for MVPPs, applied to the count of uninterrupted packet-processing completions for sensor type s to time t, denoted $\breve{N}_{i,s,t}^D$, yields

$$E\left[\breve{N}_{i,s,t}^D\right] = E\left[\sum_n \int_{\underline{t}_n^s}^{\wedge \tau_{n+1}^s} \sigma_{i,s,v} dv\right], \text{ where } N_{i,s,t}^D - \sum_n \int_{\underline{t}_n^s}^{\wedge \tau_{n+1}^s} \sigma_{i,s,v} dv \quad (1)$$

is a zero-mean $(F_t, \wp)$-martingale, given F_t is a σ-algebra of network events to time t containing the history $\{\breve{N}_{i,s,v}^D, 0 \leq v \leq t\}$. The count $\breve{N}_{i,s,t}^A$ of uninterrupted arrivals to node i of type s to time t is also represented in terms of the conditional arrival rate $\alpha_{i,s,t}$ and the sequence $(\tau_{s,n}^a)$.

3.2 Statistical Properties of QoS Classes

Class 1 applications, e.g., data from audio and live video sensors, offer CBR loads and require minimal bit rates or bandwidths, although they do not benefit from resources well beyond the minima. Audio can be modeled by a non-homogeneous, Markov-modulated, Poisson process (MMPP), with one or two selectable CBRs, α_1 and α_2, the Poisson intensities. These rates are modulated by a random "on-off" process, V_A, with a mean "on" time equal to the audio activity cycle. Sensed audio is generally time-stamped and buffered at the sensor node. Class 1 data are not blocked or dropped by the processing nodes, unless the data are found to be redundant with already buffered sensed data.

Class 2 applications correspond to sporadic detection events that occur in time bursts. The data rates for these events fluctuate between low- and high-rate requirements. Previous studies have successfully used MMPPs to represent fused voice, video and data VBR traffic.[13]

Class 3 applications encompass sensor data from sustained or long-term detection events. Packetized data can flow with required quality at varying packet rates or, equivalently, within varying bandwidths. Packets for such connection-less processing are buffered at nodes. Arrivals and processing completions for ABR sensed events have been modeled by PDFs with parameters adjustable to available resources, such as those for marked renewal processes.

3.3 Energy-Efficient Routing

The cross-layer design for routing, satisfying required QoS, also must optimize energy efficiency along routes, since available power at nodes is the primary factor limiting network lifetime. *Load distribution* seeks to balance energy use among all nodes by selecting a path with under-used nodes rather than the shortest route. This may result in longer routes, but packets are routed only through energy-rich nodes. Protocols based on load distribution do not necessarily provide the lowest-energy route, but prevent overload at selected nodes, ensuring longer network lifetime.

The load distribution routine used is the Conditional Max-Min Battery Capacity Routing (CMMBCR) protocol.[14] The energy-aware route maintenance algorithm of the CMMBCR for WSNs has been described in detail for MANETs.[1] In CMMBCR, when all nodes on a set of paths have remaining battery capacity above a threshold θ, a path with minimum total transmission power among these paths is selected. Since less total power is required to forward packets on each path, the relaying load for most nodes can be reduced and their lifetime extended. However, if all paths have nodes with battery capacity below θ, a path including nodes with the lowest battery reserves must be avoided to extend the lifetime of these nodes. Battery capacity for source-sink path P_j at

time t is $R_j(t) = \min_{\text{node } i \in P_j} C_i(t)$, where $C_i(t)$ is the battery capacity of node i at time t. In the following, let $C(t) = \left(C_1(t), \ldots, C_{M_{\max}}(t)\right)$ denote the vector of remaining battery capacities of the nodes at time t, and $\theta_i(s)$ be the minimum requirement for the residual power of nodes requested by the sink for path P_j to support packets of sensor type s.

3.4 Routing Policies

The function of the WSN routing protocol is to forward data packets from sensed events at the network edge through processing at intermediate nodes on source-to-sink paths. Packet routing from node i to node j is modeled by a random vector $u_{ij,t} = \left(u_{ij,1,t}, u_{ij,2,t}, \ldots, u_{ij,S,t}\right)$. The s'th component of $u_{ij,t}$ is the indicator of the event of packet transmission from node i to node j containing data from sensor type s. More broadly, the entries of $u_{ij,t}$ can represent changes in processing or packet rate as well as processing cessations or interruptions, due to insufficient resources along the path. In mathematical terms, the components of $u_{ij,t}$ are $\left(F_t\right)$-predictable indicator functions of random events and are Borel-measurable. The expected values of the $u_{ij,t}$, with respect to the reference probability measure $\wp$ and outcome space Ω, are probabilities of packet forwarding, data correlation, and dropping that shape the network flows.

The $M_{\max} + 1 \times M_{\max} \times S$ routing array, $U_t =$

$\left(u_{ij,s,t}; 0 \le i \le M_{\max}, 1 \le j \le M_{\max}; s = 1, \ldots, S\right)$, $t \in [0, T]$,

can be decomposed into subarrays for the K clusters covering the $M_{\max}$ nodes. The array describes the general random connectivity due to routing, resource reservation, radio path distortions, and data fusion over mission period. The sum of entries of $u_{ij,t}$ over j may be greater than 1 to model point-to-point, point-to-multipoint, and broadcast packet transmissions among nodes. The set of routing arrays are closed under concatenation in time, i.e., $U = [U_1, U_2, t]$, $t \in [0, T]$, is also in the set if U_1 and U_2 are members and the arrays satisfy a stochastic causality property. The set of arrays $U_t, t \in [0, T]$, that satisfy these defining conditions, are referred to the *class of admissible routing policies*, $\mathcal{U}$.

Sample paths of the $u_{ij,s,t}$ form the temporal evolution of source-to-sink connections. In other words, for a request to node j_1 moving to node k, $(u_{j_1 j_2, s, \tau_1}(\omega), u_{j_2 j_3, s, \tau_2}(\omega), \ldots, u_{j_n k, s, \tau_n}(\omega))$ is one history of the path traversed by the packet flow bearing data from sensor type s, where $\tau_1(\omega) < \tau_2(\omega) < \cdots < \tau_n(\omega)$ are the n data arrival times or processing completion times at nodes $j_1, j_2, \ldots, j_n$, respectively.

3.5 Network Processes and Conditional Rates

The state process must have sufficient dimensionality to account for different sensor types, nodes, data sources and sinks in order to describe all possible events of sensed data detection and processing in the WSN. Packets from Class 2 and Class 3 can be queued during periods of deep fading, high channel interference, blocking, connection instability, low-battery capacity, and other

link disturbances or to allow preemption by more delay-sensitive sensor data, while Class 1 packets are given the highest priority at processors. The queueing process of packets at node i, containing data from *all* active sensor types, including those in or awaiting processing, at time $t > 0$, is the discrete-valued vector of parallel queues, $Q_{i,t} = \left(Q_{i,1,t}, Q_{i,2,t}, \ldots, Q_{i,S,t}\right)$. Each component of $Q_{i,t}$ has a corresponding birth-and-death equation and a semi-martingale representation as the sum of a discrete-valued jump, right-continuous, zero-mean, $(F_t, \wp)$-local (or $(F_t, \wp^U)$-local) martingale and an integrated conditional rate with respect to the family of σ-algebras $\left(F_t, t \in [0, T]; F_t \subseteq F\right)$[11], generated by the evolution of observed events to time t,

$$Q_{i,s,t} = Q_{i,s,0} + \left(Q_{i,s,t} - \int_0^t (\alpha_{i,s,v} - 1(Q_{i,s,v-} > 0) \sigma_{i,s,v}) dv\right)$$
$$+ \int_0^t (\alpha_{i,s,v} - 1(Q_{i,s,v-} > 0) \sigma_{i,s,v}) dv \qquad (2)$$

At nodes with a single processor, the total number of packets at node i at time t is the sum over the number of sensor types of the components (2) of $Q_{i,t}$. The *packet flow* state $Q_t = (Q_{1,t}, Q_{2,t}, \ldots, Q_{M_{\max},t})$ is augmented with $C_t = (C_{1,t}, C_{2,t}, \ldots, C_{M_{\max},t})$, the vector of battery capacities at the nodes at time t, to form the *network state* process, $X_t = (Q_t, C_t)$. The corresponding instantaneous (F_t)–progressively measurable conditional rates for packet arrivals and departures containing sensed data of type s at node i, where node 0 is the sensed environment, are given by

$$\alpha_{i,s,t} = u_{0i,s,t-}\alpha_{s,t-}$$
$$+ \sum_{j \in \{ \text{neighborhood of node } i\}} u_{ji,s,t-} 1(Q_{j,s,t-} > 0)\ \sigma_{j,s,t-} \qquad (3)$$

and $\sigma_{i,s,t}$, respectively. The exact forms of the conditional rates depend on the PDFs of the underlying events of packet arrival and processing completion for each type s, the observed network history to time t on which the rates are estimated, and the control policy $U \in \mathcal{U}$. In (2) and (3), terms $1(Q_{j,s,t-} > 0)$ indicate that uninterrupted packet processing cannot occur when no packets of type s are in queue at node i. Under heavy traffic conditions, typical of the nodes in many WSNs, $1(Q_{j,s,t-} > 0) \equiv 1$ and the expressions (3) simplify to a set of rates depending solely on the routing control and underlying PDFs of inter-arrival and inter-processing times. In the special case of non-homogeneous Poisson arrivals with deterministic, time-varying intensities, $\alpha_{i,s,t}$, and single-stage exponential processors with deterministic transient rates, $\sigma_{i,s,t}$, at the nodes, the *form* of the rates coincide with (3). At nodes with L ($L > 1$) processors, L simultaneous processing requests can be in progress, each with packet data from up to S sensors, yielding at most $L \times S$ processing modes. Packets arrive at node j from both sensor and intermediate nodes within its cluster. At node j the random instantaneous rates of packet arrivals and processing completions of sensor type s are given by

$$\alpha_{j,s,t} = \sum_{k \in \{\text{sources in neighborhood of node } j\}} u_{kj,s,t-}\alpha_{k,s,t-} +$$
$$\sum_{i \in \{\text{intermediate nodes in neighborhood of node } j\}} u_{ij,s,t-} 1(Q_{j,s,t-} > 0)\sigma_{i,s,t-}, \qquad (4)$$

$$\sigma_{j,s,t-} = \sum_{l=1}^{L} 1\left(Q_{j,s,t-}^{l} > 0\right) \sigma_{j,s,t-}^{l} \qquad (5)$$

respectively, where $\sigma_{j,s,t}^{l}$ is the packet processing completion rate of the l'th processor for type s at node j. While the $u_{ij,s,t-}$ in (4) and (5) determine the packet flows over the links comprising source-to-sink paths, connectivity is also determined by less controllable and often unobservable events, caused by constraints on buffer size, processor speed, queueing delays, battery capacities as well as the link disturbances due to multipath reflections, shadowing and channel interference. Thus, the general structure of $u_{ij,s,}$ is a product of indicators of both controlled, observed events and constrained, indirectly observable events. As the indicators of random events, expectations of the $u_{ij,s,}$ with respect to $\wp$ and $\left(F_{t}, t \in [0,T]; F_{t} \subseteq F\right)$ are probabilities of the events.

A family of probability measures $\wp^{U}$ on the outcome space Ω is constructed from a reference measure $\wp$ and the admissible routing policies $\mathcal{U}$. An absolutely continuous change of measure is constructed in terms of the local description, i.e., the conditional rates of MVPPs that define packet-flow behavior: $dN_{ijs,t} d\wp \rightarrow dN_{ijs,t} d\wp^{U}$, and the rates, $p_{ojs} \alpha_{t} \rightarrow u_{0js,t} \alpha_{s,t}$, $p_{ijs} \sigma_{is,t} 1(Q_{is,t-} > 0) \rightarrow u_{ijs,t} \sigma_{is,t} 1(Q_{is,t-} > 0)$ for $U \in \mathcal{U}$. Details are omitted but have been previously presented as a variation of the result by Doleans-Dade applied to MVPPs created for the WSN.[12], [15]

3.6 Protocol Parameters and QoS

QoS metrics depend on parameters of the protocol layers and the functions $\left(u_{ij,s,t}\right)$. In turn, throughput, delay, queue length, packet loss rate, SINR, battery reserve, occupied bandwidth, and connectivity stability explicitly determine the conditional rates of packet arrivals and processing of data from each sensor type. The constructed MVPP models address both *QoS-based* and *energy-efficient routing*.

As a specific example, PHY-layer adaptive beam-forming within a node's coverage area can simultaneously increase gain factors for BER and reduce co-channel interference. Assuming a uniform distribution of deployed nodes, sectoring coverage areas reduces interference and increases capacity by antenna gain factor, G_{A}. In general, $u_{ij,s,t}$ is the product of indicators that include $1\left(BER \leq 10^{-n}\right)$ for sensed data of type s. The latter indicator can be factored further into a product of event indicators for parameter thresholds that ensure the required QoS for the packet flow:

$$1\left(BER_{t} \leq 10^{-n}\right) = 1\left(P_{TX,i,t} \geq P_{s}\right) \cdot 1\left(G_{\text{coding}, ij,t} \geq 10^{y/10}\right) \cdot 1\left(G_{A,j,t} \geq g\right)$$
$$\cdot 1\left(P_{\text{avg loss}, ij,t} \leq \pi\right) \cdot 1\left(C_{i}(t) \geq \theta(s)\right) \cdot 1\left(C_{j}(t) \geq \theta(s)\right) \cdot$$
$$1\left(I_{\text{co-ch.interf.}, ij,t} \leq \varsigma_{s}\right) \cdot 1\left(BW \geq B_{s}\right).$$

For link fading, the condition $P_{\text{avg loss}, ij,t} \leq \pi$ can be replaced with indicators for the events of the number of replicas received at node j from a transmission from node i, according to a discrete-event distribution, with the relative path loss on each replica obeying a Rayleigh or Rician distribution. The joint PDF for the reflected paths and the amplitude of the reflections are used to determine the expected value of multipath fading. Traffic shaping with resource constraints is represented by expressing $u_{ij,s,}$ as a product of factors, $1\left(\text{Available } NR_{j,t} \geq \rho_{s}\right)$, where $NR_{j,t}$ is the resource at node j at time t required at level ρ_{s} to provide QoS for sensed data of type s.

3.7 Performance Metrics

Throughput and queue length. Throughput, an application-layer metric, is the time average of the number of packets of sensed data from each type delivered from a source to sink per unit of time. Related metrics include the distribution of the number of packet transmissions from each node, a PHY-layer metric; the time average of packet delay, an application-layer metric; the fraction of channel capacity used for successful transmission, a network-layer metric; and the probability of successful packet transmission, or "goodput," an application-layer metric. The number of packets based on sensed data from type s, $s = 1, \ldots, S$, at time t is the sum of the individual components in the queue state Q_{t}, that is, $Q_{\text{total},s,t} = \sum_{i=1}^{M_{\text{max}}} Q_{i,s,t}$.

Link throughput of sensed data from type s over subinterval $(v,t] \subset [0,T]$ for any $v < t$ from node i to node j is given by $\tilde{N}_{ijs,t-v} = [N_{ijs,t}^{C} - N_{ijs,v}^{C}] - [N_{ijs,t}^{NC} - N_{ijs,v}^{NC}]$, the difference between the number of successfully received and lost packets containing sensed data of type s, respectively, transported on link $<i,j>$ over the subinterval. The *node i throughput of sensed data from type s* over $(v,t]$, is given by, for $i = 1, \ldots, M_{\text{max}}$;

$$s = 1, \ldots, S; \text{ and } v \in [0,T]; \quad L_{is,t-v} = \sum_{j=1}^{M_{\text{max}}} \tilde{N}_{ijs,t-v}. \text{ The } system$$

throughput of sensed data of type s over $(v,t]$ is given by

$$L_{s,t-v} = \sum_{j=1}^{M_{\text{max}}} \tilde{N}_{id,t-v}, \text{where, it is assumed that, when an}$$

information packet based on data of type s, reaches a sink node, it is not returned. Lastly, the *system throughput to time t from node i* is given by $L_{i,t} = \sum_{s=1}^{S} L_{is,t}$. Note that the derived expressions for the throughput metrics are simple linear combinations of the set of underlying MVPPs $\left(N_{ijs,t}^{C}, N_{ijs,t}^{NC}\right)$; so these metrics have the $(\wp, F_{t})$-semi-martingale structure.

Link capacity. Denote by $U_{C}\left([0,t]\right)$ the set of array values of all admissible controls U_{C} over $[0,t], t \in [0,T]$. The *mean link capacity from node i to node j* to time $t \in [0,T]$ is given by $\kappa_{ij,t} = \max_{U_{C}([0,t])} E^{U}[\tilde{N}_{ijs,t}]$. *Mean capacity at node i* to t is given by $\kappa_{i,t} = \max_{U_{C}([0,t])} E^{U}[L_{i,t}]$, where $E^{U}[\cdot]$ represents integration with respect to $\wp^{U}$. The counting processes $N_{ijs,}$ admit $(\wp^{U}, F_{t})$-rates that depend on samples $u_{ijs,t}(\omega)$, $\omega \in \Omega$ in $U_{C}([0,t])$.

Packet dropping and fusion. The expected number of dropped and

fused packets are expectations of counting processes for the corresponding network events, summed over the node indices.

The Fubini Theorem can be invoked to represent the expectations, with respect to $\wp$, as the sums, over the desired node indices, of the integrals of expected (F_t)-predictable rates of corresponding

packet flows over $[0,T]$. Thus, the number of dropped and fused packets containing sensed data of type s at node j over the $[0,T]$ is represented as the $(F_t, \wp)$-semi-martingale

$$N_{drop/fuse,js,[0,T]} = \left[\left(N_{drop/fuse,js,[0,T]} - \int_0^T (1-u_{0js,v}) \mathbb{1}(Q_{ijs,v-} \geq q_{j,s}) \alpha_{s,v} dv \right) + \int_0^T (1-u_{0js,v}) \mathbb{1}(Q_{ijs,v-} \geq q_{j,s}) \alpha_{s,v} dv \right]. \quad (6)$$

The expected value is

$$E[N_{drop/fuse,js,[0,T]}] = E\left[\int_0^T (1-u_{0js,v}) \mathbb{1}(Q_{ijs,v-} \geq q_{j,s}) \alpha_{s,v} dv \right] = \int_0^T \left(P(Q_{js,v-} \geq q_{0js,v}) - P(u_{0js,v}(Q_{ijs,v-} \geq q_{j,s})) \right) \overline{\alpha}_{s,v} dv, \quad (7)$$

where the second term in the integrand on the right-most side of (7) is a joint probability, $\overline{\alpha}_{s,t}$ is the $\wp$-mean of the arrival rate of sensed data of type s at time t, and $q_{j,s}$ is the buffer size at node j for packets of type s. Instantaneous average rates of dropped or fused packets are integrands of the average number of the corresponding network events.

Delay. Causes of delay over source-to-sink routes are node mobility, blocking, channel contention due to congestion and interference, packet retransmissions in response to unrecoverable errors, queueing latency, and power failures at nodes. Delay sources other than queueing latency are assumed to be relatively small. The queueing latency or delay at a node can be bounded to not exceed a maximum delay target, Δ_s, assigned at the application layer to ensure QoS for information derived from sensed data of type s. The average allowable delay for sensed data of type s is denoted $\overline{\Delta}_s$.

Little's formula states that the average number of customers in a queueing system in *steady-state* is equal to the arrival rate of customers to the system, times the average time spent in the system. The result makes no assumptions about arrival or processing distributions and does not depend on the number of servers or the queueing discipline in the system. In terms of the queueing model for packet flows of sensor type s, the arrival rate of packets of type s and delay limits, the delay condition can be approximated instantaneously at time t or by a time average over $[0,T)$, respectively:

$$\mathbb{1}\left(Q_{is,t} < \Delta_s \left[u_{0is,t} \alpha_{s,t} + \sum_{j \in \{neighborhood\,of\,node\,i\}} u_{jis,t} \sigma_{js,t} \right] \right), \quad (8)$$

$$\mathbb{1}\left(\int_0^T Q_{is,t} dt < \overline{\Delta}_s \left[\int_0^T u_{0is,v} \alpha_{s,v} dv + \sum_{j \in \{neighborhood\,of\,node\,i\}} \int_0^T u_{jis,w} \sigma_{js,w} dw \right] \right). \quad (9)$$

4. CROSS-LAYER OPTIMIZATION

Cross-layer protocol optimization is formulated in terms of metrics on WSN performance for $U \in \mathcal{U}$ and expressed as the constrained optimization of cost functionals formed from the MVPPs.

4.1 Energy-Constrained QoS Optimization

To each admissible $U \in \mathcal{U}$, there is a unique cost of form:

$$C(U) = E^U \left[\int_0^T c(t,U_t) \, dt + f_T \right] \quad (10)$$

In (10), for each $U \in \mathcal{U}$, the cost rate $(c(t,U_t), t \in [0,T])$ is a composite process, $c(t,U_t(\omega),\omega) = (c \cdot U)(t,\omega)$, and is (F_t)-adapted for each sample path $U_t(\omega) \in P$, the space of values taken by the routing arrays. The function c is Lebesgue-measurable with respect to t and continuous in the sample path values $U_t(\omega)$ for all $\omega \in \Omega$. Moreover, c has left-continuous sample paths with finite right-hand limits at each discontinuity for each $U_t(\omega)$ for all $\omega \in \Omega$. Function c is thus progressively measurable with respect to the family of σ-algebras, $(F_t, t \in [0,T]$; $F_t \subseteq F)$, with left-continuous sample paths for each $U \in \mathcal{U}$. Terminal cost f_T is a nonnegative, F-measurable, and $\wp^U$-integrable function for $U \in \mathcal{U}$; it represents the cost incurred at the end of the mission. The performance metrics in Section 3.7 satisfy the assumptions on the general cost.

The optimization problem is the determination of policies $U^* \in \mathcal{U}$ over $[0,T]$ that satisfy $C(U^*) = \inf_{U \in \mathcal{U}} C(U) = \min_{U \in \mathcal{U}} C(U)$, subject to the battery reserve constraints at nodes and along routes. The minimum can be achieved, since each $U \in \mathcal{U}$ takes values in a compact set and the cost rate and terminal cost are almost surely $\wp^U$-bounded for $U \in \mathcal{U}$. A policy U^*, if it exists, that achieves the minimum is termed an *optimal routing policy* for the energy-constrained cross-layer protocol.

4.2 Recursive Optimality Conditions With Complete Network Observations

Backward recursive conditions that characterize optimal routing policies for real-time MVPP models of discrete-space events are presented here without proof for the case of complete observations.[11] The local (instantaneous) structure of these conditions is similar to Hamilton-Jacobi-Bellman dynamic programming conditions. With complete observations, the *conditional cost function* $\phi(U_1, U_2, t)$ for the policy concatenated at time t from U_1 and U_2 in $\mathcal{U}$, obeys

$$\phi(U_1, U_2, t) = E^{[U_1, U_2, t]} \left[\int_t^T c(v, U_{2,v}) dv + f_T \middle| F_t \right] =$$

$$E^{U_2} \left[\int_t^T c(v, U_{2,v}) dv + f_T \middle| F_t \right] = \phi(U_2, U_2, t), \quad \text{applying the}$$

causality condition of the $U \in \mathcal{U}$. The *optimal cost-to-go function*

$$W_t^U = \min_{\hat{U} \in \mathcal{U}} \phi(U, \hat{U}, t) = \min_{\hat{U} \in \mathcal{U}} \phi(\hat{U}, \hat{U}, t), \quad \text{i.e.,} \quad W_t^U = W_t \text{ does}$$

not depend on U.

The following theorem introduces functions, w_n, that generalize the optimal cost-to-go concept used in dynamic programming conditions. The result is a special case of a theorem characterizing local optimality conditions for the control of general packet-switched radio networks, based on either partial or complete observations of the events underlying the system state.[12]

Theorem 1. Suppose the cross-layer protocol designs have complete observations of the energy-constrained network state $\left(\mathbf{X}_t = (\mathbf{Q}_t, \mathbf{C}_t), t \in [0,T]\right)$ and, for every $U \in \mathcal{U}$, the packet flow array $(\mathbf{Q}_t)$ has a local description in terms of the conditional $(\wp^U, F_t)$-rates of the network MVPPs. Then $U = U^*$ is optimal in $\mathcal{U}$ if and only if there exist functions, $w_n(t, t_0, x_0, \ldots, t_n, x_n)$, measurable in their arguments and absolutely continuous in t, such that, for τ_n the n-th state-transition time of X, e_{jks} the $M_{max} + 1 \times M_{max} \times S$ array with all zero entries except 1 in the (i,j,s) position, for $\tau_n \leq t < \tau_{n+1}$, and the energy-aware route maintenance algorithm for the battery vector C_t [14] is in effect at time t,

$$\frac{\partial w_n(t, \tau_0, \mathbf{Q}_0, \ldots, \tau_n, \mathbf{Q}_n)}{\partial t} + \min_{U \in \mathcal{U}} \left\{ \sum_s \sum_{l=1}^{M_{max}} \left[\alpha_{s,t}(u_{0ls,t}) \cdot \left(w_{n+1}(t, \tau_0, \mathbf{Q}_0, \ldots, \tau_n, \mathbf{Q}_n + e_{0ls}) - w_n(t, \tau_0, \mathbf{Q}_0, \ldots, \tau_n, \mathbf{Q}_n) \right) \right] + \right.$$

$$\sum_s \sum_{j=1}^{M_{max}} \sigma_{js,t} 1(Q_{js,t^-} > 0) \left[\sum_{k=1}^{M_{max}} (u_{jks,t}) \cdot \left(w_{n+1}(t, \tau_0, \mathbf{Q}_0, \ldots, \tau_n, \mathbf{Q}_n + e_{jks}) - w_n(t, \tau_0, \mathbf{Q}_0, \ldots, \tau_n, \mathbf{Q}_n) \right) \right] +$$

$$\left. \sum_s \sum_{m=1}^{M_{max}} \sigma_{ms,t} 1(Q_{ms,t^-} > 0) \cdot (u_{m0s,t}) \cdot \left(w_{n+1}(t, \tau_0, \mathbf{Q}_0, \ldots, \tau_n, \mathbf{Q}_n + e_{m0s}) - w_n(t, \tau_0, \mathbf{Q}_0, \ldots, \tau_n, \mathbf{Q}_n) \right) + c(t, U_t) \right\} = 0 \text{ a.s. } \wp^U \quad (11)$$

and $\quad w_n(t_0, \tau_0, \mathbf{Q}_0, \ldots, \tau_n, \mathbf{Q}_n) = f_T, \tau_n \leq T < \tau_{n+1}$ a.s. $\wp^U$, (12)

where a.s. $\wp$ means that an equation holds except on a set of zero $\wp$-measure. The minimum in (12) is attained at optimal routing policies $U_t^*(\omega)$ a.s. $\wp^{U^*}$. The optimal cost-to-go process at $t \in [0, T]$ takes the form

$$W_t = \sum_{n=0}^{\infty} 1(\tau_n \leq t < \tau_{n+1}) \cdot w_n(t, \tau_0, \mathbf{Q}_0, \ldots, \tau_n, \mathbf{Q}_n). \quad (13)$$

4.2 Optimality Conditions with Markov Assumptions

Conditions (11) - (13) simplify when the controls $U \in \mathcal{U}$ depend only on the last observed transition in the network state, i.e., $Y_t = X_{t \wedge \tau_n}, \tau_n \leq t < \tau_{n+1}$, or a limited history of observations consisting of the last m successive transitions in the state, i.e.,

$Y_t = [X_{t \wedge \tau_n}, X_{t \wedge \tau_{n-1}}, \ldots, X_{t \wedge \tau_{n-m+1}}], \quad n \geq m > 1, \tau_n \leq t < \tau_{n+1},$

so that $U_t(\omega) = U(t, Y_{t^-}(\omega))$ for each $\omega \in \Gamma, \Gamma \in F_t$, and the assumptions on the MVPPs comprising the state are such that Y is a Markov process. These assumptions are inspired by published studies of models for wireless communication networks that assume the packet queues at nodes are finite-state Markov regenerative processes, e.g., M/D/1/K and M/G/1/K, where here K denotes the finite buffers at network nodes.[16] The assumptions here yield the special case of Markov control of a Markov process. Although these assumptions do not accurately represent the "memory" requirements for the dynamics of packet flows and interdependence of network events occurring at nodes and along paths, they do facilitate a simplified structure for (11) - (13), now expressed in terms of the *optimal cost-to-go function* $W_t = V_t = V_t(Y_{t^-})$ to yield dynamic programming conditions. The infinitesimal generator for V_t depends on the conditional $(\wp^{U^*}, F_t)$-rates for the MVPPs in (3) - (6). The explicit expressions for the V_t, and, in turn, the optimal controls, can be determined by iteratively solving the recursive backward equations from final time $t = T$. In general, V_t takes the form of an product of exponential terms of the instantaneous rates established in the semi-martingale representations of the underlying MVPPs.

5. EXPLICIT SOLUTIONS VERSUS CLOSED-FORM SOLUTIONS

Iterative solutions to the explicit backward recursive equations (11) - (13) for V_t can be established via computer. In special cases of MANETs of a few nodes and traffic classes in conjunction with simple statistical assumptions on arrivals, processing, and queueing priorities, closed-form expressions for the network state and routing policies can be found. However, it is shown that similar closed-form solutions are not generally possible for *ad hoc* WSNs, in which nodes are deployed in environments containing targets at unknown positions and of unknown sensor phenomena. The following example illustrates the difference between the MANET model and *ad hoc* WSN model.

Assume that (i) the heavy traffic condition holds at all nodes: at least one packet is in processing or awaiting processing at each active node; symbolically, $1(Q_{j,s,t} > 0) \equiv 1, \forall j,s$ and $t \in [0, T]$. Equations (2) and (3) become a system of birth-death equations for the random packet flows at nodes and the conditional rates of the flows, with no implicit dependence on the packet flow elements $Q_{j,s,t}$, other than initial queue sizes, $Q_{j,s,0}$. Thus, for known conditional rates, the optimal routing values of $U \in \mathcal{U}$, can be determined from V_t in (11) - (13), based on observation(s) of Y. Also assume that (ii) the PDFs for packet inter-arrival times and inter-processing times do not depend on the stopping times n, i.e., $F_{j,s,n}(t) \equiv F_{j,s}(t)$, so conditional rates are functions of time alone.

Example: Suppose that (i) and (ii) hold and the routing controls depend only on a Markov substate, Y_t, of the packet flow state X_t .In a simple topology for both MANET and WSN, the number of nodes $M_{max} = 8$, grouped in $K = 2$ clusters, each with 4 nodes – 2 nodes at the edge, 1 intermediate node, and 1 sink node. Each "edge" node in the WSN has $S = 2$ sensor types: an acoustic sensor, generating real-time (Class 1) audio reports, and a magnetometer, generating sensed data (Class 2 or Class 3). Similarly, mobile nodes in the MANET have two traffic inputs: digitized audio (Class 1) and data file transfers (Class 3). No buffer limits are imposed at any node. All nodes but the sinks are battery-operated with 100 hours of reserve power at mission start. The two sensor nodes within each cluster are connected, without direct feedback; one sensor node in a cluster is connected to one node of the other cluster; the intermediate nodes can be connected

to any of the 4 sensor nodes, to the other intermediate node, and to one sink node. There are no connections between the two sinks. This topology yields 16 bidirectional connections, and, hence, at most $32=16\times 2$ non-zero entries in the routing control array, for each type of sensed events, to be optimized. Further, assume the inter-event times for processing at the two intermediate nodes and the two sinks are exponentially distributed with constant rates σ_5, σ_6, σ_7, and σ_8, respectively. For Poisson packet arrivals generated *internally* at the 4 mobile nodes of the MANET, with constant rates α_1, α_2, α_3, and α_4, respectively, the optimal routing control problem for the network of M/M/2 queues can be solved from the dynamic programming conditions using the methods for rate control in Chapter VII of Brémaud.[11] However, a closed-form solution for an *ad hoc* WSN of the same structure is precluded by the *absence* of a complete description of the spatial densities and sensor phenomena of *external* targets that drive the network. But complete *a priori* knowledge of the phenomena *before ad hoc* deployment is contrary to the underlying principle for the WSN.

There *are* WSNs that circumvent this limitation. In networks of a small number of stationary nodes that monitor systems of objects at known locations and with known sensor phenomena, e.g., an industrial processing plant, a commercial warehouse, etc., the model becomes analogous to the simple MANET case under Markov assumptions. For these WSNs, closed-form solutions can be determined.

Explicit solutions to conditions (11) – (13) for routing controls and the controlled state processes can be found in *ad hoc* WSNs through customized discrete-event simulations of likely target profiles. A suitable simulation environment has been implemented and maintained for wireless extensions of cross-layer designs by researchers of the Monarch Group at Carnegie-Mellon University, is based on the open ns-2 simulator created by the VINT Project at the University of California at Berkeley.[17]

6. CONCLUSIONS

A real-time QoS model for an energy-constrained *ad hoc* WSN has been derived from earlier models for MANETs. At the core of the QoS routing control model is a cross-layer design that adapts and optimizes the multiple layers of the WSN protocol to resource availability. Cross-layer optimization is based on analytical MVPP models of real-time packet flows for heterogeneous sensors on routes. Routing entries and protocol parameters influence conditional random rates in the semi-martingale decompositions of the MVPPs for these events. With complete observations, backward recursive conditions characterize routing policies to provide the best available QoS performance. Markov assumptions and the limitations to determination of closed-form versus explicit solutions for *ad hoc* WSNs have been discussed. Future work will evaluate the real-time models for *ad hoc* WSNs in discrete-event simulations to forecast cross-layer protocol optimization, based on potential statistical characteristics of stationary and mobile targets.

7. REFERENCES

[1] Hortos, W. S., "Cross-layer protocols optimized for real-time multimedia services in energy-constrained mobile ad hoc networks," *Proc. of Virginia Tech's 14th Symp. on Wireless Pers. Commun.*, Blacksburg, VA, 1-14 (2004).

[2] Hortos, W. S., "Real-time performance analysis of wireless multimedia networks based on partially observed,

multivariate point processes," *Digital Wireless Commun.II*, *Proc. SPIE*, 4045-05, (2000).

[3] Braden, R., Clark, D. and Shenker, S., "Integrated services in the Internet architecture: an overview," *IETF RFC 1633*, (1994).

[4] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and Weiss, W., "An architecture for differentiated services," *IETF RFC 2475*, (1998).

[5] Nikaein, N., C. Bonnet, C., Moret, Y., and Rai, I. A., "2LQoS - two-layered quality of service model for reactive routing protocols for mobile ad hoc networks," *Proc. SCI - 6th World Multiconf. on Systemics, Cybern. and Informatics*, (2002).

[6] Nikaein, N., and Bonnet, C., "Layered quality of service model for routing in mobile ad hoc networks," *Proc. WMAN*, (2003).

[7] Acampora, A. S. and Naghshineh, M., "Control and quality-of-service provisioning in high-speed microcellular networks," *IEEE Personal Comm. Mag.*, 36-43 (1994).

[8] Nikaein, N. and Bonnet, C., "A glance at quality of service models for mobile ad hoc networks," *Proc. DNAC 2002: 16th Conf. on New Arch. for Comm.*, (2002).

[9] Min, Y., Peng, H., Bo, Y., Xiaofeng, M., Dilin, M., and Chuanshan, G., "A light-weight multicast schema for wireless sensor network with multi-sinks," *WICOM 2006: Intern. Conf. on Wireless Commun., Netw. and Mobile Computing, 2006*, 1-4 (2006).

[10] Frost, V. S. and Melamed, B., "Traffic modeling for telecommunication networks," *IEEE Comm. Mag.*, 32(3), 70-81 (1994).

[11] Brémaud, P., [Point Processes and Queues: Martingale Dynamics], Springer-Verlag, New York, (1981).

[12] Hortos Jr., W. S., "Partially observable point processes and the control of packet radio networks." Doctoral Dissertation, University of Michigan, UMI Dissertation Information Services, Ann Arbor, (1990).

[13] Babu, T. V. J. G., Le-Ngoc, T. and Hayes, J. F., "Performance of a priority-based dynamic capacity allocation scheme WATM systems," *Proc. IEEE GLOBECOM '98*, 4, 2234-2238 (1998).

[14] Kim, D.-K., Park, J. W., Toh, C.-K., and Choi, Y.-H., "Power-aware route maintenance protocol for mobile ad hoc networks," *Proc. IEEE 10th Int. Conf. Telecom. (ICT) 2003*, 501-506 (2003).

[15] Doléans-Dade, C. and Meyer, P. A., "Intégrales stochastique par rapport aux martingales locales," *Séminaire Probabilités, IV*, [Lecture Notes in Mathematics, 124], Springer-Verlag, Berlin, 77-107 (1970).

[16] Chen, Z., C. Lin, C., Wen, P. Hao, and Yin, Hao, "An analytical model for evaluating IEEE 802.15.4 CSMA/CA protocol in low-rate wireless application," *21st Intern. Conf. Adv. Info. Netw. and Appl. Workshops (AINAW '07)*, 1-6 (2007).

[17] [CMU Monarch Project, wireless and mobility extensions to ns-2], http://www. monarch.cs.cmu.edu/cmu-ns/html.