UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

IVAN MÜLLER

GERENCIAMENTO DESCENTRALIZADO DE REDES SEM FIO INDUSTRIAIS SEGUNDO O PADRÃO *WIRELESS*HART

Porto Alegre
2012

IVAN MÜLLER

GERENCIAMENTO DESCENTRALIZADO DE REDES SEM FIO INDUSTRIAIS SEGUNDO O PADRÃO *WIRELESS*HART

Tese de doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Rio Grande do Sul como parte dos requisitos para a obtenção do título de Doutor em Engenharia Elétrica.

Área de concentração: Controle e Automação - Automação Industrial

ORIENTADOR: Prof. Dr. Carlos Eduardo Pereira

Porto Alegre
2012

IVAN MÜLLER

GERENCIAMENTO DESCENTRALIZADO DE REDES SEM FIO INDUSTRIAIS SEGUNDO O PADRÃO *WIRELESS*HART

Esta tese foi julgada adequada para a obtenção do título de Doutor em Engenharia Elétrica e aprovada em sua forma final pelo Orientador e pela Banca Examinadora.

Orientador: _____
Prof. Dr. Carlos Eduardo Pereira, UFRGS
Doutor pela Universidade de Stuttgart, Alemanha

Banca Examinadora:

Prof. Dr. Altamiro Amadeu Susin, UFRGS
Doutor pelo Institut National Polytechnique de Grenoble, França

Prof. Dr. Eric Ericson Fabris, UFRGS
Doutor pela Universidade Federal do Rio Grande do Sul, Brasil

Prof. Dr. João César Netto, UFRGS
Doutor pela Universite Catholique de Louvain, França

Prof. Dr. Leandro Buss Becker, UFSC
Doutor pela Universidade Federal do Rio Grande do Sul, Brasil

Prof. Dr. Valner João Brusamarello, UFRGS
Doutor pela Universidade Federal de Santa Catarina, Brasil

Coordenador do PPGEE: _____
Prof. Dr. João Manoel Gomes da Silva Jr.

Porto Alegre, dezembro de 2012.

AGRADECIMENTOS

Dedico este trabalho a toda minha família, colegas e professores do PPGEE.
À minha mãe, que sempre me apoia e incentiva incondicionalmente.
Ao meu pai, que me inspirou e que teria orgulho dos meus feitos.
À minha esposa, que me ama e me permite buscar o que almejo profissionalmente.
Aos meus filhos, para que tenham orgulho de mim e porque me orgulho deles.
Aos meus irmãos, que sempre acreditam na minha capacidade.
Agradeço também ao comitê gestor e participantes do Namitec, CNPq e Capes pela provisão das bolsas de mestrado e doutorado e à Finep, pela bolsa de extensão no país.

RESUMO

O uso de equipamentos sem fio em sistemas de monitoramento e controle de processos industriais vem aumentando gradativamente. Os elevados custos de cabeamento incentivam os gerentes de fábricas para que considerem o uso de sistemas industriais sem fio, uma vez que estes promovem grandes reduções nos custos de instalação finais. Entretanto, a robustez do enlace de rádio frequência e os requerimentos de comunicação em tempo real são frequentemente citados como obstáculos reais para o uso de tecnologias sem fio na indústria. À frente das demais organizações, a HART Foundation lançou em 2007 o *Wireless*HART, sendo este o primeiro padrão aberto de comunicação sem fio especificamente desenvolvido para ambientes industriais. Apesar de ser um protocolo seguro e de propiciar comunicação em tempo real, o estudo aprofundado do *Wireless*HART revela diversas possibilidades de melhorias. Nesse protocolo, o gerenciamento da rede é feito de forma centralizada, o que garante maior simplicidade e controle sobre o escalonamento, roteamento e a segurança das mensagens. Por outro lado, grandes latências, especialmente nos processos de conexão e desconexão de nós, impossibilitam o uso de dispositivos móveis ou intermitentes. A concentração do gerenciamento em um só elemento também leva à fragilidade do sistema: na falha do gerenciador central, toda rede fica inoperante. Neste trabalho, é apresentada uma arquitetura de rede *Wireless*HART inovadora, que utiliza coprocessadores gerenciadores distribuídos para possibilitar a descentralização do gerenciamento e consequentemente as suas vantagens. O sistema desenvolvido permite diferentes configurações descentralizadas, parciais ou totais, e com mínimas modificações no protocolo original. A arquitetura de rede proposta mantém os dispositivos de campo com suas características originais, de baixo custo e consumo, além de permitir a coexistência com equipamentos certificados. A metodologia para a descentralização é apresentada e o estudo de caso realizado permite a avaliação da proposta. A originalidade deste trabalho é verificada através do estudo do estado da arte. Os resultados obtidos de um estudo de caso demonstram o grande potencial da proposta, com reduções significativas nos tempos de agregação e manutenção da rede. A proposta de descentralização possibilita configurações e modos operacionais antes impossíveis no *Wireless*HART, tais como o uso de dispositivos móveis ou intermitentes e o emprego de técnicas de *handover*.

Palavras-chave: Redes sem fio industriais, Protocolo *Wireless*HART, Redes de sensores sem fio, Coprocessador para gerenciamento, Sistemas distribuídos.

ABSTRACT

The use of wireless devices in industrial systems for process monitoring and control has been gradually increasing. The high costs of cabling encourage plant managers to consider the use of such industrial wireless systems, since they promote large reductions in final costs. However, the link robustness and the real-time communication requirements are frequently cited as real obstacles to the use of wireless technologies in the industry. Ahead of other organizations, the HART Foundation launched the in 2007 *Wireless*HART, which is the first open standard for wireless communications developed specifically for industrial environments. *Wireless*HART is a reliable and secure protocol and provide real-time communications but the thorough study of this protocol reveals several possibile improvements. In this protocol, network management is done centrally, ensuring simplicity and control over the scheduling, routing and messaging security. Moreover, large latency, especially in the connection and disconnection processes of field devices, preclude the use of mobile or intermittent devices. The concentration of one element in management also leads to the system fragility: if the central manager fails the entire network fail. In this work, a innovative *Wireless*HART network achitecture is presented and implemented by means of hardware-based distributed network manager coprocessors to enable decentralization and its blessings. The developed system allows different decentralized, partial or total architectures with minimal changes to the original protocol. A hardware-based solution maintains field devices with its low cost and consumption features and allow coexistence with certified equipment. The decentralization methodology along with its originality is presented and case study allows the evaluation of the proposal. The results demonstrate the architecture has great potential with significant reductions in join and maintenance times. The proposed decentralization enables configurations and operating modes previously impossible in *Wireless*HART such as the use of mobile or intermittent devices and *handover* techniques.

Keywords: Wireless sensor networks, Industrial wireless networks, *Wireless*HART protocol, Coprocessing, Distributed systems.

SUMÁRIO

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS

ACK	*Acknowledge*
AES	*Advanced Encryption System*
AP	*Access Point*
ASN	*Absolute Slot Number*
CAP	*Contention Access Period*
CCA	*Clear Chanel Assessment*
CFP	*Contention Free Period*
CSMA-CA	*Carrier Sense Multiple Access - Colision Avoidance*
COP	Coprocessador
DCS	*Distributed Control Systems*
DLL	*Data Link Layer*
DSSS	*Direct Sequence Spread Spectrum*
FD	*Field Device*
FDAP	*Field Device - Access Point*
FDMA	*Frequency Division Multiple Access*
FHSS	*Frequency Hopping Spread Spectrum*
FFD	*Full Function Device*
FPGA	*Field Programmable Gate Arrey*
FSK	*Frequency Shift Keying*
GTS	*Guaranteed Time Slot*
GUI	*Graphical User Interface*
HCF	*Hart Communication Foundation*
ID	Identificador
IDE	*Integrated Development Environment*
IEEE	*Institute of Electrical and Electronics Engineers*
IoT	*Internet of things*

ISM	*Industrial, Scientific and Medical*
ISO	*International Standard Organization*
LR-WPAN	*Low Rate-Wireless Personal Area Network*
MAC	*Medium Access Controller*
MCU	*Micro Controller Unit*
MIC	*Message Integrity Code*
OEM	Onda Eletromagnética
OSI	*Open Systems Interconnection*
PA	*Power Amplifier*
RF	Rádio Frequência
RFD	*Reduced Function Device*
RSL	*Received Signal Level*
RSSF	Redes de Sensores sem Fio
RSSI	*Received Signal Strength Indication*
RTC	*Real Time Clock*
SoC	*System-on-a-Chip*
TCP	*Transmission Control Protocol*
TDMA	*Time Division Multiple Access*
TSMP	*Time Synchronized Mesh Protocol*
WH	*Wireless*HART
WLAN	*Wireless Local Area Network*

1 INTRODUÇÃO

O uso de equipamentos sem fio em sistemas de monitoramento e controle de processos industriais vem aumentando gradativamente (SONG et al., 2006). A principal vantagem no uso deste tipo de equipamento reside na facilidade de instalação quando comparado a dispositivos cabeados. O elevado custo de cabeamento é um grande incentivador para que muitos gerentes de fábricas considerem o uso de sistemas de controle e monitoramento sem fio, visto que eles podem promover reduções de até 80% nos custos de instalação (CARO, 2004), (BLEVINS et al., 2002), (WILLIG; MATHEUS; WOLISZ, 2005). Entretanto, a robustez do enlace de RF (Rádio Frequência) e os requerimentos de comunicação em tempo real são frequentemente citados como obstáculos reais para o uso de tecnologias sem fio na indústria. Por este motivo, grandes esforços são realizados para o desenvolvimento de redes sem fio adequadas para aplicação em ambientes industriais (JONSSON; KUNERT, 2009). Várias organizações industriais têm fomentado o uso de tecnologias sem fio para aplicações industriais. À frente das demais, a HART Foundation lançou em 2007 a norma HART 7, que já contava com as seções relativas ao *Wireless*HART (WH), sendo portanto o primeiro padrão aberto para comunicação sem fio especificamente desenvolvido para uso industrial. Desde então, vários estudos sobre o WH e outros protocolos de redes sem fio industriais têm sido feitos, promovendo amplas discussões sobre as vantagens e desvantagens de um ou de outro (SONG et al., 2008), (CHEN; NIXON; MOK, 2010). O aumento gradativo das vendas de produtos compatíveis com o WH indicam que o protocolo está em franca aceitação pela comunidade da automação e controle e isto se deve primariamente à sua robustez, sendo este fator, determinante para aceitação do seu uso em fábricas. Apesar disto, o estudo aprofundado do protocolo revela diversas possíveis melhorias. Entre elas, é citado o gerenciamento mais eficiente da rede através do emprego de diferentes técnicas de roteamento e escalonamento das mensagens. Também são citados; uso de camada física alternativa ao IEEE 802.15.4, emprego de técnicas de rádio cognitivo, o uso de dois ou mais pontos de acesso para aumento de redundância de rotas e *throughput* e a possibilidade de descentralização do gerenciamento da rede através de técnicas TDMA, objetivando o aumento de dinamicidade da mesma. Outras técnicas ainda podem ser desenvolvidas e adaptadas ao WH: saltos adaptativos entre canais, uso de modulação de potência de RF e técnicas de agregação de pacotes ao longo dos saltos na rede.

De acordo com (CHEN; NIXON; MOK, 2010), a utilização de gerenciamento centralizado em redes sem fio como no caso do WH, favorece os requisitos de tempo real necessários para aplicações industriais. Entre as vantagens citadas, está o fato de que os nós da rede não exercem tarefas de gerenciamento, e portanto necessitam hardware e software mínimo para comunicação, o que propicia equipamentos de baixo custo e consumo de energia. Como o escalonamento das mensagens é tarefa designada ao gerenciador da

rede, a tomada de decisões importantes tais como ocupação de banda, escalonamentos, escolha de rotas e segurança são simplificadas. Porém, em detrimento destas vantagens, ocorre um grande aumento na latência da comunicação de diversas atividades da rede, tais como agregação e desligamento de nós e reestruturação da topologia. A concentração do gerenciamento em um só elemento de rede também leva à fragilidade do sistema: na falha do gerenciador central, toda rede fica inoperante. Mas a maior desvantagem observada é a latência na formação e reestruturação da rede, de tal forma que o uso de nós intermitentes ou móveis não é possível. A observação desta limitação já levou à propostas de descentralização com vistas a aumentar a dinamicidade. As propostas apresentadas tratam basicamente da modificação do mecanismo TDMA empregado, o que as tornam incompatíveis com o protocolo. Aliado a isso, as modificações acarretam em grande aumento da complexidade ao protocolo, exigindo maior poder computacional dos nós da rede e levando ao aumento de consumo de energia. Neste ponto é encontrado um grande motivador para pesquisas relacionadas à redes sem fio industriais: elaborar alguma estratégia para aumentar a dinamicidade destas redes de forma a manter as características fundamentais que elas possuem, já que são seguras, confiáveis e atendem requisitos de tempo real. Esta tese apresenta uma proposta inovadora, que permite explorar as características positivas de uma rede WH agregando funcionalidades de sistemas distribuídos.

Este trabalho está organizado da seguinte forma: no Capítulo 2, são apresentados os conceitos básicos utilizados ao longo da tese. No Capítulo 3, é apresentada uma revisão bibliográfica referente a trabalhos anteriormente realizados, onde são avaliados diferentes protocolos de comunicação sem fio industrial e mais especificamente o WH. A proposta de tese é apresentada no Capítulo 4 e a sua implementação, no Capítulo 5. O trabalho é finalizado com a apresentação dos resultados obtidos (Capítulo 6) e conclusões e trabalhos futuros no Capítulo 7.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Requisitos de redes sem fio para aplicações industriais

Nesta seção são abordados os requisitos necessários para que um protocolo de comunicação de rede sem fio seja considerado adequado para aplicação em ambientes industrias. Considera-se a importância do assunto para entendimento desta tese e, para tanto, os protocolos IEEE 802.X e suas utilizações no desenvolvimento de padrões industriais são discutidos. Os principais protocolos são apresentados e suas aplicabilidades em ambiente industrial são avaliadas. São considerados os requisitos fundamentais para estas aplicações tais como determinismo, dinamicidade, segurança e robustez.

Sistemas de comunicação industriais podem fazer uso de diferentes meios para a propagação de dados tais como cabos, fibras ópticas e RF. Um dos grandes benefícios do uso de sistemas de RF é permitir a mobilidade dos nós que compõem a rede. Dentro de uma fábrica, a mobilidade não significa somente a possibilidade do emprego de dispositivos móveis mas também a facilidade de instalação, possibilidade de reposicionamento e a escalabilidade da rede, que pode ser ampliada sob demanda. Ainda, há a grande redução dos custos de instalação devido à supressão dos cabeamentos e estruturas civis para condicionamento dos cabos, tais como dutos, estruturas de suporte e calhas. A principal desvantagem dos sistemas de comunicação baseados em RF reside na susceptibilidade aos fenômenos de propagação de ondas eletromagnéticas (OEM) que interferem nos enlaces produzindo reflexões, difrações, desvanecimentos, bloqueios e interferências. Ainda há o problema da coexistência com outros equipamentos de RF que estão ou venham a ser instalados na planta, especialmente quando são utilizadas as faixas de radiação restrita, tais como as bandas ISM (*Industrial, Scientific and Medical*). Os dispositivos de campo (FD - *field devices*) que empregam RF como meio para comunicação devem ser tolerantes a estas dificuldades, especialmente as que ocorrem mais frequentemente. A fim de atingir altos níveis de confiabilidade, diversas técnicas devem ser empregadas. Estas, se corretamente implementadas, podem propiciar aos sistemas de comunicação por RF, confiabilidade tão alta quanto a de redes cabeadas (JONSSON; KUNERT, 2009), (DOHERTY; TEASDALE, 2006).

Em sistemas de comunicação industrias, os transceptores de rádio devem prover comunicação confiável concomitantemente com baixo consumo de energia, para propiciar equipamentos exclusivamente alimentados a bateria. Técnicas de espalhamento espectral, originalmente desenvolvidas para uso militar, estão presentes na maioria dos transceptores de RF modernos. Estas permitem redução de energia por bit, redução de perda de dados por fenômenos de multi caminhos e interferências ativas (*jamming*) além de dificultarem a detecção intrusiva. Tipicamente, o espalhamento espectral é obtido por saltos de frequência ou pela multiplicação do sinal de banda base por um código pseudo-aleatório.

A sequência de saltos de canais ou a semente que gera o código pseudo-aleatório é conhecida pelo receptor de modo que este pode recuperar o sinal de banda base. O emprego de técnicas de espalhamento espectral permitem elevado grau de coexistência. Ainda assim, para obtenção de elevada robustez necessária para aplicações industriais, outros mecanismos devem ser empregados. A maioria dos protocolos de comunicação sem fio faz uso de um determinado número de canais disponíveis que podem ser utilizados ao longo da banda empregada. Frequentemente, técnicas de saltos de canais são utilizadas para evitar o *jamming*.

Outro requisito importante é a segurança, uma preocupação frequente dos gerentes de fábricas. Ao contrário de cabos e fibras ópticas, a RF é um meio de comunicação que pode ser facilmente interceptado. Para dificultar ataques à rede, além de técnicas de espalhamento espectral, algum sistema de segurança e checagem de integridade das mensagens deve ser empregado. Algoritmos para encriptação de mensagens são frequentemente utilizados com o objetivo de proteger os dados de processos bem como para a verificação da integridade das mensagens. Chaves múltiplas, simétricas ou assimétricas em conjunto com contadores que não se repetem, são utilizados para evitar a descoberta de informações críticas e ataques de repetição.

O acesso múltiplo deve ser provido de modo que vários FDs possam utilizar o mesmo meio de comunicação. Dentro de uma mesma faixa de frequências, a camada de comunicação de dados pode prover acesso múltiplo através do emprego de alguma técnica, tal como as divisões de frequência (*frequency division multiple access* - FDMA), tempo (*time division multiple access* - TDMA) ou por código (*code division multiple access* - CDMA). No FDMA, múltiplos pares de rádios podem comunicar entre si ao mesmo tempo, através da atribuição de diferentes canais (frequências) para cada par. No TDMA as transações entre os diversos dispositivos ocorrem em intervalos de tempo previamente definidos, comumente chamados *time slots*. Alguns *time slots* são designados para a inclusão de novos dispositivos que queiram juntar-se à rede, outros são designados para publicação periódica de dados e outros, ainda, para controle da rede. Esta estratégia possibilita baixo consumo pela ciclagem dos dispositivos, que despertam somente no devido *slot* de tempo para comunicação. O CDMA emprega as técnicas de espalhamento espectral anteriormente mencionadas, para propiciar comunicações simultâneas entre pares de rádios. Isto é possível porque somente um receptor possui de antemão o código que irá decodificar a mensagem de um transmissor.

Os sistemas de acesso múltiplo ao meio utilizam algum mecanismo de arbitração para que o canal possa ser compartilhado. As técnicas baseadas em probabilidade de acesso, tais como o CSMA-CA (*Carrier Sense Multiple Access - Collision Avoidance*) não garantem determinismo, uma vez que colisões ocorrem com frequência, e estas aumentam com o aumento da densidade da rede. Isto se deve ao problema denominado “terminal escondido”, ilustrado na Figura 1. Nela, um dispositivo *A* transmite para *B*. *C* também deseja transmitir para *B* e portanto, inicia o processo de escuta de canal através do mecanismo CSMA. Como *A* e *C* não tem possibilidade de enlace devido ao alcance dos transceptores, não percebem as transmissões um do outro e desta forma, assumem que o canal está livre. Ambos iniciam suas transmissões o que leva à colisão entre as mensagens.

Por outro lado, as técnicas TDMA garantem determinismo uma vez que as comunicações ocorrem em tempos determinados, assim como as retransmissões que porventura sejam necessárias em função de interferências. Com o correto escalonamento dos *slots* de tempo, o gerenciamento da rede pode prever o maior e o menor atraso em uma comunicação, garantindo-se desta forma o determinismo da rede.

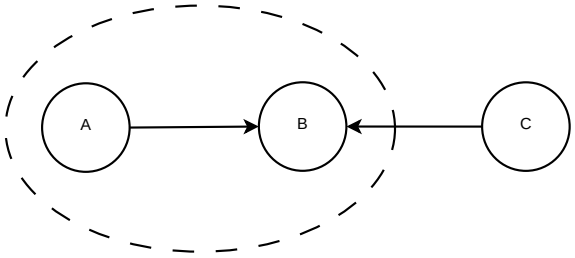

Figura 1: O problema do terminal escondido, que afeta o mecanismo MAC CSMA-CA.

As topologias de rede empregadas fazem com que alguns protocolos sejam mais sensíveis à falhas do tipo bloqueio e *jamming* do que outros. Topologias do tipo estrela e árvore são menos resistentes, pois nestas há concentração do fluxo de dados em algum ponto das mesmas. Por outro lado, os protocolos que possibilitam as chamadas redes em malha permitem diversos caminhos para comunicação. Porém, há diferentes graus de redundância em redes em malha, no que se refere à suas simetrias, balanceamentos, profundidade, entre outros parâmetros. Para cada tipo, há vantagens e desvantagens e, ao final, estes quesitos refletem na confiabilidade geral da rede.

O emprego de concentradores de rede, sejam eles coordenadores locais ou globais, implica em perda de robustez, uma vez que parte ou toda rede falha se eles falham. Desta forma, idealmente todos dispositivos da rede em malha devem ser roteadores e o gerenciamento deve ser distribuído, de forma que uma falha isolada não degrade de forma significativa o comportamento da rede.

A escolha da faixa de RF empregada também é um critério de grande importância. A realização de enlaces dentro de ambientes industrias demanda cautela, pois os fenômenos de interação da OEM com a matéria podem ser diversos. Dependendo do tipo da planta industrial podem ocorrer reflexões ou absorções das OEM, que podem ser negativas ou até mesmo positivas ao enlace. Atualmente, a maioria dos sistemas de comunicação sem fio industriais utiliza redes do tipo LR-WPAN (*low rate-wireless personal area network*) , normalmente nas faixas de 900 e 2400 MHz. Redes WLAN também são utilizadas, e deste modo, comunicações na faixa de 5 GHz também poderão estar presentes. Deve-se primariamente verificar se as faixas de frequências utilizadas são permitidas no local da instalação, já que algumas das bandas ISM não são permitidas em certos países. Posteriormente, recomenda-se a análise do espectro de frequências no local, afim de detectar a presença de sinais que poderão ser interferentes. Como exemplo, a faixa de 2,4 GHz é utilizada por diversos equipamentos, tais como telefones sem fio tipo DECT (*Digital Enhanced Cordless Telephone*), fornos de micro-ondas, WLAN e IEEE 802.15.4 (original, proprietários, Zigbee) e Bluetooth. Embora não seja comum a utilização de fornos de micro-ondas em ambiente industriais, há uma grande possibilidade de que sinais de redes WLAN estejam presentes. Os sinais interferentes podem degradar o desempenho de uma rede, levando à necessidade de retransmissões que inevitavelmente geram latência, e possivelmente impedem o determinismo da rede. Deve-se considerar também a coexistência com redes iguais, tal como frequentemente ocorre em redes WLAN.

De acordo com as leis fundamentais da propagação de OEMs no espaço, o alcance de um enlace diminui com o aumento da frequência. Deste modo, sistemas que empregam portadoras na faixa de 5 GHz apresentam alcances menores que sistemas de 900 MHz e, devido ao comprimento de onda muito pequeno, sofrem com os efeitos atmosféricos, requerendo frequentemente o uso de mais pontos de acesso o que resulta no aumento

do custo final da rede. Por outro lado, portadoras de mais alta frequência possibilitam maiores larguras de banda e por consequência, maiores taxas de dados.

Uma vez apresentada a análise dos principais requisitos que um protocolo de rede sem fio para aplicações industriais deve atender, resume-se:

a) Utilizar faixas de frequências restritas reduz o custo dos equipamentos mas pode acarretar em problemas de coexistência.

b) Utilizar portadoras de frequências elevadas permite grande largura de banda, mas reduz as distâncias máximas dos enlaces.

c) O protocolo empregado deve possuir mecanismos de segurança, afim de evitar espionagem e ataques.

d) O protocolo deve propiciar topologias com grande diversidade de caminhos para comunicação.

e) A técnica de acesso ao meio utilizada deve permitir multiplicidade de comunicação com a menor latência possível, e ainda, que esta seja controlável para permitir determinismo.

f) O consumo de energia do dispositivo de campo deve ser o mais baixo possível para permitir o uso de baterias.

g) De forma sintética, assume-se que a robustez do protocolo de rede sem fio empregado deve ser comparável à de redes cabeadas.

h) Outra forma de sintetizar as características necessárias para o emprego de determinado protocolo em ambientes industriais é a definição de três diversidades: temporal, espacial e de frequências. Se um determinado protocolo emprega estas diversidades de alguma forma, é potencialmente adequado.

2.2 Os padrões IEEE 802.X e derivados

Nesta seção são abordados os principais padrões IEEE 802.X para comunicações sem fio atualmente. A maioria dos protocolos de comunicação industriais sem fio fazem uso destas normas, integral ou parcialmente.

2.2.1 IEEE 802.15.1

O IEEE 802.15.1, popularmente conhecido como Bluetooth, é um protocolo de comunicação sem fio de baixo custo e consumo, desenvolvido para redes de curto alcance (até 10 m). Ele foi apresentado em março de 2002 e sua utilização abrange primariamente o uso pessoal e em escritório, mas a aplicação em ambiente industrial já foi fomentada (DAVIES, 2002). A técnica de espalhamento espectral utilizada neste padrão é o *frequency hopping* ou saltos de frequência realizados a uma taxa de 1600 saltos por segundo entre 79 canais de 1 MHz de largura de banda. A faixa utilizada é a de 2,4 GHz. Os saltos são realizados de forma pseudo-aleatória com uso de algoritmos combinacionais recursivos. Todos dispositivos Bluetooth compartilham a mesma banda de frequência e podem coexistir com outros dispositivos na faixa de 2,4 GHz que utilizem outras técnicas de modulação. Três classes de dispositivos são definidas, para as potências de transmissão de 1, 2,5 e 100 mW com alcances em linha de visada de 1, 10 e 100m respectivamente. A taxa máxima de transferência não efetiva é de 1 Mbps, 3 Mbps na versão 2.1 e 24 Mbps na versão 3.0, embora na prática os valores máximos sejam menores. A técnica de modulação empregada é a *Gaussian Binary Frequency Shift Keying* ou GFSK. Para estabelecer as conexões, uma pequena rede de até sete dispositivos pode ser formada, em uma topo-

logia tipo estrela denominada *piconet*. O dispositivo coordenador da rede *piconet* envia mensagens de requisição a cada 1,28 segundos, para poder descobrir novos dispositivos Bluetooth. O mestre aloca um endereço de membro da rede *piconet* para cada escravo ativo e controla as suas transmissões. O *clock*, necessário para manter a sincronização da rede, é provido pelo mestre somente. Um mecanismo de acesso ao meio TDMA é empregado, onde o mestre sempre transmite seus dados em *slots* pares do *frame* de dados, e os escravos, em *slots* ímpares, de acordo com a permissão do mestre. As requisições ocorrem sempre na ordem mestre-escravo, alternando para o próximo escravo no estilo *round-robin*. As redes *piconet* estão limitadas em sete dispositivos, mas podem ser agrupadas para formar uma rede maior, de topologia tipo árvore. Cada canal de comunicação é dividido em *slots* de 625 μs e diversas transações são feitas em cada canal.

Com base na breve explanação aqui apresentada acerca deste protocolo, procede-se a uma crítica com o intuito de avaliar a possibilidade da utilização do Bluetooth em ambientes industriais:

a) O esquema de espalhamento de frequências utiliza saltos adaptativos da portadora modulada, com largura de banda de 1 MHz. Por se tratar de um sinal de espectro estreito (frequência individual), o consumo de energia por bit não é pequeno, tornando o uso de baterias viável apenas para as classes de baixa potência (I e II). Aplicações industriais que requeiram alcances maiores, na faixa de centenas de metros, exigem potências maiores.

b) As redes piconet não apresentam topologia adequadamente redundante, uma vez que um bloqueio na comunicação entre um mestre e um escravo não pode ser resolvido pois não há outras rotas disponíveis.

c) Embora seja empregado um mecanismo de acesso ao meio do tipo TDMA, o determinismo não é garantido neste tipo de rede, pois as técnicas de escalonamento não foram originalmente desenvolvidas para aplicações industriais.

2.2.2 IEEE 802.11

O padrão IEEE 802.11, popularmente conhecido como WiFi ou WLAN, é amplamente empregada em redes de computadores domésticos, em escritórios e em alguns tipos de redes industriais. Caracteriza-se por prover alta taxa de transferência de dados e conexões de rede contínuas. Quatro variações do padrão foram popularizadas: as versões *a*, *b*, *g* e *n* (WILLIG; MATHEUS; WOLISZ, 2005). A variação *a* especifica a frequência de operação de 5 GHz (na Europa, 5,15-5,35 GHz e 5,47-5,725 GHz e as bandas ISM nos EUA, 5,15-5,35 GHz e 5,725-5,825 GHz). A camada física aplica a técnica de modulação OFDM (*Orthogonal Frequency Division Multiplexing*), um sistema multiportadoras que propicia altas taxas de comunicação, atingindo valores não efetivos de 54 Mbps. Nesta faixa de frequências, as paredes de uma casa ou escritório atenuam fortemente o sinal o que dificulta o estabelecimento de enlaces que não sejam em linha de visada). As taxas brutas de 54 Mbps possibilitam a transmissão dos pacotes Ethernet de 1500 bytes, resultando em taxas efetivas de cerca de 30 Mbps. Por outro lado, devido ao grande *overhead* neste protocolo, o envio de pacotes menores como 60 bytes por exemplo, resultam em taxas efetivas de cerca de 2,6 Mbps.

O IEEE 802.11 *b* emprega modulação do tipo DSSS utilizando a faixa ISM de 2,4 GHz. Entre onze e treze frequências centrais são disponíveis, porém na prática, somente três podem ser utilizadas simultaneamente sem que ocorram colisões, devido ao *overlap* entre os canais. Para suportar as taxas de 1 e 2 Mbps do sistema IEEE 802.11 original, o *payload* da camada física é modulado de 5,5 até 11 Mbps. A taxa máxima de transferência

é de cerca de 7 Mbps para pacotes Ethernet e 0,75 Mbps para pacotes de 60 bytes.

O IEEE 802.11 *g* é uma extensão da variante *b*, utilizando a mesma faixa de frequências, porém suportando camadas físicas diferentes. Desta forma é possível a compatibilidade com as variantes *a* e *b* anteriores. A taxa máxima efetiva é de cerca de 26 Mbps para pacotes Ethernet e 2 Mbps para pacotes de 60 bytes. Como as variantes *b* e *g* do IEEE 802.11 utilizam a banda de 2,4 GHz, estas podem sofrer interferências de equipamentos Bluetooth, telefones DECT e 802.15.4. Para evitar colisões, o IEEE 802.11 utiliza o mecanismo de arbitração CSMA-CA. O dispositivo que deseja transmitir dados verifica o estado do canal de comunicação previamente, observando-o por um tempo chamado DIFS (*Distributed Inter Frame Space*). Se o canal está ocupado, o transmissor aguarda por um período aleatório (*back-off*). Ao final da transmissão, se um sinal ACK não é recebido, o transmissor faz uma nova tentativa. Se o quadro de dados é recebido corretamente, o receptor envia um ACK de volta ao transmissor após um intervalo de tempo conhecido como SIFS (*Short Inter Frame Space*). O sistema CSMA-CA pode ser incrementado com o uso de um *handshake* RTS/CTS opcional para evitar o problema do "terminal escondido", anteriormente descrito. O usuário pode controlar o *handshake* configurando limiares para os tamanhos dos quadros. Se um quadro excede um limiar, o RTS/CTS pode ser aplicado (TANENBAUM, 2003).

Com base na breve explanação acerca deste protocolo, procede-se a uma crítica no que se refere a possibilidade de utilização do 802.11 em ambientes industriais:

a) Pelo fato de ser orientado à conexão e possibilitar altas taxas de dados, o 802.11 não permite o desenvolvimento de dispositivos de baixo consumo, que apresentem grande autonomia quando alimentados à bateria.

b) A técnica de arbitração empregada no IEEE 802.11 (CSMA-CA) não é adequada para controle de processos, uma vez que não garante determinismo em função das colisões que inevitavelmente ocorrem neste mecanismo.

c) O tamanho do pacote Ethernet é desnecessariamente grande para a maioria das aplicações industriais, onde mensagens pequenas relativas à variáveis como temperatura e pressão são corriqueiras. Excetuam-se as variáveis de diagnóstico de máquinas, tais como assinaturas de vibração por exemplo. Se o pacote Ethernet for utilizado com carga útil muito menor que 1500 bytes, a taxa de aproveitamento do protocolo é muito baixa, uma vez que o *overhead* do cabeçalho será muito maior que a carga útil.

d) Considerando aplicações industriais, a versão mais aconselhada é a *b*, em virtude de determinadas características. Entre elas, a modulação DSSS é empregada e propicia menor consumo de energia em taxas adequadas a controle de processos (1 Mbps, por exemplo). Com taxas menores, o enlace é mais eficiente, apresentando faixas dinâmicas da ordem de 105 dBs. Ainda assim, existe a necessidade de associação para conexão com um ponto de acesso (AP, *access point*), o que não é desejável para aplicações de baixo consumo.

e) Redes WLAN são aplicáveis nos níveis acima de controle de processos, onde os requisitos de tempo real não são necessários. Neste caso, as taxas mais elevadas propiciadas pelas redes WLAN são vantajosas em função do baixo custo dos equipamentos, que ocorre naturalmente pela utilização em larga escala. Nestes casos, os equipamentos são alimentados pela rede elétrica, não havendo a necessidade de utilização de técnicas de redução de consumo.

f) Embora divulgado por diversos fabricantes de equipamentos WLAN, as redes 802.11 não forma malhas mas estrelas. Desta forma, não é possível garantir diversidade de caminhos para comunicação com redes deste tipo.

g) A emenda *s* da norma 802.11 pode ser considerada promissora para aplicações industriais. Nesta, são incluídas as topologias em malha verdadeira e, para tanto, enlaces tipo *ad hoc* não orientados à conexão são possíveis. Estas características, inicialmente direcionadas para aplicações de baixo custo onde as infraestruturas de rede são mínimas, são de fato aplicáveis à redes industriais. Entretanto, o desenvolvimento de dispositivos compatíveis com esta variação (assim como o que ocorre com as demais tecnologias atuais) é altamente dependente das tendências de mercado e, desta forma, poderá ficar restrita ao campo da pesquisa somente.

h) A emenda *e* da norma 802.11 também apresenta características que tornam o padrão adequado para aplicações industriais. Nesta versão, modificações implementadas na sub camada MAC permitem melhoramentos na qualidade de serviço de modo a possibilitar comunicações adequadas em rede sensíveis a atrasos. Inicialmente desenvolvidos para aplicações domésticas e de escritório, os equipamentos que adotam o padrão *e* podem ser aplicados na indústria uma vez que um período livre de contenção é disponibilizado no quadro de dados. Desta forma é possível garantir qualidade de serviço através de comunicações exclusivas, sem disputa pelo meio, e portanto, livres de colisões. Os tipos de dados que trafegam na rede são classificados de acordo com suas prioridades, onde são definidos níveis diferentes para aplicações do tipo *voice over* IP, por exemplo. Da mesma forma que ocorre com a emenda *s*, há poucos equipamentos disponíveis no mercado compatíveis com esta variação do 802.11.

2.2.3 IEEE 802.15.4

O padrão 802.15.4 foi estabelecido pelo IEEE em outubro de 2003. Para o desenvolvimento do padrão, objetivou-se a criação de sistemas de comunicação de baixo custo e consumo, para serem utilizados em equipamentos de sensoriamento e controle. Diferentemente do Bluetooth e do WiFi, o 802.15.4 foi especificamente desenvolvido para uso em aplicações de baixas taxas de comunicação e que requerem baixo consumo de energia. Estas características vem ao encontro das especificações de sistemas industriais, onde as redes são na maioria das vezes estáticas e frequentemente transmitem somente pequenos pacotes de dados. A fim de fomentar o uso do padrão, as faixas de frequências empregadas são livres de licença (radiação restrita). Na banda de 2,4 GHz, a técnica de modulação utilizada é a DSSS, que propicia circuito de RF de menor complexidade, facilitando o desenvolvimento de transceptores de baixo custo. A taxa máxima de dados é de 250 kbit/s em um único canal. Dezesseis canais são definidos, sendo que o último canal não pode ser utilizado globalmente. Entretanto, a taxa efetiva do protocolo fica em torno de 60 kbps, devido à técnica de modulação e ao mecanismo MAC utilizado. A coexistência deste protocolo com outros tais como Bluetooth, WLAN e DECT é amplamente possível e verificada.

Como a camada física do 802.15.4 é utilizada nos protocolos WH e ISA 100.11a, apresenta-se aqui uma explanação mais aprofundada do funcionamento do mesmo.

O protocolo 802.15.4 contempla dois tipos fundamentais de dispositivos: os do tipo FFDs (*full-function devices*) e os RFDs (*reduced-function devices*). Um FFD pode ser um coordenador de rede e propicia a comunicação com outros FFDs em modo ponto a ponto. Por outro lado, os RFDs são sempre associados a FFDs e estão limitados à troca de dados com estes somente. Não há possibilidade de comunicação entre RFDs. Todos dispositivos tem endereço único de 64 bits, além da possibilidade de obtenção de endereços curtos, de 16 bits, fornecidos pelo FFD. Com relação ao protocolo MAC utilizado no padrão, há duas formas diferentes de operação. No modo não balizado (*non beaconed*), todos os

nós da rede utilizam a variante CSMA sem *slots* de tempo. Nesta variante, o nó inicia a transmissão de pacotes sem efetuar CCA (*Clear Channel Assessment*) imediatamente. Ele introduz um tempo aleatório chamado *back-off* com o intuito de evitar colisões. No modo balizado (*beaconed*), o coordenador da rede impõe uma estrutura chamada *superframe*. O coordenador da rede transmite os *beacons* periodicamente, escolhendo um número de períodos configuráveis entre 15,36 ms e 251,65 s. O restante do *superframe* começa com um período de contenção de acesso na qual os RFDs acessam o meio de acordo com uma variação do CSMA-CA *slotted* que agrega mais *overhead* que a variante *unslotted*. Além disto, para estes dois modos de operação, existe um período de inatividade onde todos os nós, incluindo o coordenador são postos em estado de dormência com o intuito de conservar energia da fonte de alimentação. Os pacotes de dados são reconhecidos através de ACKs e suportam retransmissões.

Após desenvolvimentos e pesquisas chegou-se a conclusão de que a taxa efetiva do 802.15.4 na faixa de 2,4 GHz varia entre 38 e 70 kbps em um enlace entre dois dispositivos.

Por questões de segurança, o 802.15.4 provê serviços (não obrigatórios) de autenticação e encriptação das mensagens enviadas e recebidas.

O padrão possibilita a identificação de cada nó na rede por um número único além do método e formato da comunicação entre estes nós, mas não especifica, além de uma ligação do tipo ponto a ponto entre os nós, a topologia de rede bem como os esquemas de roteamento ou adesão e reparação das conexões entre os nós. Estas características de rede ficam a cargo das camadas superiores, como feito no padrão Zigbee.

O mecanismo MAC controla o acesso ao canal de comunicação provendo o controle do fluxo de dados através de reconhecimentos (ACKs) e retransmissões. Também é responsável pela validação dos dados recebidos e sincronização além de proporcionar serviços para as camadas superiores. As topologias de rede do 802.15.4 são geralmente classificadas como redes em estrela ou ponto a ponto. Topologias mais complexas como as em malha necessitam adição de código nas camadas superiores da pilha de comunicações e são empregadas em redes onde uma maior robustez é requerida. Somente dispositivos FFD podem ser coordenadores de rede para estabelecer e manter uma rede e coordenar a transmissão entre os diversos nós. A descrição das topologias é apresentada a seguir:

Topologia de rede do tipo estrela: nesta topologia, há um coordenador e um ou mais dispositivos RFD que são os nós finais ou dispositivos FFD que enviam mensagens diretamente a um coordenador. O número máximo (teórico) de dispositivos de rede é 65535.

Topologia de rede do tipo árvore: nesta topologia, há um coordenador que age como uma raiz e nós RFD ou FFD são conectados a ele com o intuito de aumentar a abrangência da rede. Os dispositivos tipo RFD podem ser as “folhas” de uma rede em árvore enquanto os FFD atuam com os “ramos” da estrutura.

Topologia de rede do tipo malha: nesta topologia, qualquer nó fonte pode comunicar com qualquer nó destino. Os nós roteadores e o coordenador estão conectados uns com os outros dentro do limite de suas faixas de alcance para que possam rotear os pacotes de dados. O rádio receptor no nó coordenador necessita estar ativo todo o tempo para que possa coordenar o roteamento dos pacotes. Existe a possibilidade de sincronizar os diversos nós roteadores através de relógios de tempo real em cada nó, sincronizados com o nó mestre para que despertem ciclicamente para realizar a comunicação, independente de delimitadores controlados (*beacons*), tal como é feito nas redes ISA 100.11a e WH. Isto leva a redução de consumo, mas requer precisão dos temporizadores utilizados. O padrão

Zigbee emprega uma versão simplificada do protocolo de roteamento AODV (*ad hoc on-demand distance vector*) cuja rota de comunicação é formada quando uma requisição é gerada por um nó fonte. Através da troca de informações entre fonte e destino, a rota pode ser reservada por nós intermediários apenas para garantir tabelas de roteamento de modo que a comunicação seja garantida.

Para a formação da topologia da rede, a norma define um mecanismo para suportar um coordenador LR-WPAN que verifica a disponibilidade de uso dos canais de RF. O procedimento inicial de formação da rede é chamado procedimento de associação que permite que diversos nós juntem-se à rede. A busca por canais disponíveis é feita através da detecção da energia de RF em cada canal. A norma provê o mecanismo de detecção de energia mas não especifica a lógica da varredura, ou seja, a sequência de canais, o tempo de permanência de escuta e qual canal deve ser o escolhido para o enlace.

Geração de um coordenador de LR-WPAN. Durante o procedimento de inicialização de uma rede de topologia estrela, o primeiro passo é a geração de um coordenador. Qualquer nó FFD pode ser um coordenador. Primeiramente é feita a detecção de energia com o objetivo de escolher um canal adequado para a transmissão dos pacotes. Após, é feita a varredura por dispositivos ativos. Uma varredura ativa permite a localização de qualquer coordenador transmitindo quadros *beacon* em seu espaço operacional. Durante a varredura ativa, o nó primeiramente envia uma requisição de comando *beacon* e logo a seguir ajusta o parâmetro de duração da varredura. Se um *beacon* não puder ser detectado durante a varredura, o dispositivo FFD assume que não há coordenador no seu espaço operacional e pode começar a construir a sua própria rede enviando *beacons* periódicos na forma *broadcast*.

Associação de dispositivos à rede. Após a construção bem sucedida de uma LR-WPAN, os outros dispositivos dentro da área de cobertura desta rede podem comunicar com o coordenador para associar-se à mesma de modo a tornar-se um dispositivo comum após certos passos. Os passos para a associação são:

a) Varredura passiva: dispositivos comuns localizam qualquer coordenador transmitindo quadros *beacon* no seu entorno. Se um *beacon* é detectado durante o período de varredura (o dispositivo pode detectar vários *beacons*), a varredura passiva cessa. Caso contrário, o dispositivo começa outra varredura passiva após determinado período.

b) Envio de um comando de requisição de associação: um dispositivo envia um comando de requisição de associação para o coordenador utilizando o algoritmo CSMA-CA. O coordenador replica com um comando ACK imediatamente após a recepção e determina se vai permitir a associação de acordo com seus recursos e espaço de armazenamento. Se permitir, deverá alocar um endereço LR-WPAN para o nó que requisitou a associação.

c) Espera pelo processamento do coordenador: no protocolo 802.15.4, após a recepção de um comando ACK de requisição de associação oriundo do coordenador, o dispositivo ajusta um temporizador para aguardar pelo processamento da requisição de associação.

d) Enviando um comando de requisição de dados: quando o temporizador expira, o dispositivo envia um comando de requisição de dados para o coordenador utilizando novamente o mecanismo CSMA-CA. O coordenador replica com um comando ACK imediatamente após a recepção do comando e a seguir envia o comando de resposta de associação para o dispositivo.

e) Replicando com um ACK: quando um dispositivo recebe o comando de associação ele imediatamente replica com um ACK para o coordenador. Após o recebimento do ACK, o coordenador fecha o canal. Após todos estes procedimentos, o dispositivo está

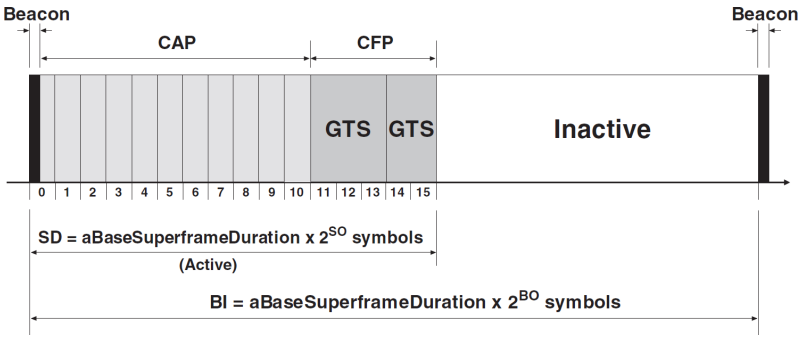

Figura 2: Em detalhe, a estrutura de um quadro IEEE 802.15.4 do tipo balizado (IEEE 802.15.4).

efetivamente associado a uma LR-WPAN e tem seu próprio endereço para que possa se comunicar com os outros.

Estratégia de comunicação. Objetivando atender os requerimentos de baixo consumo e qualidade de serviço (QoS), as redes que atendem a norma 802.15.4 podem ser divididas em *beaconed* e *unbeaconed* LR-WPAN. Em uma rede com *beacons* habilitados, o coordenador utiliza um *superframe* para regular a transmissão de pacotes. Um *superframe* começa em um *frame beacon* e termina no início do próximo *frame beacon*. Os quadros *beacon* enviados pelo coordenador provêm a informação relacionada ao comprimento do *superframe* e a quadros pendentes e sincroniza todos os dispositivos pelo coordenador. Um *superframe* possui períodos ativos e inativos. O período ativo possui 16 *slots* de tempo de mesmo tamanho. Um coordenador pode se comunicar com outros nós somente durante o período ativo e coordenadores e nós podem manter-se inativos ou se comunicar com outras LR-PAN no período inativo. Os *beacons* provêm portanto, a informação necessária para um novo dispositivo que tem a intenção de juntar-se à rede. Quando transmitido periodicamente (rede com *beacons* habilitados), estes podem ser utilizados para sincronizar os dispositivos previamente reconhecidos, para identificar a LR-WPAN e para descrever a estrutura do *superframe*. A fim de prevenir colisões, duas técnicas podem ser empregadas, dependendo dos requisitos de tempo real da aplicação. No período de contenção de acesso (*contention access period* - CAP), cada nó pode transmitir de acordo com o mecanismo de acesso CSMA-CA utilizando um determinado algoritmo de espera probabilístico (*backoff*). Diferentemente, no período livre de contenção, somente nós com *slots* de tempo reservados podem tentar transmitir pacotes de dados, evitando portanto, prováveis colisões. A fim de propiciar a transmissão segura de dados, os *slots* de tempo garantido (*guaranteed time slots* - GTS) podem ser reservados para nós que o requeiram. Nesta porção de tempo, somente os nós que requisitaram podem transmitir, encontrando para tanto, um canal de RF livre. As dimensões da janela de dados, as durações de tempo da fase ativa (também chamada de *superframe duration*, SD) e a duração do GTS são definidas por dois parâmetros trocados no sinal de *beacon*. Este sinal é periodicamente enviado pelo coordenador para sincronizar todos os nós remotos da rede e sinalizar o início da janela *beacon*, como mostra a Figura 2 (INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, 2003).

Quando os *beacons* estão habilitados, a camada MAC do 802.15.4 propicia grande flexibilidade ao protocolo, sendo esta, a forma de utilização a mais corriqueira. Quando os *beacons* são desabilitados, todos os nós da rede devem coordenar o acesso à rede utilizando o mecanismo CSMA-CA. A vantagem deste modo de operação é a maior pos-

sibilidade de escalonamento de mensagens e a possibilidade de auto-organização da rede. Por outro lado, ao desabilitar os *beacons*, não há mais garantia de entrega de dados, a não ser que os relógios dos nós da rede sejam sincronizados para cadenciar a transmissão de mensagens. Isto exige, além da grande precisão dos temporizadores de cada nó, o emprego de alguma técnica de disseminação de relógio pela rede.

O modo *beaconed* permite a alocação e liberação de *slots* de tempo em um *superframe*, criando os GTS e provendo garantia mínima de serviços. Deste modo, é possível predizer o desempenho temporal da rede no pior caso de funcionamento. O conceito de GTS é uma forma ao acesso múltiplo por divisão de tempo (TDMA) para a alocação de *slots* de tempo, porém, de tamanho limitado. Uma quantidade de largura de banda é periodicamente reservada para garantir determinado fluxo de dados. A quantidade de largura de banda disponível é proporcional à duração dos *slots* de tempo e de sua periodicidade.

Pelo breve estudo da norma aqui apresentado, verifica-se que, apesar de propiciar baixo consumo de energia, o 802.15.4 não consegue prover determinismo na sua forma original utilizando o mecanismo MAC CSMA-CA. Em função disto, diversas tentativas foram feitas com o intuito de se utilizar o período livre de contenção com os GTS, através dos quais é possível garantir determinismo. Infelizmente, o espaço de tempo ali disponível é pequeno, de forma que não é possível atender redes compostas por muitos nós. Verifica-se ainda, que na possibilidade de se utilizar o período de contenção, o tempo necessário para sincronismo dos dispositivos durante o processo de escuta pelos *beacons* faz com que muita energia seja gasta. Ainda, as colisões que ocorrem naturalmente no 802.15.4 na sua forma integral, invariavelmente aumentam com o aumento do número de nós. Embora o número de nós em redes para controle de processos industriais não seja grande (na faixa de 50 a 100 nós), ainda assim o número de colisões poderá ser grande de forma a gerar um elevado número de retransmissões e consequente aumento no consumo de energia.

2.2.4 IEEE 802.15.4, emenda *e*.

O desenvolvimento da norma 802.15.4 propiciou o surgimento de rádios de espectro espalhado de baixo consumo e custo, para emprego direto em LR-WPANs. As diversas revisões foram feitas com o objetivo de contemplar aplicabilidade global, em função das especificações locais acerca dos padrões de radiação restrita. Poucas alterações foram feitas em relação às técnicas de acesso ao meio mas sim, algumas feitas na camada física, incluindo o uso de técnicas do tipo *chirp*. Porém, recentemente um novo grupo de trabalho reuniu-se com o objetivo de produzir uma nova emenda, denominada *e*. A análise desta emenda é válida para este trabalho, pois reflete o quão importante foi o desenvolvimento do protocolo WH para a indústria, uma vez que esta emenda claramente busca atender os requisitos implementados no WH e 100.11a. Especificamente, as alterações propostas pela emenda para a camada MAC vão de encontro com as necessidades de um protocolo de redes sem fio para aplicações industriais. Com a aceitação e implementação da emenda *e*, boa parte das subcamadas MAC e LLC poderão ser implementadas em hardware, diretamente no transceptor, o que permitirá maior *footprint* para as camadas de rede e aplicação no MCU (este é, de fato, um problema atualmente enfrentado pelos desenvolvedores de WH / ISA 100.11a). As principais alterações propostas na emenda *e* são:

Modo de extensão síncrono e determinístico multi canais. Este modo foi concebido para dar suporte à aplicações industriais que necessitam latência determinística e enlaces

de alta confiabilidade. Uma rede operando neste modo utiliza *beacons* para o sincronismo de *superframes* múltiplos. O uso de vários *superframes*, também propicia melhora na operação através de GTS no 802.15.4-2006. Para o período CFP, o salto de canais é agora suportado. Os saltos podem ser adaptativos ou diretos. Um par de dispositivos pareados despertam em um determinado *slot* GTS reservado para a troca do quadro de dados e de ACK.

Modo *Time Slotted* com saltos de canais. Este modo foi desenvolvido para aplicações em automação de processos, com foco especial em monitoramento. Em operação, todos dispositivos na mesma rede são sincronizados através de *slotframes*. Todos *slots* de tempo estão contidos em um *slotframe* e são repetidos ao longo do tempo. As comunicações em um *slot* incluem TX, RX e ACK.

Modo de rede determinístico, de baixa latência. Neste modo, o foco de aplicação reside em automação de fábricas, tais como automotivas, maquinário de usinagem e robôs. São utilizados *beacons* e *slots* de tempo para prover determinismo. Para obter baixa latência, os quadros são pequenos, assim como a densidade da rede. O *overhead* do mecanismo MAC é reduzido substancialmente de modo a prover *superframes* pequenos que, por consequência, reduzem a latência. Além disso, uma função de ACK de grupo foi incluída, para reduzir os requisitos totais de banda.

Modo de baixa energia. Este modo foi desenvolvido com o foco em rede LR-WPAN baseadas em IP (chamadas IP-cêntricas, para utilização em IoT). Ele cria a ilusão de que o nó está sempre disponível, enquanto permite baixa latência, capacidade de *multicast* e ACKs síncronos. Os receptores adquirem amostras periódicas dos canais para tráfego de rádio, na faixa dos ms. O transmissor transmite um quadro de sequência de despertar (*wakeup frame*) antes de cada quadro de carga útil. Na primeira ocasião de envio, o processo de despertar é assíncrono com o receptor. Uma sequência de despertar curta é utilizada para sincronismo com o receptor.

Neste modo, há um compromisso entre latência e baixo consumo, o que o torna mais adequado para emprego em RSSF. Outra possibilidade suportada é a transmissão iniciada pelo receptor. O receptor envia periodicamente pacotes de requisição de dados do tipo *broadcast* e escuta por um curto período de tempo em busca por transmissões (na casa dos ms). O transmissor espera a recepção de um pacote de requisição de dados oriundo de um receptor e imediatamente envia o quadro de carga útil. Desta forma, o transceptor funciona melhor quando altas latências são toleradas (tipicamente o caso de RSSF).

Comportamento de baliza (*beacon*) melhorado. Neste modo, informações adicionais são incluídas nos *beacons*, de modo a propiciar filtragens. Assim, são incluídos elementos que permitem a adição de informações no formato do quadro existente sem a necessidade de um novo tipo de quadro para esta tarefa.

Outras características importantes da emenda *e* são:

a) Agregação rápida, que permite a um dispositivo enviar uma resposta de associação em modo de comunicação direta, eliminando atrasos na formação de rede, como ocorre tipicamente no WH.

b) Métrica de rede, que inclui diversos parâmetros para medição de QoS; *macRetryCount*, *macMultipleRetryCount*, *macTXFailCount*, *macTXSuccessCount*, relacionado a transmissão dos quadros de dados. Além destes, são adicionados os atributos *macFCSErrorCount*, *macSecurityFailure*, *macDuplicateFrameCount*, *macRXSuccessCount*, re-

lacionados à recepção de quadros de dados.

c) Melhoramentos em quadros de reconhecimento (ACK), que podem ser seguros, incluir cargas úteis e, com atraso programável, de modo a permitir tempo suficiente tempo para decriptação (um problema recorrente em implementações de segurança baseadas em software).

Além disso, é importante ressaltar que a diversidade de canais é agora suportada, sendo esta, uma das três diversidades consideradas fundamentais.

De forma resumida, pode-se afirmar que a emenda *e* do protocolo 802.15.4 foi concebida para atender requisitos que antes só poderiam ser feitos desabilitando-se a MAC original do 802.15.4. Nesta emenda, vários modos são possíveis, e foram desenvolvidos para aplicações diferentes tais como redes IP-cêntricas, redes de baixa latência, redes determinísticas, coexistindo e consumindo menos energia. Pode-se vislumbrar a possibilidade de que o comportamento de um rádio seja modificável de acordo com a necessidade, permitindo que um mesmo dispositivo possa apresentar baixa latência, na ocasião em que é móvel (ou está em movimento), baixo consumo, quando é estacionário e comunicando em baixa taxa de dados, ou determinístico, para o caso de aplicações industriais críticas.

2.2.5 ZigBee

O ZigBee é um conjunto de protocolos de comunicação de alto nível desenvolvidos pela Zigbee Alliance que utiliza o padrão 802.15.4 como base (camadas física e de dados) e adiciona camadas superiores para roteamento e funcionalidades de rede. O ZigBee foi estabelecido como um conjunto de especificações para redes LR-WPAN que podem ser utilizadas em uma variedade de aplicações comerciais. As topologias de rede possíveis incluem as básicas do 802.15.4 (estrela e árvore) e a topologia em malha, de grande aplicabilidade em RSSF. O mecanismo de acesso ao meio utilizado não é modificado, sendo portanto, o CSMA-CA padrão do 802.15.4. Como o ZigBee foi concebido para aplicações de baixas taxas e baixo consumo, ele pode ser facilmente incorporado a sistemas embarcados tais como linha branca, medidores de energia e controles remotos.

O Zigbee apresenta algumas vantagens com relação a outros protocolos graças à padronização. Entre as vantagens estão a interoperabilidade entre fabricantes, diversos transceptores disponíveis comercialmente, baixo custo facilidade de implementação. Graças à camada física do 802.15.4, os dispositivos Zigbee podem operar em ambientes de baixa relação sinal-ruído, estabelecendo enlaces de até 100 metros a uma taxa da ordem de 60 kbps. Outra vantagem é a utilização da faixa de 2,4 GHz que é mundialmente disponível. O protocolo pode ser implementado em MCUs de oito bits com memória ROM tipicamente com 64 kB e RAM de até 10 kB.

O Zigbee Aliance regulamenta a norma e é formado por mais de 200 empresas (VARCHOLA; DRUTAROVSKY, 2007). Através do consórcio, são desenvolvidas soluções para as áreas residencial, industrial e comercial. Embora seja teoricamente possível a formação de redes com até 65535 nós, os primeiros desenvolvedores reportaram grande instabilidade na rede com muitos nós operando em números muito menores que o máximo teórico. Isto se deve ao fato de que a estrutura de rede básica suportada tem topologia do tipo árvore, que restringe o número de endereços dos nós a números bem menores que os especificados teoricamente. Em 2006, um esquema de endereçamento aleatório com resolução automática de conflitos melhorou o desempenho de redes com grande número de nós. Foram feitas adições na norma ao longo dos últimos anos com o objetivo de aumentar a confiabilidade de dispositivos para aplicações industriais tais como a inclusão

da possibilidade de saltos entre os canais de RF.

Apesar dos esforços dos grupos de trabalho do padrão Zigbee na tentativa de torná-lo aplicável a indústria, este não pode ser considerado adequado, uma vez que o mecanismo CSMA-CA não foi alterado na sua forma original. Uma vez que as camadas inferiores do 802.15.4 são inalteradas, pouco pode ser feito nas camadas superiores com o intuito de tornar o protocolo robusto.

2.2.6 *Wireless*HART

O protocolo WH foi desenvolvido com o objetivo de estabelecer um padrão de comunicação sem fio para uso em aplicações industriais (SONG et al., 2008) (CHEN; NIXON; MOK, 2010). O WH é uma extensão do protocolo cabeado HART e permite compatibilidade total com sistemas legados, suportando aplicações de ciclos da ordem de 250 ms. Por ser um protocolo seguro, sincronizado em tempo e de baixo consumo, é adequado a controle de processos industriais. A compatibilidade é definida basicamente pela estrutura de comandos DDL (*device description language*), anteriormente desenvolvida pela organização HART. A coordenação de uma rede WH é feita pelo estabelecimento dos *slots* de tempo e em que frequência ocorrerá comunicação. As ligações entre os nós da rede são feitas através de *superframes* que são repetidos periodicamente para proporcionar tráfego de comunicações cíclicas e acíclicas. De acordo com as camadas OSI, o WH contém claramente as camadas física, enlace, rede, transporte e camada de aplicação (a mesma do HART original). As camadas são brevemente descrita a seguir:

Camada física: IEEE 802.15.4, somente na faixa de 2,4 GHz. O último canal (26) não é utilizado por questões de internacionalização.

Camada de dados, MAC e LLC: suporte a saltos de canais e TDMA. O CSMA-CA é utilizado em conexões compartilhadas (assíncronas).

Camadas de rede e transporte: topologia em malha e roteamento estático por grafos ou nas próprias mensagens.

As transações no WH ocorrem em *slots* de tempo de 10 ms cada. Os *slots* podem ser dedicados ou compartilhados entre vários nós da rede, disputados por meio do mecanismo CSMA-CA. Os requisitos de latência são estabelecidos pelo escalonamento da comunicação de modo que os pacotes alcançam os seus destinos em tempo conhecido, considerando o número máximo de saltos na rede e possíveis retransmissões. As mensagens podem ser múltiplas, fim a fim, com notificação espontânea de exceções, requisição e resposta *ad hoc* e em rajada (modo *burst*). A segmentação e reagrupamento de mensagens é feita de forma automática.

O protocolo especifica três elementos principais para a formação da rede: os dispositivos de rede sem fio (*Field Devices*, FD), o gateway e o gerenciador de rede. Os dispositivos de rede são conectados diretamente ao processo ou planta sendo, portanto, os nós sensores ou atuadores da rede. O gateway possibilita a comunicação entre o *host* da planta e os dispositivos de campo, possuindo para tanto, um ou mais pontos de acesso à rede (quanto maior o número de pontos de acesso, maior a redundância e vazão de dados). O gerenciador de rede é responsável pela configuração da mesma através do agendamento das comunicações entre os dispositivos, pela configuração dos *superframes* e dos links, gerenciando as tabelas de roteamento de mensagens e reportando o estado geral da rede. A estrutura geral de uma rede WH pode ser visualizada na Figura 3. Nela, é possível identificar os elementos fundamentais (FDs, gerenciador, gateway e ponto de acesso) além do *handheld* (com ou sem fio, para comissionamento e inspeção) e da ligação com o barra-

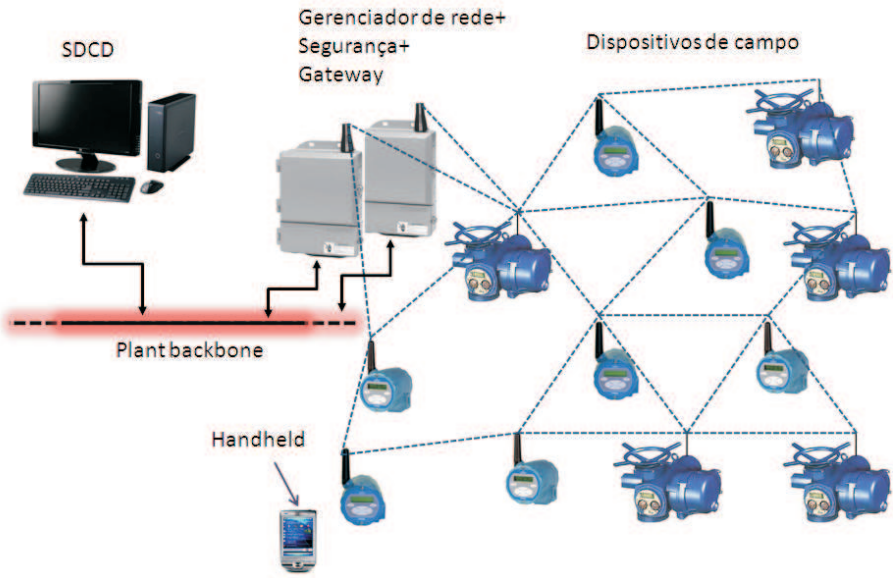

Figura 3: Esquema geral de uma rede *Wireless*HART.

mento de dados da planta. Nesta Figura, o gerenciador de rede, o gateway e o ponto de acesso estão contidos em um único equipamento, o que tem sido comercialmente usual.

As variáveis de processos podem ser publicadas nas combinações diversas (e/ou): periodicamente, quando há uma mudança de valor ou quando um determinado limiar é ultrapassado. Requisições e repostas de tráfego *ad hoc* são possíveis para a realização de rotinas de manutenção, verificação, configuração e calibração de dispositivos. As características das camadas física e de dados do protocolo são derivadas do 802.15.4 com alguns detalhes importantes que diferenciam o WH dos demais protocolos. Estas diferenças tornam o WH de fato adequado para aplicação em redes industriais. Entre elas, estão a utilização de potência de saída mais elevada (10 dBm) prevista para proporcionar alcances maiores, o uso de mecanismo de arbitração de meio determinístico e roteamento de mensagens redundante. As maiores diferenças em relação aos outros protocolos do tipo LR-WPAN residem na camada MAC. A arbitração do barramento de dados é feita por mecanismo TDMA, onde *slots* de tempo são organizados em *superframes*. Todos os dispositivos suportam múltiplos *superframes* para comunicações rápidas (desde 250 ms), lentas (na faixa de minutos), além de tráfego acíclico. Os *superframes* podem ser habilitados ou desabilitados pelo gerenciador conforme a demanda de comunicação. As comunicações ocorrem em *slots* de tempo determinados para transmissão ou recepção, e em cada uma destas, um canal RF diferente é utilizado. Os pacotes de *handshake* (ACK) incluem a informação de temporização para que o mecanismo TDMA entre os nós da rede esteja sincronizado. Desta forma, o relógio é distribuído pela rede. As eventuais colisões que ocorram durante o tráfego de mensagens em links compartilhados são tratadas pelo mecanismo CSMA-CA da mesma forma que no 802.15.4 original. Estes são utilizados para comunicações assíncronas ou para agregação de novos dispositivos na rede. Os canais são trocados a cada mensagem e a sinalização de *offset*, que indica o canal a ser utilizado, está no link onde a mensagem será transmitida ou recebida. Canais não suportados (provavelmente por problemas de coexistência) podem ser listados para que não sejam mais utilizados. A definição dos canais habilitados é feita somente na inicialização da rede.

No protocolo WH, as mensagens são priorizadas de acordo com quatro níveis distin-

tos: comando, dados de processo, normais e alarme. A priorização é feita para que o gerenciamento de latência e do fluxo dos dados na rede possa ser realizado. As mensagens do tipo comando têm a mais alta prioridade e são encontradas em qualquer pacote que contenha dados oriundos do gerenciador de rede. As mensagens de dados de processo estão em qualquer pacote que contenha os comandos tradicionais do protocolo HART como os comandos 3 e 9. Estas são as mensagens utilizadas normalmente durante o controle e monitoramento de um processo automático. As mensagens do tipo alarme são de baixa prioridade e contém informações normalmente acíclicas, tais como falha de caminho, falhas de hardware e outras. Mensagens caracterizadas como normais são as que não se enquadram em nenhuma das categorias anteriores.

Com relação ao roteamento de mensagens, em uma rede WH todos os dispositivos são do tipo FFD, para que possam formar a rede em malha completa. Qualquer dispositivo ou nó da rede deve fornecer ou repassar pacotes a fim de realizar o roteamento das mensagens dentro da rede. A rede é por consequência formada por caminhos de comunicações múltiplos e redundantes, e devem suportar roteamento para todos os nós vizinhos. Logo, os caminhos de comunicação possíveis devem ser verificados continuamente, de modo que o estado da rede é sempre conhecido e adaptado pelo gerenciador de rede. Segundo a organização HART, a robustez de uma rede WH bem constituída é tipicamente maior que 3 σ (99,73 $\%$) (DOHERTY; TEASDALE, 2006). Cabe ressaltar que a definição de rede bem constituída é imprecisa, mas neste caso refere-se a dois parâmetros: condição de radioenlace e rotas múltiplas. Considera-se o caso em que todos os enlaces ponto a ponto são ótimos e que existe pelo menos uma rota redundante para as comunicações entre os dispositivos de campo. As transmissões são gerenciadas através de tabelas de rotemento e realizadas nos modos *broadcast*, *multicast* ou *unicast*. Cada dispositivo mantém estatísticas a respeito de seus vizinhos, tais como níveis de intensidade do sinal recebido e contagem de pacotes, sendo estas informações, constantemente encaminhadas ao gerenciador de rede. A descoberta de novos vizinhos também é reportado pelos nós da rede além de desconexões e perdas. Na camada de aplicação, o WH utiliza o mesmo sistema HART, com os tipos de dados e procedimentos padrões da norma legada.

A segurança em redes WH é obtida pela encriptação das mensagens através do mecanismo de segurança AES-128, utilizando quatro chaves de criptografia simétricas. Em uma rede WH, diferentemente do Zigbee, o sistema de criptografia é obrigatório e desta forma está sempre ativo. Uma chave de acesso é utilizada para que novos nós possam agregar-se à rede e, uma vez que o novo dispositivo seja aceito pelo gerenciador da rede, as sessões de comunicação utilizarão outras chaves. Somente dispositivos conhecidos podem ingressar em uma rede WH, e para tanto, estes devem possuir a chave de entrada na rede (*Join Key*). Ataques de repetição também são evitados através de criptografia, com a verificação da integridade das mensagens. Estas, não são baseadas em uma chave apenas, mas em um contador único (que não se repete) e que impossibilita a replicação de mensagens reproduzidas externamente à rede.

O gateway de uma rede WH provê a integração com outros padrões de comunicação industriais, tais como Profinet e Profibus-DP. Os adaptadores de campo WH permitem que dispositivos HART do tipo ponto a ponto ou *multidrop* possam utilizar a rede sem fio.

Com a introdução da tecnologia sem fio ao protocolo HART, dois tipos de DLLs (*data link layers*) são possíveis: passagem de Token (*wired*) e o TDMA (*wireless*). Ambos suportam a mesma camada de aplicação comum do HART original. O modelo OSI de sete camadas para o HART pode ser visto na Figura 4.

Toda comunicação em uma rede WH passa pelo *gateway*, exceto as possíveis cone-

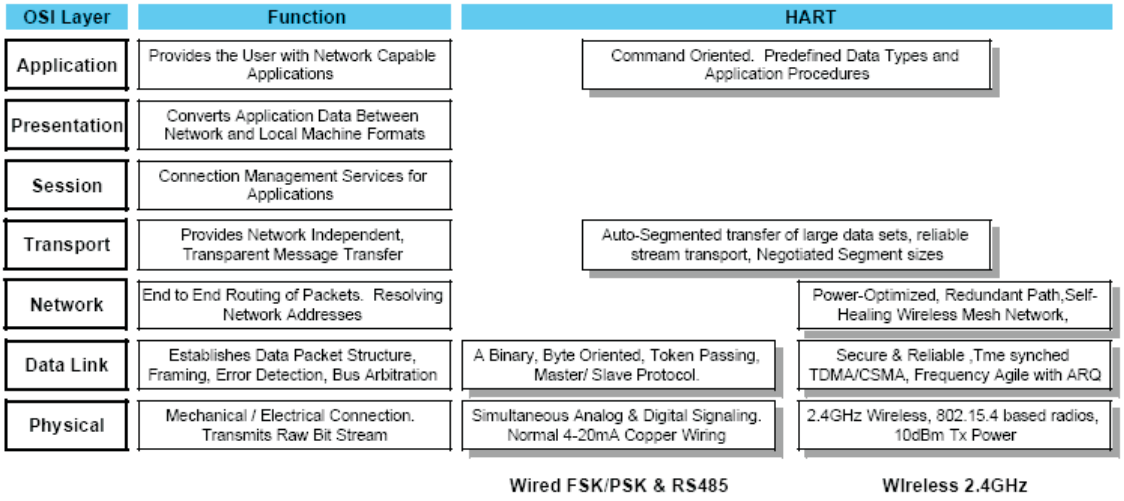

Figura 4: Modelo OSI para o HART e *Wireless*HART (CHEN; NIXON; MOK, 2010).

xões dedicadas (*ad hoc*), previstas na norma, mais ainda não implementadas. Desta forma, o WH não pode ser considerado um protocolo de rede de sensores sem fio pois estas são em geral descentralizadas. O gateway endereça os pacotes aos destinos específicos que podem ser os dispositivos de campo, a aplicação central (no SDCD) ou o gerenciador de rede. Os comandos HART padrão são utilizados para comunicação com os destinos específicos. A rede de automação da planta pode ser de qualquer tipo usualmente empregado, como as baseadas em TCP, Modbus ou Profibus DP.

No processo de inicialização da rede, o gerenciador de rede cria um *superframe* e configura o gateway e o ponto de acesso. As regras para arbitração do barramento garantem que as transmissões para os dispositivos ocorrem de forma ordenada, sem colisões. O gerenciador de rede é sempre o responsável por especificar quando um dispositivo está autorizado a transmitir e receber mensagens, e de que forma fará.

Na camada de dados do WH, são definidos:

a) Sincronização dos *slots*;
b) Identificação dos dispositivos que desejam acessar o meio;
c) Propagação das mensagens recebidas pela camada de rede;
d) Detecção de pacotes propagados por vizinhos;

A camada de dados é responsável portanto, pela propagação das informações de enlace (DLPDUs) através dos links *ad hoc*. Para que isto seja possível, o dispositivo deve possuir tabelas de vizinhos, *superframes*, links, e grafos que configuram a comunicação entre ele mesmo e o vizinho com o qual deve comunicar. Estas tabelas são produzidas pelas informações enviadas pelo gerenciador de rede. A tabela de vizinhos é preenchida na medida em que os vizinhos são descobertos, sem a intervenção do gerenciador, porém, qualquer comunicação com vizinhos é sempre ordenada pelo gerenciador. Um escalonador de mensagens avalia as tabelas de dispositivos e, conforme a necessidade de comunicação, escolhe o próximo *slot* de tempo a ser utilizado na escuta ou envio de um pacote. Máquinas de estado controlam a propagação dos pacotes através da camada MAC que deve manter os *slots* de transações normais e de serviço. A operação da MAC é fundamentalmente baseada em eventos e respostas a primitivas evocadas pelas camadas superiores.

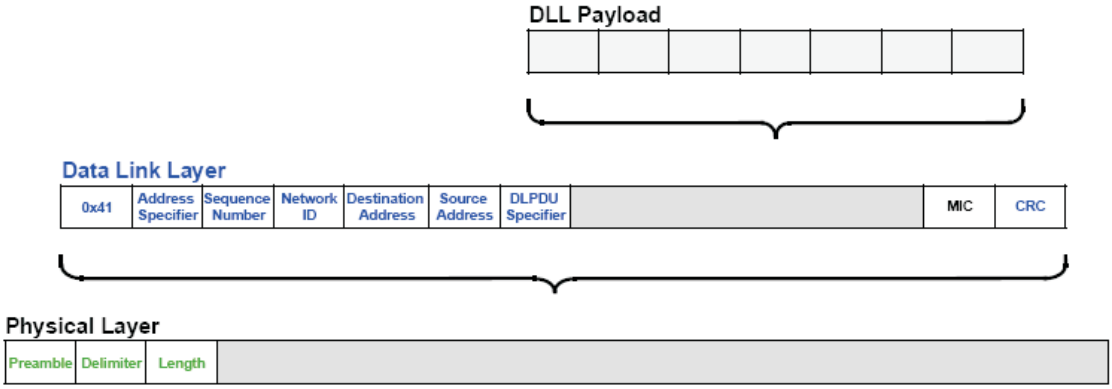

Figura 5: Estrutura da DLLPDU do WH (HCF spec075 r1.1).

Camada de enlace (DLPDUs).

O formato dos pacotes de DLLPDU pode ser visto na Figura 5. Cada pacote é consistido pelos seguintes campos:

a) Um byte com o valor fixo 0x41 ;
b) Um byte especificador de endereço;
c) Um byte com o número sequencial;
d) Dois bytes referentes ao identificador de rede;
e) Endereços de destino e origem: podem ser 2 ou 8 bytes, conforme o identificador utilizado (UID ou *nickname*);
f) Um byte especificador da DLLPDU;
g) O *payload* DLL;
h) O código de integridade de mensagem (MIC), em quatro bytes;
i) Dois bytes de CRC16 ITU-T;

O tamanho total do pacote pode chegar a 127 bytes.

Acesso ao meio por TDMA.

O WH utiliza o mecanismo TDMA e a troca sucessiva de canais para arbitrar o acesso à rede. O TDMA é uma técnica de controle de acesso ao meio largamente utilizada, que provê comunicação determinística e livre de colisões dentro da mesma rede. São utilizados *slots* de tempo onde as comunicações ocorrem. Uma série de *slots* de tempo formam um *superframe*, conforme pode ser visto na Figura 6.

Na Figura 7, é apresentado o detalhe da divisão de tempo que ocorre em um *slot* de tempo no mecanismo TDMA do WH. As diversas faixas de temporização são implementadas na sub camada MAC, geralmente através da reprogramação de módulos de temporização do MCU empregado.

O mecanismo TDMA para acesso múltiplo apresenta a vantagem de propiciar conexões livre de colisões, exceto em casos de coexistências. Também há possibilidade de uso de *slots* compartilhados e reuso de *slots*. Como as transações são programadas para que ocorram em tempos absolutos, o mecanismo propicia determinismo, um requisito fundamental em aplicações industriais. As dificuldades apresentadas na implementação destes algoritmos incluem a necessidade de sincronismo entre todos nós, que leva a necessidade de técnicas de disseminação de relógio pela rede e a eventuais problemas relacionados ao *drift* dos geradores de relógio (TJOA et al., 2004.). Todos os dispositivos de uma rede WH

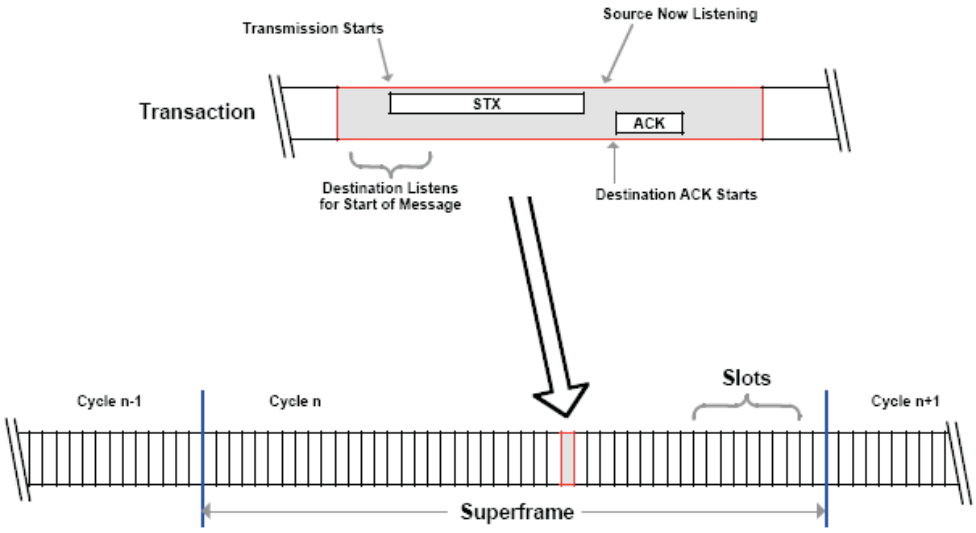

Figura 6: Estrutura de um *superframe* WH (HCF spec075 r1.1).

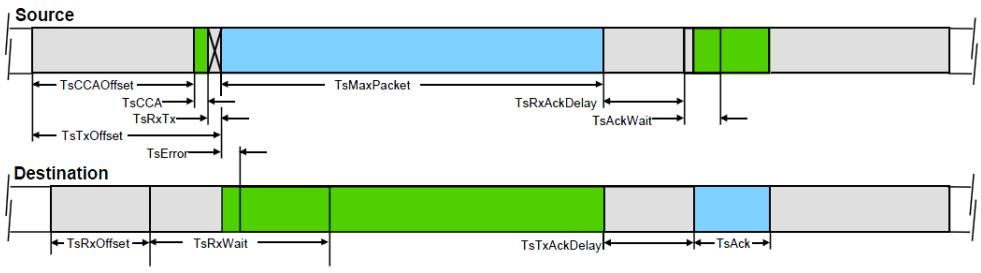

Figura 7: Estrutura de um *slot* de tempo WH (HCF spec075 r1.1).

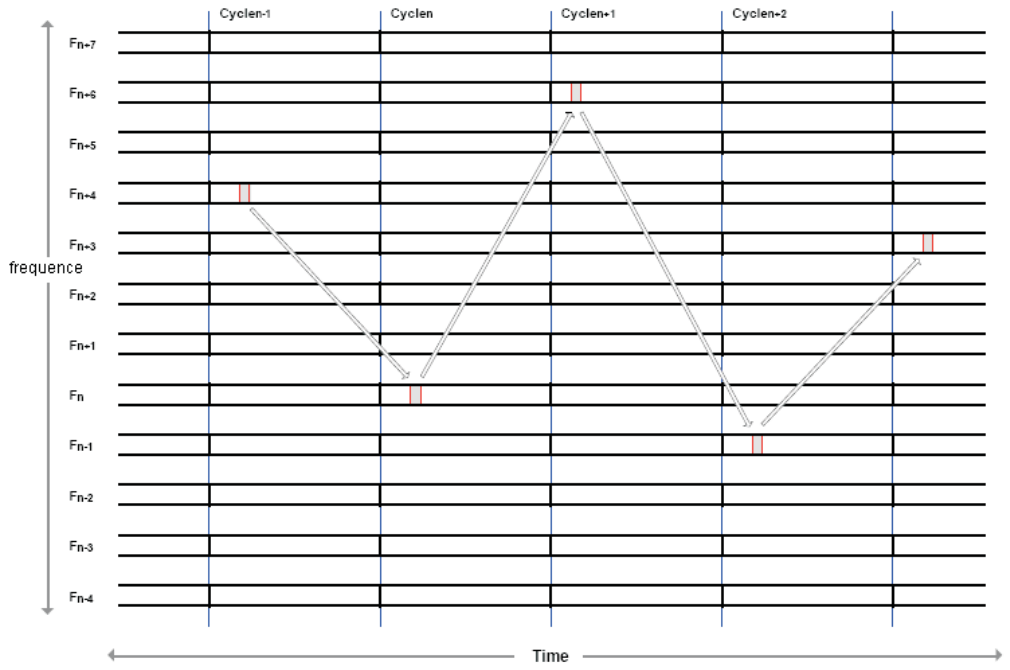

Figura 8: Saltos entre os canais da MAC do WH (HCF spec075 r1.1).

devem suportar diversos *superframes*. O tempo dos *slots* e o comprimento dos *superframes* em número de *slots* são fixos, e formam um ciclo de rede com uma taxa de repetição fixa. Consequentemente, as tolerâncias de tempo são especificadas no protocolo para garantir sincronização em toda rede. É imperativo que os dispositivos de campo saibam exatamente onde começa determinado *slot* de tempo que pretenda usar. Dentro do *slot*, a transmissão da mensagem começa em determinado instante de tempo, especificado após um certo intervalo de atraso aceitável. Durante este atraso, os dispositivos que vão comunicar tem tempo para ajustar o canal de RF a ser utilizado e comutar os transceptores para mudar dos modos TX para RX e vice-versa. Este atraso é necessário devido à tolerância dos *clocks*. Por conta disto, o receptor deve começar a escutar o canal antes do tempo ideal. Uma vez que a transmissão está completa, a direção de comunicação é revertida e o dispositivo destino indica, pela transmissão de um ACK, se recebeu corretamente os dados do transmissor. Todo esse processo ocorre dentro de um *slot* de tempo.

O salto de canais é adicionado ao TDMA para aumentar a robustez do protocolo (ver Figura 8) provendo diversidade de frequências e diminuindo a chance de que ocorram *jammings*. Para que os dispositivos possam comunicar na rede, são endereçados *superframes* e links contendo os *slots* de tempo e os *offsets* de canal.

***Slots* compartilhados.**

O WH permite o compartilhamento de *slots* de tempo entre dispositivos, que podem tentar transmitir mensagens, tais como anúncio de rede e repostas para agregação na mesma. Isto pode levar a colisões, logo o mecanismo de acesso ao meio nativo do 802.15.4 é empregado. No caso de uma colisão ocorrer, o dispositivo destino pode não receber a mensagem e por consequência não produzirá um ACK. Um atraso aleatório do tipo *back-off* é empregado devido a um *slot* que não recebeu ACK. Um dispositivo mantém duas variáveis para cada vizinho: o expoente de *back-off* (BOExp) e um contador de *back-off*. Ambas variáveis são inicializadas com valor zero. Quando uma transmissão em um *slot* de tempo compartilhado falha, o período de resguardo é calculado baseando-se

Tabela 1: Valores possíveis de BOExp no CSMA-CA do 802.15.4.

BOExp	Valores possíveis
1	{0,1}
2	{0,1,2,3}
3	{0,1,2,3,4,5,6,7}
4	{0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15}

no BOExp. Para cada tentativa mal sucedida, o BOExp é incrementado e um conjunto de números é calculado. Este conjunto é {0,1,2...L}, onde L é:

$$L = (2^{BOExp} - 1) \quad (1)$$

Os valores possíveis de um a quatro do BOExp são mostrados na Tabela 1. A partir destes valores calculados baseados no BOExp, um valor aleatório para o BOCntr é selecionado. Para cada link compartilhado subsequente, o BOCntr deve ser decrementado e somente quando o BOCntr correspondente é zero, o dispositivo fonte pode transmitir em um *slot* compartilhado. O valor do BOExp não poderá exceder o expoente MaxBackoff. Mensagens do tipo broadcast não poderão ser transmitidas em *slots* compartilhados.

Desenvolvimento do algoritmo MAC.

Como descrito anteriormente, o WH utiliza o TDMA e troca de canais para controlar o acesso ao meio. Cada dispositivo tem uma tabela na qual todas as informações para uma comunicação é especificada. Quando um dispositivo deseja transmitir uma mensagem, ele deve fazer uma chamada da função MAC que lê a tabela do dispositivo e verifica se está apto a transmitir no *slot* atual (se é reservado para ele). Se a resposta é positiva, a transmissão é permitida, caso contrário bloqueada. A técnica de troca sucessiva de canais permite que diferentes dispositivos transmitam no mesmo *slot* de tempo porém em frequências diferentes (ainda não implementado em dispositivos comerciais). Como o uso de *slots* compartilhados é possível, o sistema de detecção de colisões deve ser implementado. Quando um dispositivo necessita transmitir em um *slot* compartilhado, a função MAC verifica o estado do canal (CCA). Se o canal está ocupado, o *back-off time* é computado, conforme já explanado.

Tabelas de comunicação.

Todos os dispositivos de uma rede WH mantém uma série de tabelas que controlam as comunicações entre eles. Para o controle da rede, são definidas as seguintes tabelas:

a) *Superframes* e *links*: múltiplos *superframes* podem ser configurados pelo gerenciador da rede e múltiplos links em um *superframe* são configurados para especificar comunicação com um vizinho em especial ou *broadcast*.

b) Tabela de vizinhos: a tabela de vizinhos é uma lista de todos os dispositivos vizinhos com os quais o dispositivo tem a possibilidade de se comunicar.

c) Tabela de grafos: grafos são utilizados para encaminhar as mensagens desde a fonte até o destino e vice versa.

As tabelas de comunicação e as suas inter-relações podem ser vistas na Figura 9.

Sincronismo das tarefas do dispositivo.

Para que uma comunicação TDMA seja eficiente, o sincronismo dos *clocks* entre os dispositivos da rede é fundamental. Consequentemente, as tolerâncias são específicas

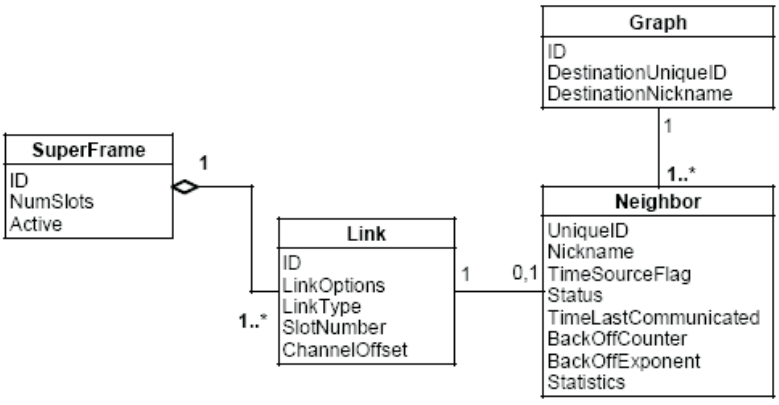

Figura 9: Tabelas de comunicação do WH (HCF spec075 r1.1).

para garantir o sincronismo da rede. É fundamental que o dispositivo saiba o instante de tempo em que um *slot* começa. Por esta razão, a primeira operação na função MAC é a leitura do valor do RTC com o qual, o dispositivo estará apto a calcular o tempo do *slot*:

$$ActualSlotNumber = 1 + ((\frac{ActualSimtime + exectime}{SlotSize})\%Superframe Size) \quad (2)$$

onde o *exectime* é o tempo de execução da tarefa do dispositivo que evocou a função MAC. No WH, a precisão do RTC também é importante para manter o cálculo do *nonce* (contador que não se repete), utilizado em conjunto com a chave de rede no sistema de autenticação de mensagens. Caso ocorra uma defasagem entre os tempos de dois rádios numa comunicação ponto a ponto, uma falha de autenticação ocorrerá e o pacote será descartado.

Transceptor de RF para WH.

No WH, o transceptor de RF emprega a camada física do 802.15.4 na frequência de 2,4 GHz. Nesta faixa, o projeto do circuito e da placa impressa demandam cautela, especialmente com relação ao casamento de impedâncias. A 2,4 GHz, uma trilha de circuito impresso, mesmo que curta, poderá apresentar uma indutância considerável de modo à levar o circuito à instabilidade. Desta forma, o projeto de transceptores WH demandam o uso de boas referências, pois descasamentos de impedâncias produzirão reflexões, reduzindo o desempenho dos circuitos (BOWICK, 2008). Devido às restrições temporais do TDMA, dispositivos compatíveis com o WH devem incluir um RTC com tolerância de frequência tão boa quanto 3 ppm, caso contrário, perderá o sincronismo, conforme explanado anteriormente. Outros componentes são selecionados de acordo com o tipo de dispositivo de campo a ser desenvolvido, relacionando-os com outros parâmetros tais como fonte de alimentação e interface com o usuário.

2.2.7 ISA 100.11a

O ISA 100.11a, assim como o WH, é um protocolo de comunicação sem fio verdadeiramente adequado para uso industrial. A norma foi desenvolvida para ser confiável e segura, abrangendo uma ampla gama de especificações em comparação com WH. Diferentemente do WH, diversos detalhes não foram deixados de fora na norma, tais como o projeto do gerenciador de segurança, especificado com políticas diversas. O sistema de saltos de canais também é bem mais amplo, com três técnicas distintas (o WH apresenta

uma apenas). Diversos protocolos podem ser tunelados, incluindo o HART. No rádio, o mecanismo TDMA é também empregado, mas o 100.11a pode apresentar intervalos de tempo dos *slots* variáveis, o que permite a utilização em diferentes tipos de processos industrias, de diferentes dinamicidades. Todos os dispositivos são roteadores, permitindo topologias em malha plena. A camada de rede permite vários tipos de comunicações, incluindo solicitação-resposta e mensagens de rajada. Mesmo sendo superior ao WH em diversos aspectos, o 100.11a não teve até o presente momento grande aceitação na indústria. Desta forma, o questionamento principal a cerca deste protocolo é: o que está (ou deu) errado no 100.11a? A negativa do 100.11a foi devida a grandes atrasos de definição e lançamento, ocorridas ao longo do tempo. Isto ocorreu devido à ampla faixa de cobertura do padrão, o que exigiu muito tempo para desenvolvimento e lançamento. Diferentemente do WH, que entrou no mercado de forma rápida, o projeto ambicioso do 100.11a acabou por fazê-lo complexo demais, deixando de ser uma opção factível para muitos desenvolvedores. Embora não tenha sido feito um levantamento preciso, percebe-se que os poucos equipamentos comercialmente disponíveis são simplificados, de forma que não diferem muito dos WH pois o 100.11a sobrepõe as características do WH. Ainda assim, o número de equipamentos comerciais disponíveis está crescendo a cada ano. Uma comparação abrangente entre WH e ISA 100.11a pode ser encontrada em (WANG, 2011).

2.2.8 WIA-PA

O protocolo WIA-PA (*Wireless Networks for Industrial Automation-Process Automation*) é o padrão chinês para arquiteturas de comunicação sem fio especificamente desenvolvido para aplicações industriais. O protocolo foi homologado em 2008 pela IEC e tornou-se o segundo padrão de comunicação sem fio industrial, depois do WH. O WIA-PA provê serviços de comunicação sem fio para dispositivos de campo empregados em medições e laços de controle em processos industriais. O protocolo adota uma topologia de rede de dois níveis (estrela e malha), suportando dispositivos de campo, *handhelds*, roteadores e gateways. As redes em estrela são constituídas pelos dispositivos de campo e as malhas, pelos roteadores. O modelo OSI aplicado ao protocolo inclui as camadas física, dados, rede e de aplicação. A camada física é baseada no protocolo 802.15.4, em duas bandas de rádio possíveis: 868/915 MHz e 2,4 GHz. A sub camada MAC de dados é a nativa do 802.15.4, ou seja, CSMA-CA balizado com períodos de contenção e livre de contenção. A parte ativa do quadro é empregada para a comunicação entre os dispositivos de campo e a parte inativa, para comunicação inter e intra *clusters*. No período ativo, a porção de contenção é utilizada para agregar dispositivos à rede, gerenciamentos intra *cluster* e retransmissões. Já a porção livre de contenção, é utilizada para comunicação entre os dispositivos de campo e os *cluster heads*. A camada de rede inclui o gerenciamento da formação de rede, descoberta e manutenção de rotas e o roteamento de pacotes em multi saltos. A técnica de roteamento empregada é do tipo estática. A Figura 10 apresenta um diagrama geral de uma rede do tipo WIA-PA.

Com base nas premissas apresentadas no início deste Capítulo, verifica-se que o WIA-PA pode atender parte dos requisitos necessários para uso em aplicações industrias, tais como baixo consumo de energia e uso de diferentes topologias de rede. Por outro lado, há uma série de detalhes que parecem desfavorecer este protocolo. Entre eles, o uso do CSMA-CA como forma primária de acesso ao meio na ponta extrema da rede, onde estão os FDs. Tomando-se uma visão mais ampla, pode-se chegar a duas conclusões importantes. Primeiro, do ponto de vista técnico, tem-se a impressão que os desenvolvedores tiveram interesse em utilizar o 802.15.4 na sua forma original com o intuito de explorar ao

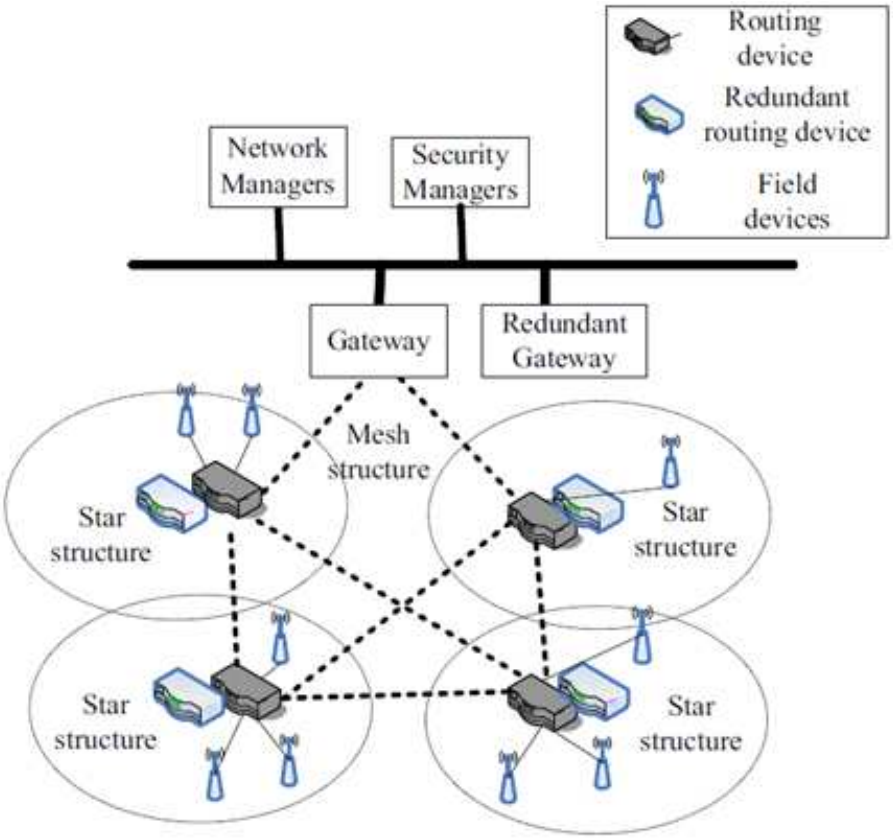

Figura 10: Visão geral de uma rede WIA-PA (MENG; XIAOJIE, 2011).

máximo suas capacidades inerentes. A utilização de redes em estrela permite a obtenção de dados de processo de forma simplificada e rápida, sem necessidade de modificação da camada MAC. Isto propicia não somente um curto período de desenvolvimento do padrão como o uso de rádios de baixo custo. Segundo, a não adoção de um padrão já estabelecido como o WH propicia o surgimento de um novo mercado, fomentando desenvolvedores locais. Esta prática é comum no mercado chinês e não necessariamente leva ao estabelecimento de padrões mais adequados.

2.2.9 Outros protocolos

Os protocolos de comunicação anteriormente apresentados são os mais amplamente utilizados, considerando aplicações industriais ou não. Entretanto, há várias outras tecnologias, oriundas de desenvolvimentos legados da telecomunicação tradicional, sem fim específico para aplicação em processos industriais. Nesse sentido, podem ser incluídas tecnologias diversas, como as que empregam rádio enlaces de longo alcance ou até mesmo a telefonia celular. Como exemplo, um estudo mais amplo pode ser realizado se forem incluídas outras faixas de frequências operacionais tais como a faixa dos 315, 434 ou 900 MHz (ISM). Nestas faixas, as atenuações na propagação da OEM permitem enlaces maiores, se comparados com a faixa de 2,4 GHz, mais utilizada atualmente. Para a mesma faixa dinâmica de enlace, um rádio de 900 MHz pode apresentar alcance até sete vezes maior que um de 2,4 GHz. Por outro lado, as dimensões das antenas aumentam, a banda disponível diminui e o espectro está ainda mais poluído. As frequências mais baixas, necessitam posicionamento mais elevado dos transceptores em relação ao solo, por consequência efeito das zonas de Fresnel. Mesmo assim, a faixa de 900 MHz é amplamente

utilizada na indústria, especialmente através de rádio modems que empregam frequentemente técnicas de espalhamento espectral, criptografia e saltos de canais. Por outro lado, a utilização de um transceptor de 900 MHz implica em menor banda disponível e dispositivos de consumo mais elevado, desfavorecendo aplicações alimentadas à bateria. Por conta destes fatores, a faixa de 2,4 GHz é a mais utilizada em ambientes industriais atualmente.

2.3 Escalonamento TDMA

Redes de sensores sem fio são caracterizadas por apresentarem muitos nós sensores e alguns nós coletores. As topologias variam, mas em geral, um fluxo de dados do tipo *convergecast* ocorrerá no momento em que as informações coletadas são direcionadas ao nó coletor. Em redes industriais, esta topologia de rede é frequente, uma vez que os dados de sensoriamento e controle da planta convergem para um gateway. Se o gerenciamento de rede é centralizado, há também fluxo *broadcast*, embora na maior parte do tempo, o gerenciamento exige fluxos *unicast*.

A técnica de arbitração do meio é altamente influente em diversas das características fundamentais de uma RSSF, tais como consumo de energia, determinismo e latências. Basicamente, pode-se dividir em duas técnicas: as baseadas em contenção e as TDMA. As primeiras, não proporcionam baixo consumo de energia, uma vez que as colisões que ocorrem frequentemente, acarretam na necessidade de retransmissões. Ainda, a necessidade de longos períodos escutando a rede em busca de sincronismo também impacta no consumo de energia. Um transceptor 802.15.4 consome quase tanta energia no modo RX quanto no TX, se a potência de RF é no entorno dos 0 dBm. Os protocolos TDMA por outro lado, garantem links livres de colisão (com a própria rede) e podem determinar os atrasos mínimos e máximos nas comunicações. Quando a arbitração do meio é feita por técnica TDMA, o problema do escalonamento das mensagens consiste basicamente em determinar o endereçamento dos links numa estrutura tipo *superframe* de modo que não ocorram conflitos temporais e que os requisitos de latência e determinismo sejam atendidos. A latência pode ser mínima, no caso de links contíguos ou muito próximos, e o determinismo pode ser garantido pela escolha dos links para retransmissão. O escalonamento TDMA ainda tem comportamento diferenciado no consumo de energia, que pode ser adaptativo, em função da energia remanescente que irá definir um novo espaçamento entre os links. Longos espaçamentos levam a baixos consumos, e por consequência, longa duração da bateria.

A alocação de links deve é feita em função das periodicidades necessárias e da topologia da rede. Para um esquema de gerenciamento distribuído, dois requisitos são importantes: agilizar ao máximo os processos de agregação e manutenção e garantir um escalonamento livre de conflitos hierárquicos.

O escalonamento pode ser adequado para redes de salto único, como as rede de telefonia celular, ou multi saltos, como no caso do WH. O escalonamento TDMA multi saltos é um problema mais complexo, uma vez que há mais variáveis a serem consideradas. A literatura define os chamados conflitos primários e os secundários (ERGEN; VARAIYA, 2010), (DJUKIC, 2008). O primeiro ocorre quando um nó é programado para transmitir e e receber no mesmo *slot* de tempo ou recebe mais de uma transmissão designada a ele no mesmo tempo. O segundo ocorre quando um nó recebe uma transmissão ao mesmo tempo em que outro nó recebe outra transmissão, mas ambas são detectadas concomitantemente.

2.3.1 Algoritmos para roteamento em redes TDMA

Para a modelagem de uma rede TDMA, considera-se o caso do WH, com um único ponto de acesso e nós sensores e atuadores cujos períodos de *burst* são diversos. Os links ocorrem em um *slot* de tempo de 10 ms e são bidirecionais, com o período maior do *slot* em modo TX e o menor em RX para o nó transmissor, e o inverso para o nó receptor. O período menor refere-se ao ACK que sempre ocorre numa rede WH, exceto nos links de tipo *broadcast*.

Um grafo de rede é representado como *G = (V,E)* onde *V* é o conjunto de nós totais da rede e *E*, os vértices. O ponto de acesso é definido como nó numero um. Cada mensagem é roteada para o nó vizinho em direção ao ponto de acesso, num esquema ponto a ponto e *convergecast*, na maior parte do tempo. Mensagens *broadcast* também ocorrem e portanto, grafos de *uplink* e *downlink* são usuais. Os vértices do grafo são definidos por:

$$E \subset V \times V \tag{3}$$

Como um nó pode interferir em outro, atribui-se um grafo para as interferências, *C = (V,I)*. Para este grafo, *I* é definido por:

$$I \subset V \times V, (u, v) \in I \tag{4}$$

O conjunto *(u, v)* são os nós interferentes, cujas comunicações não são intencionais. Assim, se *u* transmite, *v* recebe, mas não deveria, porque a transmissão tinha outro destino. Deste modo, o escalonador deverá evitar colocar *v* em modo recepção no mesmo tempo em que *u* é colocado no modo transmissão.

Correspondente aos grafos anteriores, o grafo de conflitos *GC* é produzido, definido por: *GC = (V, EC)*. O link é definido por:

$$(i, j) \subset E \tag{5}$$

Cada nó *i* e *j* de *GC* tem link correspondente. *EC* são os vértices do grafo *G* que não devem transmitir ao mesmo tempo. Desta forma, pelo fato de que dois nós não podem transmitir ao mesmo tempo, constata-se:

$$Se(i, j) \in E, (i, j) \in EC \tag{6}$$

Se *i* transmite e outro nó transmitir ao mesmo tempo, *j* poderá escutar ambos, caracterizando interferência.

Um *superframe* tem a coleção de todos os links que um nó da rede utilizará periodicamente. Um quadro de escalonamento tem duração que começa quando cada nó da rede gera uma quantidade de pacotes e termina quando todos os pacotes alcançam o ponto de acesso. A distância entre dois nós *u* e *v* é o número de vértices no caminho entre eles no grafo. Este número de vértice é também chamado de nível, em relação ao ponto de acesso. Cada nó *G* gera um número de pacotes no início do quadro de escalonamento. Dado o grafo de interferência *C* o problema do escalonamento é achar um tamanho mínimo de quadro durante o qual todos os nós podem enviar seus pacotes ao ponto de acesso.

Uma das técnicas utilizadas para produzir o roteamento e escalonamento TDMA é a coloração de grafos. Em (ERGEN; VARAIYA, 2010) e (DJUKIC, 2008), os autores provam que o problema do escalonamento é NP-completo. O número cromático de um grafo *G* é o menor número *k* para colorir *G*. *G* é *k*-colorável se os seus vértices podem ser coloridos utilizando *k* diferentes cores de tal modo que os vértices adjacentes tem

cores diferentes. Se *GP = (VP, EP)* com $VP = v_1, v_2, ...v_n$ é a instância do grafo cujo número cromático necessita ser encontrado, inicialmente constrói-se *GC = (V, EC)*. Primeiramente, *GC* inclui todos os nós e vértices de *GP*. Após, para cada nó v_i, outro nó w_i é adicionado e então, os vértices $(w_i, w_j), (v_i, w_j)$ pertencentes a *EC* para todo *i, j*. Por fim, adiciona-se o nó referente ao ponto de acesso e vértices (AP, w_i) para todo *i*.

Diversos tipos de algoritmos são empregados para geração de rotas em grafos. O Bellman-Ford é um dos algoritmos clássicos utilizado para descoberta do menor caminho em um grafo com arestas valoradas. Os valores ou pesos são somados ao longo da rota *P*, ou seja, é definido um custo para cada sessão entre um nó da rede e o ponto de acesso (GOODRICH; TAMASSIA, 2002) (ver equação abaixo).

$$w(P) = \sum_{k-1}^{i=0} w(e_i) \tag{7}$$

No caso do WH, os pesos podem ser equivalentes à viabilidade do enlace entre dois nós. Os parâmetros fundamentais a serem considerados são o RSL (*Received Signal Level*) e o tipo de alimentação do dispositivo. Também podem ser considerados os dados da resposta do comando 777 (*Read Wireless Device Capabilities*) além de outros, fornecidos por outros comandos. Aliado a estes parâmetros, os próprios algoritmos de escalonamento e roteamento devem levar em consideração fatores como o desbalanceamento da malha e saturação de vértices.

A distância de um vértice *v* a um vértice *u* em *G* é denotada por *d(v, u)* é o menor caminho de *v* a *u*, se existir. Se *v* não tem rota para *u*, $d(v, u) = \infty$. Também há casos em que existe uma rota mas ela não é viável por outros fatores e sendo portanto, desconsiderada por limiar.

Os grafos podem ser divididos em dígrafos com pesos, um para mensagens de fluxo *downlink* e outro para *uplink*. O conceito dos dígrafos é mais adequado ao WH, uma vez que um nó pode ser pai de outro mas o inverso pode não ocorrer, pois as rotas *uplink* e *downlink* podem ser assimétricas. O algoritmo também deverá controlar se há diferenças entre os números de saltos entre os dígrafos, para controle estrito de latência.

Para desenvolvimento do algoritmo de roteamento, define-se *D[u]*, um rótulo para sinalizar cada vértice de G do qual quer-se saber a distância relativa a outros vértices. Inicializa-se $D[v] = 0$ e $D[u] = \infty$ para cada $u \neq v$. *C* é definido como o conjunto de vértices do grafo, inicialmente com valores nulos. A cada iteração do algoritmo seleciona-se um vértice *u* não pertencente a *C* com o menor rótulo $D[u]$, e *u* é carregado em *C*. Na primeira iteração, *v* é carregado em *C*. A cada novo vértice *u* carregado em *C*, o rótulo $D[z]$ é atualizado para cada vértice *z*, adjacente a *u* e fora de *C*, para refletir o fato de que poderá haver uma nova e melhor rota para chegar a *z* através de *u*. Esta operação é chamada de relaxamento de vértices. O problema final a ser resolvido é de otimização, pois quer-se encontrar o mínimo dos pesos das arestas (rotas numa rede).

O algoritmo Bellman-Ford realiza *n-1* iterações para cada vértice no dígrafo. O pseudo código para este algoritmo é apresentado a seguir:

1: *Entrada*: um grafo *G* com *n* vértices.
2: *Saída*: um rótulo $D[u]$ para cada vértice *u* de *G* tal que $D[u]$ é a distância de *v* para *u* em *G* ou uma indicação de que *G* tem um ciclo de peso negativo.
3: $D[v]$ é inicializado com zero.
4: *Para* cada vértice $u \neq v$ de *G* faça:
5: $\quad D[u] = \infty$

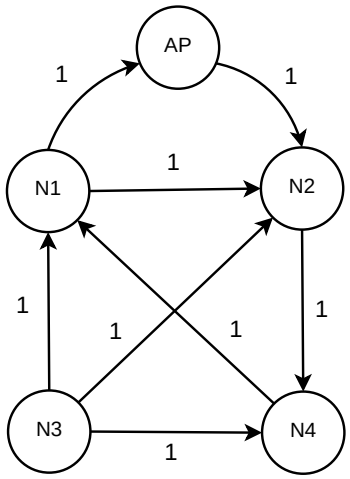

Figura 11: Grafo de exemplo de uma rede WH.

Tabela 2: Tabela com a matriz de pesos e conexões entre os nós do grafo e os respectivos custos totais das rotas.

Nó	AP	N1	N2	N3	N4	$ AP	$ N1	$ N2	$ N3	$ N4
AP	0	0	1	0	0	0	3	1	∞	2
N1	1	0	1	0	0	1	0	1	∞	2
N2	0	0	0	0	1	3	2	0	∞	1
N3	0	1	1	0	1	2	1	1	0	1
N4	0	1	1	0	0	2	1	2	∞	0

6: *Para* $i = 1$ até $n - 1$ faça:
7: *Para* cada aresta de $G(u, z)$ saindo de *u*, faça:
8: *Se* $D[u] + w(u, z) < D[z]$, então *//operação de relaxamento de vértices*
9: $D[z] = D[u] + w(u, z)$
10: *Se* não há mais vértices com possibilidade de operação de relaxamento, então:
11: Retorna o rótulo $D[u]$ para cada vértice u
12: *Senão*
13: Retorna a informação de que *G* contém um ciclo de peso negativo.

A Figura 11 apresenta um grafo de exemplo para entendimento do algoritmo de roteamento. Nele, há quatro nós numa rede que devem realizar comunicação em fluxo *uplink*. Os pesos atribuídos às arestas são unitários, para simplificar a conferência do resultado do algoritmo.

A matriz de links é definida na Tabela 2. O algoritmo Bellman-Ford é alimentado com os valores das rotas e os valores dos custos finais são obtidos e indicados na tabela pelo caracter $. É possível acompanhar na figura do grafo as trajetórias necessárias para comunicação de um dispositivo *x* com um *y*, e verificar o custo final na tabela apresentada.

2.3.2 Algoritmos para escalonamento em redes TDMA

A teoria de grafos é utilizada para gerar os algoritmos que definem as rotas da rede em malha. Os resultados dos algoritmos empregados são utilizados para alimentar os algoritmos de escalonamento que, em conjunto com outros parâmetros tais como as requisições de banda dos dispositivos de campo, formam o escalonamento da rede. Os algoritmos de

escalonamento acabam por definir os tamanhos dos *superframes*, as alocações e os tipos de links designados aos nós da rede.

Dentre as técnicas usuais para escalonamento em tempo real, citam-se as mais comumente empregadas.

O escalonador de tipo **Taxa Monotônica** (RM) baseia-se na periodicidade das mensagens. Quanto menor a periodicidade, maior a prioridade da mensagem dentro do conjunto de mensagens a serem escalonadas. A viabilidade do uso de um escalonador RM exige que:

a) As mensagens sejam sempre periódicas;

b) O *deadline* para transmissão de uma mensagem (D_i) seja igual ao seu período (P_i);

A equação a seguir foi anteriormente elaborada para verificação da garantia de escalonabilidade de um conjunto de mensagens.

$$U = \sum_{i=1}^{n} \frac{C_i}{P_i} < (\sqrt[n]{2} - 1) \tag{8}$$

Onde *U* é a taxa de utilização do escalonador e *C* o tempo de computação da mensagem *i*. *P* é a periodicidade da mensagem e *n* é o número de mensagens a serem escalonadas. O escalonamento é factível se a taxa de utilização do escalonador é menor que o valor calculado.

No WH, as mensagens tem característica periódica (variáveis de processo periódicas, relatórios de saúde de rede (*health reports*) e aperiódicas (comandos, alarmes e outras). Desta forma, o escalonador RM não é adequado para todo tipo de mensagem a ser escalonada.

No escalonador tipo ***Deadline* Monotônico** (DM), o *deadline* da mensagem é quem regra a escalonabilidade. Quanto menor o *deadline*, maior a prioridade da mensagem. Um conjunto de tarefas tem escalonamento factível se e somente se:

$$R_i^{n+1} = C_i + \sum_{\forall k \in H_i} \left\lceil \frac{R_n^i}{P_k} \right\rceil C_k \tag{9}$$

Onde R é o tempo máximo de alocação da mensagem. Para que uma mensagem seja escalonável, seu tempo de resposta máximo obtido pelo processo iterativo, deverá ser menor que seu *deadline*.

O ***Earliest Deadline First*** (EDF) é um esquema de agendamento capaz de ordenar as mensagens dinamicamente. As mensagens são agendadas baseando-se nos seus *deadlines* absolutos. O cálculo que verifica a viabilidade do escalonamento de um conjunto de mensagens é:

$$U = \sum_{i=0}^{n} \frac{C_i}{P_i} < 1 \tag{10}$$

Na prática, o EDF demanda muito recurso computacional, uma vez que toda a fila de mensagens deve ser reordenada a cada nova demanda.

O ***Least Laxity First*** (LLF) escalona as mensagens pertencentes às tarefas de menor relaxamento (*laxity*). O relaxamento é definido como o tempo restante das transmissões remanescentes, ou tolerância. A tolerância é calculada como a diferença entre o tempo entre o *deadline* e o tempo de computação restante. O agendamento de uma mensagem

que apresente relaxamento negativo, não terá seu prazo cumprido. Deste modo, o relaxamento fornece a detecção precoce de falhas temporais, de forma que ações posteriores podem ser tomadas. O escalonador LLF, assim como o EDF, é um algoritmo ótimo.

Diversos tipos de algoritmos para escalonamento foram desenvolvidos e aprimorados e novos tem sido criados. Alguns exemplos são o algoritmo de máxima urgência primeiro (MUF), *Deadline* monotônico proporcional (PD) e *Deadline* proporcional mais recente primeiro (EPD). Mais de um algoritmo pode ser utilizado para o desenvolvimento de um escalonador WH. Alguns candidatos potenciais são o DM e o LLF e já foram propostos em trabalhos científicos. De qualquer forma, várias tarefas, atribuições e testes devem ser feitos antes da utilização do algoritmo escalonador escolhido. Desta forma, o escalonador utilizado em redes WH é na verdade uma composição de algoritmos.

3 ANÁLISE DO ESTADO DA ARTE

3.1 Introdução

Neste capítulo é apresentada a revisão bibliográfica referente a trabalhos anteriormente realizados, cujo tema principal tenha sido julgado relevante para a realização desta tese. Primariamente, são analisados artigos sobre o protocolo WH, pois este foi o escolhido para o desenvolvimento deste trabalho em função das características apresentadas no Capítulo anterior. A diversidade de objetos de estudo encontrados revela a importância do protocolo WH e apresenta algumas de suas desvantagens. Dentre as limitações do protocolo, cita-se a baixa dinamicidade destas redes, que ocorre devido à característica centralizada, onde as mensagens devem trafegar em direção ao ponto central da rede e de volta à origem. Uma das possibilidades para aumento de dinamicidade é a distribuição do gerenciamento. Neste contexto, são encontrados artigos que apresentam propostas para aumento de dinamicidade, mas nenhuma semelhante com esta tese.

3.1.1 Trabalhos sobre o desenvolvimento do protocolo *Wireless*HART

Um dos primeiros trabalhos relativos à redes LR-WPAN empregando TDMA e saltos de frequência desenvolvido com interesse especial em aplicações industriais foi apresentado em 2006 por (DOHERTY; TEASDALE, 2006). Os conceitos fundamentais do protocolo já estavam sendo estudados e os trabalhos desenvolvidos resultaram no primeiro *chipset* dedicado para WH, produzido e comercializado pela empresa Dust Networks, atualmente Linear Technologies. No artigo, foi proposta uma RSSF baseada em TDMA que, segundo os autores, apresentou níveis de confiabilidade de 99,99% em regime permanente em uma rede simulada composta por 50 nós. A estratégia de gerenciamento empregada foi centralizada, provavelmente em função do propósito da aplicação (industrial) onde a maioria dos nós sensores e atuadores são fixos. Na simulação realizada, a magnitude dos erros, quantificados como perdas de pacotes é tão pequena que as técnicas de checagem tradicionais são avaliadas como tendo pouca importância para a proposta. Com os altos níveis de confiabilidade apresentados, as falhas são reduzidas à uma razão de 1/10000. Com cada falha estimada como tendo periodicidade de um dia, esta razão significa que cada uma poderá ocorrer a cada 27 anos. Na Figura 12 é apresentado o gráfico de confiabilidade obtido após 17 dias de operação de uma rede empregando o protocolo proposto. No entanto, uma avaliação mais criteriosa deve ser feita, porque o que se percebe na prática do uso de redes sem fio, é diferente. Se o que se entende como falha é a perda de um único pacote, estas são vistas muito mais frequentemente, na razão de dezenas de falhas por semana, em função de bloqueios e/ou problemas de coexistência. Por outro lado, se o protocolo atende os requisitos de tempo real, como no caso do WH, as retransmissões são computadas na dinâmica da rede, e o que se vê na prática, são as perdas de atualiza-

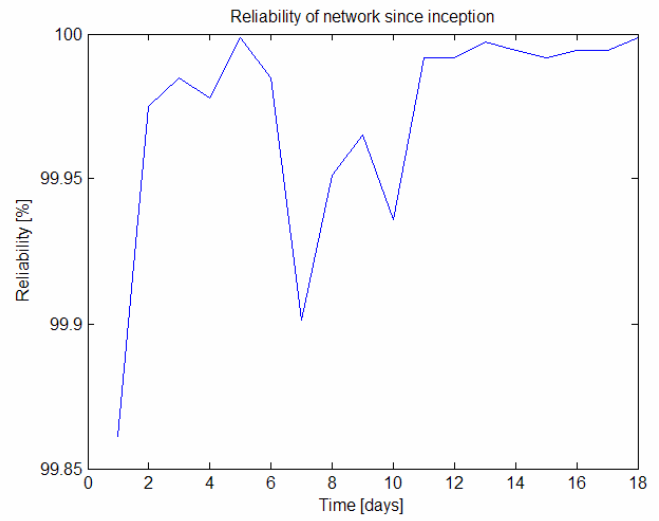

Figura 12: Resultado da confiabilidade obtida após 17 dias de uso da rede TSMP (DOHERTY; TEASDALE, 2006).

ção das variáveis de processos, onde todas as tentativas de retransmissão falharam. Neste processo influi o mau gerenciamento da rede, que na eventualidade de um escalonamento mal feito, os pacotes podem ser perdidos em função da perda de um *deadline*.

Outros trabalhos evidenciam a tendência dos estudos iniciais, onde adaptações eram propostas com o objetivo de tornar as RSSF robustas o suficiente para aplicações industriais. Os autores (ZAND; SHIVA, 2008) propuseram um novo formato para o quadro de dados do 802.15.4, na qual foram introduzidas novas funcionalidades para formar uma rede em malha sincronizada no tempo. Foi sugerido o uso do *Beacon Only Period*, BOP na primeira região do quadro para envio de *beacons* para sincronização. Propuseram também trocas dinâmicas de canais ao longo do tempo, para prover agilidade de frequências. Neste trabalho, há diversas semelhanças com algumas das técnicas empregadas no WH, mas foram mantidas algumas características originais da MAC do 802.15.4. O quadro original foi dividido em quatro regiões: primeira, a região dos *beacons*, segunda, a região CAP onde os nós utilizam o CSMA-CA para disputar o acesso, e as duas últimas regiões chamadas TDMA/FDMA1 e TDMA/FDMA2, onde os dados são gerados ciclicamente e enviados a uma entidade centralizadora. Percebe-se que o autor modificou o período de GTS da MAC do 802.15.4 para aplicar a técnica proposta (Figura 13). Os autores simularam as modificações propostas para a MAC no simulador ns-2 e concluíram que o novo quadro de dados propiciou melhorias no controle de acesso ao meio do 802.15.4, no que se refere à redução do número de colisões com o aumento do número de nós. Por outro lado, percebe-se que a utilização do CFP original do 802.15.4 não é satisfatória, tendo em vista o pequeno tamanho disponível que resulta em uma baixa densidade de rede. Neste contexto, outros autores (HUANG; PANG; HUNG, 2008), (KOUBAA; ALVES; TOVAR, 2006) realizaram adaptações semelhantes, com a utilização dos chamados *mini time slots* ou a utilização de algoritmos adaptativos com o objetivo de melhor aproveitar os GTS.

No trabalho de (NIXON; BLEVINS; MOK, 2008) são abordados aspectos que comprometem o uso de redes sem fio em aplicações industriais. Os autores, abordam a latência e o *jitter* de redes sem fio e como estes problemas são contornados através de algoritmos para redes de topologia em malha, com enlaces sincronizados no tempo.

No trabalho de (SONG et al., 2008), os autores abordaram o protocolo WH neste que

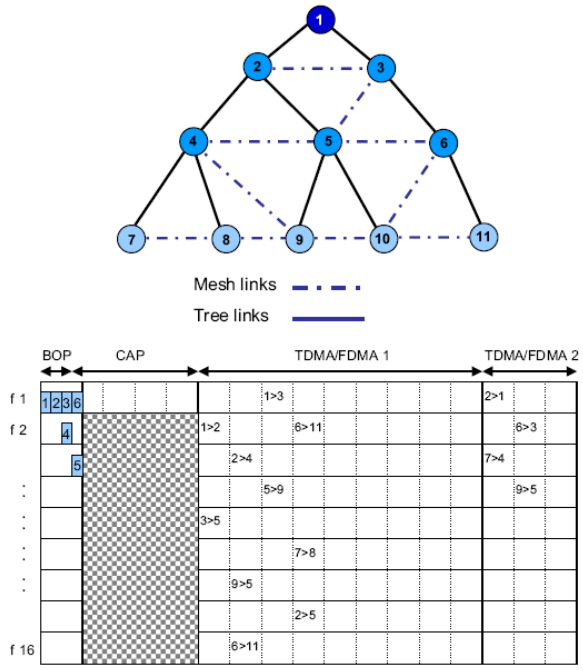

Figura 13: Novo *frame* de dados proposto para o 802.15.4 (ZAND; SHIVA, 2008).

foi um dos primeiros artigos sobre o protocolo. Foi desenvolvido um sistema composto por uma unidade centralizadora (*gateway*) e dois protótipos de dispositivos de campo. Foram utilizados sistemas de desenvolvimento da Freescale, compostos por MCUs de 32 bits (JM128) e um transceptor 802.15.4 (MC13129). O protocolo não foi completamente desenvolvido, mas sim algumas partes tais como as temporizações para geração dos *time slots*, que foram posteriormente utilizadas para a produção de dispositivos certificadores (HAN et al., 2009). Foram implementados também algoritmos de segurança, testes de roteamento e técnicas de gerenciamento de rede centralizadas. A cada etapa foram providos os detalhes e dificuldades encontrados de modo a possibilitar uma visão geral dos desafios que um desenvolvedor de dispositivos WH encontra. A análise feita pelos autores revelou que, devido às exigências temporais do protocolo, algumas tarefas deveriam ser feitas em hardware para que não fossem perdidas as janelas de comunicação do mecanismo TDMA. Os autores ainda atentaram para o detalhe do escalonamento de mensagens feito no gerenciador de rede centralizado. Segundo eles, é vantajoso implementar e testar diversas estratégias de agendamento de tarefas por meio de algoritmos que podem minimizar as latências e o *jitter* da rede. Abordaram também a necessidade de que fossem realizados estudos sobre coexistência com outros dispositivos que operam na faixa de 2,4 GHz como ZigBee e Bluetooth.

Em (ZHU et al., 2011) os autores ressaltam as características necessárias para o desenvolvimento de dispositivos WH, abordando especialmente questões como ocupação de memória, velocidade de processamento e funcionalidades do transceptor de rádio. Com relação à segurança, a velocidade de processamento necessária para a aplicação do algoritmo AES-128 é abordada. Foram utilizadas autenticações CCM (CTR e CBC-MAC), através de várias chamadas de função feitas para produzir os 4 bytes ao final de cada mensagem, como uma técnica para compensação do tempo necessário para encriptação.

Após a recepção de um pacote, um dispositivo WH deve efetuar o cálculo da autenticação dentro de um período máximo de 1 ms, existente entre o tempo de chaveamento TX/RX e o ACK. Dentro deste período, deve-se:

a) Autenticar a mensagem recebida, que pode ser de até 127 bytes;
b) Processar todas as tarefas na camada MAC relativas à mensagem recebida;
c) Gerar o pacote ACK;
d) Gerar a autenticação para o pacote ACK a ser transmitido.

Posto isto, chega-se à conclusão de que a implementação de um algoritmo AES-128 em um dispositivo WH implica, ou na utilização de um processador com elevado MIPS ou em um módulo dedicado de hardware. A implementação em hardware é em geral mais vantajosa, tendo em vista a redução de consumo propiciada e velocidade elevadas na execução de tarefas específica.

Com relação à capacidade de memória, de acordo com as especificações HCF SPEC-75 e 85, o tamanho mínimo de memória para a DLL é de 6112 bytes para um dispositivo de campo e 116356 bytes para um gateway. Isto se deve ao número mínimo de vizinhos, *superframes*, links, grafos e buffers de pacotes para cada dispositivo. Para a camada de rede devem ser previstos 1106 e 36508 bytes para os dispositivos e gateway, respectivamente. Os autores atentam ainda para o tipo de memória utilizada pelo MCU, já que os dados acima se referem às tabelas somente. Com o acréscimo da pilha de comunicação e aplicação, estes números aumentam significativamente. Ainda, MCUs baseados em ARM7 não executam firmware a partir da memória Flash, mas sim da RAM, que recebe um espelho do conteúdo da Flash. Isto faz com que sobre menos espaço para dados na RAM. Atenta-se para este fato, uma vez que a grande maioria das implementações da pilha WH atualmente é baseada em um MCU ARM7 (MC1322X, utilizado pelas empresas Nivis, Awiatech e neste trabalho).

A questão do gerenciamento do sinal de relógio, que deve ser preciso em função do mecanismo TDMA empregado também é abordado neste trabalho. Pela norma 802.15.4, uma incerteza de +/- 40 ppm de frequência do MCU é tolerável, mas para o WH, são necessários 10 ppm no mínimo. Desta forma, é imperativo empregar algoritmos para correção de tempos bem como fontes de relógio mais precisas. Usualmente, TCXOs de 3 ppm são utilizados, tanto com o objetivo de manter a precisão do mecanismo TDMA e de criptografia quanto com a necessidade de manter sincronismo em longos períodos de inatividade do rádio. Como exemplo, a janela TsRxWait do quadro TDMA (ver Figura 7) dura 2,2 ms, tempo limite para um dispositivo reconhecer que recebeu uma mensagem. Para um RTC de 10 ppm de precisão, um dispositivo escravo deve ser atualizado pelo mestre a cada 50 segundos aproximadamente. Se existem saltos no caminho de comunicação, deve-se atualizar a cada 50/n, onde n é o número de saltos. Usualmente, é utilizado um relógio de alta frequência e menor precisão e outro de mais baixa frequência e alta precisão para realinhamento (arquitetura *two tier*). Os inícios das janelas de tempo do TDMA são ajustados periodicamente, através do realinhamento em função das informações de outros rádios que são fontes de relógio para a rede. Para os longos períodos de inatividade, que garantem o baixo consumo de um dispositivo WH, o relógio externo é programado para gerar uma interrupção que reative o MCU.

No trabalho, são utilizados quatro chips diferentes, onde são verificados os parâmetros memória, potência e velocidade de processamento. Os autores foram pioneiros no desenvolvimento do protocolo, que culminou com um dispositivo posteriormente incorporado à HCF, como certificador WH. O artigo é finalizado com uma lista de desejos de requi-

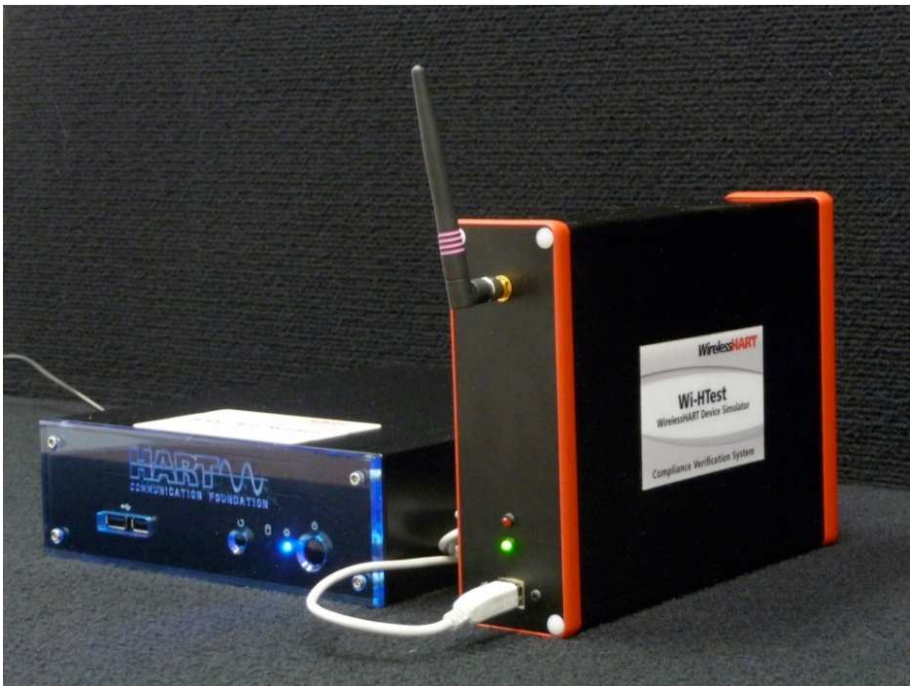

Figura 14: O Wi-HTest tool (HAN et al., 2009).

sitos de hardware para o protocolo WH, que inclui o tamanho mínimo de memória, os requisitos de precisão de relógio, o dispositivo de criptografia AES-128 e até mesmo uma sugestão de implementação da pilha de protocolo WH (ou de parte dela). Esta observação vem do exemplo do MC1322X, que implementa a MAC IEEE 802.15.4 em ROM.

No trabalho (FIORE et al., 2009), o co-simulador TrueTime (*plug-in* para o Matlab) foi utilizado para simular uma rede WH, porém modificada. Os autores propuseram um sistema com dois pontos de acesso para explorar a capacidade de saltos de canal no WH de outra forma. Através deste sistema, o protocolo pode fazer uso de mais de um canal simultaneamente. O artigo descreve ainda um algoritmo de escalonamento que gerencia a rede de saltos e canais múltiplos baseada no WH. Os resultados obtidos com as simulações, demonstram aumento de robustez da rede em função da redundância de rotas para os pontos de acesso, o que é comprovado em (HAN et al., 2011). Outros autores também utilizaram o TrueTime para simular redes WH, objetivando apresentar um simulador de uso genérico para o protocolo (SHAH; SECELEANU; GIDLUND, 2010).

Outra importante contribuição para o desenvolvimento do WH foi feita por (HAN et al., 2009). Neste artigo, uma ferramenta para teste de compatibilidade de dispositivos, chamada Wi-HTest *suite* é apresentada (Figura 14). Também é descrita outra ferramenta, o Wi-Analys que é utilizada em conjunto com a primeira. O equipamento é capaz de adquirir pacotes de dados da rede, analisá-los e gerar relatórios de compatibilidade com o protocolo. É feita uma demonstração do processo de agregação à rede de um dispositivo em desenvolvimento. A verificação diz respeito tanto à funcionalidade do dispositivo sob teste quanto a sua acuidade temporal, requisito importante no protocolo.

No artigo, é feita uma introdução à especificação do protocolo e dos métodos de testes. Os autores apresentam uma tradução dos casos de testes definidos na especificação para os scripts, empregados na ferramenta de análise. A arquitetura do Wi-HTest tool é apresentada com a descrição dos detalhes críticos, como a geração de pacotes de teste com controle de acuidade temporal e manipulação de pacotes para injeção de faltas, con-

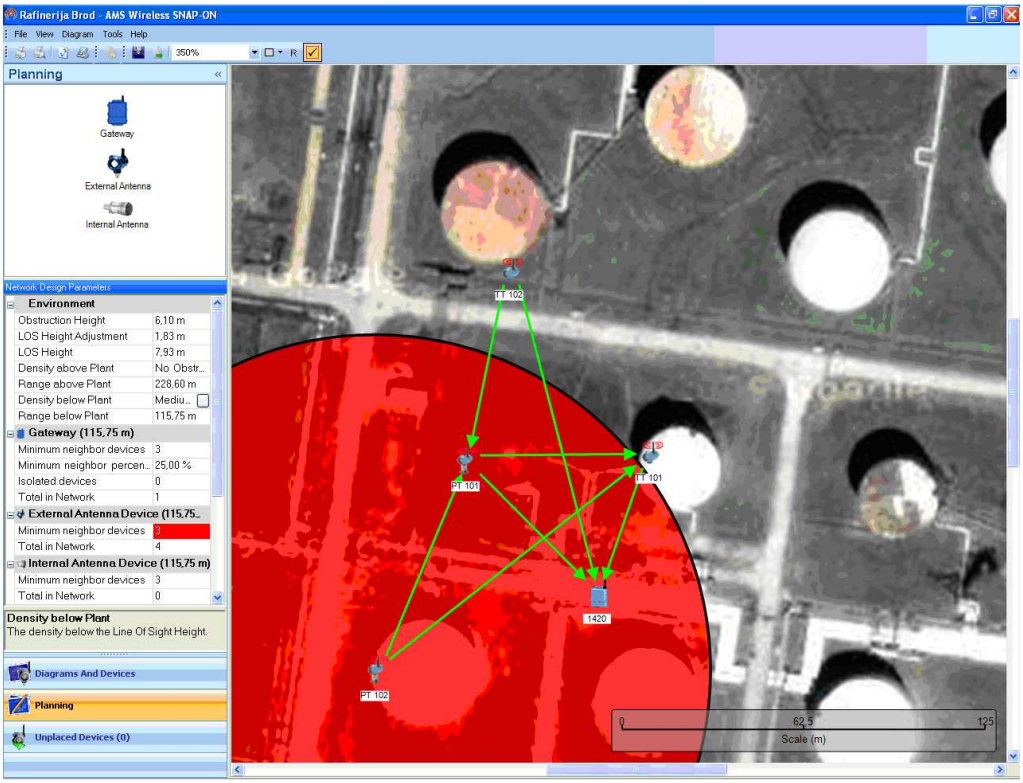

Figura 15: Mapa do local de instalação da rede, introduzido no software de gerenciamento AMS (KOSTADINOVI; BUNDALO; BUNDALO, 2010).

forme descrito na norma. São apresentadas também a captura e análise em tempo real dos pacotes de dados de uma rede. Este trabalho resultou no sistema oficial da organização HART para certificação de dispositivos WH.

No trabalho de (KOSTADINOVI; BUNDALO; BUNDALO, 2010) são explorados o planejamento e gerenciamento da distribuição dos dispositivos de campo e ponto de acesso de uma rede WH composta por dispositivos da marca Emerson Rosemount. Foi feito uso do AMS *Wireless Configurator* com o plug-in SNAP-ON, que permitem o planejamento e gerenciamento de uma rede no ambiente industrial. A aplicação permite o uso de mapas dos locais de instalação de modo que uma rápida visualização dos obstáculos e suas interações com os rádios podem ser observadas, como pode ser visto na Figura 15. São inseridas as localizações dos nós da rede no plano de coordenadas bem como suas alturas em relação ao solo. Vários parâmetros são adicionados no software, tais como tipo de antena, características do ambiente e número mínimo de vizinhos, para que este possa calcular a viabilidade de comunicação na disposição física pretendida.

3.1.2 Comparações, usabilidade e integração com o *Wireless*HART

No trabalho de (LENNVALL; SVENSSON; HEKLAND, 2008) são exploradas as dificuldades do protocolo Zigbee para uso em aplicações industriais, o que motivou o desenvolvimento de outros protocolos LR-WPAN para tais aplicações. O WH é brevemente apresentado e comparado ao ZigBee. Várias deficiências do Zigbee são apresentadas assim como as vantagens do WH. Os autores resumem uma comparação geral entre o Zigbee e o WH na forma de uma tabela que evidencia as diferenças entre os protocolos.

Em outro trabalho de comparação (HAYASHI; HASEGAWA; DEMACHI, 2009), são introduzidos os protocolos ZigBee, WH e ISA 100.11a. Além das comparações, também

Tabela 3: Tabela comparativa entre o WH e o ISA 100.11a (LENNVALL; SVENSSON; HEKLAND, 2008)

Item	WH	ISA 100.11a
Norma contém:	Extensão WH da HART	Especificação de sistema industriais sem fio, família ISA100, novas tecnologias (RFID, automação de fábrica, Alimentação, *Trustworthy wireless*, etc.
PHY e MAC	IEEE 802.15.4, 2.4GHz DSSS Channel hopping, TDMA, Channel Blacklisting	IEEE 802.15.4, 2.4GHz DSSS, Channel hopping, TDMA, CSMA, Channel Blacklisting
Camada de rede	Extended HART address, rede em malha	IPv6 (6LowPAN), rede em malha
Camada superior	HART protocol, Comando-resposta, modo Burst	ISA100.11a Publish-Subscribe, Client-Server, Bulk, Alert, Object Mapping, tunelamento (HART, FOUNDATION Fieldbus, Profibus, Modbus, etc.)
Segurança	AES-128, chave simétrica pública, Join Key, Network ID, segurança fim a fim	AES-128, chave simétrica pública, Join Key, Network ID, segurança fim a fim

são comentados alguns esforços internacionais para padronização de RF na indústria. As técnicas empregadas nos protocolos industriais, tais como saltos de canais, topologias de rede em malha e listagem de canais proibidos também são explanadas. A Tabela 3 compara os principais aspectos dos protocolos WH e ISA100.11a.

A integração do WH com outras redes industriais também tem sido estudada. (AKERBERG et al., 2010) descrevem um método para integrar redes WH em sistemas DCS (*Distributed Control Systems*) por meio do Profinet IO. É feito a modelagem de uma rede WH num arquivo de descrição tipo *Generic Station Description* que descreve um dispositivo Profinet IO. Uma vez feito isto, as configurações relativas ao WH podem ser distribuídas entre as estações da planta. Desta forma, a configuração do controle de processo e a configuração da rede WH é gerenciada e mantida por uma estação central. Além disto, os autores alegam que o usuário não necessita aprender a utilizar nenhuma ferramenta específica adicional. A proposta é testada no que os autores chamaram de prova de conceito, onde a configuração dos dispositivos WH é realizada.

Uma comparação recente e mais abrangente entre os principais protocolos de redes sem fio para aplicações industriais foi feita por (ZAND et al., 2012). No trabalho, são estudados os padrões WISA, WH, Zigbee PRO, ISA 100.11a, WIA-PA e os protocolos TSMP e 802.15.4-*e*. Uma grande tabela é apresentada comparando diversos parâmetros tais como as técnicas utilizadas nas camadas das pilhas de comunicação, topologias possíveis, sistema de gerenciamento e escalabilidade da rede. As conclusões sobre as comparações revelam vários aspectos dos padrões, listados a seguir.

Aspectos positivos:

TSMP: resolve problemas de sincronização de forma eficiente do ponto de vista do

consumo de energia. Transmissão de ACKs para troca de informação de *offset* de tempo. Roteamento redundante;

802.15.4-*e*: baixo consumo, confiabilidade, robustez, melhora e adiciona funcionalidades ao 802.15.4-2006, MAC adequada para equipamentos industriais;

WISA: altas taxas de dados, mínimos atrasos, robustez;

Zigbee PRO: escalabilidade;

WH: Baixo consumo, confiabilidade, robustez, roteamento redundante, coexistência;

WIA-PA: confiabilidade, utiliza saltos de canais adaptativos nas comunicações infra-cluster, agrega dados em pacotes em dois níveis diferentes da rede;

ISA 100.11a: escalabilidade, confiabilidade, robustez, coexistência, integração com Internet, baixo consumo, níveis de potência de transmissão ajustáveis, roteamento redundante, interoperabilidade.

Aspectos negativos:

TSMP: dependente de um gerenciador centralizado. Para o caso de descentralização, não há um mecanismo que permita nós da rede agregarem-se uns aos outros;

802.15.4-*e*: em princípio, mesmo aspecto negativo do TSMP;

WISA: interoperabilidade, diversidade espacial;

Zigbee PRO: confiabilidade, roteamento, robustez, transmissão de balizas;

WH: escalabilidade, troca de canais "cega", gerenciamento centralizado, indefinições na norma;

WIA-PA: sincronização por balizas, topologias híbridas, indefinições na norma;

ISA 100.11a: troca de canais "cega", carência de definições na norma, falta de modulação de potência de transmissão.

3.1.3 Propostas de melhorias no *Wireless*HART

No trabalho de (KIM et al., 2008), o protocolo WH é novamente introduzido. Alguns aspectos ainda não explorados são apontados pelos autores, tais como o uso de nós móveis, características dos canais de RF que variam no tempo, o *handover*, que deve ocorrer quando um dispositivo móvel tem a possibilidade de conectar com outro ponto de acesso, e estudos sobre as constantes trocas de topologia da rede. Os autores ainda citam a possibilidade do uso do protocolo em fábricas autônomas, através do uso de robôs. Outras possibilidades, tais como o emprego de sistemas MIMO (*multiple input-multiple output*) na camada física também são citadas.

Um sistema de localização utilizando o protocolo WH foi proposto por (ZHU et al., 2009). Foi proposto um sistema determinístico baseado em software que analisa a intensidade do sinal recebido e o aplica a um algoritmo de trilateração. Segundo os autores, uma maior precisão do sistema proposto pode ser obtida pelo sensoriamento de mais de uma rota e pelo treinamento do sistema antes da utilização. Os erros médios previstos e obtidos estão na faixa de quatro metros, o que pode ser considerado adequado para a maioria das aplicações industriais. A possibilidade de uso de dispositivos móveis, que coletam

informações da localização e trajetória de um funcionário também é aventada. O melhor par de coordenadas é encontrado pelo uso dos mínimos quadrados com simulações implementadas no Matlab. Os autores apontam ainda, que há muito a ser explorado, como a escolha de um modelo de propagação mais adequado, o uso de filtros para a intensidade do sinal recebido.

3.1.4 Trabalhos sobre descentralização

Os estudos previamente apresentados versam sobre generalidades acerca do protocolo WH, onde diferentes possibilidades de melhorias são apresentadas além de enfatizar sua aplicabilidade. Esta subseção trata da possibilidade de descentralização de redes WH, com vistas a aumentar sua dinamicidade. Os poucos trabalhos apresentados até o presente momento enfatizam a possibilidade de incremento na dinamicidade de redes WH pela descentralização do gerenciamento.

Uma importante constatação sobre áreas de pesquisa em aberto relativas a redes sem fio industriais é apresentada em (ZAND et al., 2012). Em concordância com esta tese, o autor alerta para a necessidade de que sejam realizados estudos sobre a utilização de sistemas distribuídos para melhorar as demandas de tempo real. No artigo, é apresentado o seguinte parágrafo acerca do gerenciamento de rede distribuído:

"WH, WIA-PA (no nível da malha) e ISA100.11a utilizam técnicas de gerenciamento de rede centralizado para escalonamento e roteamento das comunicações. Esta abordagem pode ser mais simples em termos de implementação mas leva uma série de desvantagens. Sistemas centralizados apresentam tempos de reação pobres, pois todas atualizações devem ser primeiramente enviadas ao sistema de gerenciamento que reside em um único ponto da rede. O gerenciador realiza os cálculos e dissemina as informações atualizadas para os nós relevantes da rede. Como os tempos de ida e volta para a tomada de ações de decisão são elevados (especialmente quando a contenção de rede é alta), as abordagens centralizadas não podem lidar com situações de alta dinamicidade (muitos dados em rajada, variações de qualidade de link e mobilidade de nós). Este problema aumenta na medida em que a densidade da rede aumenta, levando em casos extremos, a perda de pacotes e atrasos fora das especificações de tempo real que acabam por aumentar o consumo de energia. Por outro lado, as abordagens descentralizadas permitem que o sistema adapte-se rapidamente às perturbações na rede. Entretanto, as tecnologias atuais de controle de redes sem fio que utilizam abordagens distribuídas apresentam baixo desempenho em termos de confiabilidade, eficiência e robustez."

O parágrafo anterior sintetiza constatações sobre as redes WH. A justificativa para utilização da abordagem descentralizada proposta pelo autor é amplamente verificada na prática sendo portanto, o grande motivador desta tese. Atenta-se ainda para o fato de que as redes sem fio descentralizadas (RSSF, na sua maioria) não apresentam os mecanismos que as tornam seguras e robustas, como o WH.

A norma HART na sua versão 7.3, também apresenta uma apreciação sobre a centralização do gerenciamento de rede (HART COMMUNICATION FOUNDATION, 2010). Segundo esta, é possível implementar uma sessão que conecte dois dispositivos arbitrários em uma rede WH, independentemente das ações posteriores do gerenciador. Entretanto, segundo a mesma, isto pode resultar em comunicações não detectáveis e não monitoráveis entre dispositivos, o que consequentemente resulta em uma ameaça à segurança. Também é citado no texto da norma que se o gerenciador de rede em questão suportar sessões ponto a ponto entre os dispositivos de campo, as comunicações entre eles devem ser encaminhadas ao gateway e, que desta forma, pode-se detectar através do gerenciador

algum comportamento malicioso na rede. Neste caso, percebe-se a centralização da rede com o intuito de mantê-la segura, pois com a permissão de comunicações não monitoradas entre dispositivos da rede, abrem-se brechas para comportamentos não autorizados. Entretanto, verifica-se na prática que a rede pode ser prejudicada de formas muito mais simples, e que desta forma, os esforços para descentralização não são de fato um impedimento. Argumenta-se ainda, que se um dispositivo autônomo é intrinsecamente seguro pelo fato de ser certificado, ele implementa os mecanismos de segurança da rede. Desta forma ele não irá propiciar um evento de comunicação passível de falha de segurança de forma autônoma.

No trabalho (TINKA et al., 2010), os autores evidenciam as vantagens dos protocolos TDMA aplicados em RSSF, especialmente por proporcionar determinismo e evitar colisões. Uma vez que os sistemas TDMA são na maioria das vezes centralizados, os autores atentam para o fato de que isto leva à grande latência na formação da rede, o que dificulta a dinamicidade da rede. Desta forma, cientes de que a centralização do escalonamento de mensagens é um complicador para a dinamicidade, os autores propõem um algoritmo TDMA distribuído, para redes com saltos de canais. São propostos dois algoritmos, um baseado em Aloha e outro baseado em reserva, e ambos são comparados ao final. Os algoritmos propostos são testados em uma RSSF constituída por nós que flutuam na água, o que propicia um amplo cenário de testes no que se refere à mobilidade de nós. O algoritmo baseado em Aloha faz uso de um canal fixo para anúncios do tipo *broadcast* para os vizinhos (canal 11 no 802.15.4). Os pacotes de anúncio são alocados em uma área inicial do *superframe*. Uma vez detectados por um novo vizinho, os *slots* de anúncio são convertidos para *slots* de conexão e se a transação é efetivada com sucesso, tornam-se links de comunicação normal, caso contrário, voltam a ser *slots* do tipo Aloha. Este mecanismo permite que a entidade centralizadora seja dispensada, uma vez que os *slots* Aloha podem ser atribuídos localmente de forma aleatória pelo nós da rede. Diferentemente do que é feito em uma rede WH, onde cada tipo de *slot* é criado pelo gerenciador central e programado em cada nó da rede, neste algoritmo os *slots* de anúncio são reusados e acabam por criar o escalonamento da rede. O algoritmo baseado em reserva é um melhoramento do Aloha, proposto pelos autores. Neste, o anúncio é feito em qualquer canal, baseado no mecanismo de diversidade de frequências através do uso do ASN (*Absolute Slot Number*, número de 32 bits referente ao *slot* corrente). O pacote de anúncio está presente no *slot* zero, somente. No anúncio, os nós incluem informações sobre os vizinhos com os quais estão conectados, incluindo os canais utilizados pelos mesmos no *slot* zero. O anúncio feito por um nó inclui a informação de *slots* ímpares somente, o que propicia a criação de um link completo, de transmissão e recepção (no *slot* par adjacente). Os algoritmos são simulados em uma rede composta por 25 nós e superframe de 17 *slots* (um número a mais que os 16 canais do 802.15.4, para promover uso de todos os canais no escalonamento). Modelos de propagação de RF, de interferência em mesmo canal e de mobilidade de nós são programados em um ambiente de simulação próprio, baseado em Python. O teste físico é feito com nós de uma RSSF flutuante em água. Os resultados obtidos revelaram que o algoritmo baseado em reserva é superior em desempenho quando comparado ao Aloha, mantendo links por mais tempo, requisito necessário para o uso em redes composta por nós móveis. Não foram feitas análises referentes à latência e tempo de formação da rede.

A escolha das rotas mais adequadas na técnica de roteamento empregada num protocolo como o WH é um requisito fundamental, pois através do algoritmo empregado obtém-se a menor latência e menor consumo de energia dos nós da rede. O algoritmo Bellman-Ford é adequado e pode ser facilmente implementado, sendo portanto um ponto

de partida. Porém, outros mais eficientes já foram desenvolvidos, visando atender requisitos como redundâncias e latências mínimas.

Em (SAIFULLAH et al., 2010), os autores formulam o problema do escalonamento de tempo real para comunicações fim a fim baseado nas características do WH e provam que o problema é NP-completo. Um algoritmo livre de conflitos é proposto e simulado com topologias aleatórias. Segundo os autores, o algoritmo proposto é eficiente para redes WH, ultrapassando os comumente utilizados.

Em (ZHANG; SOLDATI; JOHANSSON, 2009), o processo de agregação em fluxo *convergecast* com saltos de canais, como nas redes WH é explorado. São apresentadas as políticas de escalonamento de tempo e canal ótimas levando em consideração capacidades de bufferização de pacotes. É estabelecido um limite inferior para o número de canais ótimo sob *convergecast* para diferentes capacidades de buffer. Segundo os autores, é demonstrado que, dado qualquer número fixo de canais, os algoritmos são capazes de gerar programações eficientes de modo a minimizar o tempo *convergecast*, permitindo explorar as vantagens e desvantagens entre o número de intervalos de tempo e os canais necessários para completar as comunicações *convergecast*. São feitas simulação e experiências e os resultados confirmam que os sistemas propostos podem fornecer tempos de *convergecast* muito rápidos em redes WH.

Em (HAN et al., 2011), os autores apresentam vários algoritmos eficientes para construir grafos e técnicas de recuperação para casos de falhas de roteamento. Os agendamentos de comunicações na camada DLL são posteriormente gerados baseadas nos grafos para propiciar comunicação fim a fim em tempo real. Os algoritmos desenvolvidos são testados no sistema WH completo desenvolvido pelos autores, integrado ao gerenciador de rede. Os resultados demonstram que os algoritmos propiciam latência inferior e comunicação em tempo real e estável com custo e *overhead* modesto.

3.2 Comentários Finais

Pela análise dos estudos prévios sobre redes sem fio industriais, percebe-se que há muito espaço para melhorias, especialmente no que se refere à dinamicidade de rede. Dos poucos trabalhos realizados neste sentido, nenhum aborda o uso de soluções de gerenciamento distribuído, que encontram aplicabilidade na descentralização de redes mantendo-se a integridade da pilha de comunicações, como será exposto no próximo Capítulo.

4 PROPOSTA DE TESE

4.1 Introdução

Neste capítulo são apresentadas a motivação e a proposta fundamental desta tese. Trata-se do incremento da dinamicidade de redes WH pelo uso de coprocessadores para descentralização do gerenciamento. Justifica-se a proposta em função dos benefícios que apresenta em comparação com soluções existentes, baseadas em gerenciamento centralizado. A solução proposta permite diversas configurações desde a centralização parcial até a descentralização total do gerenciamento de uma rede WH, o que possibilita explorar os benefícios de ambas as arquiteturas. Nenhuma proposta semelhante foi encontrada na literatura até o momento da escrita desta tese. São discutidas as possibilidades de melhoramentos no WH em função de suas desvantagens no que se refere à dinamicidade de rede. A descentralização parcial é proposta e os benefícios possibilitados são inicialmente discutidos e, posteriormente, são apresentadas as técnicas empregadas que levaram à escolha pela implementação.

4.2 Motivação

Pela análise apresentada no Capítulo 2, é possível afirmar que o protocolo WH foi de fato elaborado para lidar com os problemas relativos ao uso de RF na indústria e desta forma, até o presente momento tem sido considerado o único protocolo capaz de atender plenamente aos requisitos de aplicações industriais. Entretanto, o estado da arte revela várias possibilidades de melhorias neste protocolo. Dentre as diversas possibilidades, citam-se: emprego de sistemas MIMO, camadas físicas adaptativas, sistemas de localização por RF, uso de técnicas múltiplas de modulação, rádio cognitivo, uso de escalonadores para sistemas distribuídos, distribuição do gerenciamento de rede, roteamento distribuído, outras técnicas de diversidade de frequências e sistemas para a compensação de latências.

Todas as possibilidades de melhorias citadas são factíveis e diversas delas já fazem parte das linhas de pesquisas de diferentes grupos. Entretanto, as características do WH que chamam mais a atenção são aquelas relacionadas com a dinamicidade de rede, com a eficiência do escalonamento e roteamento de mensagens. Pelas características fundamentais do protocolo, especialmente as referentes à centralização e segurança, verifica-se que a dinamicidade da rede no processo de formação e manutenção é extremamente baixa. Esta característica não apresenta grande inconveniente em redes estáticas, como as utilizadas para o monitoramento e controle de processos, mas torna impossível o emprego de dispositivos móveis ou em outras aplicações industriais, tais como controle de manufatura e processos fabris.

Os algoritmos empregados para gerar o escalonamento e roteamento das mensagens

também merecem especial atenção no processo de melhoria do protocolo. O escalonamento mais eficiente das mensagens pode reduzir latências. O protocolo garante determinismo pela forma como foi concebido, mas este fator é extremamente relacionado com as técnicas de escalonamento e roteamento implementadas no gerenciador de rede. Como a norma não especifica estes algoritmos, os fabricantes implementam os seus próprios, o que acarreta em grandes diferenças de desempenho em redes compostas por gerenciadores de diferentes marcas.

O gerenciamento descentralizado é um ponto importante na tentativa de melhoramento da dinamicidade e confiabilidade de redes WH. Pelo fato de ser centralizado, o WH apresenta um ponto de fraqueza na confiabilidade, dependendo de sistemas redundantes de complexidade elevada a fim de manter um gerenciador em estado de espera. Ainda, como todo o fluxo de mensagens deve sempre passar pelo gateway, o processo de formação e manutenção da rede é lento, uma vez que o gerenciador é centralizado e comunica-se com a rede sem fio através do gateway.

Pelos motivos aqui apresentados, propõe-se a descentralização do gerenciamento de redes WH (extensível à outras do mesmo tipo, como o ISA 100.11a) através do uso de gerenciadores de rede distribuídos que operam conjuntamente com o principal, conectado ao gateway. A fim de manter as características fundamentais do protocolo, bem como propiciar um sistema de baixo consumo, a implementação dos gerenciadores pode ser feita em hardware, através de coprocessadores que comunicam com os dispositivos de campo especiais sob demanda. Nas subseções a seguir, são apresentados mais detalhes das justificativas que levaram à esta proposta de tese.

4.2.1 Dinamicidade de redes *Wireless*HART

A dinamicidade do processo de formação e manutenção de redes WH tem sido observada e medida há um longo tempo, em função de diversos testes realizados com gerenciadores de diferentes marcas. Para entendimento da dinamicidade de uma rede WH, o processo de agregação (*join*) de um novo dispositivo de campo à rede é apresentado a seguir. Os passos deste processo são padronizados pela norma e, desta forma, podem servir como métrica para comparações entre equipamentos de marcas diferentes. O processo de manutenção da rede também é centralizado e, portanto, sua dinamicidade pode ser compreendida extrapolando-se o processo de agregação, uma vez que também é baseado no envio de comandos e suas respostas.

O processo de agregação inicia logo após as etapas iniciais de formação de rede, que ocorrem no âmbito dos elementos centrais, ou seja, o gateway, o gerenciador de rede e o ponto de acesso. A partir do momento em que estes três elementos estão operacionais, o ponto de acesso inicia a propagação dos pacotes de anúncio de rede, conforme pode ser visto na Figura 16. Na imagem à esquerda da Figura, os pacotes de anúncio são emitidos pela unidade centralizadora, na ocasião da agregação do primeiro dispositivo de campo. Na imagem à direita da Figura, os pacotes são emitidos por todos elementos de uma rede WH já formada, que tenham sido autorizados a fazê-lo pelo gerenciador. Posteriormente, uma rede já plenamente formada poderá não mais propagar anúncios, impedindo a entrada de novos dispositivos à mesma. Isto leva à redução do consumo de energia dos dispositivos de campo, ao aumento de banda disponível e ainda dificulta a espionagem da rede.

O processo de agregação pode ser dividido em quatro etapas: anúncio, requisição, primeira resposta e segunda resposta. Ao término da segunda resposta, um dispositivo já faz parte da rede WH e, desta forma, tem suas variáveis de processo disponíveis. A

Figura 16: Formação de uma rede WH: à esquerda, a propagação dos anúncios a partir do ponto de acesso, e à direita, a partir dos elementos de uma rede já formada

.

seguir, a explanação de cada etapa.

Anúncio: no pacote de anúncio são enviadas informações para que um dispositivo que deseja ingressar na rede possa fazê-lo. A mensagem é criptografada com a chave de agregação (*join key*) de dezesseis bytes. Esta chave é confidencial e definida no gerenciador pelo usuário, assim como no dispositivo de campo, através da porta de manutenção. A chave, em conjunto com o identificador de rede de dois bytes, são os dois parâmetros necessários para que um dispositivo possa ser agregado à rede. No pacote de anúncio são enviadas a identificação de *superframe* e links que o dispositivo poderá utilizar para enviar a requisição de agregação. Também são enviados o mapa de canais, previamente configurados como válidos no gerenciador de rede e o identificador de prioridade de agregação (para o dispositivo anunciante). O dispositivo de campo calculará os canais a serem utilizados em função do *offset* de canal, indicado nos links recebidos.

Requisição: a resposta ao anúncio de agregação é a requisição do dispositivo de campo. Nesta, são enviadas as respostas de três comandos gerados internamente no dispositivo de campo: 0, 20 e 787. O comando 0 retorna o identificador único de oito bytes (UID - *unique* ID, com três bytes de identificação de organização, dois bytes de tipo e três de identificação única), além de diversos outros identificadores do dispositivo, tais como número de preâmbulos da mensagem HART e revisões de software e hardware. O comando 20, retorna a etiqueta (*long tag*), ou denominação do aparelho. O comando 787 indica as intensidades de sinal de RF e o apelido de dispositivos vizinhos. Com base nesses dados, o gerenciador escolherá por meio de qual vizinho as mensagens de agregação serão trocadas com o novo dispositivo de campo.

Primeira resposta: na primeira resposta, o gerenciador envia três comandos de formação de conexão com o novos dispositivo. Os comandos são enviados diretamente ao novo dispositivo, em caso de comunicação direta, ou através de proxy, em comunicação indireta. Os comandos enviados são o 961, 962 e 963. A autenticação das mensagens ponto a ponto é feita utilizando uma chave pública conhecida e fim a fim utilizando a chave de agregação. O comando 961 define uma nova sessão *unicast* entre o dispositivo e o gerenciador de rede, que será utilizada enquanto a rede existir, a fim de permitir a comunicação encriptada entre estes dois elementos da rede. O comando 962 define um apelido para o novo dispositivo, definido pelo gerenciador. O apelido de dois bytes, será utilizado futuramente no lugar do UID, na ocasião do envio e respostas de novos comandos. O comando 963 define uma nova chave de rede em substituição da chave bem conhecida. A nova chave é gerada pelo gerenciador e será utilizada para geração do código de integridade de mensagem na camada de dados, para verificação da autenticidade das mensagens trocadas ponto a ponto. Desta forma, eventuais ataques de repetição são detectados e evitados. Esta chave substitui a chave pública conhecida, previamente utilizada, tornando a

rede mais segura.

Segunda resposta: na segunda resposta, o novo dispositivo de campo recebe do gerenciador os comandos de escalonamento de rede que definem as rotas, os *superframes* e o links a serem utilizados. A partir deste ponto, o novo dispositivo já faz parte da rede e pode ter suas variáveis de processos lidas (ou escritas) pela aplicação através do gateway. Porém, na prática, diversos outros comandos são enviados antes da requisição das variáveis de processo. Quais outros comandos são enviados e em que ordem, têm dependência com a implementação do gerenciador de rede, que não é definida na norma.

O processo de agregação foi definido de forma a garantir o correto funcionamento do protocolo, tanto em termos de determinismo e diversidade, providas pelo mecanismo TDMA, quanto pela segurança, na forma de encriptação das mensagens. Porém, o processo é naturalmente lento, uma vez que os diversos comandos devem trafegar pela rede na direção direta e reversa da localização do ponto de acesso, conectado ao gateway. No melhor caso, o novo dispositivo tem comunicação direta com o gerenciador e no pior, através de diversos saltos de comunicações entre os outros nós da rede. Sendo assim, para garantir os requisitos de tempo real e menor latência possível, o gerenciador centralizado provavelmente empregará algoritmos de escalonamento e roteamento de mensagens que gerarão topologias centralizadas, no caso extremo, na forma de estrela, onde o elemento central é o ponto de acesso. Para se ter uma ideia mais realista, os seguintes comandos foram capturados de uma rede formada entre dois dispositivos somente (as entidades centralizadoras e um dispositivo de campo):

CMD 963: *write session*;
CMD 961: *write network key*;
CMD 962: *write device nickname*;
CMD 965: *write super frame*;
CMD 967: *write link*;
CMD 971: *write neighbor flags*;
CMD 777: *read device capabilities*;
CMD 64512: *read wireless transceiver module revision*;
CMD 963: *write session*;
CMD 805: *enable / disable CCA mode*;
CMD 795: *write timers interval*;
CMD 793: *write UTC time and date*;
CMD 808: *read time to live interval*;
CMD 973: *write service*;
CMD 974: *write route*;

A partir do último comando (974), as variáveis de processo do dispositivo de campo são lidas.

Levando-se em conta uma rede WH típica, onde o número de nós pode chegar a 100 unidades, intui-se que as latências no processo de formação são muito elevadas, visto que o tráfego de mensagens diverge do ponto de acesso e converge de volta a ele. De fato, tempos na casa de dezenas de minutos são comumente obtidos em experimentos de formação de redes WH. Extrapolando-se o raciocínio, verifica-se que o mesmo ocorrerá no processo de manutenção da rede, em maior ou menor escala dependendo das ações a serem tomadas. Estas podem ser mais simples, como no caso de solicitação de banda feita por um determinado dispositivo da rede, ou mais complexa, como no caso de uma

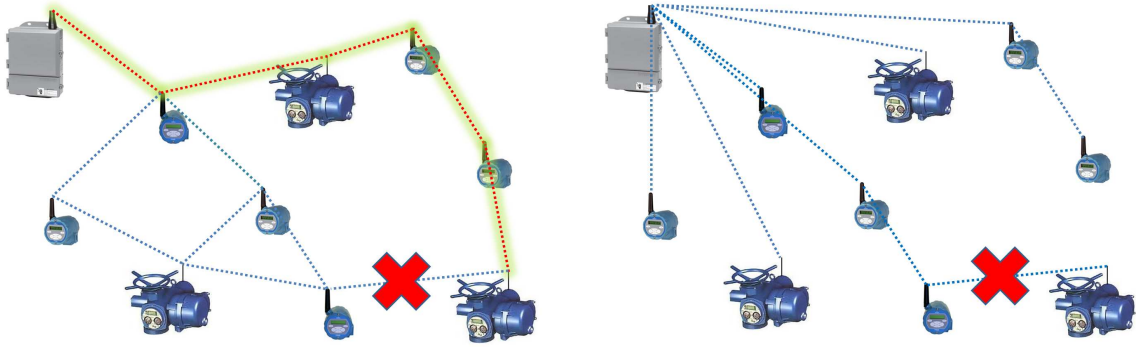

Figura 17: Topologia centralizadora do WH: à esquerda, uma rede em malha, onde um bloqueio é resolvido pela adoção de outra rota para encaminhamento das mensagens, e à direita, um bloqueio que não pode ser resolvido devido à rede centralizada.

desconexão ou bloqueio permanente que resulte na necessidade de reescalonamento de toda a rede.

Os tempos de agregação e manutenção medidos demonstram a total incapacidade de redes deste tipo para operarem com dispositivos móveis, tais como robôs ou veículos autônomos. Mesmo que o dispositivo móvel seja capaz de responder a um anúncio da forma mais rápida possível, as mensagens e repostas trafegarão pela rede na direção do gerenciador e de modo reverso, de forma que o processo todo será muito mais lento do que a passagem do nó móvel. A centralização da rede ainda acarreta em outros problemas. A Figura 17 à esquerda ilustra uma rede WH bem formada, com rotas estabelecidas em redundância. Na ocasião de um bloqueio, outra rota pode ser utilizada, como mostra a Figura. Porém na prática, as topologias mais observadas são as do tipo estrela, como ilustra à direita na Figura 17. Isto se deve à natureza centralizada do protocolo, que leva a escolhas de gerenciamento centralizado para que os requisitos de tempo sejam atendidos. Desta forma, uma rede como esta não poderá lidar com um bloqueio conforme ilustrado. Isto leva à conclusão de que a natureza centralizada do protocolo reduz a confiabilidade da rede, na medida em que a diversidade de caminhos é prejudicada.

4.2.2 Utilização de múltiplos pontos de acesso

Dentre as possibilidades de modificação do protocolo WH visando melhoria na dinamicidade da rede, são citados na literatura o uso de múltiplos pontos de acesso e o emprego de algoritmos de escalonamento distribuídos. A arquitetura fundamental de uma rede WH é composta pelas entidades centrais (gerenciador de rede, gateway e ponto de acesso) e os dispositivos de campo, conforme observado na Figura 18. De um lado, o ponto de acesso comunica com a rede em malha através de um transceptor de RF e do outro, com o gateway, através de um *host* que basicamente realiza o encapsulamento e verificação da integridade das mensagens através de algum protocolo não definido pela norma. O *host* comunica-se com o gateway, assim como o gerenciador de rede e o barramento de campo. As aplicações de controle da planta comunicam-se com a rede WH através do gateway. Desta forma, o gateway é responsável pela conversão de diversos protocolos e interpretação dos comandos WH recebidos. Basicamente, o gateway possui sistemas de gerenciamento de tarefas de comunicação para poder lidar com os diversos fluxos de dados. O gerenciador de rede provê o funcionamento do ponto de acesso, fornecendo a este um apelido e definindo um mapa de canais a serem utilizados. O provisionamento ainda inclui o estabelecimento de um canal de comunicação entre o gateway e

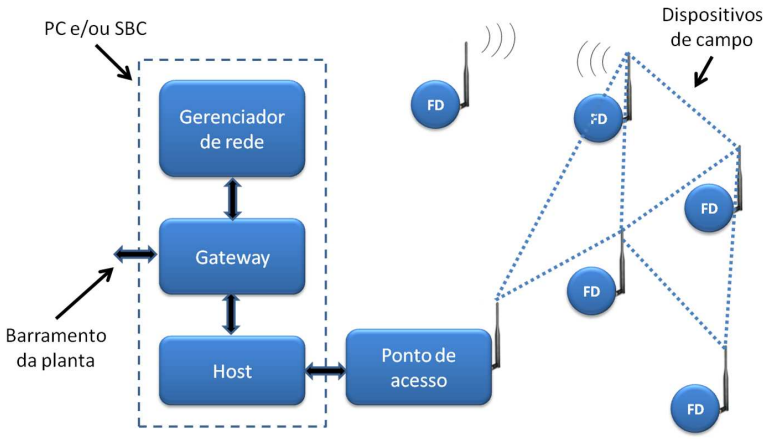

Figura 18: Arquitetura de rede WH conforme concebida na norma.

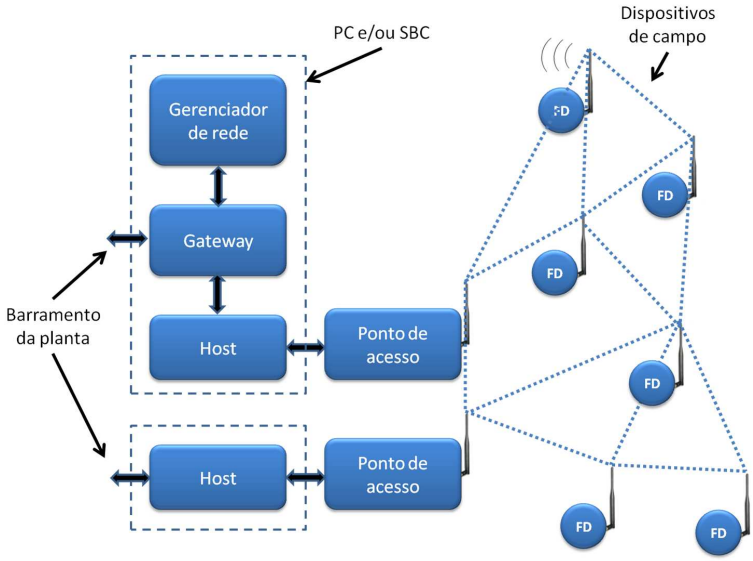

Figura 19: Arquitetura de rede WH com múltiplos pontos de acesso.

o *host*, para a troca de informações entre o ponto de acesso e o restante da rede. Sendo assim, o estabelecimento da rede sempre começa pela inicialização do gateway, que irá prover todos os canais de comunicação entre os dispositivos fundamentais.

O gerenciador de rede pode prover mais de um endereço a mais de um ponto de acesso. Os pontos de acesso podem ser localizados de forma estratégica ao longo da rede de modo que mais de um enlace com a malha possa ser estabelecido, aumentando a confiabilidade da rede. Esta é uma técnica amplamente empregada em redes do tipo WLAN, que são orientadas à conexão e não formam redes em malha. Ainda, com a utilização de mais de um ponto de acesso, o rendimento da comunicação pode ser elevado, com o uso de mais de um canal de RF concomitantemente, transformando a rede num sistema com capacidade de comunicações paralelas. O gateway pode ser virtualizado e a comunicação com os pontos de acesso pode ser feita pela rede Ethernet utilizando TCP/IP ou UDP. Desta forma, uma arquitetura como apresentada na Figura 19 é possível.

Outra vantagem, demonstrada por (HAN et al., 2011), é o fato de que uma rede WH somente apresentará roteamento plenamente redundante se utilizar pelo menos dois pon-

tos de acesso, permitindo múltiplos caminhos até o gateway. Porém, há diversos pontos críticos que devem ser observados nesta arquitetura. Primeiramente, o escalonador de mensagens centralizado deverá prover os links para todos os dispositivos da rede levando em consideração dois ou mais concentradores de rota, ou seja, os pontos de acesso. Isto implica na necessidade de algoritmos de escalonamento ainda mais complexos, que deverão levar em conta o paralelismo dos *superframes* no tempo. O algoritmo deverá provisionar os dois ou mais pontos de acesso com diferentes tabelas de roteamento e links. Isto em si não é um impedimento mas certamente acarreta em um grande aumento na complexidade e utilização de memória no gerenciador centralizado. Este aumento em complexidade é automaticamente espelhado ao sistema redundante, no caso da falha do gerenciador central. No que se refere ao aumento de dinamicidade, esta ocorrerá na medida em que o número de saltos de comunicação no processo de agregação diminuir, se considerarmos que o gerenciador consiga realizar o escalonamento das mensagens no mesmo tempo que numa rede com um único ponto de acesso. A utilização de canais de RF simultaneamente acarreta no aumento de vazão de dados da rede. Porém, deve-se tomar cuidado para que a ocupação de banda seja feita de forma adequada. O surgimento de imagens do canal modulado laterais à frequência central não é incomum, e leva à colisões entre mensagens. Em função destas dificuldades, não existem até o momento unidades centralizadoras comerciais que utilizem mais de um ponto de acesso.

4.2.3 Gerenciamento de rede: centralização versus descentralização

O assunto gerenciamento de rede centralizado versus descentralizado tem sido discutido ultimamente. Os argumentos variam bastante, mas pode-se resumir em alguns pontos principais. No que se refere a redes centralizadas, são apresentados os seguintes pontos favoráveis:

a) Redes centralizadas são mais adequadas para atender os requisitos de tempo real para aplicações industrias, porque nelas não existem conflitos entre solicitações geradas por dispositivos distribuídos. WH e ISA 100.11a são protocolos que adotam a estratégia de centralização do gerenciamento.

b) Com o gerenciamento centralizado, o hardware dos dispositivos de campo é simplificado a um mínimo necessário apenas para comunicação no protocolo (além da camada de aplicação de instrumentação). Isto permite soluções de baixo custo e consumo.

c) Numa rede centralizada, os nós não tem capacidade de agendar e rotear suas próprias mensagens. Estas são programada pelo gerenciador central, o que facilita o gerenciamento de decisões importantes como melhor ocupação de banda e atendimento dos requisitos de tempo real.

d) Para lidar com questões de segurança, as mensagens devem ser encriptadas, e as chaves, geridas por um gerenciador de segurança. Este tipo de ação deve ser centralizado para garantir acesso restrito.

Por outro lado, em defesa do gerenciamento descentralizado (aplicado especialmente em redes de sensores sem fio), antagonizando com os itens anteriores, argumenta-se:

a) Com algoritmos corretamente desenvolvidos, capazes de resolver conflitos de hierarquia, gerenciadores locais são muito mais rápidos para resolver demandas assíncronas da rede.

b) O atual estágio de desenvolvimento da microeletrônica permite dispositivos cada vez menores e de baixo custo e consumo, tais como SoCs que integram MCU, transcep-

tores de rádio e memória Flash. Se hoje em dia não é possível manter um SBC (*single board computer*) alimentado à bateria por longo tempo, isto deve ser tratado como um entrave e não como um impedimento tecnológico, visto que o desenvolvimento constante da microeletrônica permite o desenvolvimento de dispositivos que consomem cada vez menos energia.

c) Numa rede descentralizada, as decisões de roteamento de mensagens podem ser variadas, de acordo com diferentes parâmetros, tais como necessidade de latência mínima, confiabilidade espacial (múltiplas rotas), taxa de dados e gestão de energia. Por outro lado, nas redes centralizadas, o roteamento é direto, regrado pelo gerenciador. Isto leva a grandes latências devido aos tempos de planejamento e programação de rotas, o que torna esta abordagem adequada somente para redes estáticas.

d) Mecanismos de encriptação e autenticação podem ser gerenciados localmente de forma segura. Uma vez definida a política para geração de chaves de forma distribuída, de modo que não existam duplicidades, a segurança será equivalente ao sistema centralizado.

A descentralização do gerenciamento de uma rede WH certamente leva ao aumento da dinamicidade de rede, uma vez que diversos dos processos de rede podem ser resolvidos localmente. Porém sempre restará um ponto de convergência na rede: o gateway, que está sempre no extremo da rota de entrada/saída de dados da rede. A multiplicidade de gateways não está prevista na norma, pois leva a uma série de dificuldades. O sincronismo da rede é mantido na máxima instância pelo gateway, que é provido de um relógio preciso. Mesmo que seja utilizada outra técnica de disseminação de relógio, a rede cabeada, no caso de multiplicidade de gateways, trará desvantagens, uma vez que o objetivo primário é reduzir a utilização de cabos. Por outro lado, a multiplicidade de gerenciadores apresenta desafios mais simples, que se corretamente tratados, produzirão uma rede mais dinâmica e versátil.

4.3 Tese: descentralização de redes *Wireless*HART

Baseada nas ideias anteriormente apresentadas, esta tese propõe uma arquitetura parcial ou totalmente descentralizada, com capacidade de coordenar uma rede WH de forma distribuída. Objetiva-se diminuir a latência na entrada/saída de dispositivos e manutenção mais rápida, agendamento de mensagens mais eficiente e aumento geral da confiabilidade. Resumidamente, a proposta visa o aumento de dinamicidade e redundância. Por outro lado, não se pode ignorar que alguns problemas antes inexistentes são agora criados, tais como possíveis conflitos hierárquicos e a necessidade de algoritmos eficientes para a distribuição da configuração da rede. Como as soluções baseadas em software (algoritmos de descentralização na própria camada de rede da pilha de comunicação) impõem grande *overhead* aos sistemas atuais e necessitam de uma total reestruturação do sistema TDMA do WH, propõe-se a implementação da descentralização baseada em uma entidade à parte. O coprocessador de rede local (COP) permite a descentralização para solução de eventos assíncronos, independentemente do gerenciador central. A solução proposta mantém a pilha de comunicação WH praticamente inalterada, assim como o hardware dos dispositivos de campo. A arquitetura proposta para descentralização de uma rede WH pode ser vista na Figura 20.

Analisando a Figura, verifica-se que a arquitetura original da rede WH é pouco modificada, mas as modificações são importantes. Os dispositivos de campo dotados de coprocessador são os responsáveis pela descentralização do gerenciamento e, para que

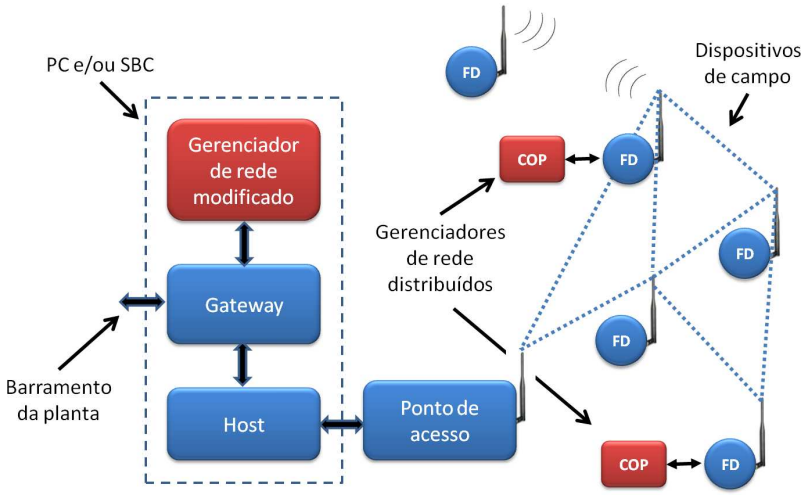

Figura 20: Arquitetura de rede WH modificada, com a inclusão de gerenciadores de rede distribuídos.

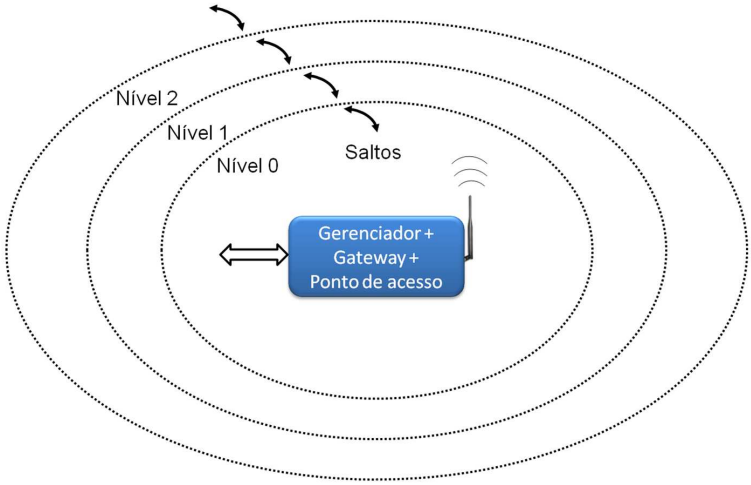

Figura 21: Definição de níveis numa rede WH, em função do número de saltos entre os nós da rede nos quais uma mensagem deve percorrer.

produzam efeito no aumento da dinamicidade da rede, devem ser estrategicamente posicionados na mesma, conforme será explanado adiante no texto. Os outros dispositivos de campo são convencionais, sem quaisquer alterações. O gerenciador central deve ser modificado de tal modo que esteja apto a aceitar comandos de reescalonamento produzidos pelos gerenciadores distribuídos.

4.3.1 Topologia de rede distribuída e critérios de decisão

Para entendimento da proposta de distribuição de rede, definem-se os níveis topológicos, conforme apresentado na Figura 21. Os níveis dizem respeito a quantos saltos uma mensagem deve realizar até atingir as entidades centrais, através do ponto de acesso. O nível zero corresponde a um salto apenas, o nível um corresponde a dois saltos e assim por diante. Uma vez definido o conceito dos níveis na rede, passa-se à analise da distribuição dos FDs especiais, através dos quais o gerenciador local é realizado.

A topologia de rede apresentada na Figura 22 leva em conta a existência dos geren-

ciadores distribuídos. A distribuição destes deve ser feita de forma que estes estejam posicionados em níveis maiores que zero, uma vez que não há vantagem (como será mostrado no estudo de caso), que o gerenciamento distribuído seja feito no nível zero. Isto se deve ao fato de que o ingresso de um novo dispositivo irá consumir mais tempo do que feito diretamente pelo gerenciador central, uma vez que o processo final consiste no envio de um comando para o gerenciador central. Após o escalonamento dos links para o novo dispositivo, um comando especial é enviado ao gerenciador central para que este tome conhecimento do novo dispositivo, o que acarretará num tempo maior. Desta forma, os gerenciadores distribuídos são posicionados estrategicamente, em níveis mais afastados e/ou em pontos críticos da rede, onde exista demanda por dinamicidade.

Um novo FD que queira agregar-se à rede, pode receber mensagens de anúncio de três classes diferentes de dispositivos: do gerenciador central (através do ponto de acesso), de um dispositivo dotado de coprocessador e de um FD comum, já agregado à rede. Os critérios para discernimento de prioridade são aqui definidos por três parâmetros: o RSL, a prioridade de agregação e o tipo de dispositivo expandido.

O RSL deve ser o maior possível, ou seja, na comparação entre os RSL das mensagens de anúncio recebidas, ganhará maior crédito a que for mais intensa. O limiar de -85 dBm é definido como o limite viável, embora os links físicos possam ocorrer a intensidades bem menores. Este valor vem da norma 802.15.4, que define a sensibilidade mínima de um transceptor para que seja compatível com o padrão.

A prioridade de agregação é definida pela variável de controle (*join control* e *join flag*, em um byte, enviado no pacote de anúncio). Quanto menor o valor, de um máximo de 15, maior a prioridade de agregação através daquele dispositivo. Desta forma, definem-se as prioridades máxima para o gerenciador central (valor zero), média para os gerenciadores distribuídos e menor para os dispositivos comuns, que ainda variam uma vez que o valor depende de critérios diversos tais como alimentação (bateria ou rede) e taxa máxima de pacotes suportada. Neste ponto, o gerenciador central deve ser modificado a fim de proporcionar prioridade de agregação adequada aos dispositivos, uma vez que pode fazê-lo através do comando 811. Caso um gerenciador decida reduzir a prioridade de um dispositivo em função do estado de carga de sua bateria, poderá fazê-lo, em concordância com as características de robustez e confiabilidade do WH. Por outro lado, deverá proporcionar aos gerenciadores distribuídos uma prioridade maior que a dos dispositivos comuns para que a descentralização seja efetiva. Este controle pode ser obtido através do tipo de dispositivo expandido.

O tipo de dispositivo expandido (*expanded device type*) é um número de dois bytes definido pela HCF para cada dispositivo homologado. Alguns valores são pré-definidos para os elementos fundamentais (entidades centrais) e outros, para ferramentas de desenvolvimento oficiais da HCF, conforme a lista a seguir:

0xF980: É o código do *Network Manager*, o gerenciador de rede único em uma rede WH;

0xF981: É o código do gateway, responsável pela conversão de diversos protocolos para a rede WH;

0xF982: É um código de dispositivo genérico. Quando um dispositivo não é homologado, deve utilizar esta especificação;

0xF983: Diz respeito ao SDC625, *Smart Device Communicator*, software oficial da HCF para checagem de descrições de dispositivos;

0xF984: É o código do Wi-Analys, o *sniffer* para WH oficial da HCF.

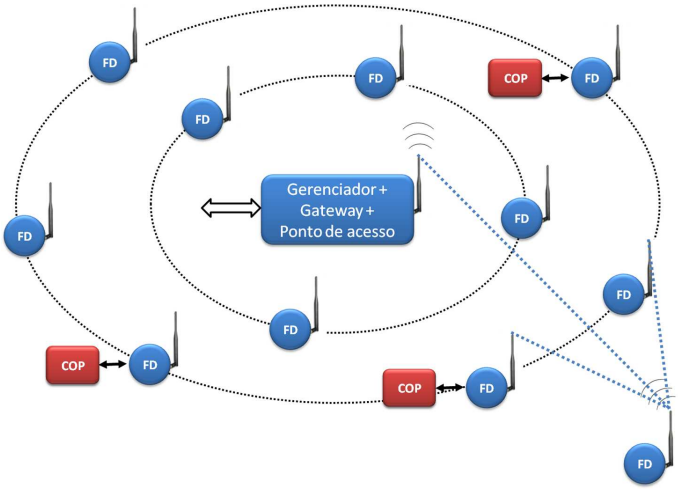

Figura 22: Topologia de uma rede WH composta pelos dispositivos convencionais e pelos alterados, com capacidade para gerenciamento distribuído.

Um valor acima destes poderia ser utilizado para definir um dispositivo dotado de coprocessador para gerenciamento. Deste modo, a própria aplicação na pilha de protocolo WH do FD diferenciado pode ser controlada para que evoque o coprocessador, uma vez que está identificada como um dispositivo de tipo especial. O mesmo ocorre com o gerenciador central. Uma vez que o comando zero enviado na requisição de agregação identifica o número correspondente ao tipo de dispositivo expandido, o gerenciador central tratará aquele dispositivo de forma diferente. Isto inclui a possibilidade de pré-escalonamento com *superframes* e links específicos para aquele elemento da rede, de forma que ele possa usufruir de links especificamente criados para a solução de eventos localmente. Cabe ressaltar que esta é uma das possíveis estratégias para a solução de conflitos hierárquicos entre os gerenciadores da rede, conforme é visto adiante no texto.

Uma vez definida a topologia adequada para distribuição do gerencimaento e os critérios de decisão, passa-se à análise dos cenários possíveis, no que se refere ao processo de agregação, que será utilizado no estudo de caso. Na Figura 23, um FD que pretende juntar-se à rede responderá preferencialmente a um anúncio propagado pelo ponto de acesso, uma vez que este se encontra no nível zero. Na Figura 24, a agregação ocorrerá através de um dispositivo com coprocessador, uma vez que o FD que quer agregar-se à rede está localizado em um nível maior que zero. Neste caso, mesmo que o novo FD perceba anúncios com oriundos de um dispositivo comum RSL maior, deverá responder ao anúncio de um dispositivo com coprocessador, desde que o RSL seja maior que o limiar para boa conexão (definido no *firmware* e obtido pelo comando 777). Outros cenários são possíveis, tal como apresentado na Figura 25. Neste caso, a escolha entre a requisição de agregação em função dos anúncios percebidos de dois dispositivos dotados de coprocessador é feita em função do RSL. Finalmente, no caso da Figura 26, o novo FD recebe dois anúncios de dispositivos com coprocessador, mas em níveis distintos. Uma vez que o coprocessador tem a capacidade de calcular em que nível da rede ele se encontra, é possível manipular a variável de prioridade de agregação, tornando mais eficiente o processo de descentralização. Considera-se ainda que o critério de escolha para agregação depende da relação confiabilidade / latência. Um novo FD pode perceber dois anúncios, um de baixo RSL emitido pelo ponto de acesso e outro de RSL elevado mas emitido por um dispositivo com coprocessador. Neste caso, pode-se optar por um link de baixa latência mas pouco

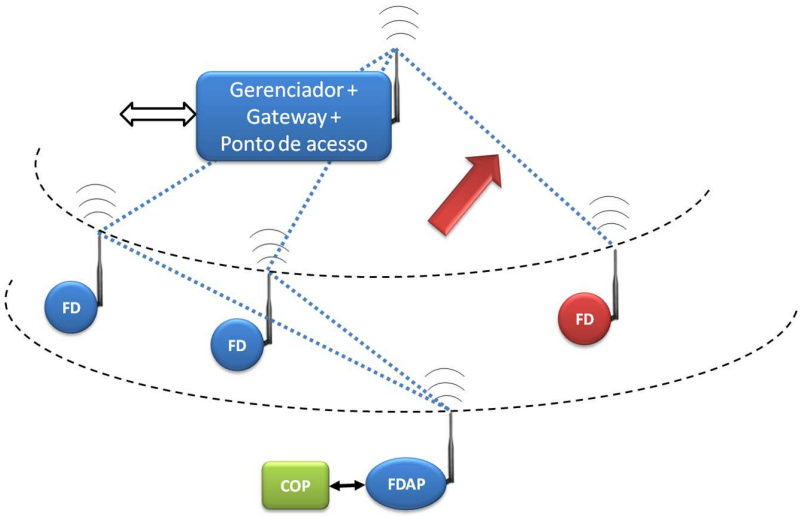

Figura 23: Possível cenário para agregação de um novo FD através do gerenciamento distribuído (I).

confiável, utilizando o enlace direto com o ponto de acesso. Caso contrário, se a opção for por um link mais confiável mas com latências a serem consideradas, utiliza-se o outro acesso. Este impasse pode ser resolvido em função dos valores de RSL e dos níveis de prioridade.

4.3.2 Possibilidade de *Handover*

O *handover* é uma técnica que foi desenvolvida para propiciar que dispositivos de comunicação móveis possam manter enlace com estações centrais sem que sejam necessárias altas potências de RF, de forma a permitir a troca dinâmica de conexão com os diferentes pontos de acesso da rede.

O *hard handover* é a técnica em que o enlace feito em um canal dentro dos limites de uma célula é liberado somente no momento em que outro canal na célula vizinha é utilizado. Neste caso, ocorre a desconexão física (mas mantém-se a lógica) no momento que a nova conexão é feita. Para que se possa manter a conectividade, o *hard handover* deve ser feito de forma rápida, para que não sejam perdidos dados durante a transição. Diferentemente, o *soft handover* mantém a conexão física com as duas estações base durante a transição para que se mantenha o link. Neste caso, a utilização de recursos da rede é maior mas por outro lado, garante-se a conectividade. Para que seja possível o *soft handover*, dois canais devem ser utilizados concomitantemente. Se a técnica de acesso ao meio for do tipo TDMA, o conceito de canal é convertido em tempo, e neste caso, o recurso utilizado é a banda de rede disponível ou seja, os *slots* de tempo designados para realizar o *handover*.

No WH, o *handover* é aventado como uma técnica para propiciar o emprego de dispositivos móveis que trocam de conexão com os diferentes pontos de acesso. A Figura 27 apresenta o cenário em que um dispositivo móvel realiza *handover* através dos pontos de acesso de uma rede WH. O dispositivo móvel está conectado ao ponto de acesso à esquerda e o *handover* ocorrerá na medida em que o RSL do enlace com aquele transceptor fique menor do que o RSL de um eventual enlace com o segundo ponto de acesso.

O estudo do *handover* como uma técnica para uso de dispositivos móveis no WH revela diversos problemas para uma possível implementação prática.

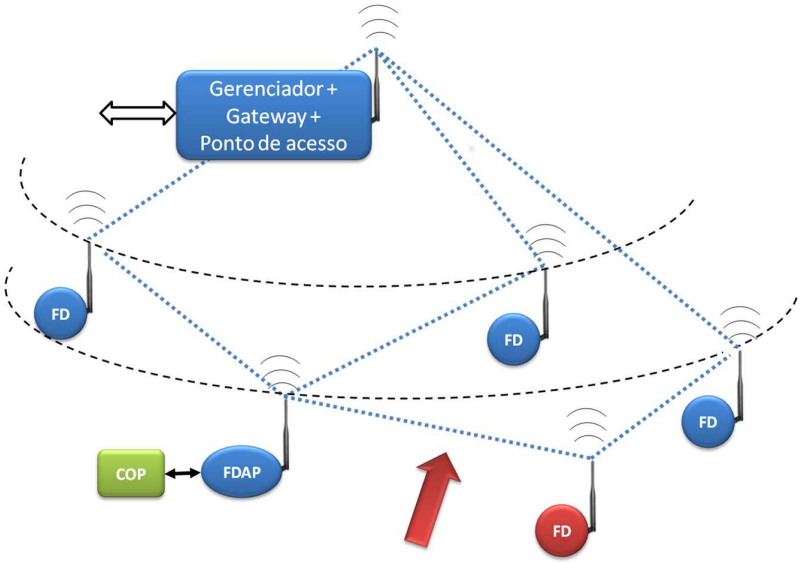

Figura 24: Possível cenário para agregação de um novo FD através do gerenciamento distribuído (II).

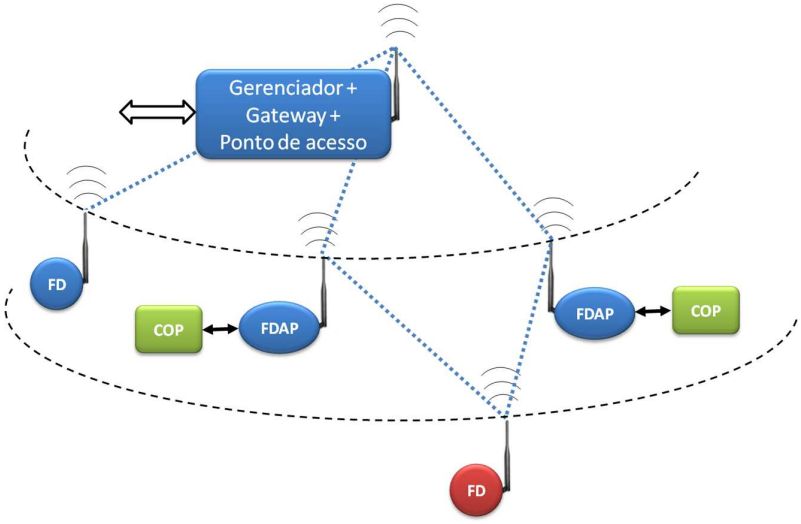

Figura 25: Possível cenário para agregação de um novo FD através do gerenciamento distribuído (III).

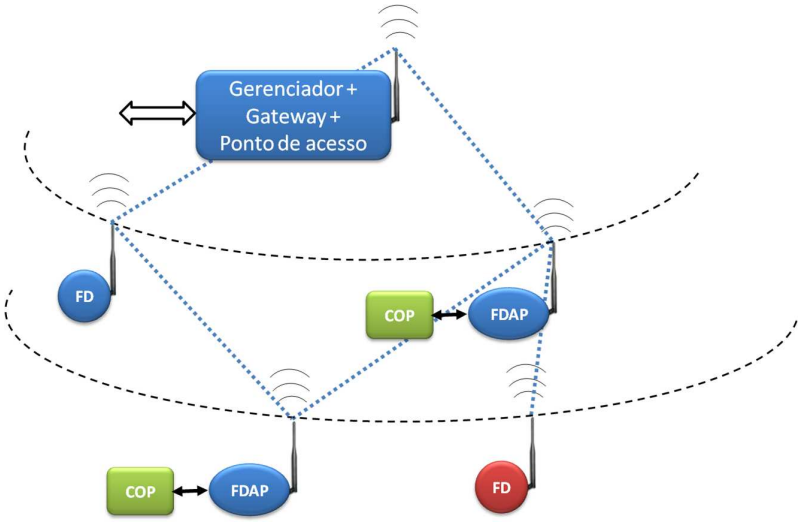

Figura 26: Possível cenário para agregação de um novo FD através do gerenciamento distribuído (IV).

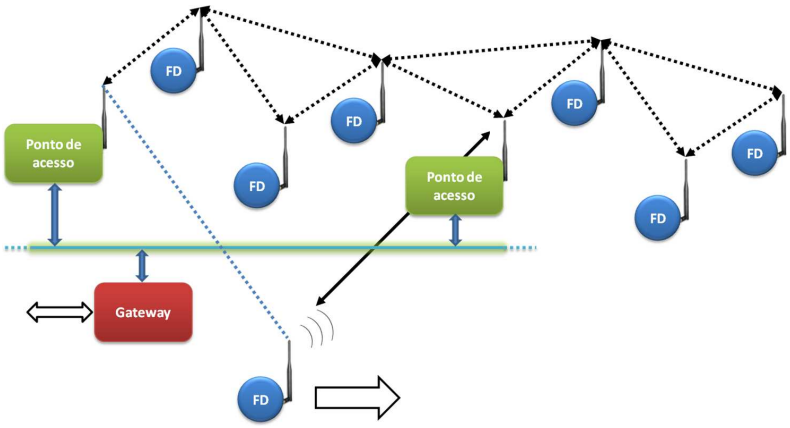

Figura 27: Avaliação da possibilidade de uso da técnica de *handover* em uma rede WH convencional.

a) Numa rede WH convencional, cada ponto de acesso tem um apelido, e, para que um dispositivo móvel possa realizar *handover*, ele deverá possuir links com todos os pontos de acesso da rede. Isto demanda a programação prévia de links para o dispositivo móvel de modo a antever o caminho físico que ele percorrerá.

b) Mesmo com um grande número de pontos de acesso, o dispositivo móvel poderá tomar um caminho físico em que saltos entre os nós da rede devam ocorrer para que a mensagem atinja o ponto de acesso. Neste caso, o dispositivo móvel também deverá ser programado com links para os demais nós da rede, que porventura venha a encontrar. Isto se deve ao fato de que conexões dedicadas entre nós vizinhos não podem ser estabelecidas sem anuência do gerenciador.

Pela análise anterior, verifica-se que o *handover* no WH convencional é impraticável, tendo em vista a natureza centralizada da rede. Para que esta técnica seja possível de ser empregada na prática, pode-se considerar uma arquitetura de gerenciamento distribuído, como apresentado na Figura 28. Neste caso, os dispositivos dotados de coprocessador podem realizar comunicação com o dispositivo móvel independentemente do gerenciador central. Numa rede WH convencional, as requisições de agregação podem até ser atendidas muito rapidamente, mas a agregação à rede pode não se concretizar, uma vez que o dispositivo móvel poderá estar fora da área de alcance quando as mensagens retornarem a ele. No caso do atendimento de um anúncio feito por um dispositivo dotado de coprocessador, na eventualidade de que uma agregação completa não seja possível, o coprocessador poderá tomar decisões totalmente distintas das que ocorrem em uma rede WH convencional. Uma das sequências possíveis inclui os seguintes eventos:

a) O dispositivo móvel recebe o anúncio de um dispositivo de campo com coprocessador.

b) A resposta à requisição é prontamente atendida, com o escalonamento da comunicação estabelecida com o dispositivo móvel.

c) Nenhuma rota para o dispositivo móvel com outros elementos da rede é estabelecida, exceto se houver outro dispositivo com coprocessador no entorno.

d) O RSL é constantemente avaliado com fator de amortecimento baixo, para que se identifique a tendência do comportamento do sinal e por consequência, do movimento do dispositivo.

e) Com o RSL aproximando-se do limiar de boa conexão, o dispositivo móvel é desconectado da rede através do comando 960, deixando-o apto a realizar nova requisição de agregação que não será novamente atendida pelo dispositivo que o desconectou. Utiliza-se histerese no nível do sinal para restabelecer conexão, caso o dispositivo móvel retorne à origem.

f) Caso seja possível a comunicação entre outros dispositivos com coprocessador incorporado, o *handover* realizado é do tipo soft, uma vez que o vizinho gerenciador pode ser programado para que tenha links com o dispositivo móvel. Neste caso, a desconexão não é feita, mantendo-se o dispositivo móvel agregado com a rede. Apenas os links anteriormente utilizados são apagados, e o dispositivo com coprocessador que originou a conexão não será mais vizinho do dispositivo móvel.

g) Em ambos os cenários, os dados do FD capturados pelos dispositivos com coprocessador, são enviados ao gerenciador central ou ao ponto de acesso ligado ao gateway sem a necessidade de estabelecimento de outros links e rotas para o dispositivo móvel.

A técnica de *handover* é aqui apresentada como mais uma possibilidade de melhoria na dinamicidade de redes WH. O uso do gerenciamento distribuído, através de dispo-

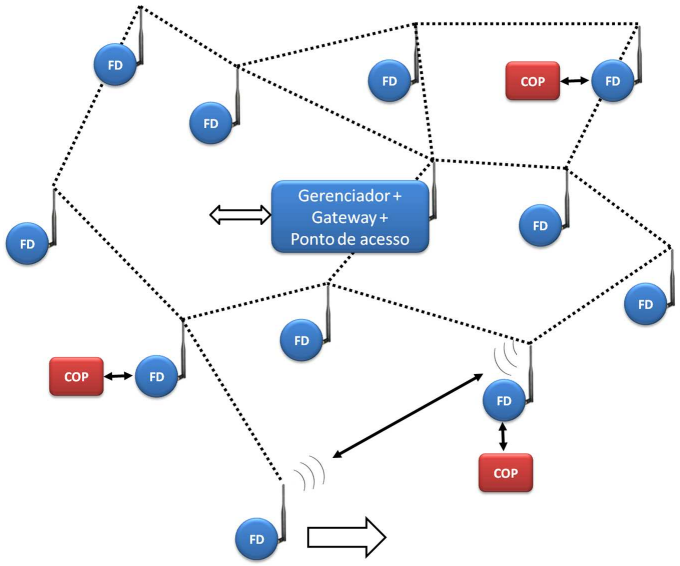

Figura 28: Técnica de *handover* em uma rede WH com dispositivos especiais, com capacidade de gerenciamento distribuído.

sitivos especiais dotados de coprocessadores estrategicamente posicionados ao longo da rede, permite o desenvolvimento de técnicas específicas para lidar com o *handover*. A agregação do FD móvel com os diferentes gerenciadores distribuídos é possível uma vez que estes tem a capacidade de criar sessões com o dispositivo móvel localmente, o que não é possível em redes WH convencionais. A estratégia aqui apresentada não é implementada, mas os resultados dos estudos de caso apresentados no Capítulo 6, servem como balizador para a avaliação de um futuro desenvolvimento do *handover* no WH.

5 IMPLEMENTAÇÃO

5.1 Introdução

Neste capítulo é apresentada a implementação da proposta de descentralização do gerenciamento de uma rede WH. O método é definido para resolver localmente um processo de agregação de um nó à rede como prova de conceito. A manutenção de rede também é possível, embora não faça parte do estudo de caso, sendo considerado um trabalho futuro. Além da descentralização parcial, o método utilizado permite a formação de uma rede a partir de um coprocessador local, com a vantagem do aumento de dinamicidade.

5.2 Dispositivo de campo modificado

Nesta seção são apresentadas as modificações implementadas num dispositivo de campo WH, para que este possa fazer parte do esquema descentralizado, de modo que o coprocessador para gerenciamento local possa utilizá-lo para atuar na rede. O FD WH possui a pilha de protocolo para comunicação sem fio (camada física 802.15.4) e com fio, através de uma porta serial conectada a um modem FSK ou RS-485. A fim de tornar possível o gerenciamento de uma rede WH localmente através de um FD, é necessário modificar sua arquitetura interna. O processo de comunicação convencional de um FD sempre inicia e termina pelo canal RF, ou seja, as mensagens oriundas de outros FDs ou do ponto de acesso são recebidas pela antena, processadas nas camadas de dados, rede e aplicação e retornam à antena na ordem reversa. As mensagens recebidas pelo canal serial através da porta de manutenção nunca são enviadas ao canal RF. Para a implementação do gerenciamento distribuído, é necessário fazer com que o fluxo de mensagens possa ser controlado, de modo que estas possam ser endereçadas do canal RF ao canal serial e vice versa. Por outro lado, pela proposta apresentada, o FD deve apresentar funcionamento convencional, quando a rede não é escalonada pelo gerenciador local. Isso demanda um comportamento dualístico, de maneira que um FD WH modificado possa ser controlado externamente para comportar-se como FD ou como AP, conforme a necessidade. A Figura 29 apresenta uma arquitetura típica de um AP. Nela é possível verificar as estruturas internas da pilha implementadas no transceptor, comumente chamado de coprocessador de rádio, e a comunicação feita com as demais entidades centrais. A pilha WH é parcialmente implementada, uma vez que o processo de formação de pacotes, encriptação, decriptação e validação de mensagens é feito externamente pelo gateway. A comunicação entre o coprocessador de rádio e o gateway é realizada através de um software hospedeiro simples que tem a função de traduzir o protocolo de comunicação do canal serial do rádio com o gateway. Embora não especificado na norma, normalmente é utilizado um soquete do tipo UDP ou TCP/IP de um lado e algum protocolo de comunicação serial do outro. O

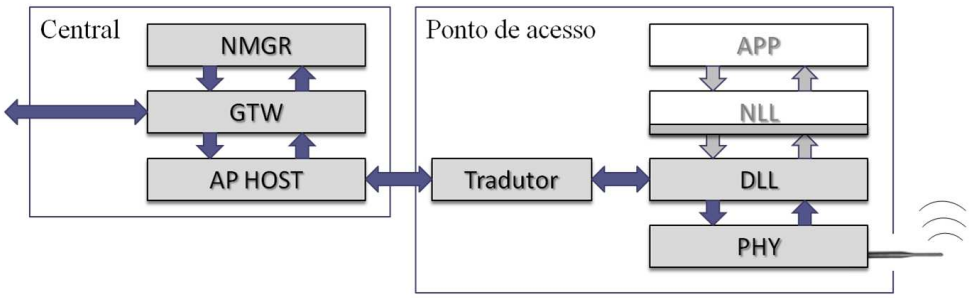

Figura 29: Arquitetura de um ponto de acesso para redes WH.

ponto de acesso WH foi desenvolvido e é utilizado neste trabalho e em pesquisas diversas. A referência (HAHN et al., 2012) pode ser utilizada para compreensão do assunto.

O ponto de acesso WH funciona de forma adequada como parte do dispositivo dotado de coprocessador para gerenciamento local. Por outro lado, é necessário implementar o mecanismo de criptografia e montagem de pacotes externamente, o que torna o coprocessador mais complexo. Ainda, a opção de funcionamento como FD WH convencional, exige a implementação das demais camadas do protocolo WH. Desta forma, a modificação de um dispositivo de campo convencional, que já contém todas as camadas e funções necessárias para uma pilha WH é a escolha mais adequada.

A arquitetura do dispositivo de campo modificado, agora apelidado de FDAP (*Field Device - Access Point*) é apresentada na Figura 30. Nesta arquitetura, a pilha WH é agora completa e a porta de manutenção é utilizada como canal de comunicação com o coprocessador externo, para gerenciamento local. A porta de manutenção é obrigatória em FDs compatíveis com o WH, e portanto, deve estar presente em todos equipamentos desenvolvidos com este propósito (LIMA et al., 2012). O comissionamento é realizado através da porta, para que o dispositivo de campo possa solicitar ingresso numa rede WH. A primeira modificação realizada é a transformação da porta de manutenção em uma porta genérica para comunicação serial com o coprocessador externo. Para tanto, o modem FSK, se presente, é removido e a comunicação tem velocidade aumentada de 1200 para 115200 bps. A principal motivação para utilização da porta de manutenção com o outro propósito é justamente tirar proveito do que já existe implementado, ou seja, o encaminhamento de mensagens HART para a camada de aplicação da pilha. Uma função de *loopback* retorna a resposta de comandos enviados pela porta de manutenção, normalmente utilizada para o comissionamento e monitoramento do processo de agregação e subsequentes atividades do FD na rede. Porém, para que se possa controlar o dispositivo, é necessário fazê-lo aceitar comandos restritos, que são normalmente aceitos somente se enviados pelo gerenciador de rede. Desta forma, é possível por exemplo, enviar comandos de escalonamento diretamente ao FD através da porta de manutenção. Com esta modificação, o controle do FD conectado ao coprocessador para gerenciamento local é garantido. Porém, estas modificações não permitem que uma mensagem enviada pela porta de manutenção seja enviada à camada física do rádio. Para que isto seja possível, são feitas diversas modificações na pilha de comunicação além da implementação de comandos específicos, ou seja, que pertencem àquela classe de dispositivo somente. São implementados os comandos 136, 137 e 138 descritos a seguir.

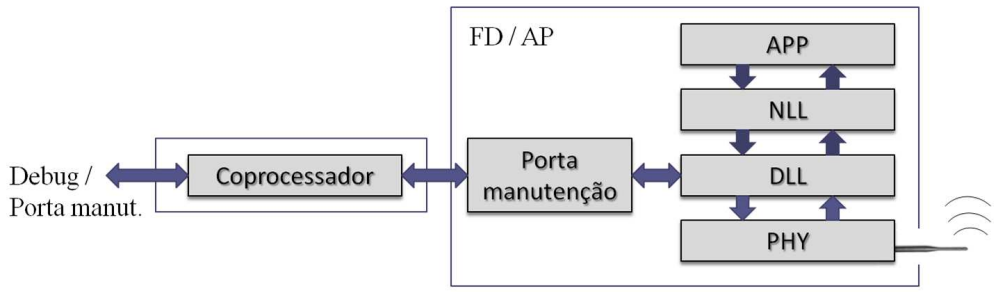

Figura 30: Arquitetura de um FDAP (*Field Device - Access Point*).

5.2.1 Comando 136

O comando implementado 136 tem como função controlar a transmissão de mensagens para, ou através do FDAP. O comando aceita os seguintes parâmetros:

a) IS_NAP: este parâmetro indica ao FDAP para que se transforme em AP para recepção de mensagens. Quando recebidas, uma variável interna na pilha é ajustada para o controle de diversas funções do firmware. A ação de controle mais importante é feita na função de processamento de recepção de mensagens, quando uma mensagem de requisição de agregação é recebida e desviada para o coprocessador. Isto é feito através de uma caixa de mensagens e da evocação do coprocessador, feita através de interrupção serial.

b) ACK: este parâmetro indica a confirmação do recebimento de uma mensagens pela DLL (DLL_Transmit_Confirm), conforme será explicado a seguir.

c) IS_XMIT: este parâmetro sinaliza o FDAP para que se transforme em AP para transmissão de mensagens. Quando o parâmetro é recebido, uma variável interna na pilha é ajustada para que o fluxo das mensagens recebidas pela porta de manutenção sejam encaminhadas em direção à camada física de RF. Junto com este parâmetro, são enviados mais dois: o UID e o apelido do dispositivo para o qual a mensagem é endereçada. Se o apelido inválido (0xFFFF) é enviado, o endereçamento é feito através do UID. Isto é necessário para o caso do escalonamento de comunicação com um novo FD, no processo de agregação do mesmo. No início do processo, o FD não tem apelido e portanto, toda comunicação deve ser feita com UID.

d) IS_FD: este parâmetro indica ao FDAP para que se comporte como um FD WH convencional. Neste caso, todas as funcionalidades de um FD compatível com WH são restauradas, o coprocessador é ignorado, e o processo de agregação e manutenção em uma rede WH ocorre da forma original.

e) JOIN: o parâmetro JOIN permite que uma mensagem de anúncio seja diretamente enviada ao FDAP. Isto permite que ele possa agregar-se à uma rede formada pelo coprocessador de gerenciamento local e pelo próprio FDAP através da porta de manutenção. Uma rede assim formada, a partir de um processo local, pode ser estendida para a formação de redes totalmente descentralizadas, tornando possível diversas configurações distribuídas. Também é útil para desenvolvimento e pesquisa de redes WH, uma vez que os algoritmos de escalonamento e roteamento são implementados no coprocessador, independentemente do gerenciador central. Porém, como não há um gateway, a rede formada localmente não consiste em uma rede WH de fato.

5.2.2 Comando 137

O comando 137 realiza a leitura de uma caixa de mensagens (*mailbox*) interna, de recepção de mensagens pelo FDAP quando este está no modo AP. Caso o FDAP esteja no modo FD, uma mensagem recebida pelo canal RF é tratada de forma normal, passando pela camada de dados, de rede, transporte e de aplicação (se for o caso) e realizando o caminho de volta com a inclusão do código de resposta. Por outro lado, caso esteja no modo AP, uma função chamada NAP_Transmit_Request é evocada. Esta função salva a mensagem recebida quando esta atinge a camada de rede, ou seja, já desencapsulada e descriptografada. Neste caso, a mensagem não faz o caminho de volta, permanecendo à disposição na *mailbox* até que seja lida. Concomitantemente, uma *flag* de indicação de recepção de mensagem dispara outra mensagem para evocar o coprocessador de gerenciamento através da porta de comunicação serial. Esta mensagem gera uma interrupção no coprocessador, indicando que há uma *mailbox* para ser lida no FDAP. A leitura da *mailbox* é feita através do comando 137. Como a mensagem está desencapsulada, o coprocessador pode interpretá-la imediatamente, porém, antes de realizar quaisquer ações, o coprocessador envia um comando 136 com o parâmetro ACK que irá indicar à camada de dados, que os *buffers* e *flags* referentes à mensagem recebida já podem ser esvaziados e limpos.

A Figura 31 apresenta os quatro possíveis fluxos de dados dentro de uma FDAP, de acordo com o controle efetuado pelo coprocessador. Em *a* e *b*, o FDAP está no modo FD e portanto, os comandos enviados pela porta de manutenção (*a*) seguem o fluxo determinado em *a* e os comandos recebidos e enviados pelo canal RF, seguem o fluxo *b*. Notar a bidirecionalidade das setas em *b*, devido à origem da mensagem, que pode ser interna ou externa ao dispositivo. Em *c* o FDAP está no modo AP e recebeu a instrução IS_XMIT do coprocessador para enviar o comando até a pilha WH, de forma a ser propagado pelo canal RF. Em *d*, o FDAP está no modo AP e redireciona as mensagens recebidas pelo canal RF até a porta de manutenção.

5.3 Gerenciador central modificado

Para composição do esquema de descentralização parcial proposto nesta tese, é necessário modificar o gerenciador central de modo que este possa aceitar comandos especiais enviados pelos gerenciadores distribuídos através dos FDAPs. No que se refere às técnicas de escalonamento descentralizado, há diversas estratégias para que o esquema de descentralização funcione da melhor forma possível, evitando conflitos hierárquicos que antes não existiam no esquema centralizado. Aliado a isto, o escalonamento central impacta diretamente na técnica de descentralização, descrita na subseção 5.4.3. Três abordagens são descritas a seguir, onde são apresentadas algumas de suas vantagens e desvantagens:

Escalonamento livre. Nesta abordagem, o escalonamento distribuído é feito pelo coprocessador independentemente do gerenciador central. Na ocasião de um evento de escalonamento local, o coprocessador entra num estado para aprendizado através do envio dos comandos de leitura do escalonamento corrente na rede na sua área de abrangência. Os comandos de leitura de links, *superframes*, canais proibidos, grafos e arestas são primariamente os necessários para conhecimento do estado atual da rede. Outros comandos que revelam estatísticas diversas da rede poderiam ser utilizados, mas são restritos ao gerenciador principal e desta forma não podem ser evocados por um FD. A análise das respostas dos comandos de escalonamento revela como a comunicação do FDAP foi feita pelo gerenciador central, mas não garante conhecimento total da rede. A partir deste

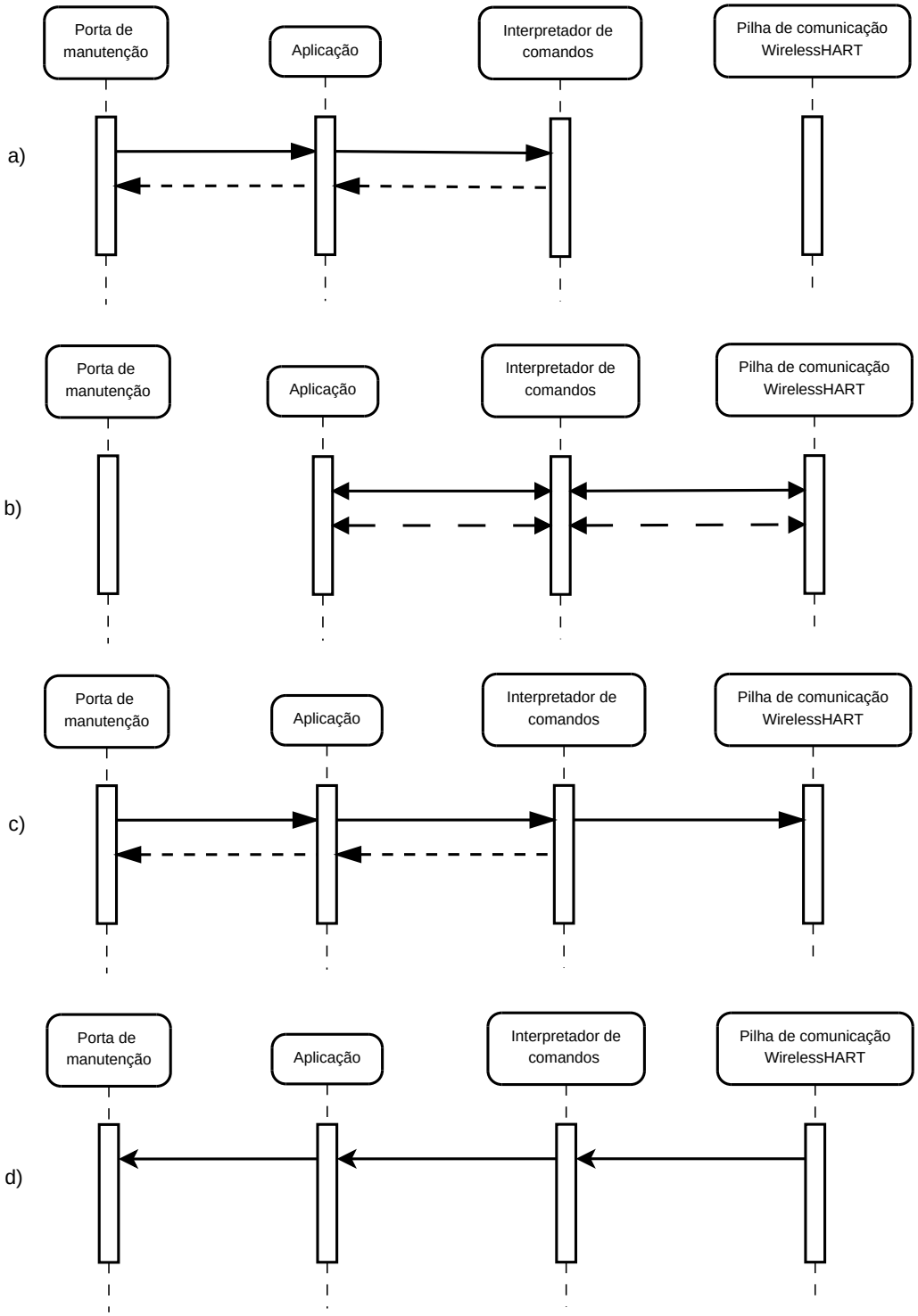

Figura 31: Possíveis fluxos de dados dentro de um FDAP.

ponto, o coprocessador gera o escalonamento para o novo dispositivo baseando-se no conhecimento da rede herdado do FDAP. Neste esquema, conflitos de atribuição podem ocorrer, uma vez que não é possível conhecer-se a rede toda. A possibilidade de conflitos é relacionada com a quantidade de nós e com o número de ocorrências de escalonamento distribuído e que são relativas à densidade de rede.

Escalonamento condicionado. O escalonamento condicionado consiste na liberação de recursos fornecidos pelo gerenciador principal. Este, previamente configurado para operar em conjunto com gerenciadores distribuídos, garante aos mesmos banda para propiciar que novos dispositivos possam ser agregados de forma distribuída. Deste modo, não ocorrem conflitos hierárquicos, uma vez que os gerenciadores distribuídos não utilizam recursos que não lhes foram concedidos. Como dificuldade adicional, o gerenciador central deve ter seus algoritmos de escalonamento reprogramados para que disponibilizem recursos para os gerenciadores distribuídos.

Reuso de *slots*. Uma das formas mais simples de se evitar conflitos entre os gerenciadores é o escalonamento distribuído através do reuso de *slots*. No WH, isto pode ser feito através dos links para anúncio, convertendo-os convenientemente em links do tipo *unicast* para comunicação com o novo FD. A estratégia consiste em transformar os links no momento em que ocorra uma requisição de agregação feita por um novo FD, uma vez que este já conhece através do próprio pacote de anúncio, quais são os *slots* de transmissão e recepção passíveis de serem utilizados para a comunicação com o proxy (FDAP, no caso descentralizado). O gerenciador centralizado nunca utilizará os links para anúncio com outro propósito, e desta forma não ocorrerão conflitos hierárquicos. Por ser a mais simples, esta foi a abordagem utilizada na implementação do FDAP. Por outro lado, dois problemas podem ocorrer:

a) O gerenciador central pode simplesmente não atribuir links para anúncio, de forma que não existam *slots* para escalonamento local. Isto pode ocorrer no caso em que o dispositivo tenha restrições de alimentação ou o gerenciador não queira atribuir links de anúncio, pelo fato de ter reconhecido alguma limitação no roteamento de pacotes através daquele dispositivo. Uma forma de resolver este problema é novamente através do identificador de tipo de dispositivo expandido. Uma vez reconhecendo que o dispositivo possui um coprocessador agregado, o gerenciador central irá programa-lo com links de anúncio. Ainda, isto pode ser feito de forma a maximizar o incremento de dinamicidade, pela atribuição de um grande número de *slots* de anúncio, que podem ser utilizados para agregação distribuída de diversos novos FDs. Esta estratégia tem como desvantagem a diminuição de banda disponível na rede como um todo.

b) Os *slots* para anúncio são compartilhados e portanto, colisões podem ocorrer. Como o mecanismo CSMA-CA é utilizado, há ainda a possibilidade do problema do terminal escondido. Caso não ocorram colisões, os links são convertidos para tipo *unicast* não-compartilhado, prevenindo definitivamente a ocorrência de colisões.

5.3.1 Comando 138

A funcionalidade de integração de um dispositivo à rede por meio do gerenciador distribuído é realizada por meio de um novo comando específico (Comando 138). Todo o processo de agregação e provisionamento é responsabilidade do gerenciador distribuído. Estas entidades provisionam o novo dispositivo com seus próprios recursos de rede (*superframes* e links), disponibilizados anteriormente pelo gerenciado principal (caso do escalonamento condicionado) durante o seu processo de provisionamento ou sob requisição de serviços. O recebimento do Comando 138 sinaliza ao gerenciador principal a integra-

Tabela 4: Parâmetros do comando 138 implementado.

Campo	bytes	Nome	Função
1	5	UID	Código identificador único
2	2	Apelido	Para uso após a resposta do anúncio
3	1	*Superframe* ID	*Superframe* que foi utilizado
4	1	Número de links	Quantidade de links atribuídos ao novo FD
5	2	*Time slot*	Número do TS utilizado
6	1	Tipo e opção	Tipo e opção de link para o TS
7	1	ID do grafo	Identificador do grafo para o gateway
8	1	ID da rota	Identificador da rota para o gateway

ção do novo dispositivo e contém informações sobre os links e *superframe* provisionados ao FD, além dos identificadores de grafo e rota a que o nó de rede está associado. O UID e apelido atribuídos ao dispositivo também são lidos do *payload* do Comando 138, para que o gerenciador principal crie um objeto *netdevice* respectivo na sua base de dados e mantenha a consistência com os gerenciadores distribuídos.

O comando 138 é implementado no FDAP para que os resultados do escalonamento, bem com outros parâmetros oriundos do novo FD que ingressou na rede, sejam imediatamente conhecidos pelo gerenciador central. Isto é fundamental para que o caminho até o gateway seja estabelecido e para que as variáveis de processo estejam imediatamente disponíveis, levando a um grande aumento da dinamicidade da rede, conforme sugerido na proposta e apresentado no estudo de caso. O comando 138 recebe como argumento os links e os seus parâmetros, que foram atribuídos ao novo FD. São eles o superframe utilizado, o UID e apelido atribuído, as variáveis de processo e algumas informações acerca do tipo do dispositivo. Posteriormente, outros comandos podem ser requisitados pelo gerenciador principal no processo normal de manutenção da rede. O formato do comando 138 é apresentado na Tabela 4.

O gerenciador central é modificado e utilizado para que se possa validar o funcionamento com dispositivos de campo compatíveis (RECH, 2012). O funcionamento do FDAP como FD é verificado através do sistema completo composto pelo gateway, ponto de acesso e gerenciador desenvolvidos, assim como através de gerenciadores comerciais. A modificação completa do gerenciador principal não é concluída nesse trabalho, restando como uma das tarefas futuras. Porém, a validação da arquitetura proposta é feita com base nas medições dos tempos de comunicação do FDAP com gerenciador, desde o escalonamento local até o envio do comando 138, conforme apresentado no estudo de caso.

5.4 Coprocessador para gerenciamento local

O coprocessador para descentralização é implementado inicialmente em plataforma PC, com hardware conectado ao FDAP através de uma porta USB e conversor RS-485. O software é desenvolvido de forma modular, com a implementação dos comandos (incluindo os especiais) em arquivos fonte separados, escritos em ANSI C. Apenas a interface gráfica é feita em C++, porém, o ambiente gráfico não é de fato necessário numa implementação final, sendo útil apenas para a realização de comissionamento através da interface com o coprocessador, uma vez que a porta de manutenção recebeu outro propósito.

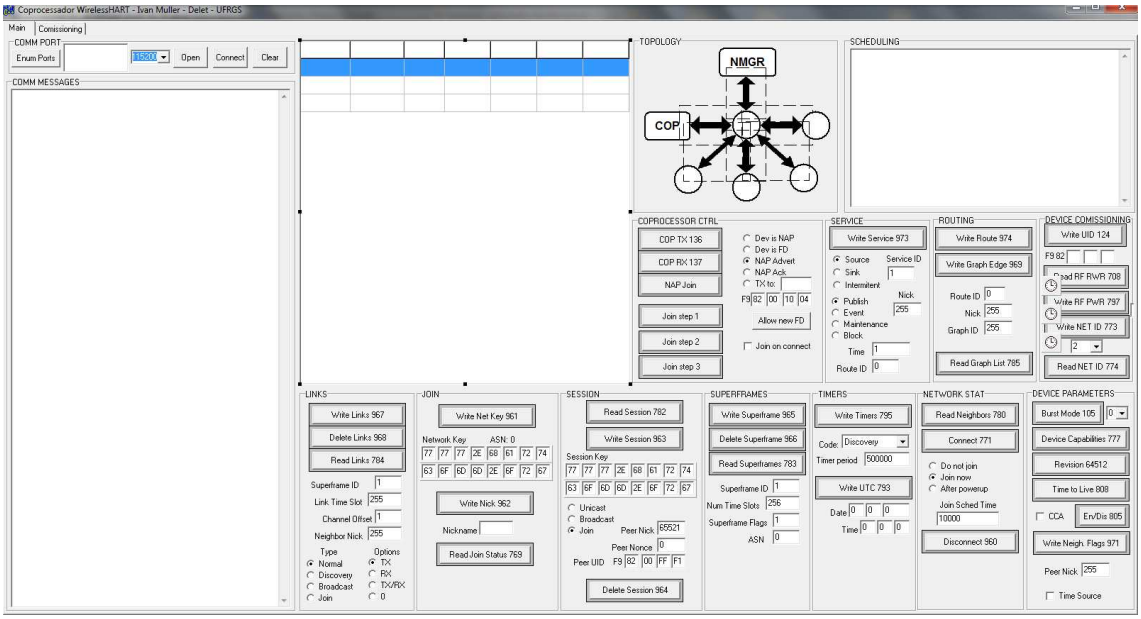

Figura 32: Interface gráfica do coprocessador implementado em software para PC.

Os principais comandos HART são implementados, excetuando-se os comandos restritos para gateway e gerenciador. Através dos comandos, é possível extrair dados do estado da rede, modificar parâmetros do FDAP e programá-lo adequadamente. Através dos comandos especiais, o FDAP é utilizado pelo coprocessador para enviar comandos ede sem fio. A interface gráfica do coprocessador é apresentada na Figura 32. A interface é desenvolvida para propiciar controle e inspeção dos comandos enviados ao FDAP. Os comandos podem ser enviados de forma sequencial, de modo a facilitar a depuração das implementações. Posteriormente, com o porte do coprocessador para um MCU, a interface gráfica poderá servir como visualizador de configurações do coprocessador e como comissionador do FD, uma vez que a porta de manutenção não é mais disponível diretamente no FDAP, embora funcione como tal a uma taxa de 115,2 kbps, fora da especificação da norma (1200 bps). A tela é apresentada inicialmente na etapa de desenvolvimento. Os painéis são utilizados para verificação de mensagens, sendo o maior à esquerda utilizado para visualização das mensagens, comandos e respostas do FDAP e o menor, no topo à direita, apresenta as mensagens referentes aos resultados dos escalonamentos locais realizados. O painel central é uma tabela que informa os links programados no FDAP, e o painel menor à direita, informa a topologia em torno do FDAP para facilitar a visualização e compreensão das atividades correntes. Este painel revela a abrangência da implementação até o momento, onde são previstas a agregação de até quatro FDs e o envio direto do comando 138 para o gerenciador central. Os demais painéis contém botões e campos de dados para troca de parâmetros dos comandos, que podem ser enviados manualmente. No modo automático o coprocessador faz uso da interface, modificando os parâmetros e basicamente evocando a método *OnClick* dos botões implementados. A Figura 33 apresenta em detalhe das mensagens obtidas no processo de agregação local de um FD.

Com o intuito de facilitar a implementação da proposta de descentralização, o coprocessador é desenvolvido para permitir a inicialização da rede de forma independente de um gerenciador central. Esta característica, além de facilitar o desenvolvimento do trabalho, permite a realização de estudos na direção da descentralização total. O processo de

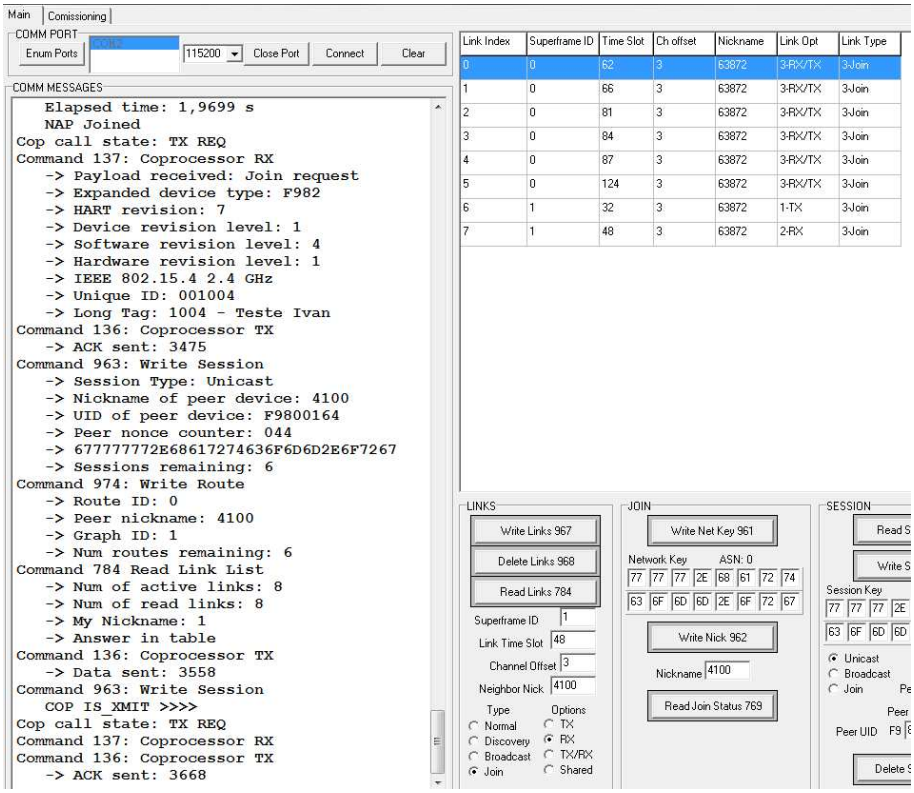

Figura 33: Detalhe da interface gráfica do coprocessador mostrando o painel de mensagens coletado após o processo de agregação de um FD localmente.

inicialização local é explicado na subseção a seguir.

5.4.1 Inicialização local de rede

A inicialização de rede de forma local é feita inicialmente através da interpretação do comando 769 (*Read Join Status*), que permite a obtenção do estado atual de agregação do próprio FDAP. Uma vez lidas as *flags* que indicam o estado, pode-se tomar a decisão de iniciar a rede localmente ou não, sendo esta característica configurável. As *flags* são ajustadas pela máquina de estados da pilha WH no FDAP e são definidas a seguir:

a) modo de operação, que fornece o estado geral do transceptor, e que pode ser: *idle, active mode, negotiating, quarentined, operational, suspended* e *deep sleep*;

b) modo de agregação, que fornece o estado geral do dispositivo de campo em relação à rede, e que pode ser: *network packets heard, ASN acquired, sinchronized to time slots, advertisement heard, join requested, join retrying, join failed, authenticated, network joined, negotiating properties* e *normal operation started*.

O estado de agregação do dispositivo é obtido pela interpretação do comando 769 e posteriormente, a máquina de estados de inicialização da rede é ativada. A troca de mensagens pode ser compreendida pelo diagrama de sequências da Figura 34.

Com o término da programação a partir do envio dos comandos de temporização (793 e 795), o FDAP inicia a propagação dos anúncios da rede que poderão ser utilizados para agregação de novos FDs localmente. A partir deste ponto, o FDAP pode ser configurado para atuar como FD comum ou como um FDAP que possui coprocessador incorporado.

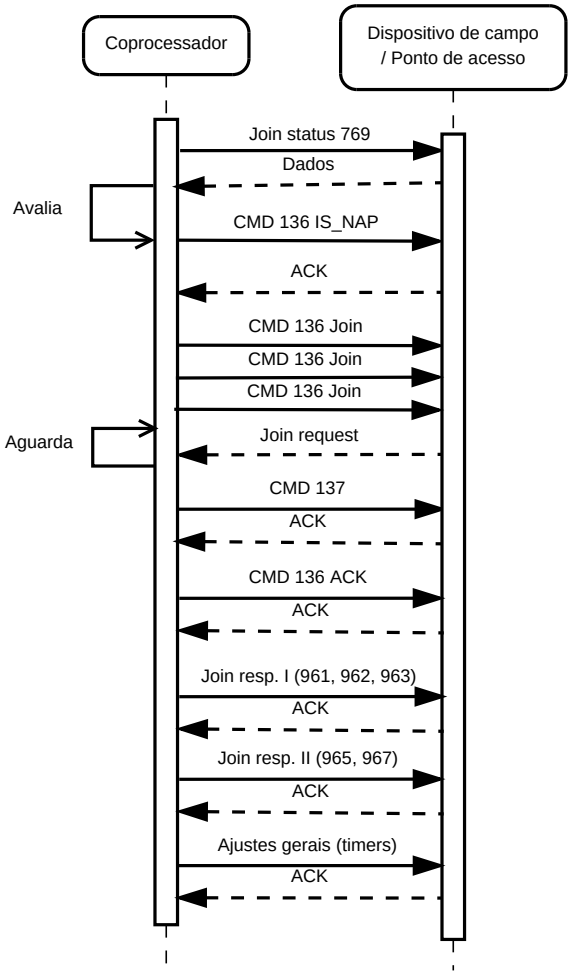

Figura 34: Diagrama de sequência do processo de inicialização local de rede.

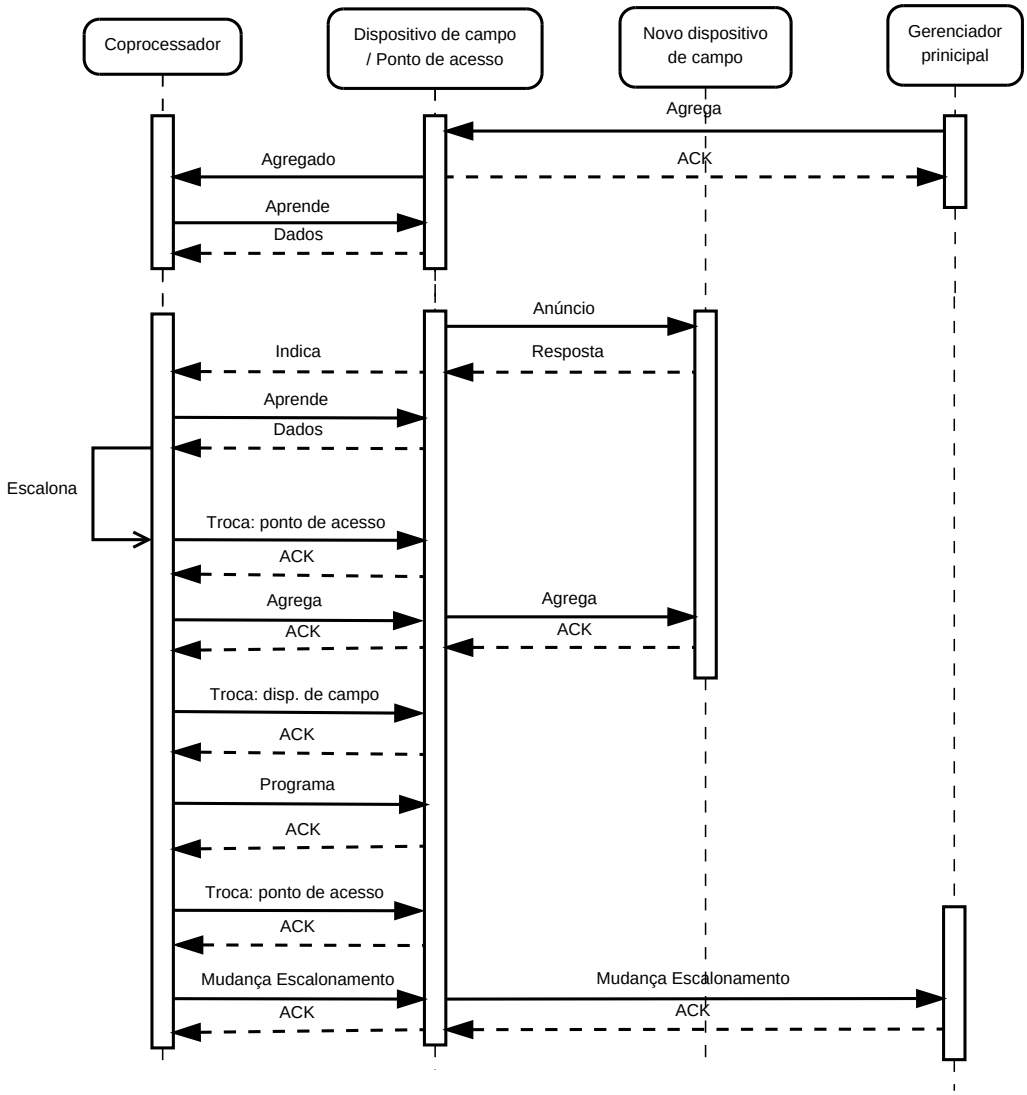

Figura 35: Diagrama de sequência do processo de agregação de um FDAP localmente, através de um coprocessador.

5.4.2 Gerenciamento distribuído a partir de uma rede já existente

O estágio final de desenvolvimento deste trabalho engloba a integração do FDAP e coprocessador com FDs comuns, com as unidades centralizadoras e com o gerenciador principal modificado. Neste último caso, a rede não é inicializada localmente, mas sim pelo gerenciador principal modificado. Esta arquitetura é a que menos transgride a norma do protocolo. O diagrama de sequência da Figura 35 indica o processo para agregação de um novo FD através de um FDAP em uma rede já formada.

As atividades de rede são iniciadas pelo gerenciador principal, que, após a inicialização do gateway, realiza o provisionamento do ponto de acesso que passa a propagar mensagens de anúncio. O FDAP é agregado a partir da resposta à requisição de agregação recebida e envia uma chamada ao coprocessador para que este tome conhecimento do escalonamento realizado. A partir deste ponto, o FDAP já envia suas próprias mensagens de anúncio, que eventualmente são recebidas por um novo FD. A resposta à requisição de agregação feita pelo novo FD aciona novamente a chamada ao coprocessador para que este aprenda novamente o estado de escalonamento da rede, uma vez que com o passar do

tempo, a rede pode ter sido modificada pelo gerenciador central. Neste ponto, o coprocessador escalona os links para o novo FD e os envia através do FDAP em modo AP. O FDAP também é reprogramado conforme a necessidade, uma vez que links devem ser estabelecidos com o novo FD. Por fim, o comando 138 é enviado ao gerenciador central através da rota que o FDAP já mantém com ele, para que tome conhecimento do escalonamento realizado para o novo FD.

5.4.3 Escalonador para descentralização da rede sem fio

O desenvolvimento do gerenciamento descentralizado requer a implementação de um escalonador TDMA local, para que este possa fornecer os links para os FDs que queiram ingressar na rede através dele. Neste contexto, dois tipos de escalonadores são necessários, conforme a abordagem de descentralização empregada. Os dois tipos de escalonadores são elaborados e denominados tipos I e II.

Um **escalonador tipo I** é empregado para o caso da descentralização parcial, baseando-se na arquitetura híbrida, com uso do gerenciador central modificado. Considera-se a ilustração de uma rede conforme apresentada na Figura 36. Neste caso, a rede possui três níveis e é previamente formada pelo gerenciador central e pelo FDAP, que encontra-se no nível dois. Todos os FDs que necessitam ser agregados à rede estão no nível dois, e desta forma o escalonamento é feito pelo coprocessador ligado ao FDAP. Os demais nós da rede, incluindo o FDAP já estão escalonados pelo gerenciador central.

O esquema de reuso de *slots*, conforme explanado na seção 5.3, é empregado de modo que os links utilizados para início de transação com os novos FDs são os links de anúncio, previamente fornecidos pelo gerenciador central. Inicialmente, no estágio II da resposta ao pedido de agregação (comandos 965 e 967), os links de anúncio são convertidos em links normais a fim de escalonar as mensagens para o novo FD. Posteriormente, os links são reconvertidos para links de anúncio, uma vez que novos links normais são atribuídos ao novo FD. Uma vez que o novo FD é programado, o comando 138 é enviado ao gerenciador central através dos links que já possui com este. Com o passar do tempo, o gerenciador pode propiciar novos links de anúncio ao FDAP para garantir mais recursos para que novos FDs possam ser agregados através do coprocessador.

O escalonador tipo I diferencia-se ainda pela necessidade de aprendizado da rede, fazendo uso os dados obtidos a partir da interpretação dos comandos que revelam como foi feito o escalonamento do FDAP.

Um **escalonador tipo II** é considerado para o caso da descentralização total, onde o elemento gerenciador centralizado é suprimido e as entidades centralizadoras são o ponto de acesso e gateway somente. A análise é feita considerando-se novamente a Figura 36, porém, desconsiderando-se o gerenciador central. Este escalonador difere do tipo I na medida em que o caminho para o ponto de acesso deve ser descoberto, para que haja fluxo de dados com o meio externo à rede sem fio. Neste caso, há necessidade do emprego de um coprocessador ligado ao ponto de acesso, e este deverá reportar a presença do gateway para que o apelido dado ao ponto de acesso seja o número um de modo a sinalizar o ponto de entrada e saída da rede. Em princípio não há necessidade de que a rede a ser formada comece pelo gerenciador conectado ao ponto de acesso, mas nesse caso, a comunicação com o meio exterior não ocorre. Uma vez que o ponto de acesso é agregado à rede, links que o incluem devem ser programados nos outros nós para que as partes distribuídas da rede encontrem caminhos para ele.

O tipo de escalonador considerado mais adequado é obtido através do estudo das diversas técnicas já desenvolvidas. A referência (DJUKIC, 2008) é utilizada por sintetizar

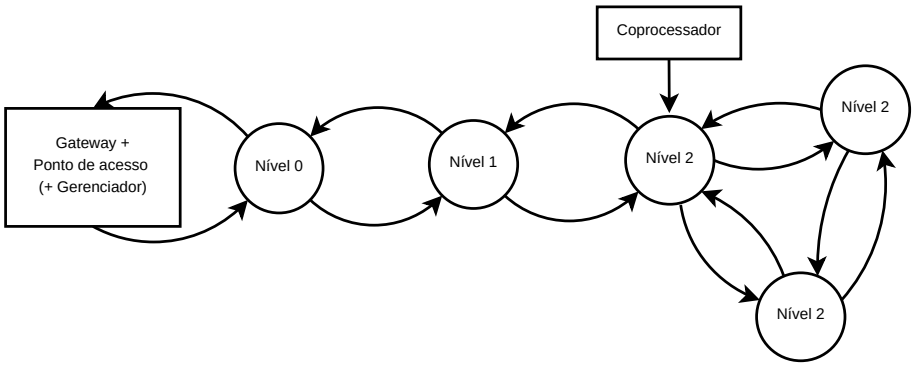

Figura 36: Ilustração de uma rede para análise da estratégia de escalonamento empregada para distribuição de gerenciamento.

os principais algoritmos, pois aborda especificamente os tipos de escalonadores desenvolvidos para redes sem fio tipo TDMA. Embora o estudo de caso não seja o protocolo WH mas sim em 802.11 e 802.16, as técnicas ali apresentadas são extensíveis ao WH. O desenvolvimento de um novo escalonador não é o foco deste trabalho, mas sim o emprego de algum tipo consolidado para validação da proposta, e que refira-se especificamente à arquitetura e técnicas para descentralização. Porém, com o desenvolvimento do sistema distribuído, diversas técnicas de escalonamento podem ser futuramente estudadas e comparadas na prática. Para a prova de conceito, o algoritmo Bellman-Ford (apresentado na seção 2.3.1) é escolhido para roteamento em conjunto com técnicas de escalonamento.

O algoritmo deve ser alimentado com os pesos das arestas do grafo. Inicialmente são escolhidos os seguintes parâmetros:

a) Intensidade do sinal recebido (RSL). O RSL é obtido a partir do indicador de qualidade do link, fornecido por serviço da MAC 802.15.4. O byte lido tem a seguinte correspondência: 0x00 = -15 dBm, 0xFF = -100 dBm. Considera-se que enlaces melhores que -30 dBm são ótimos e que piores que -85 dBm sejam péssimos (a norma 802.15.4 exige que um transceptor compatível tenha sensibilidade melhor que -85 dBm). Divide-se as intensidades de sinal em seis faixas, cada qual com sua valoração conforme apresentado na Tabela 5.4.3 apresenta. O transceptor utilizado (MC13224) tem sensibilidade garantida de -96 dBm.

b) Alimentação. O comando 777 (*Read Wireless Device Capabilities*) retorna uma série de características do FD, entre elas, o tipo de alimentação do dispositivo (0 = alimentação pela linha, 1 = alimentação por bateria, 2 = bateria recarregável ou alguma forma de coleta de energia). Para cada um dos valores, é feita uma soma para composição do peso final da rota (vide Tabela).

c) Fonte de relógio. Embora não seja uma obrigatoriedade, se um vizinho é fonte de relógio ele deve ser preferencialmente utilizado como elemento roteador em detrimento de outro que não seja fonte. Isto permite uma melhor disseminação de relógio pela rede, garantindo sincronismo entre os nós. A vinculação como fonte de relógio ou não é feita pelo comando 971 (*Write Neighbor Properties*) e obtida pelo comando 786 (*Read Neighbor Property Flag*).

Objetiva-se escolher a melhor rota em função das capacidades dos dispositivos que as compõem. Uma variável indicando o peso da rota é inicializada com dois e tem seu valor incrementado ou decrementado de acordo com os parâmetros da tabela.

Tabela 5: Critérios para valoração das rotas para alimentação do algoritmo escalonador.

Parâmetro	Peso
RSL entre -15 e -29 dBm	+4
RSL entre -30 e -44 dBm	+3
RSL entre -45 e -59 dBm	+2
RSL entre -60 e -74 dBm	+1
RSL entre -75 e -84 dBm	0
RSL Menor que -85 dBm	-1
Fonte de relógio	+2
Alimentação linha	+2
Alimentação bateria	+1
Alimentação alternativa	-1

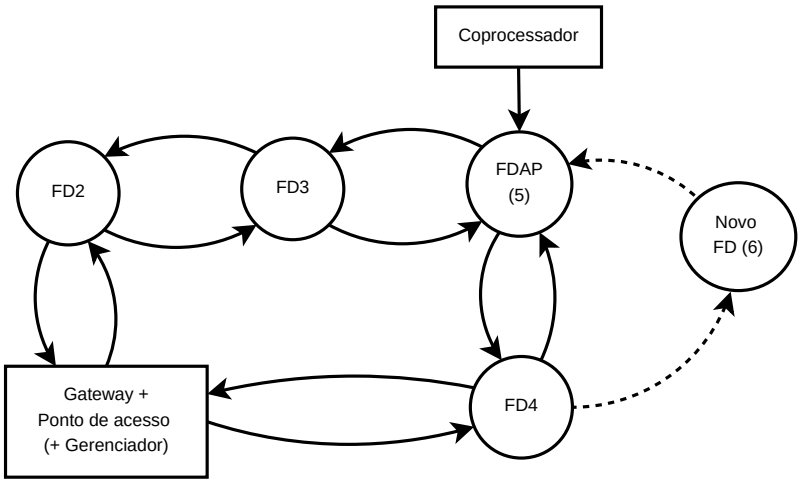

Figura 37: Estabelecimento de uma nova rota para o FD numa rede hipotética (I).

Como um exemplo para entendimento do processo de escalonamento que o algoritmo deve realizar no caso em que há um gerenciador central, considera-se a Figura 37. A Figura apresenta uma rede já estabelecida com quatro nós, sendo um deles um FDAP e FD3 (localizados no nível 1) juntamente com FD2 e FD4 (localizados no nível zero). Por hipótese, FD2 e FD4 têm periodicidade de uma unidade e FD3, duas. Um novo FD quer agregar-se à rede e de acordo com o critério de prioridade para agregação, responde preferencialmente ao anúncio feito pelo FDAP.

Na Figura 38 a rede alterada já com a inclusão do novo FD é apresentada, onde as rotas foram definidas pelo menor custo, ou seja, através de FD4. Considera-se que o FDAP não é um sensor nem atuador e, portanto, não possui mensagens periódicas para publicação. Para fins de ilustração, o novo FD requer periodicidade de duas unidades.

A Figura 39 apresenta um escalonamento para esta rede hipotética, onde três *superframes* são definidos, sendo SF1 e SF2, os *superframes* com as publicações periódicas e SF3, para eventos assíncronos. As setas verticais indicam a repetibilidade dos *superframes*, a cada 10 e 20 *slots* de tempo, para os *superframes* designados para publicação de variáveis de processo. O SF3 tem tamanho maior ou igual que SF2. Em azul, os links pertencentes ao grafo de *downlink*, em amarelo, *uplink* e em verde, links para anúncio. Para efeito de ilustração, o tamanho do *superframe* é menor do que o mínimo regulamentado pela norma, que define periodicidades mínimas de 250 ms. Na prática, *superframes* de no

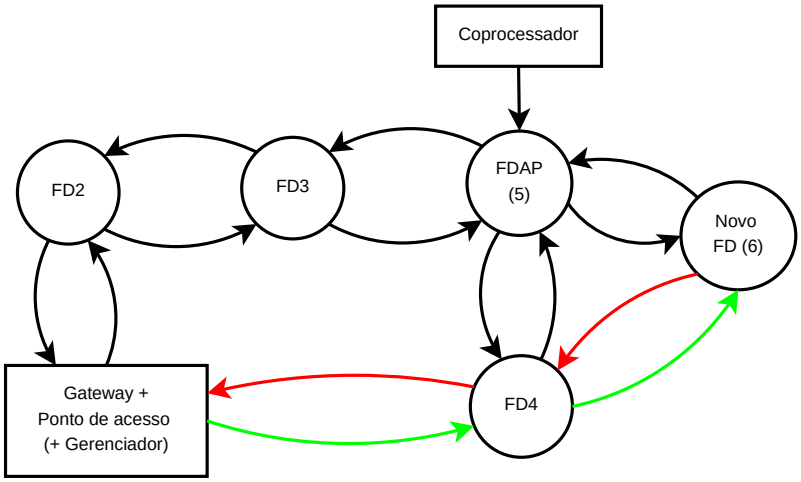

Figura 38: Estabelecimento de uma nova rota para o FD numa rede hipotética (II).

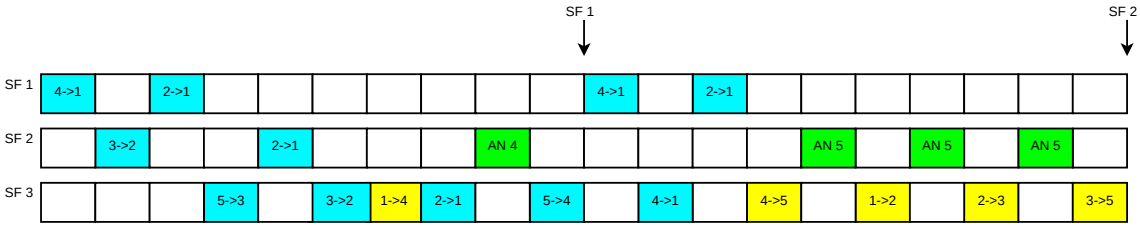

Figura 39: *Superframes* para exemplificar a rede apresentada na Figura 37.

mínimo 100 *slots* são utilizados.

O resultado do escalonamento nos *superframes* para inclusão do novo FD pode ser observado na Figura 40. Os links de anúncio, primariamente designados ao FDAP, são convertidos em links normais para o novo FD. Na sequência do SF3, novos links de anúncio poderão ser designados ao FDAP, até o limite de dispositivos que possam ser agregados à rede através dele.

O resultado do algoritmo de escalonamento é a inclusão do novo FD, com rota ótima até o gateway através de FD, e não pelo próprio dispositivo que escalonou as suas comunicações. Na sequência dos *superframes*, os links para os demais dispositivos podem ser estabelecidos, com o intuito de aumentar a redundância de caminhos e, por consequência, a confiabilidade da rede como um todo. O processo de escalonamento da rede a partir de um FDAP e coprocessador é apresentado a seguir.

1: *Aprendizado*: FDAP envia comandos para os vizinhos (FD3 e FD4) para descoberta de rotas e links já estabelecidos. Também, o próprio escalonamento é consultado, para

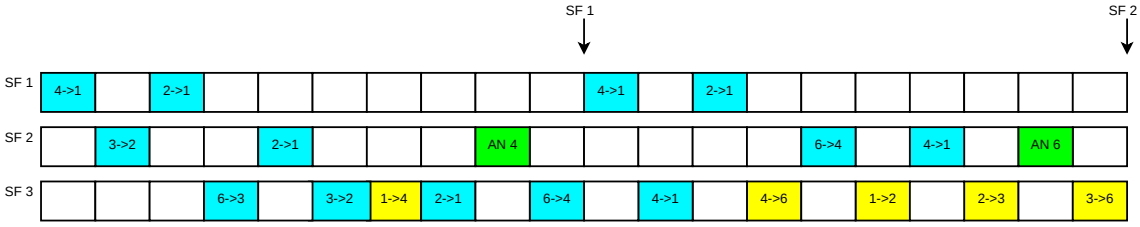

Figura 40: Resultado do escalonamento nos *superframes* para inclusão do novo FD.

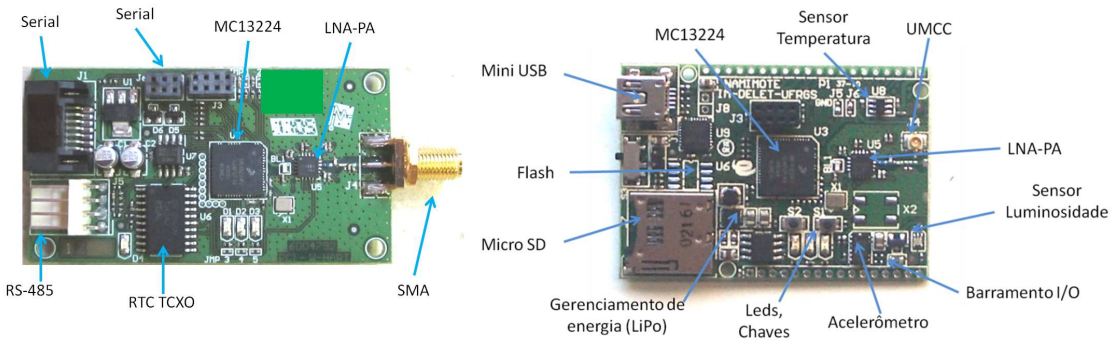

Figura 41: Transceptores WH desenvolvidos neste trabalho: à esquerda, placa adaptadora para redes WH e à direita, o Namimote

que se encontrem as rotas existentes do FDAP até o gateway.

2: *Roteamento*: o algoritmo Bellman-Ford, ou outra variante é utilizado para a definição de rotas mais valoradas, que serão preferencialmente utilizadas para agregar o novo FD. O novo FD reporta FD4 como um vizinho de bom RSL, através da requisição de agregação (comando 787).

3: *Escalonamento I*: uma vez encontrada a melhor rota até o gateway, procede-se ao escalonamento. São inicialmente utilizados os links de anúncio, para colocar o novo FD em estado agregado.

4: *Escalonamento II*: os links de anúncio são convertidos em links normais. Como o novo FD requer serviço de publicação (de periodicidade dois, no exemplo), são designados links no SF2. No SF3, restam os *slots* para *downlink* e aperiódicos. Eventualmente, outros links são utilizados para anúncio pelo FDAP, até o limite de dispositivos para aquele espaço de rede.

5.5 Hardware para descentralização

O hardware necessário para a proposta de descentralização envolve o desenvolvimento de transceptores compatíveis com o WH, de um ponto de acesso WH e do coprocessador para gerenciamento local, que inicialmente é desenvolvido em plataforma PC.

A Figura 41 apresenta dois transceptores compatíveis com o WH desenvolvidos ao longo deste trabalho. O dispositivo apresentado à esquerda da Figura foi concebido para ser uma placa adaptadora WH para atuadores, sem sensores integrados. A alimentação é obtida diretamente do atuador e desta forma não há gerenciamento de energia eficiente para uso com bateria, nem sistema de recarga para a mesma. Uma descrição completa dos requisitos para desenvolvimento de um transceptor compatível com o WH pode ser obtida em (MULLER et al., 2010), que originou o desenvolvimento desta placa adaptadora de rede. O transceptor da direita da Figura foi elaborado para utilização em redes de sensores sem fio e portanto, é primariamente alimentado à bateria. Ele possui sensores integrados (temperatura, luminosidade e aceleração em três eixos), unidades de memória para operação como coletor de dados e gerenciamento de energia para uma bateria de lítio-íon de 3,7 Volts. O espaço de placa é de 38 x 55 mm. Uma descrição mais detalhada deste nó para RSSF (denominado Namimote) pode ser obtida em (MULLER et al., 2012).

Qualquer um dos dispositivos anteriormente descritos pode ser utilizado em conjunto

Tabela 6: Estatísticas de implementação (*excluído o código da GUI).

Dispositivo	Linhas	Processos	SO	Linguagem	Código	Dados
AP	54980	10	CMX	C	66,3 kB	37,5 kB
FD	237871	14	CMX	C	73,1 kB	46,2 kB
FDAP	238657	12	CMX	C	73,3 kB	45,3 kB
Porta manut.*	9081	1	Windows	C++	106,5 kB	32,8 kB
Coprocessador*	9508	2	Windows	C++	114,6 kB	53,3 kB

com o coprocessador, de modo a formar um FDAP. Basta que seja escolhida uma das portas de comunicação serial disponíveis (TTL e RS-485 no transceptor para o atuador, TTL e USB no Namimote). Independentemente de qual seja a escolha, uma das portas é utilizada para a comunicação com o coprocessador e a outra, para depuração através de algum tipo de programa terminal. Esta escolha é feita no momento da compilação do FDAP.

O coprocessador é implementado em um software para PC e um canal de comunicação serial é aberto e utilizado a uma taxa de 115,2 kbps. A implementação do coprocessador em hardware dedicado poderá ser feita de forma simples e rápida, uma vez que todas as funções são escritas em ANSI C, exceto as relativas à interface gráfica com o usuário. O futuro porte do código do coprocessador, que poderá ser feito para um MCU dedicado ou como um processo da camada de aplicação da pilha do protocolo, é descrito na seção seguinte.

5.6 Software para descentralização

Os trabalhos realizados durante a implementação desta tese envolvem o desenvolvimento de software para diferentes plataformas (PC e MCU). O hardware utilizado nos dispositivos de campo e o FDAP emprega o MC13224 da Freescale . As estatísticas de software são apresentadas na Tabela 6, objetivando apresentar a abrangência da implementação e comparar o acréscimo devido ao FDAP e o coprocessador. Parte dos códigos utilizados é herdada de desenvolvimentos anteriores.

No coprocessador, em comparação com o software da porta de manutenção anteriormente desenvolvido, foram implementados todos os comandos de escrita (ou manipulação), necessários para programação do FDAP e do FD associado à ele durante o processo de agregação. Aliado a isto, os códigos das máquinas de estado do coprocessador são implementadas de forma a possibilitar as diferentes arquiteturas distribuídas.

No FDAP, são alteradas as partes do código do FD (pilha WH) referentes às camadas de rede e dados. Também são incluídos os comandos especias e manipulação da interface de comunicação serial para troca de dados como o coprocessador.

São considerados códigos totais, incluindo arquivos de cabeçalho e de configuração. Para obtenção das contagens de linhas, é utilizado um software comercial que apresenta estatísticas gerais, e para obtenção de *footprint* de memória, os resultados dos processos de compilação do próprios IDEs.

Das estatísticas apresentadas na Tabela 6, as que demandam mais atenção são as relativas ao coprocessador, que uma vez completamente operacional poderá ser portado para um MCU (hardware dedicado) ou como um processo (na camada de aplicação da pilha do protocolo).

6 ESTUDO DE CASO

6.1 Introdução

Neste capítulo são apresentados os estudos de caso realizados para validação desta tese. As avaliações de desempenho são feita através de medições de tempo realizadas no software do coprocessador e através do Wi-Analys, que produz marcas de tempo a cada amostra coletada, que são utilizadas para comparações.

6.2 Verificação de funcionalidade

Com o objetivo de validar a proposta apresentada nesta tese, são feitas comparações entre três equipamentos, dois comerciais, das marcas Emerson Rosemount modelo 1420 e Nivis Versa Router 810, e um protótipo desenvolvido ao longo da tese. A métrica empregada é a medida do tempo necessário para a agregação de um dispositivo, considerando o envio da requisição pelo FD e a última resposta, com os comandos de escalonamento. Este procedimento é regulamentado pela norma e consiste nos passos para evolução da variável interna de um FD que indica o estado de agregação, conforme explicado anteriormente. Os FDs utilizados nos testes são os transceptores desenvolvidos anteriormente, descritos na seção 5.5.

Em função da complexidade e abrangência da implementação da proposta, são feitas algumas restrições, que eventualmente serão suplantadas em desenvolvimentos futuros. O cerceamento do estudo de caso leva aos cenários de testes descritos a seguir:

a) A rede é inicializada pelo FDAP, por motivo de simplicidade e controle geral.

b) O algoritmo escalonador distribuído é implementado e testado. A técnica empregada para escalonamento distribuído é baseada no reuso de *slots*.

c) Como o gerenciador centralizado não é utilizado, o término do escalonamento é definido pela captura do comando 138, enviado pelo FDAP.

d) Os tempos de envio do comando 138 devidos a saltos sucessivos até o gerenciador principal (nos casos em que o FDAP está localizado em níveis maiores que zero) são obtidos pela extrapolação de outros testes realizados em redes convencionais. Considera-se uma aproximação válida, pois o comando 138 é enviado normalmente pelo FDAP e é um evento de comunicação de tipo comando, de alta prioridade no WH. Os envios e recepções de comandos são corriqueiramente observados em laboratório através do *sniffer* e seus tempos foram extensamente medidos em trabalhos anteriormente realizados.

e) O esquema de segurança de rede é simplificado, utilizando-se a mesma chave para agregação, enlace ponto a ponto e sessões. A simplificação é devida à não necessidade da segurança completa para prova de conceito e pela facilidade na implementação do có-

digo. Por outro lado, devido ao fato de que o esquema de segurança no WH é um dos seus grandes atrativos, não são negligenciados os aspectos fundamentais: todas as mensagens propagadas pela antena são criptografadas (carga útil e verificação de mensagem), com as sementes geradas da forma correta (utilizando as chaves e *nonce*), que previnem ataques de repetição e descoberta de conteúdo. Porém, como a geração local de chaves é um problema da descentralização, uma solução para futura implementação é aqui apresentada.

As chaves utilizadas para criptografia no WH são quatro: a chave de agregação, a chave "bem conhecida", a chave de rede e a chave de sessão. A chave de agregação é previamente programada nos elementos da rede, nos dispositivos de campo através das portas de manutenção e na GUI do gerenciador de rede. A chave bem conhecida é utilizada inicialmente para verificação da integridade das mensagens trocadas com um novo FD, pois este não possui uma chave de rede. Posteriormente, a chave de rede é passada ao FD através do comando 961 no processo de agregação. O mesmo ocorre com a chave de sessão. O problema da descentralização está justamente na geração das chaves de rede e sessão de forma distribuída. Numa rede centralizada, isto não consiste um problema, uma vez que todas as chaves são criadas no gerenciador de segurança central, que as mantém numa tabela, relacionando-a com o dispositivo afim. Na rede descentralizada, as chaves de sessão e rede são criadas localmente pelo gerenciador local, e desta forma, poderão ser criadas em duplicidade. A segurança esperada em uma rede WH é de que exista uma chave de sessão para cada comunicação fim a fim e uma chave de rede para cada enlace ponto a ponto. Considerando o exemplo de uma rede com 50 nós, em que existam 100 sessões (50 entre os nós e o gerenciador de rede e 50 entre os nós e o gateway) e cerca de 250 a 300 links ponto a ponto, será necessário que o gerenciador de rede crie até 400 chaves diferentes. Na prática, a chave de rede pode ser única para toda a rede pois o contador *nonce* garante exclusividade na semente de encriptação, embora isto reduza o nível final de segurança. As chaves são números de 16 bytes, e desta forma, dificilmente serão gerados números repetidos após 400 criações aleatórias (0,61% de chance). Ainda assim, propõe-se o seguinte esquema para a geração local de chaves:

a) Utiliza-se um gerador de número aleatório, baseado no algoritmo LFSR (*Linear Feedback Shift Register*), por exemplo.

b) A semente utilizada é o próprio UID, que não pode ser repetido numa rede.

c) As chaves são geradas localmente pelos gerenciadores distribuídos, no momento em que coordenam um processo de agregação.

d) Todos os gerenciadores devem implementar o esquema, que não falha uma vez que um novo FD é agregado à rede através de um único gerenciador.

e) Para que a chave de rede seja diferenciada da chave de sessão criada, é feita uma operação aritmética única na semente, que irá produzir uma nova sequência aleatória.

Mesmo que o algoritmo produza uma sequência diferente, existe a chance de que a chave de rede produzida seja igual à outra já existente (de sessão), em função da simples operação aritmética realizada. Isto não constitui uma falha de segurança na medida em que os conjuntos de chaves serão únicos mesmo que apresentem interseção não nula.

A funcionalidade do protótipo é verificada pelos resultados das capturas feitas com o *sniffer* Wi-Analys, através do monitoramento dos dispositivos pelas portas de depuração e através da GUI do coprocessador. A Figura 42 apresenta a tela do *sniffer*, capturada no momento da inicialização da rede feita pelo coprocessador, onde é possível visualizar os pacotes de anúncio, a chegada da requisição de agregação e o envio do comando 963.

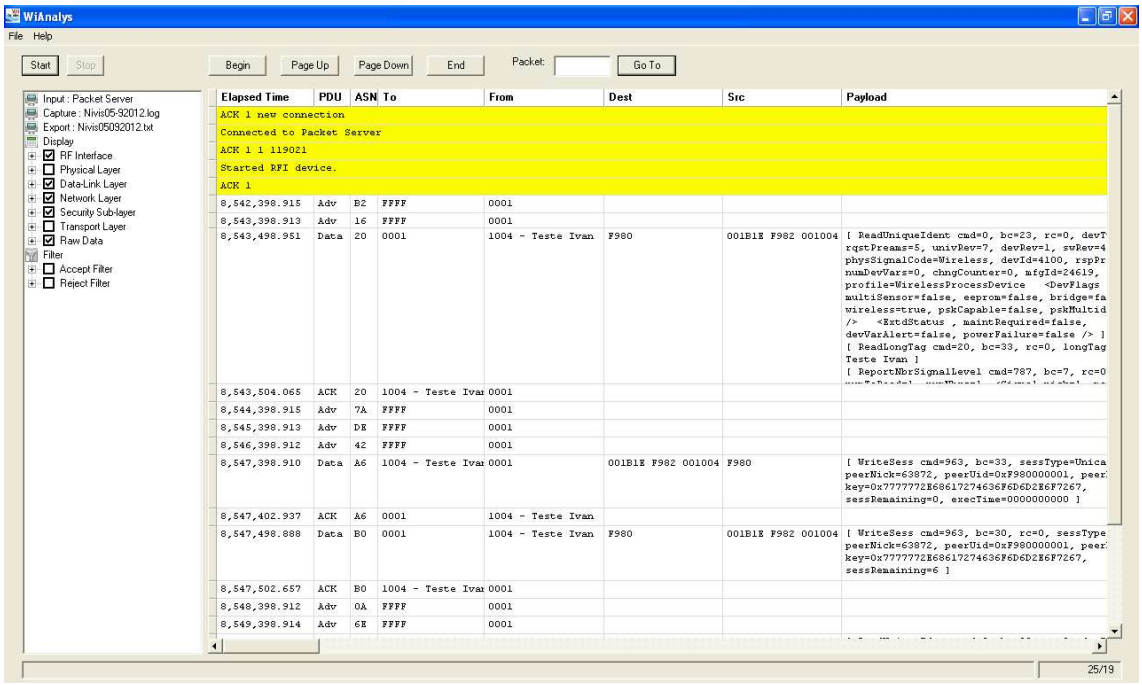

Figura 42: Tela capturada da GUI do *sniffer* Wi-Analys, onde é possível visualizar o funcionamento do FDAP realizando comunicação direta com um novo FD sem a intervenção de um gerenciador central.

6.3 Comparativos

Os dados para comparações são obtidos pelo *sniffer* Wi-Analys, utilizado em conjunto com as ferramentas desenvolvidas ao longo da tese. O próprio software do coprocessador também realiza medições de tempo através do relógio do PC. São utilizados os temporizadores UTC (*Coordinated Universal Time*) e de geração de ASN. Outros dados para comparação são obtidos pelo software de análise de roteamento de redes WH (WINTER et al., 2011). Este software comunica com o gateway WH através do protocolo HART-*over*-UDP e permite o envio de comandos diretamente ao gateway. Com o software, é possível realizar a análise posterior da topologia de rede e inferir quais dispositivos estão realizando comunicação em saltos múltiplos e em que tempo as mensagens foram enviadas e recebidas pelo gerenciador / gateway.

Os resultados obtidos podem ser visualizados na Figura 43. A Figura apresenta o tempo médio de agregação obtido em três cenários, compostos pelas entidades centralizadoras e um dispositivo de campo WH. São avaliados os tempos de agregação nos níveis zero, um e dois. Os equipamentos gerenciadores de rede empregados são um Emerson Rosemount modelo 1420, um Nivis VersaRouter 810 e o protótipo composto pelo ponto de acesso, gateway e script de testes, desenvolvidos ao longo desta tese. Este último, atualmente ainda não é capaz de realizar agregação via proxy. Ainda assim, o tempo para agregação em nível um é obtido pela medida do tempo de envio da mensagem de requisição e chegada até o ponto de acesso.

Os dados referentes ao FDAP (I) dizem respeito ao tempo transcorrido entre a recepção da requisição de agregação (comandos 0, 20 e 787) e o último comando de escalonamento, que coloca o FD no estado *joined*. Para o FDAP este tempo é sempre referente ao nível zero, ou seja, são considerados os FDs que se agregam à rede através dele por proxy. Como a agregação é feita no nível zero, os tempos se referem à comunicações

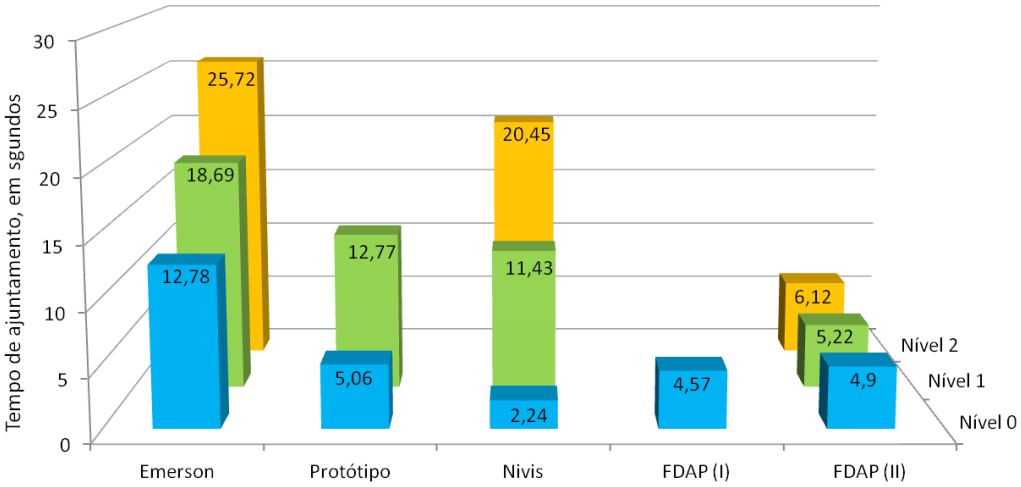

Figura 43: Comparações de tempos entre os diferentes dispositivos WH.

diretas, medidas no software do coprocessador. Cabe aqui ressaltar que há vários melhoramentos possíveis, que podem reduzir os tempos obtidos. Um deles, diz respeito aos comandos enviados pelo FDAP ao novo FD, que são feitos individualmente, sem agrupamento. Isto ocorre atualmente por motivo de simplificação, devido ao uso do mecanismo de comunicação HART. Na sequência do desenvolvimento do FDAP e coprocessador, o agrupamento de comandos irá reduzir drasticamente os tempos envolvidos.

Os tempos referentes ao FDAP (II) levam em consideração o envio do comando 138 ao gerenciador central, e portanto, devem ser considerados como métrica para avaliar a proposta de descentralização. Estudos realizados recentemente, avaliam as latências comumente observadas no envio de um pacote através dos diversos saltos entre os nós de uma rede WH. Estes valores são altamente dependentes das estratégias de escalonamento adotadas pelo gerenciador. Os requisitos de banda, feitos por FDs que publicam dados (sensores, basicamente), são muito diferentes de atuadores, exceto nos casos em que o monitoramento é requisito fundamental. Avaliando-se os tempos propiciados pelos links corriqueiramente fornecidos para dispositivos do tipo sensor, que publicam suas variáveis de processo a cada minuto, por exemplo, chega-se aos tempos apresentados na Figura 43 (FDAP II). Estes tempos referem-se ao estabelecimentos dos links físicos e lógicos do novo FD com o gerenciador central, e em implementações futuras, com o FDAP que está conectado ao gateway num sistema totalmente distribuído. Novamente, ressalta-se que estes tempos podem ainda ser em muito reduzidos, especialmente com o emprego da técnica apresentada na seção 5.3.

A topologia gerada no estudo de caso é apresentada na Figura 44. A imagem é obtida da GUI do coprocessador. O dispositivo FDAP tem apelido e 0x7B11 e possui links com quatro FDs que são agregados à rede localmente através dele. Nos testes, não são criados links entre os FDs e também com o gerenciador central. Estes links poderão ser criados ao longo do tempo da rede, em implementações futuras. A seta que representa a comunicação entre o FDAP e o coprocessador está invisível na figura pois o escalonamento já foi realizado, levando ao estado de dormência do coprocessador.

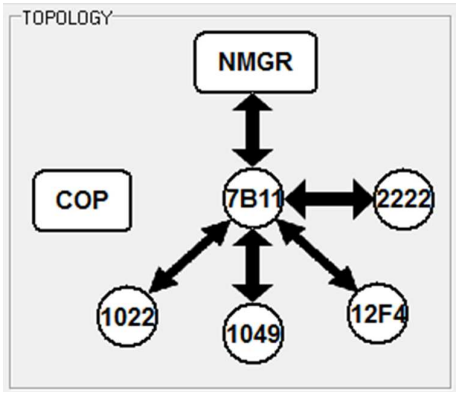

Figura 44: Topologia gerada no estudo de caso de gerenciamento distribuído.

Tabela 7: Tempos para início de propagação de anúncios (desconsiderando inicialização do *Linux e **Windows).

Nivis	Emerson	Protótipo*	Coprocessador**
71 s	52,78 s	32,64 s	2,35 s

6.4 Contribuições da tese

No intuito de melhorar a dinamicidade de redes WH são estudadas e desenvolvidas outras técnicas não restritas somente aos tempos de inicialização de rede. Uma das abordagens envolve o tempo de resposta ao anúncio de rede, que tanto foi explorado no desenvolvimento do FDAP quanto no *stack* do FD. Aliado a isto, o processo de inicialização e formação de uma rede WH foi estudado e modificado. Embora não seja uma característica fundamental em redes WH convencionais, o processo de formação de rede apresenta dinamicidade extremamente baixa, sendo verificados tempos na casa de minutos em equipamentos comerciais. Isto em princípio pode não ter grande importância, uma vez que uma rede WH é primariamente estática no que se refere à mobilidade dos nós e desta forma, é formada uma única vez, idealmente. Por outro lado, o encurtamento do tempo de formação da rede está intimamente ligado a melhorias na dinamicidade e propicia redes intermitentes, além de permitir o desenvolvimento de técnicas de *handover* para uso com FDs móveis.

6.4.1 Redução do tempo de resposta ao anúncio

Para que se tenha uma ideia dos tempos de inicialização comumente observados, a Tabela 7 apresenta os resultados médios obtidos dos equipamentos comerciais WH disponíveis. A métrica utilizada foi o tempo desde a ativação do equipamento até o primeiro anúncio de rede, detectado pelo *sniffer* Wi-Analys. Pode-se argumentar que este teste não produz de fato um termo de comparação justo, uma vez que os meandros dos processos de inicialização não são conhecidos e são altamente dependentes da implementação, devido a parâmetros tais como a inicialização do sistema operacional embarcado, transceptor WH, *scripts*, etc. Ainda assim, os tempos observados revelam a incapacidade do uso destes equipamentos para propósitos diferentes do inicial. Prevê-se que o conjunto coprocessador e FDAP apresentarão tempos na casa de poucos segundos, uma vez que não há inicialização de sistema operacional envolvida.

No FDAP, a alteração feita para redução do tempo de agregação é obtida através do envio de pacotes de anúncio falsos, mas sincronizados no tempo, o que leva à agregação

Tabela 8: Tempos de agregação do FDAP, medidos em segundos.

Medida	FDAP 1 vizinho	FDAP 3 vizinhos
Valor médio	31,04303	1,48224
Valor máximo	31,5376	1,7557
Valor mínimo	30,9531	1,3725
Desvio padrão	0,13422	0,07758

Tabela 9: Tempos de agregação do FD, medidos em segundos.

Medida	FD 1 vizinho	FD 3 vizinhos, 1 link	FD 3 vizinhos, 2 links
Valor médio	32,78	13,1	13,6
Valor máximo	38,1	25,8	52,4
Valor mínimo	26,9	4,8	1,2
Desvio padrão	4,41	8	21,83

com o coprocessador de forma muito rápida. Três pacotes são necessários, e a resposta ao vizinho correto (o coprocessador) é possível pela atribuição dos níveis de RSL dos vizinhos “falsos”. A Tabela 8 apresenta a média dos tempos obtidos após a coleta de dez amostras. A redução de tempo é significativa e permite o estudo de formação de redes WH totalmente descentralizadas, a partir da iniciativa dos coprocessadores ligados aos FDAPs.

Na pilha WH do FD constata-se que uma simples alteração permite o rápido envio da requisição de agregação, após a recepção de uma única mensagem de anúncio, conforme a Tabela 9. O número necessário de mensagens é reduzido a uma somente, provocando a mudança de estado no FD para envio imediato dos comandos 0, 20 e 787. Embora muito mais eficiente em termos de velocidade de resposta (como mostra a Tabela 9), esta alteração pode levar a situações de sincronismo entre os dispositivos de comunicação imediata. Mas o maior problema é devido ao fato de que o gerenciador central necessita saber quem são os vizinhos do novo FD para que possa estabelecer os links e as rotas para agregação do mesmo. Por este motivo, o novo FD não requisita imediatamente a agregação e passa algum período em modo recepção na tentativa de encontrar novos vizinhos. Uma vez encontrados, ou pelo término do período de tempo, o novo FD faz a requisição de agregação e envia a lista de vizinhos através do comando 787. Com o uso do gerenciamento distribuído não há necessidade de que sejam descobertos os vizinhos através de um novo FD, uma vez que o coprocessador, através do FDAP, tem capacidade de reconhecer a rede a sua volta ao longo do tempo. Isso leva à possibilidade de redução do número de vizinhos necessários para geração da requisição de agregação a um somente, o que torna a requisição instantânea após a recepção do anúncio. Novamente, um tipo de dispositivo expandido dedicado ao FDAP poderá ser o diferenciador para o que um FD saiba através de quem está sendo agregado à rede, e neste caso, os tempos serão reduzidos.

A grande variabilidade percebida nos tempos de agregação do dispositivo de campo se deve a diversos fatores, especialmente a defasagem entre o canal de RF do transmissor e do receptor. Cabem aqui estudos mais aprofundados, entre eles o desenvolvimento de técnicas para sincronização inicial entre os rádios da rede.

Pela análise dos valores máximos e mínimos obtidos, fica claro que a grande dependência no tempo de chegada da requisição de agregação está no fato da defasagem de canais entre o dispositivo anunciante e o novo FD.

6.4.2 Coletor WH rápido

Um dispositivo para utilização de nós intermitentes, aqui denominado coletor WH rápido, é um tipo de adaptador WH modificado, para que possa atuar como um coletor de dados oriundos de um FD móvel e/ou intermitente, sem detrimento das características fundamentais do protocolo tais como segurança e tempo real. Diversas abordagens são possíveis, entre elas, o uso de um canal fixo para anúncio (tanto no coletor quanto no FD móvel), a modificação da requisição de agregação, excluindo-se o comando 787 em troca de um comando especial que informe dados fundamentais do FD móvel (entre eles, as suas variáveis de processo) e marcas de tempo relacionadas com a resposta ao anúncio e com o evento de perda de link físico, baseados no contador UTC.

Da mesma forma que um FDAP, um coletor rápido WH pode ser identificado pelo tipo de dispositivo estendido diferenciado, de modo a ser reconhecido pelo(s) gerenciador(es) de rede.

O coletor rápido permite a utilização de nós móveis e intermitentes de forma robusta e segura. Os FDAPs são modificados para poderem realizar a tarefa de coletor rápido, identificando o dispositivo intermitente e realizando o manejo dos links físico e lógico.

Como prova de conceito, uma implementação é feita com a modificação da máquina de estados do coprocessador. A máquina é alterada de modo a enviar o comando 1 (*Read Process Variable*) logo após o estabelecimento de uma sessão com o dispositivo intermitente (Comando 963).

Para a implementação de uma modalidade de coletor rápido, a sequência de comandos representada no diagrama da Figura 45 é enviada ao FDAP pelo coprocessador. O coprocessador programa uma sessão no FDAP (que está no modo FD) com o UID do FD que requisitou agregação. A seguir, o esquema de reuso de *slots* é feito, através da deleção de anúncio e criação dos links normais para comunicação com o FD intermitente. Já no modo IS_XMIT, o FDAP é utilizado pelo coprocessador para que uma sessão entre o FD intermitente e o FDAP. Desta forma, o comando 1 pode ser enviado ao FD intermitente e sua variável de processo é lida. O funcionamento do esquema para coleta de dados de um dispositivo móvel é verificado, tendo proporcionado comunicação segura, através do esquema de segurança nativo do WH, e robusta, através do escalonamento de links sem possibilidade de colisões na própria rede. Os tempos medidos para o processo de coleta ficam na casa de 1,5 segundos desde o momento da requisição de agregação feita pelo FD até a leitura da variável de processo. Novamente, estes tempos serão reduzidos a partir da implementação do agrupamento de comandos na camada de transporte do FDAP.

Além desta implementação, várias modificações podem ser feitas facilmente, uma vez que o FDAP é totalmente controlado pelo coprocessador.

O algoritmo anterior foi implementado e testado com um FD WH convencional, sem quaisquer alterações. Esta abordagem apresenta o inconveniente de que o FD nunca mudará seu estado para agregado e o processo de leitura intermitente sempre dependerá da resposta ao anúncio feito pelo FDAP. Para o desenvolvimento de um dispositivo WH dedicado ao uso intermitente, o algoritmo apresentado seguir poderá ser implementado.

a) A requisição ao anúncio feito pelo coletor rápido inclui um comando especial, e detrimento do comando 787, uma vez que o dispositivo intermitente não irá de fato ser agregado à rede. Este comando inclui as características do dispositivo além das variáveis de processo.

b) O FDAP registra uma marca temporal quando a requisição é feita.

c) O FDAP envia os comandos 961, 962 e 963, modificando o estado de agregação do

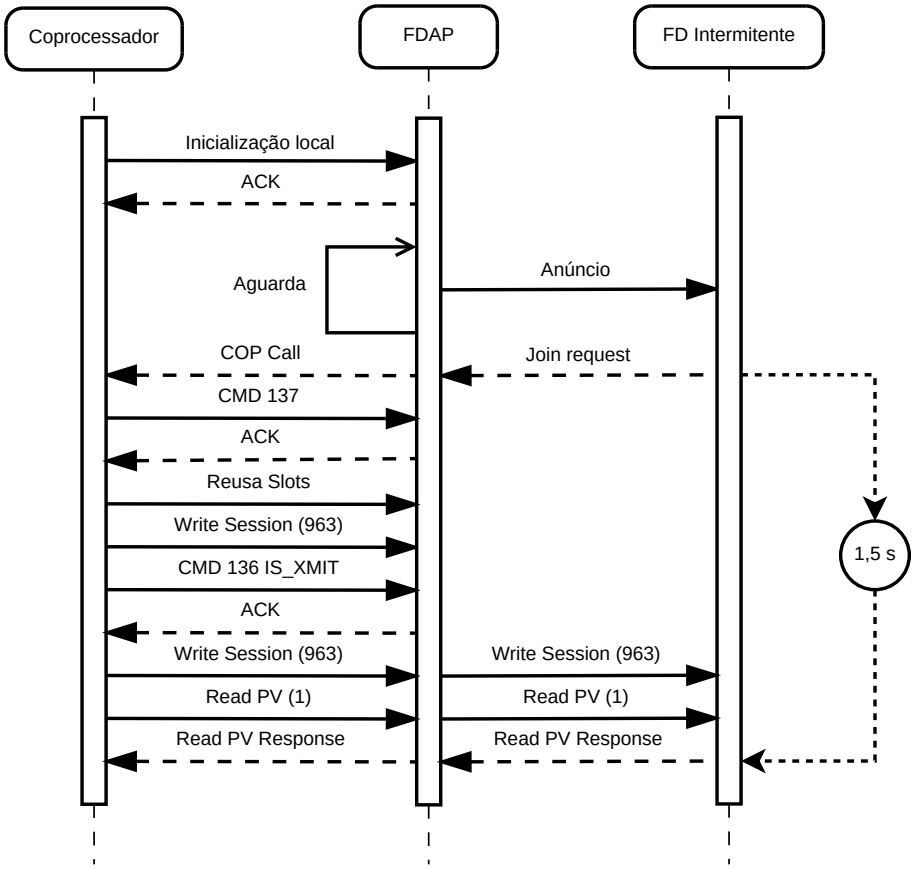

Figura 45: Diagrama de sequência do processo de inicialização local de rede.

dispositivo intermitente.

d) É enviado um comando de escrita de *superframe* e links normais. A periodicidade é mínima, para permitir o rastreamento do dispositivo móvel.

e) O FDAP envia o comando 795 (Write Timers, com parâmetro de ajuste do *Heath report*) para monitoramento do RSL. Quando o RSL atinge o limiar de boa conexão, o comando 960 é enviado, para desconectar o dispositivo móvel.

f) No momento da desconexão, uma marca temporal é salva.

g) A qualquer momento e de posse dos dados (RSL iniciais e finais, tempo de início e fim e variáveis de interesse), um comando especial revela o estado corrente do dispositivo móvel.

O algoritmo anteriormente descrito revela uma das aplicabilidades do coprocessador e FDAP, que podem ser programados para realizar as tarefas anteriormente descritas. O gerenciador central deve ser programado para que aceite e saiba lidar com as informações do coprocessador, que enviará os dados do FD intermitente através do FDAP.

6.4.3 Outras topologias e aplicações

As arquiteturas descentralizadas para redes sem fio industriais podem ainda permitir topologias diversas, formadas por subredes, ou com pontos de acesso coordenados pelos gerenciadores distribuídos no entorno da rede. A Figura 46 apresenta uma topologia descentralizada, sem a presença de um gerenciador central. O gerenciamento é desta forma feito em modo totalmente distribuído mas com o fluxo de mensagens convergente para o ponto de acesso conectado ao gateway, que provê caminho para o meio externo.

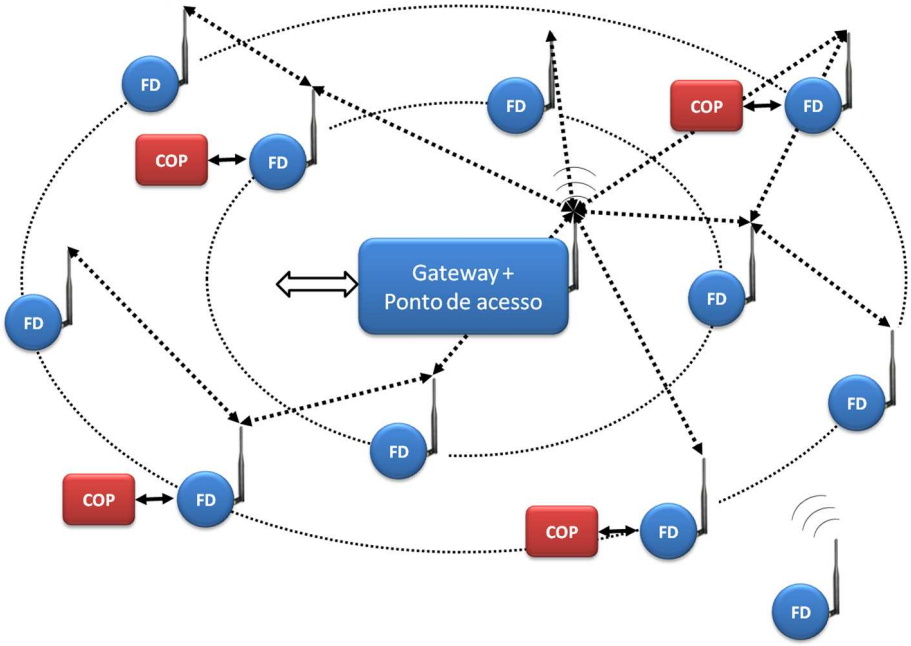

Figura 46: Topologia de rede com gerenciamento descentralizado e ponto de acesso único ao gateway.

Neste caso, os desafios são relativos à descoberta de rotas para o gateway, uma vez que a rede não necessariamente inicia a partir do ponto de acesso central.

Evoluindo nas possibilidades propiciadas pela descentralização do gerenciamento, a Figura 47 apresenta uma topologia descentralizada composta por sub redes com pontos de acesso que conectam-se a um gateway através de rede cabeada.

Com o desenvolvimento do gerenciamento distribuído, o sistema WH modificado passa a possibilitar outras configurações de rede. O *handover* no WH somente será factível se os eventos de transferência de conexão forem coordenados por um gerenciador local, uma vez que os tempos devidos ao gerenciamento centralizado (*uplink* e *downlink*) inviabilizam qualquer tentativa. Os testes iniciais com o sistema desenvolvido, mesmo que passíveis de grandes melhoramentos, revelam a possibilidade do uso do coprocessador e FDAP como elementos de controle de transferência das conexões entre os dispositivos da rede. O controle é feito localmente, sem necessidade de intervenção de um gerenciador central e com a possibilidade de programação de dispositivos WH comuns para que aceitem o dispositivo móvel como vizinho, no momento em que há link físico disponível. São inúmeras as possibilidades de desenvolvimento nesta configuração distribuída.

Os algoritmos de escalonamento e roteamento são implementados na sua forma mais simples, com o objetivo de validar a proposta desta tese. Os tempos das comunicações fim a fim podem ser minimizados de acordo com o escalonamento adaptativo da rede, que pode ser modificado dinamicamente. Com o gerenciamento centralizado, não existe esta possibilidade, uma vez que toda a programação deve ser enviada desde a unidade central, tornando a manutenção da rede lenta.

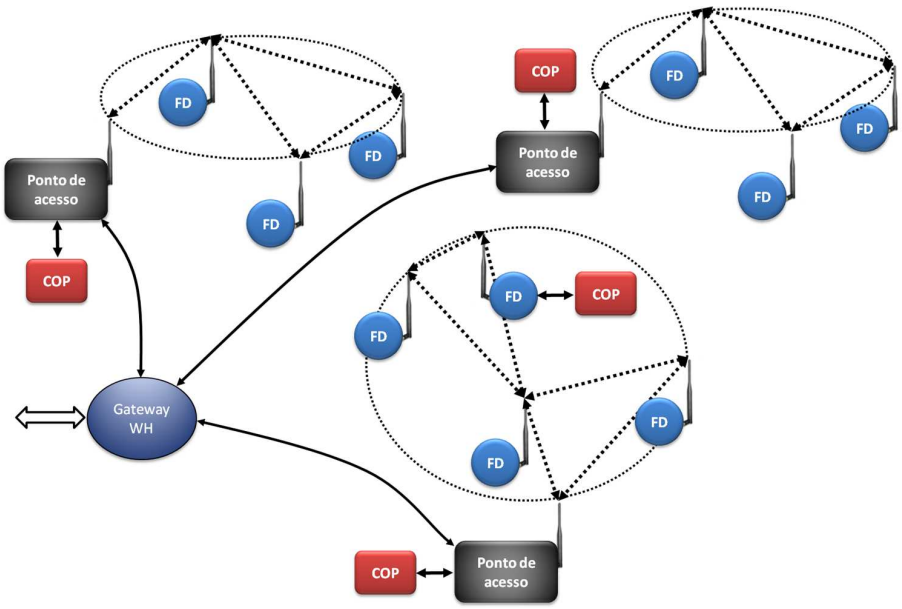

Figura 47: Topologia de subredes com gerenciamento descentralizado e pontos de acesso múltiplos.

7 CONCLUSÕES

Lançado no final 2007, o *Wireless*HART vem sendo considerando desde então o protocolo mais adequado para aplicações industriais em função de suas características inerentes de segurança e robustez. Porém, devido em grande parte a sua arquitetura de gerenciamento centralizado, o WH apresenta baixa dinamicidade no gerenciamento de rede o que o torna incapaz para uso com dispositivos móveis e intermitentes. Os poucos trabalhos apresentados que tratam especificamente deste problema dizem respeito ao emprego de técnicas de escalonamento distribuído. Estas implementações requerem modificações profundas no mecanismo TDMA do WH e eventualmente acabam por resolver as restrições dinâmicas mas criam diversos outros problemas, tais como o aumento da complexidade do sistema de enlace, redução de banda disponível e a exigência de maior poder computacional dos nós da rede.

Neste trabalho, propôs-se a utilização de coprocessadores gerenciadores para permitir a coordenação de eventos de rede, tais como agregação de dispositivos e manutenção de forma distribuída, cooperativa ou independentemente. Objetivou-se melhorar a dinamicidade de redes WH, de modo que estas pudessem atender outras aplicações, entre elas, o uso de equipamentos móveis, eventos de comunicação intermitente e rastreabilidade. As comunicações fim-a-fim também são passíveis de melhorias através da arquitetura proposta, uma vez que é possível modificar os agendamentos das mensagens dinamicamente. O esquema de gerenciamento distribuído permite uma ampla gama de desenvolvimentos futuros tanto no âmbito dos algoritmos de escalonamento quanto nas variações das arquiteturas e topologias. A solução proposta pouco modifica o protocolo original e permite a manipulação da rede de forma rápida, simples e eficiente, permitindo manter o gerenciador central para aproveitamento as vantagens do sistema centralizado. Também é possível a distribuição total do gerenciamento e o uso dos gerenciadores para a coordenação de operações intermitentes.

O coprocessador para descentralização da rede foi implementado aproveitando-se ao máximo a pilha WH bem como a arquitetura original, de modo que pôde ser comprovado em prática. Os estudos de caso, mesmo que restritos em função da complexidade total dos sistemas que compõem uma rede WH real, comprovaram o incremento da dinamicidade através do escalonamento local. As reduções de tempos de agregação são possíveis devido aos processos de gerenciamento local, que desoneram o gerenciador central no sistema misto. Os tempos de resposta aos anúncios de rede são drasticamente reduzidos em função do conectividade local dos novos dispositivos de campo com os gerenciadores distribuídos. Aliado a isso, ocorre grande redução da banda utilizada para agregação e manutenção centralizada, que podem corresponder a até 30 % do disponível.

Um importante efeito colateral deste desenvolvimento é a enorme possibilidade de estudos futuros, uma vez que a solução de eventuais conflitos hierárquicos em função da

arquitetura distribuída demanda o estudo e implementação de técnicas de roteamento e escalonamento distribuído.

As arquiteturas distribuídas propostas incluem redes parcialmente distribuídas, totalmente distribuídas, uso de dispositivos móveis e intermitentes e *handover*. Esta ampla gama de possibilidades poderá ser totalmente implementada na forma de serviços fornecidos pelos gerenciadores distribuídos, colocando o WH em outra categoria de protocolos, diferentemente da atual, limitada a controle e monitoramento de processos industriais.

REFERÊNCIAS

AKERBERG, J. et al. Integration of WirelessHART networks in Distributed Control Systems using Profinet IO. In: IEEE INTERNATIONAL CONFERENCE ON INDUSTRIAL INFORMATICS (INDIN), 8., 2010, Osaka. **Proceedings...** New York: IEEE Press, 2010. p.154–159.

BLEVINS, T. L. (Ed.). **Advanced Control Unleashed**: Plant Performance Management for Optimum Benefit. North Carolina: ISA Press, 2002.

BOWICK, C. (Ed.). **RF Circuit Design**. 2nd.ed. Oxford: Newnes Publishing, 2008.

CARO, D. (Ed.). **Wireless Networks for Industrial Automation**. North Carolina: ISA Press, 2004.

CHEN, D.; NIXON, M.; MOK, A. (Ed.). **WirelessHART Real-Time Mesh Network for Industrial Automation**. New York: Springer, 2010.

DAVIES, A. An overview of Bluetooth Wireless Technology and some competing LAN Standards. In: IEEE INTERNATIONAL CONFERENCE ON CIRCUITS AND SYSTEMS FOR COMMUNICATIONS, 1., 2002, Xiamen. **Proceedings...** New York: IEEE Press, 2002. p.206–211.

DJUKIC, P. **Scheduling Algorithms for TDMA Wireless Multihop Networks**. 2008. 219p. Doutorado em Ciências da Computação — Department of Electrical and Computing Engineering, University of Toronto, Toronto, 2008.

DOHERTY, L.; TEASDALE, D. A. Towards 100% Reliability in Wireless Monitoring Networks. In: ACM INTERNATIONAL WORKSHOP ON PERFORMANCE EVALUATION OF WIRELESS AD HOC, SENSOR AND UBIQUITOUS NETWORKS, 3., 2006, Torremolinos. **Proceedings...** New York: ACM, 2006. p.132–135.

ERGEN, S. C.; VARAIYA, P. TDMA scheduling algorithms for wireless sensor networks. **Journal Wireless Networks**, Hingham, v.16, p.985–997, 2010.

FIORE, G. et al. Multihop multi-channel scheduling for wireless control in WirelessHART networks. In: IEEE CONFERENCE ON EMERGING TECHNOLOGIES & FACTORY AUTOMATION, 2009, Mallorca. **Proceedings...** New York: IEEE Press, 2009. p.1–8.

GOODRICH, M. T.; TAMASSIA, R. **Algorithm Design Foundations, Analysis, and Internet Examples**. New Jersey: John Wiley & Sons, 2002.

HAHN, D. H. et al. Desenvolvimento de um Ponto de Acesso para Redes WirelessHART. In: CONGRESSO BRASILEIRO DE AUTOMÁTICA, 2012, Campina Grande. **Anais...** Campinas: SBA, 2012. v.1, p.3700–3707.

HAN, S. et al. Wi-HTest: compliance test suite for diagnosing devices in real-time wirelesshart network. In: IEEE REAL-TIME AND EMBEDDED TECHNOLOGY AND APPLICATIONS SYMPOSIUM, 15., 2009, San Francisco. **Proceedings...** New York: IEEE Press, 2009. p.327–336.

HAN, S. et al. Reliable and Real-Time Communication in Industrial Wireless Mesh Networks. In: IEEE REAL-TIME AND EMBEDDED TECHNOLOGY AND APPLICATIONS SYMPOSIUM (RTAS), 17., 2011, Chicago. **Proceedings...** New York: IEEE Press, 2011. p.3–12.

HART COMMUNICATION FOUNDATION. **HCF SPEC-085, Rev. 1.1**. Austin: HCF, 2010. Parte de norma.

HAYASHI, H.; HASEGAWA, T.; DEMACHI, K. Wireless technology for process automation. In: ICROS-SICE INTERNATIONAL JOINT CONFERENCE, 2009, Fukuoka. **Proceedings...** New York: IEEE Press, 2009. p.4591–4594.

HUANG, Y.-K.; PANG, A.-C.; HUNG, H.-N. An Adaptive GTS Allocation Scheme for IEEE 802.15.4. **IEEE Transactions on Parallel and Distributed Systems**, New York, v.19, p.641–651, 2008.

INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. **Part 15.4, Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs)**. New York: IEEE Computer Society, 2003. Parte de norma.

JONSSON, M.; KUNERT, K. Towards Reliable Wireless Industrial Communication with Real-Time Guarantees. **IEEE Transactions On Industrial Informatics**, New York, v.5, n.4, p.1, Nov. 2009.

KIM, A. et al. When HART goes wireless: understanding and implementing the wirelesshart standard. In: IEEE INTERNATIONAL CONFERENCE ON EMERGING TECHNOLOGIES AND FACTORY AUTOMATION, 2008, Hamburg. **Proceedings...** New York: IEEE Press, 2008. p.899–907.

KOSTADINOVI, M.; BUNDALO, Z.; BUNDALO, D. Planning and management of WirelessHart network. In: INTERNATIONAL CONVENTION, MIPRO, 33., 2010, Opatija. **Proceedings...** New York: IEEE Press, 2010. p.567–571.

KOUBAA, A.; ALVES, M.; TOVAR, E. GTS Allocation Analysis in IEEE 802.15.4 for Real-Time Wireless Sensor Networks. In: INTERNATIONAL WORKSHOP ON PARALLEL AND DISTRIBUTED REAL-TIME SYSTEMS, 20., 2006, Rhodes Island. **Proceedings...** New York: IEEE Press, 2006. p.176–176.

LENNVALL, T.; SVENSSON, S.; HEKLAND, F. A Comparison of WirelessHART and ZigBee for Industrial Applications. In: IEEE INTERNATIONAL WORKSHOP ON FACTORY COMMUNICATION SYSTEMS, 2008, Dresden. **Proceedings...** New York: IEEE Press, 2008. p.85–88.

LIMA, C. P. et al. Porta de Manutenção para Comissionamento e Análise Local de Redes WirelessHART. In: CONGRESSO BRASILEIRO DE AUTOMÁTICA, 19., 2012, Campina Grande. **Anais...** Campinas: SBA, 2012. v.1, p.4923–4929.

MENG, L.; XIAOJIE, D. Research of Industrial Wireless Network WIA-PA Multi-path routing protocol WMDSR. In: INTERNATIONAL SYMPOSIUM ON COMPUTER SCIENCE AND SOCIETY, 2011, Kota Kinabalu. **Proceedings...** New York: IEEE Press, 2011. p.51–54.

MULLER, I. et al. Development of a WirelessHART Compatible Field Device. In: IEEE INTERNATIONAL INSTRUMENTATION AND MEASUREMENT TECHNOLOGY CONFERENCE, 2010, Austin. **Proceedings...** New York: IEEE Press, 2010. p.1430–1434.

MULLER, I. et al. Namimote: a low-cost sensor node for wireless sensor networks. In: SERGEY ANDREEV SERGEY BALANDIN, Y. K. (Ed.). **Internet of Things, Smart Spaces, and Next Generation Networks. Internet of Things, Smart Spaces, and Next Generation Networks**. St. Petersburg: Springer, 2012. v.1, p.391–400.

NIXON, M. D. C.; BLEVINS, T.; MOK, A. Meeting Control Performance over a Wireless Mesh Network. In: IEEE INTERNATIONAL CONFERENCE ON AUTOMATION SCIENCE AND ENGINEERING, 2008, Washington, DC. **Proceedings...** New York: IEEE Press, 2008. p.540–547.

RECH, J. R. **Desenvolvimento de um gerente de rede WirelessHART**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2012.

SAIFULLAH, A. et al. Real-Time Scheduling for WirelessHART Networks. In: IEEE REAL-TIME SYSTEMS SYMPOSIUM (RTSS), 31., 2010, San Diego. **Proceedings...** New York: IEEE Press, 2010. p.150–159.

SHAH, K.; SECELEANU, T.; GIDLUND, M. Design and implementation of a WirelessHART simulator for process control. In: INTERNATIONAL SYMPOSIUM ON INDUSTRIAL EMBEDDED SYSTEMS, 2010, Trento. **Proceedings...** New York: IEEE Press, 2010. p.221–224.

SONG, J. et al. Improving PID control with unreliable communications. In: ISA EXPO TECHNICAL CONFERENCE, 2006, Houston. **Proceedings...** North Carolina: ISA Press, 2006. p.1–12.

SONG, J. et al. WirelessHART: applying wireless technology in real-time industrial process control. In: IEEE REAL-TIME AND EMBEDDED TECHNOLOGY AND APPLICATIONS SYMPOSIUM, 2008, St. Louis. **Proceedings...** New York: IEEE Press, 2008. p.377–386.

TANENBAUM (Ed.). **Computer Networks**. Indianapolis: Prentice Hall, 2003.

TINKA, A. et al. A decentralized scheduling algorithm for time synchronized channel hopping. **ADHOCNETS'10**, Victoria, v.1, p.201–216, 2010.

TJOA, R. et al. Clock Drift Reduction for Relative Time Slot TDMA-based Sensor Networks. In: IEEE SYMPOSIUM PIMRC, 15., 2004., Barcelona. **Proceedings...** New York: IEEE Press, 2004. v.2, p.1042–1047.

VARCHOLA, M.; DRUTAROVSKY, M. Zigbee Based Home Automation Wireless Sensor Network. **Acta Electrotechnica et Informatica**, London, v.7, p.1–8, 2007.

WANG, G. **Comparison and Evaluation of Industrial Wireless Sensor Network Standards ISA100.11a and WirelessHART.** 2011. 97p. Mestrado em Engenharia de Comunicações — Chalmers University of Technology, Gothenburg, 2011.

WILLIG, A.; MATHEUS, K.; WOLISZ, A. Wireless Technology in Industrial Networks. In: IEEE, 2005, New York. **Proceedings...** New York: IEEE Press, 2005. v.93, p.1130–1151.

WINTER, J. M. et al. WirelessHART Routing Analysis Software. In: BRAZILIAN SYMPOSIUM ON COMPUTING SYSTEM ENGINEERING, 2011, Florianópolis. **Anais...** Porto Alegre: SBC, 2011. p.96–98.

ZAND, P. et al. Wireless Industrial Monitoring and Control Networks: the journey so far and the road ahead. **Journal of Sensor and Actuator Networks**, [S.l.], v.1, p.123–152, 2012.

ZAND, P.; SHIVA, M. Defining a new frame based on IEEE 802.15.4 for having the synchronized mesh networks with channel hopping capability. In: IEEE INTERNATIONAL CONFERENCE ON COMMUNICATION TECHNOLOGY, 11., 2008, Chengdu. **Proceedings...** New York: IEEE Press, 2008. p.54–57.

ZHANG, H.; SOLDATI, P.; JOHANSSON, M. **Efficient Link Scheduling and Channel Hopping for Convergecast in WirelessHART Networks**. [S.l.]: Royal Institute of Technology (KTH), 2009.

ZHU, X. et al. A Location-Determination Application in WirelessHART. In: IEEE INTERNATIONAL CONFERENCE ON EMBEDDED AND REAL-TIME COMPUTING SYSTEMS AND APPLICATIONS, 15., 2009, Beijing. **Proceedings...** New York: IEEE Press, 2009. p.263–270.

ZHU, X. et al. Hardware Challenges and Their Resolution in Advancing WirelessHART. In: IEEE INTERNATIONAL CONFERENCE ON INDUSTRIAL INFORMATICS (INDIN), 9., 2011, Lisbon. **Proceedings...** New York: IEEE Press, 2011. p.416–421.