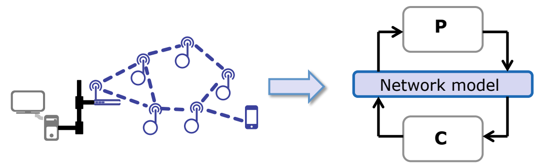
Wireless networking for control: technologies and models

Mikael Johansson
KTH · Stockholm · Sweden

Based on joint work with R. Jäntti, P. Soldati, Z. Zou, H. Zhang and O. Landsiedel

What this lecture is about



Aim

- Give a basic understanding of fundamentals of industrial wireless
  - Propagation and fading
  - Interference, bit errors and packet errors
  - Medium access control
  - Routing and transport control
- Illustrate realistic models for losses and latencies
  - Assumptions, time-scales, ball-park numbers
- Review current standards for industrial low-power wireless control

Outline

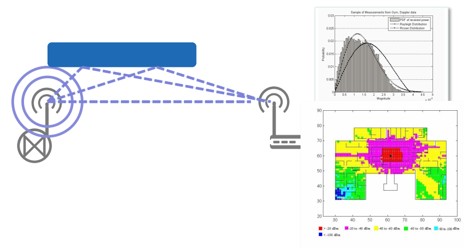
1. The single link:
   - From wireless propagation to packet error rates
2. Sharing the wireless medium
   - Medium access control
3. From single hop to networking
   - Routing, forwarding, and transport
4. Standards for industrial wireless
   - Focus on *Wireless*HART

Part I: The single link

Aim: to understand
- basic properties of wireless propagation
- models of wireless channels and packet reception ratios

(and to know what's why the illustration above is misleading)

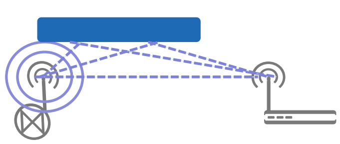
What was wrong with the picture?

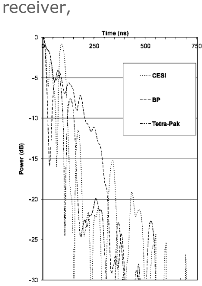


Radio waves behave like all other waves
- subject to reflection, diffraction, absorption, scattering
- a multitude of paths, each with different gain and phase shift

Small-scale effects

Multi-path reflections “smear out” channel impulse response
- symbols sent too closely in time interfere at receiver,
  (➔ **inter-symbol interference**)





Many theoretical models (Rayleigh, Rician, Nakagami, …)

Multi-path in practice

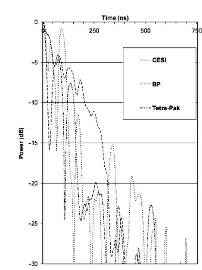
Characterized by RMS **delay spread**

$$\sigma_\tau = \sqrt{\frac{1}{P_{tot}}\left(\sum_k (\tau_k - \bar{\tau})^2 P(\tau_k)\right)}$$

In frequency domain: **coherence bandwidth**

$$B_{C,50} \approx \frac{1}{5\sigma_\tau}$$

Reported values in literature: 16-85 ns



Low-power radios could tolerate up to 50ns (rough calculations) or 900ns delay spread (simulations, limited range), so most likely OK.

Movements and Doppler spread

Moving objects (transmitters, obstacles, ...) shift signal frequency

- termed **Doppler shift**
- depends on carrier frequency, velocity of movement

$$B_D = \frac{v}{c} f_c$$

- Time-domain quantity: **coherence time**

$$T_C \approx \frac{0.423}{B_D}$$

Industrial measurements [S] have reported coherence times of 100ms

[S] D. Sexton et al., "Radio Channel Quality in Industrial Wireless Sensor Networks", 2005

Large-scale effects

Frii's equation describes received signal power **propagation**

$$P_{rx}^{dBm} = P_{tx}^{dBm} + G_{tx}^{dBi} - L_p^{dB} + G_{rx}^{dBi}$$

Powers in mW, G are antenna gains, and L is **path loss**

$$L_p^{dB} = L_0^{dBm} + n10\log_{10}(d)$$

Must be tuned to practical conditions:

- In free space n=2, but it is typically larger (up to 4)
- L_0^{dBm} depends on carrier frequency used.

Extra term often added to account for **shadowing** objects

Signals, interference and noise

Decoding: essentially to separate signal from noise

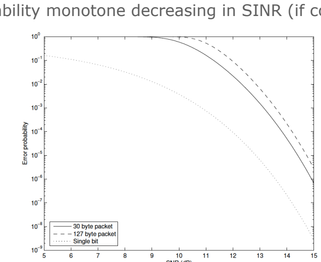
Easier at high **signal-to-interference and noise ratios** (SINRs)

$$\gamma(t) = \frac{g(t)P_{Tx}}{N_0 B + I_{Rx}}$$

- Expected path gain $E\{g(t)\}$ determined by large-scale effects
- Path gain distribution determined by small-scale effects
- Thermal noise and communication bandwidth
- Interference from other wireless equipment, microwaves, …

From SINR to packet error rates

Bit error probability monotone decreasing in SINR (if coherence time large)

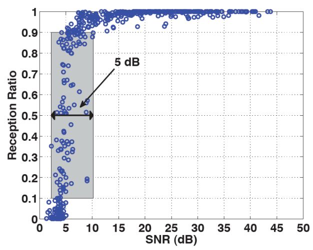


Simple **outage** model: packets lost when $\gamma \le \gamma^{th}$

Receive sensitivity: receive power needed for γ^{th}(under no interference)

Packet reception ratios in practice

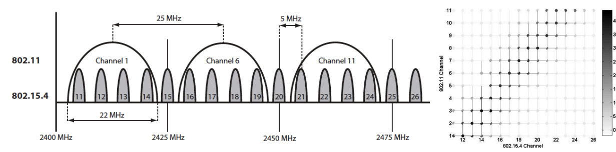
Real measurement display very similar characteristics



(note: packet reception ratios rather than loss probabilities)

Interference from WiFi

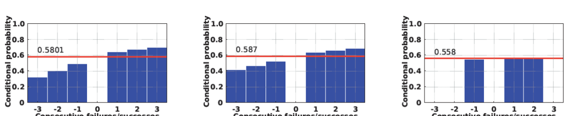
Dominating interference source is often WiFi (802.11)



Avoided by frequency planning or frequency hopping.

Packet reception ratios in practice

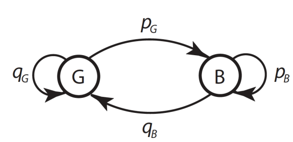
Packet losses typically occur in bursts (are correlated in time)
- More pronounced with short inter-packet-intervals (IPI)



(a) IPI=10ms (b) IPI=500ms (c) IPI=2s

Markov models

Simplest one with correlated losses is due to Gilbert and Elliot



Note that $q_G + p_G = 1$ so model essentially has two parameters

Can be estimated from real traces by observing average holding times

$$T_B = \frac{1}{q_B} \qquad T_G = \frac{1}{1-q_G}$$

More states allow to better model holding time distributions.

Means of increasing reliability

The raw packet error probabilities can often be improved.

- Coding (forward error correction)
  - Include extra bits that allow to repair bit errors
- Retransmissions
  - Re-send packets that are not acknowledged (possibly with increased amount of coding)
- Diversity techniques
  - Frequency hopping: avoid consistently bad channels
  - Multiple antennas: multiple independent signal paths

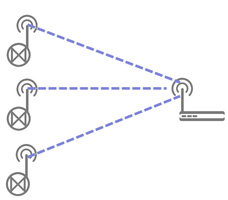
Packets in 802.15.4 standard

Packet format: header, footer and up to 127 byte payload

Octets:2	1	4 to 20	variable	2
Frame control	Data sequence number	Address information	Data payload	Frame check sequence
MAC header			MAC Payload	MAC footer

Ballpark figure for packet transmission time: 5ms

Part II: Sharing the medium



Aim: to understand
- the most important medium access control principles
- their pros and cons, and how they can be analyzed

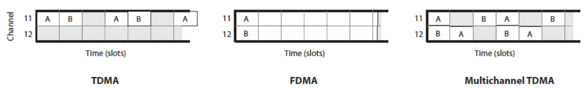
Medium access protocol

Medium access control (MAC) used for sharing the spectrum

Broadly classified into
- **Scheduled access**: reserve communication resource to one link
  - Examples: time/frequency/code division multiple access
  - Efficient in high-volume deterministic traffic
- **Contention-based approaches**: resolve collisions when they appear
  - Examples: Aloha, CSMA, 802.11, 802.15.4
  - Efficient in light, uncorrelated traffic

Scheduled access

Spectrum divided into disjoint communication resources
- Time slots, frequency bands, codes, or combinations
- Allocate fixed number of resources to each user (link)

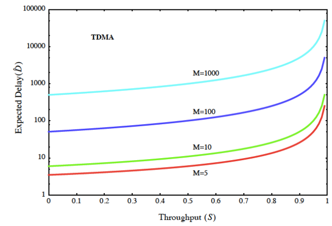


Advantages
- Deterministic behavior, high resource utilization

Disadvantages:
- Overhead to allow conflict-free scheduling (central management)
- Overhead to create disjoint resources (guard bands, guard times, …)

Delay-throughput analysis

Typical delay-throughput analysis in books on MAC protocols
- Assume traffic generated according to some stochastic process
- Characterize expected delay and throughput



Q: Is this analysis relevant for us? Why/why not?

Contention-based medium access

Basic idea introduced in the **Aloha protocol:**
- Try to transmit packet
- If collision, then re-try at a random future time

Slotted Aloha: nodes synchronized, time slot allows one packet transmission
- Immediately try to transmit packet
- If collision, nodes become backlogged
- Backlogged nodes transmit in each slot w. probability p until success

Tuning parameter: transmission probability p.
- In saturated traffic, should let p<1/M (suboptimal in transient traffic)

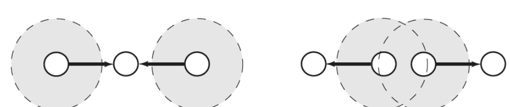
Q: can you see any problems with Aloha?

Carrier sensing

Problems with Aloha: tries to transmit even if other nodes are transmitting

Better solution:
- Listen to medium to make sure that it is free before transmitting
- Procedure called **carrier sensing**

Very useful, but hard to implement perfectly in practice:



Hidden node problem Exposed node problem

CSMA variants

Slotted Aloha+carrier sensing → **p-persistent CSMA**
- Listen to medium, if busy then refrain from transmitting
- If medium is idle, transmit with probability p, refrain with 1-p

Alternative: **non-persistent CSMA**
- Listen to medium, if idle then transmit
- If medium is busy, reschedule for some future time
- Implemented using **back-off counters:**
  - Draw random number in [0, CW]
  - Decrement counter in each time slot, transmit when counter=0

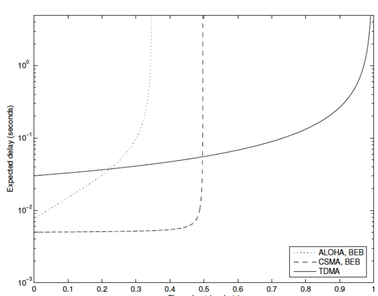
Q: How to choose CW to get similar performance as p-persistent?

Adjusting to number of users

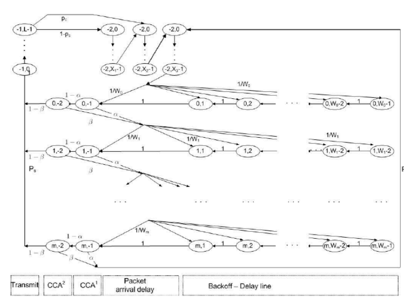
The proper value of CW depends on number of contenders
- not known in most ad-hoc applications (cf. WiFi AP)
- natural to try to adapt contention window size

Contention window adapted using **binary exponential backoff**
- CW doubled at each collision (up to a maximum value)
- Reset to nominal value at success

Latency-throughput



CSMA in 802.15.4 standard



Q: can you note carrier-sensing, non-persistent CSMA, BEB?

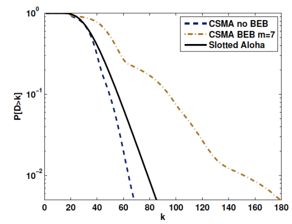
Accounting for traffic

The vast majority of performance analyses
- assume independent stochastic traffic sources
- analyze stationary quantities (e.g. expected delay)

But for control applications
- traffic is typically deterministic, or at least correlated
- we need to consider transient quantities

Latency under correlated traffic

Delay distributions for M=10 nodes generating packets at the same time



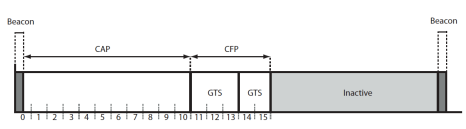
$$\Pr(D=d)=\sum_{i=1}^{M} a_i \sum_{j=1}^{M} \frac{(a_j-1)^M}{\prod_{k\neq j}(a_j-a_k)}(1-a_j)^d, \quad a_k = k(1-p)p^{k-1}$$

Dynamic scheduling & hybrid MACs

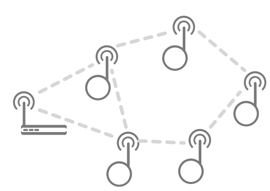
Dynamic scheduling: allocate time slots only when needed

Hybrid MACs: combine scheduled and contention-based access

Example: 802.15.4 frame structrure



Part III: From links to network



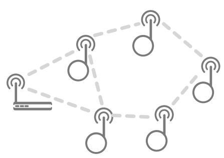
Aim: to understand
- Advantages and challenges of multi-hop communications
- Basic principles of routing and transport protocols

Adressing, routing and forwarding

Addressing: associating nodes with unique IDs

Routing: decide along which paths (routes) data traverse network

Forwarding: selecting the next hop for packet towards its destination



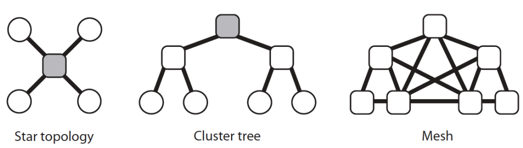
Routing protocols

Again, several hundred options.

Normally classified into
- Reactive (on-demand): set up paths when needed
- Proactive: continuously find and maintain paths for all traffic

Our focus: routing topology and its impact on end-to-end performance

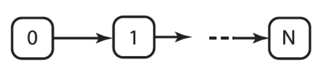
Routing topologies



Star: limited range, reliability
Cluster tree: better range, still reliability issues
Mesh tree: range and reliability

Multi-hop latency and reliability

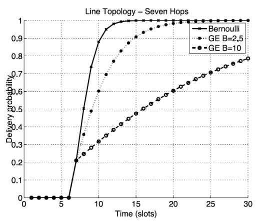
Multi-hop communication impacts both latency and reliability
- typical low-power radios cannot receive and transmit at the same time



Assume TDMA MAC:
Q: what is the minimum latency?
Q: what is the success probability if links fail independently w. prob p?

Retransmissions and reliability

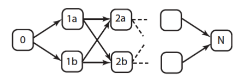
Retransmissions allow to increase reliability (at the price of longer delay)

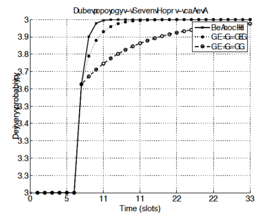


Less effective when channels are correlated (why?)

The multi-path advantage

Corresponding reliability in “braid” network

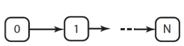


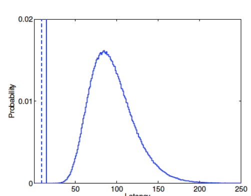


Q: What happened and why?

Interaction with MAC

Consider line when all nodes have packet to send at time zero (N=10)





Aloha, TDMA and multi-channel TDMA (which is which?)

Shortest path routing

Most efficient paths are “short” in the appropriate metric
- latency: number of hops
- end-to-end reliability: reliability of individual links

Shortest paths computations efficient, can be distributed
- Foundations for routing in the (fixed) Internet
- Dijkstra’s algorithm, distance vector protocols
- Key challenge: making sure routes are loop-free

Routing metrics

Example: end-to-end reliability

$$r_P = \prod_{l \in P} (1 - p_l)$$

Not additive in link reliabilities. But log-transform

$$\log(r_P) = \sum_{l \in P} \log(1 - p_l)$$

Allow us to find max-reliability paths using Dijkstra.

More popular alternative: expected transmission count, ETX

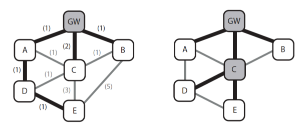
$$ETX = \frac{1}{d_f d_r}$$

Constrained shortest path

Useful to exclude certain nodes or links from routing topology

Example: battery-powered nodes do not need to forward traffic

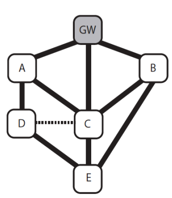
Typical implementation:
- Remove links that do not satisfy requirement from topology
- Perform shortest path routing on reduced topology



Shortest path Constrained shortest path

Graph-based routing

Sets up destination-oriented directed acyclic graph and forwards data on this



Graph

Multiple paths, forward decision now critical:
- which parent to use? retransmit on new channel or to new parent?

Transport layer protocols

Transport layer: provides end-to-end communication services to applications

Most well-known protocols in fixed Internet
- UPD: simple forwarding without acknowledgements or retransmissions
- TCP: acknowledgements, retransmissions and flow control

TCP has many issues in a wireless setting:
- Flow control misinterprets link losses for congestion
- Interaction between BEB and flow control gives raise to starvation

Most industrial protocols use UDP or TCP without flow control.

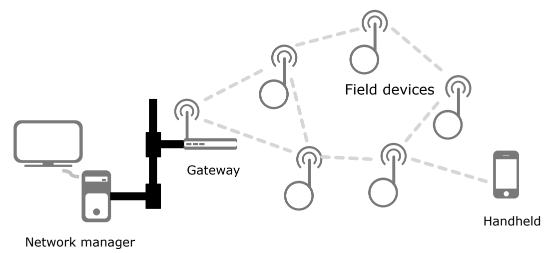
Wireless standards

Many standards and specifications for 802.15.4-compliant radios:

- **Zigbee PRO:** only specification, some reliability issues.
- ***Wireless*HART:** Standardized since 2007.
- **ISA100:** Still ongoing.
- **IETF RPL/ROLL, 802.15.4e:** Still in draft form

We will focus on *Wireless*HART (only existing standard)

*Wireless*HART network organization



*Wireless*HART communication stack

Application	Command Oriented. Predefined Data Types and Application Procedures	
Presentation		
Session		
Transport	Auto-Segmented transfer of large data sets, reliable stream transport, Negotiated Segment sizes	
Network		Power-Optimized Redundant Path, Mesh to the edge Network
Data Link	A Binary, Byte Oriented, Token Passing, Master/Slave Protocol	Secure, Time Synched TDMA/ CSMA, Frequency Agile with ARQ
Physical	Simultaneous Analog & Digital Signaling 4-20mA Copper Wiring	2.4 GHz Wireless, 802.15.4 based radios, 10dBm Tx Power

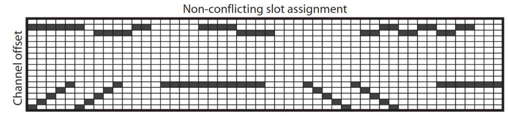
*Wireless*HART data link layer

Centrally managed, globally synchronized, multi-channel TDMA
- 10ms time slots
- 16 parallel channels
- employs channel hopping with black listing

Time slots can either be dedicated or shared

Schedule organized in multiple (superimposed) superframes
- different data streams/classes, management info, …

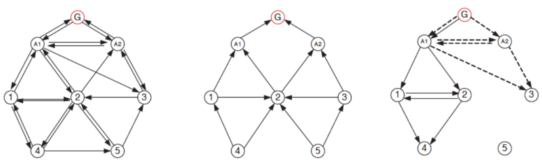
Non-conflicting slot assignment



*Wireless*HART routing

Routing is either source-based (single path) or graph-based

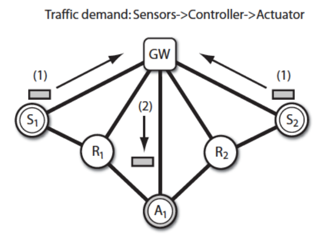
All data must pass gateway (uplink graph/downlink graphs)



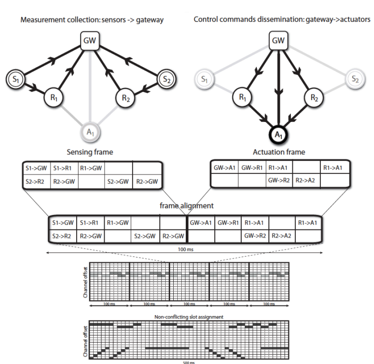
(a) Original network topology (b) Uplink graph (d) Downlink graph to Dev 3 and Dev 4

Scheduling example

Simple control example: two sensors, one actuator.



Scheduling example cont'd



Loss and latency of *Wireless*HART

Latency and reliability analysis tools for *Wireless*HART begin to appear:

- Reliable links
  - Latency-optimal unicast and convergecast schedules, bounds of performance
  - Heuristics for scheduling of multiple streams
- Uncorrelated link losses
  - Delay-constrained maximum reliability unicast schedules
  - Analysis of unicast and multicast latency distributions
- Correlated link losses
  - Achievable delay-constrained reliability for single unicast stream
  - Analysis of unicast and multicast latency distributions
- Link losses vs. end-to-end losses.

Additional *Wireless*HART features

Transport control:
- reliable stream transfer

Security:
- AES 128-bit encryption, data and network protection/integrity

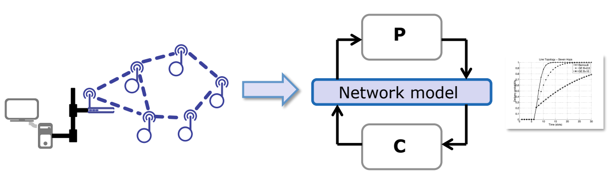
Application layer
- Commands and device specifications (e.g. PID)

and much more…

Summary

- A basic understanding of fundamentals of industrial wireless
  - Propagation and fading
  - Interference, bit errors and packet errors
  - Medium access control
  - Routing and transport control
- Realistic models for losses and latencies
  - Assumptions, time-scales, ball-park numbers
- A review of current standards for industrial wireless control

So what *is* the right model?



Almost always: latency+loss
- The single link: packet loss probability, Markovian
- Multiple links: latency distribution + loss, (MAC, traffic, link)
- Network: latency+loss (link, routing, scheduling, traffic, MAC)

References

Most of the material comes from the book chapter

M. Johansson and R. Jäntti, “Wireless networking for control: technologies and models”, in Bemporad et al., Eds. Networked Control Systems, Springer Verlag, 2010.

and the references therein.

Extensive experimental measurements of 802.15.4 in

K. Srinivasan et al., “An empirical study of low-power wireless”, ACM Transactions on Sensor Networks, 2010.