



Diseño de una arquitectura de aviónica tolerante a fallas basada en componentes COTS para vehículos satelitales de nueva generación

Por ***Arias Emmanuel***

Presentado ante la Universidad Nacional de La Matanza y la Unidad de Formación Superior de la CONAE
como parte de los requerimientos para la obtención del grado de

MAGISTER EN DESARROLLOS INFORMATICOS DE APLICACION ESPACIAL

UNIVERSIDAD NACIONAL DEL CÓRDOBA

Junio, 2017

©UFS-CONAE 2017

©UNLAM 2017

DIRECTOR

Gustavo Wiman

INVAP, Bariloche, Provincia de Rio Negro

*Dedicado a
alguien*

Abstract

Keywords:

Resumen

Esta tesis se trata de

Agradecimientos

Muchas gracias

Tabla de Contenidos

1. Introducción	2
1.1. Motivación	3
1.2. Hipótesis	4
1.3. Objetivo del trabajo y preguntas de investigación	4
1.3.1. Objetivo	4
1.3.2. Objetivos Específicos	4
1.3.3. Preguntas de investigación	5
2. Marco Teórico	6
2.1. Terminología	6
2.2. La fiabilidad en el software	8
2.3. Impedimentos de la confiabilidad	8
2.3.1. Orígenes de la falla	9
2.3.2. Modos comunes de fallas	9
2.3.3. Fallas en el Software	9
2.4. Medios de fiabilidad	10
2.4.1. Evitación de Fallas	10
2.4.2. Tolerancia a Fallas	10
2.4.3. Eliminación de Fallas	11
2.4.4. Predicción de Fallas	11

TABLA DE CONTENIDOS

2.5. Atributos de la fiabilidad	11
2.5.1. Confiabilidad	11
2.5.2. Disponibilidad	12
2.5.3. Seguridad	13
2.6. Tolerancia a falla	13
2.7. Clasificación de un sistema de control tolerante a fallas	15
2.8. Redundancia en el software	15
2.8.1. Técnicas single version	15
2.8.2. Técnicas multi-version	16
2.8.3. Técnicas de detección de fallas	16
2.8.4. Técnicas de recuperación de fallas	17
2.9. Técnica de evaluación de fiabilidad	18
2.10. Medidas comunes de fiabilidad	19
2.10.1. Failure rate	19
2.10.2. Tiempo medio medio de falla	20
2.10.3. Tiempo medio de reparación	20
2.10.4. Tiempo medio entre fallas	21
2.10.5. Cobertura de fallas	21
2.11. Métodos de cálculos de fiabilidad	21
2.11.1. Diagramas de bloques de confiabilidad	21
2.11.2. Utilización de procesos de Markov	22
2.12. Modelos de falla	22
2.12.1. Tiempo hasta la falla	23
2.12.2. Función de confiabilidad	23
2.12.3. Tasa de falla	24
2.12.4. Tiempo medio hasta la falla	24
2.12.5. Vida restante media	24
2.12.6. Distribución binomial	24
2.12.7. Distribución exponencial	25

TABLA DE CONTENIDOS

2.13. Protocolos de comunicación de tiempo real	25
2.13.1. Sistemas de tiempo real	25
2.14. Estrategia de acceso al medio	26
2.14.1. CSMA	26
2.14.2. TDMA	27
2.14.3. Minislotting	27
2.15. Revisión de protocolos	27
2.15.1. CAN	27
2.15.2. byteflight	27
2.15.3. ARINC 659 o SAFEbus	28
2.15.4. TTP/C	28
2.16. Posibles fallas en una red	28
2.17. Protocolo CAN	28
2.17.1. Consideraciones previas	28
2.17.2. Introducción	28
2.17.3. Elementos necesarios	30
2.17.4. Capa física	30
2.17.5. Capa de Enlace	31
2.17.6. Formato del mensaje	32
2.17.7. CANAerospace	34
3. Estado del arte	36
3.1. Árboles binarios	36
3.1.1. Esquema de árbol binario con backups	37
3.1.2. Esquema de árbol binario con degradación de performance	38
3.2. Sistemas Hypercube	38
3.3. Redes distribuídas	39
3.3.1. Algoritmo de ruteo	39
3.4. Redes Ethernet en aviónica	40

TABLA DE CONTENIDOS

3.4.1. Experiencia de vuelo	41
3.5. Arquitectura de red basada en BUS	41
3.5.1. Evaluación de la confiabilidad de arquitecturas basadas en BUS	42
3.6. Métrica y modelado de la confiabilidad de sistemas	43
3.7. CAN en la actividad espacial	45
4. Análisis y desarrollo de arquitectura tolerante a fallas	46
4.1. Introducción	46
4.2. Requerimientos para el análisis	46
4.3. Nomenclatura	47
4.4. Estudio de topologías de arquitecturas	48
4.4.1. Árbol binario	48
4.4.2. Red distribuida	49
4.4.3. Red hypercube	50
4.5. Topología utilizada en la arquitectura a diseñar	51
4.6. Topología propuesta	53
4.7. Protocolo de comunicación	53
5. Arquitectura propuesta	57
5.1. Introducción	57
5.2. Árbol de requerimientos	57
5.3. Casos de Uso	60
5.4. Diseño Estructural	62
5.4.1. Bridge tolerante a fallas	62
5.4.2. Bloques Internos	63
5.4.3. Nodo Monitor	65
5.5. Modelo dinámico	65
5.5.1. Máquina de estado	66
5.5.2. Diagrama de secuencia	66
5.5.3. Diagrama de actividades	67

6. Conclusión	76
6.1. Cumplimiento de los objetivos	78
6.2. Trabajos a futuros	78
Bibliografía	84
A. Protocolo de comunicación: CANae 0.1 Alpha Version	85
A.1. Introducción	85
A.1.1. Revisiones	85
A.2. Capa de aplicación	86
A.3. Tipos de servicios de la capa de aplicación	86
A.4. CANae Application Layer	87
A.5. CMS	88
A.5.1. Prioridades de los objetos	88
A.5.2. Objeto CMS	88
A.5.3. Servicios de CMS	88
A.5.4. Eventos	90
A.6. NMT	91
A.6.1. Objetos NMT	92
A.6.2. Servicios NMT	92
A.6.3. Protocolos NMT	94
A.7. DBT	97
A.7.1. Objetos y servicios de DBT	98
A.7.2. Descripción de los servicios DBT	98
A.7.3. Objetos DBT	99
A.7.4. Servicios DBT	99
A.7.5. Protocolos DBT	99
A.8. Gestor de mensajes	102
A.8.1. Receiver Manager	102
A.8.2. Prepared Message	103

TABLA DE CONTENIDOS

A.8.3. Sorter Message	104
A.8.4. Buffer Messages	104
A.8.5. Buffer Message Receive y Buffer Message To Send	105
A.9. Gestor de Nodo	105
A.9.1. Tabla primaria y secundaria de ruteo	106
A.9.2. Servicios	107
A.10.Formato de mensajes	108
A.10.1. Start of Frame	109
A.10.2. Priority	109
A.10.3. Node-ID	109
A.10.4. Remote transmission request (RTR)	109
A.10.5. Type of Message (TOM)	109
A.10.6. Bit reservado	109
A.10.7. Data length	109
A.10.8. Data field	110
A.10.9. CRC	110
A.10.10CRC delimiter	110
A.10.11ACK	110
A.10.12ACK delimiter	110
A.10.13End-of-Frame (EOF)	110
A.11.High Application Layer CANae	111
A.11.1. Network Management	111
A.11.2. Task Management	112
A.11.3. List of Tasks	114
A.11.4. Task	115
A.12.Arquitectura completa del CANae	115
A.13.Estados de nodos y red	115
B. Especificación de Casos de Uso	118

TABLA DE CONTENIDOS

B.1. Enviar Mensaje	118
B.2. Recibir Mensaje	119
B.3. Procesar Mensaje	120
B.4. Rutear Mensaje	120
B.5. Gestionar Red	121
B.6. Gestionar Tareas	122
B.7. Dividir Tareas	123
B.8. FDIR	123
B.9. Reconfigurar Arquitectura	124

Índice de figuras

2.1. Fiabilidad (Hitt y Mulcare, 2070)	8
2.2. Failure rate de HW vs tiempo	19
2.3. Failure rate SW vs tiempo	20
2.4. Confiabilidad vs tiempo	20
2.5. Arquitectura estándar propuesta por ISO898	29
2.6. Tráfico en el BUS CAN	30
2.7. Frame de mensaje del CAN estándar	33
2.8. Frame del mensaje del CAN extendido	34
2.9. Formato del mensaje de CANAerospace	35
3.1. Árbol binario de 4 niveles	37
3.2. Confiabilidad con respecto al tiempo de una arquitectura de árbol binario de 4 niveles	38
3.3. Arquitectura básica TTEthernet	40
3.4. Frame de mensaje TTEthernet	41
3.5. Arquitecturas stack-trees	42
3.6. Arquitecturas que no responden al modelo stack-trees	42
3.7. Esquema CST_D	42
3.8. Modelo de sistema satelital (Hoque et al., 2015)	45
4.1. Confiabilidad con respecto al tiempo de árbol binario de 4 niveles	48
4.2. Red distribuida	49

ÍNDICE DE FIGURAS

4.3. Confiabilidad de red distribuida	50
4.4. Confiabilidad de red distribuida con 4 nodos fallando	50
4.5. Red Hypercube	51
4.6. Confiabilidad de red hypercube	51
4.7. Comparación de confiabilidad	52
4.8. Arquitectura propuesta utilizando topología de red distribuida	54
4.9. Conexión entre la red y el subsistema	54
4.10. Estructura de la capa de aplicación de CANae en alto nivel	55
5.1. Diagrama de requerimientos de la arquitectura propuesta	59
5.2. Diagrama de Casos de Uso General de la arquitectura Propuesta	60
5.3. Diagrama de Casos de Uso de arquitectura propuesta	61
5.4. Diagrama de bloques de la arquitectura completa	63
5.5. Posibilidad de interconexión de nodos.	64
5.6. Posibilidad de interconexión de nodos.	64
5.7. Conexión entre la red y el subsistema	65
5.8. Diagrama de bloques interno del nodo.	69
5.9. Máquina de estado de la arquitectura completa	70
5.10. Máquina de estado de los nodos.	70
5.11. Diagrama de secuencia del inicio de la arquitectura.	71
5.12. Diagrama de secuencia de CA	71
5.13. Crear Objeto Red, Objeto Nodo y Objeto Nodo Remoto	72
5.14. Preparar nodo para la conexión a la red CAN	72
5.15. Preparar nodo para la conexión a la red CAN	72
5.16. Conectar el nodo a la red CAN	73
5.17. Diagrama de actividades del inicio de nodos	73
5.18. Diagrama de actividades del inicio de nodos	74
5.19. Diagrama de actividades del inicio de nodos	75
6.1. Comparación de confiabilidad	77

A.1. Estructura de la capa de aplicación de CANae en alto nivel	86
A.2. CANae Application Layer	87
A.3. Definición del Objeto CMS	89
A.4. Definición de la entidad NMT (bloques internos)	91
A.5. Definición de la entidad NMT	92
A.6. Protocolo NMT	95
A.7. Crear Red	95
A.8. Crear Objecto Red, Objeto Nodo y Objeto Nodo Remoto	96
A.9. Preparar nodo para la conexión a la red CAN	96
A.10. Conectar el nodo a la red CAN	96
A.11. Comenzar la comunicación de nodos	97
A.12. Arquitectura de la entidad DBT	98
A.13. Protocolo create_node_table_configuration	100
A.14. Protocolo enable_distribution	100
A.15. Protocolo disable_distribution	101
A.16. Protocolo create_node_definition	101
A.17. Protocolo delete_node_definition	101
A.18. Protocolo get_checksum	102
A.19. Arquitectura de <i>Message Management</i>	103
A.20. Diagrama interno del <i>Message Management</i>	103
A.21. Arquitectura del <i>Node Management</i>	105
A.22. Frame de datos CANae	108
A.23. Diagrama de bloques del High Application Layer de CANae	111
A.24. Diagrama de bloques internos del High Application Layer de CANae	112
A.25. Diagrama de interacción del comportamiento del <i>Network Management</i>	113
A.26. Diagrama de interacción del comportamiento del <i>Network Management en caso de errores</i>	113
A.27. Diagrama de secuencia de la ejecución de tareas del Task Management	114
A.28. Diagrama de secuencia para detener procesos por parte del Task Management	114
A.29. Arquitectura completa del CANae	116

ÍNDICE DE FIGURAS

A.30. Definición de la entidad NMT	117
A.31. Definición de la entidad NMT	117

Índice de tablas

2.1. Disponibilidad en relación con su baja de servicio por año. Tabla modificada de Dubrova (2013)	13
4.1. Comparación de confiabilidad de topologías	53
5.1. Tabla de Requerimientos	58
6.1. Tabla de objetivos cumplidos	79
A.1. Revisiones de CANae	85
A.2. Prioridad eventos	91
A.3. Prioridades de mensajes	109

Lista de acrónimos

SW	Software
HW	Hardware
FT	Tolerancia a Fallas
FA	Evitación de Fallas
FR	Eliminación de Fallas
FF	Predicción de Fallas
MTBF	Tiempo Medio Entre Fallas
MTTF	Tiempo Medio de Fallas
MTTR	Tiempo Medio de Reperación
CMF	Modo Común de fallas
FDIR	Detección, Aislación y Recuperación de Fallas
CONAE	Comisión Nacional de Actividades Espaciales
UNLAM	Universidad Nacional de La Matanza
INVAP	Investigación Aplicada
NASA	National Aeronautics and Space Administration
COTS	Commercial Off-The-Shelf
TT	Time-Trigged
CST	Complete Stack-Tree
CAN	Controller Area Network
CPU	Central Proceesing Unit

Introducción

En el marco del Plan Espacial Nacional, desarrollado por la Comisión Nacional de Actividades Espaciales (CONAE) de Argentina, y con el propósito de llevar a cabo actividades de investigación y aplicación, provenientes de la Universidad Nacional de La Matanza (UNLAM) se presenta este plan de tesis con el fin de ampliar los conocimientos y la participación de la CONAE y UNLAM, en el campo del Desarrollo Informático y Ciencias de la Computación.

Las actividades desarrolladas para este trabajo de tesis son realizadas, en su mayor proporción, en la Unidad de Desarrollo Investigación Aplicada (INVAP), ubicada en San Carlos de Bariloche, Provincia de Río Negro. Este trabajo se encuentra orientado a brindar un nuevo conocimiento, que ayude en cierta medida, en el desarrollo de los diferentes proyectos con los que cuenta actualmente esta empresa, agregando un grado de innovación en el resultado que se obtenga.

INVAP tiene como visión ser un referente en proyectos tecnológicos a nivel mundial (INVAP, 2016), por lo tanto, debe asegurarse que cada uno de los productos que se lleven a cabo sean competitivos. Para lograr cumplir con esto, es necesario que tales proyectos se encuentren a la vanguardia tecnológica y científica.

El desarrollo de proyectos satelitales conlleva costos de importante magnitud, y dependen de cada misión. Una parte importante de los costos está conformado por el desarrollo¹ y sobre todo los materiales que se utilizan para su fabricación. Esto es debido a que se utilizan componentes que son exclusivos para el ámbito espacial, en otras palabras que se encuentran “calificados para volar”. Estos componentes son fabricados especialmente para soportar el ambiente hostil del espacio.

Si se considera al ámbito espacial como una industria, algo que ha sido demostrado en los últimos años; y si se tiene en cuenta las intenciones de crecimiento y competitividad de la empresa INVAP, de permitir el ingreso de nuestro país en el mercado satelital INVAP (2016), resulta de gran importancia lograr reducir los costos en fabricación y desarrollo de vehículos satelitales.

La National Aeronautics and Space Administration (NASA) tiene un enfoque de desarrollo bajo el lema “faster, cheaper, better” Forsberg y Harold (1999), lo cual busca desarrollar sus proyectos y misiones de forma rápida, barata y mejor. Bajo este enfoque se han realizado diversos estudios e investigaciones dando resultados sumamente positivos Tai et al. (1999), Chau et al. (1999), Schneidewind y Nikora (1998), Forsberg y Harold (1999). En estos trabajos se utilizan componentes que no se encuentran “calificados para volar”, los cuales también son llamados componentes Commercial Off-The-Shelf (COTS), o de estantería. Debe mencionarse, que también hubo algunos fracasos en su aplicación.

¹Nota: entiéndase por desarrollo al proceso de planificación, análisis, diseño e implementación.

A simple vista, la utilización de estos componentes ayudaría a reducir costos. Sin embargo, esto no es tan directo. Los componentes COTS al no estar calificados, se les deben realizar tareas de calificación adicional. Además deben ser aplicados a un ambiente, que asegure que no fallarán durante la misión; o si fallan, no será motivo de pérdida de la misma.

Los componentes COTS suelen tener un costo de compra entre 100 y 1000 veces menores que aquellos que está calificados para volar. Por lo que el aumento en la utilización de estos componentes, aplicados al desarrollo de diferentes tipos de satélite, **permitiría reducir los costos y ahorrar algunos millones de dólares del proyecto satelital**. Esto facilitaría el ingreso de Argentina en un mercado altamente competitivo.

El desafío de este trabajo de tesis es analizar y estudiar arquitecturas que sean tolerantes a fallas, que permitan una correcta comunicación entre los diferentes subsistemas de un vehículo espacial de nueva generación, y que tenga como característica principal un cierto grado de confiabilidad, de modo tal que pueda ser aplicado con componentes COTS.

1.1. Motivación

Los costos de un proyecto satelital se pueden clasificar, a grandes rasgos, en 5 grupos:

- Desarrollo
- Materiales
- Ensamblado, integración, y tests
- Lanzamiento
- Operaciones

Este trabajo de tesis se centrará principalmente en el desarrollo (proceso de planificación, análisis, diseño e implementación.), y en los materiales utilizados en la fabricación de vehículos satelitales.

No se puede mencionar a ciencia cierta cuál es el costo “verdadero” de desarrollar un satélite. Este depende exclusivamente del tipo de satélite y de la misión. Lo que si se debe tener en claro es que las tareas de desarrollo representan una parte muy importante del costo total del proyecto.

Desarrollar un vehículo espacial con componente COTS, en un principio podría representar costos adicionales, ya que se le deben realizar tareas de calificación adicional, debido a que no están “preparados” para resistir las condiciones hostiles del espacio.

Uno de los puntos positivos, y que motivan la aplicación de componentes COTS, es que a la hora de desarrollar varios satélites en base a la misma ingeniería, se puede ahorrar en gran medida en los materiales que se utilizan. Los componentes COTS suelen tener un costo de compra entre 100 y 1000 veces menores que aquellos que están calificados para volar. **Esto ayudaría a ahorrar algunos millones de dólares de los proyectos satelitales.**

Otra de las ventajas de utilizar componentes COTS, es que la mayoría cuentan con una tecnología más avanzada que aquellos que son calificados para volar. Esta tecnología permite:

- Aumentar prestaciones, mediante el incremento de las capacidades de procesamiento, memoria, velocidades de procesamiento, etc.
- Implementar funciones que son imposibles de aplicar en tecnologías viejas.

- Reducir tiempos de desarrollo.
- Reducir volumen, masa y consumo

El último punto mencionado anteriormente es de especial interés, ya que al reducir volumen y masa, permite reducir costos adicionales como el de lanzamiento.

Esta reducción de costos de proyectos satelitales tienen ventajas directas a la hora de introducir a Argentina en un mercado altamente competitivo, donde la mínima reducción de estos, representa ganancias económicas importantes.

Uno de los puntos en contra de la utilización de componentes COTS es que al no ser calificados para volar, es necesario llevar a cabo tareas y estrategias inteligentes, con el fin de hacer frente a esa “deficiencia”. Por ello, se exige realizar una investigación y análisis de diferentes arquitecturas de aviónica, que puedan ser utilizadas para lograr que el sistema sea tolerante a fallas, y así, cumplir con los requerimientos de una misión satelital.

El estudio de arquitecturas tolerantes a fallas, no solamente tiene aplicación en el ámbito espacial, si no que también puede ser extendido a cualquier sistema crítico, los cuales necesitan ser robustos y tolerantes a fallas, como es el caso de aviones comerciales, plantas nucleares, automóviles, etc.

1.2. Hipótesis

La hipótesis de esta tesis es la siguiente: “Una arquitectura de aviónica basadas en componentes COTS, robusta y tolerante a fallas, es totalmente aplicable y utilizable en vehículos espaciales, con un alto nivel de confiabilidad, lo cual permite disminuir la complejidad de los sistemas actuales de aviónica”.

1.3. Objetivo del trabajo y preguntas de investigación

En esta sección se mencionan el objetivo principal y específicos; como así también las preguntas de investigación, que guiarán la elaboración de esta tesis.

1.3.1. Objetivo

El objetivo de este trabajo es investigar y analizar arquitecturas de comunicación de los subsistemas de aviónica tolerante a fallas basada en componentes COTS para vehículos satelitales de nueva generación.

1.3.2. Objetivos Específicos

1. Realizar un estudio del estado de la cuestión sobre arquitecturas tolerantes a fallas para sistemas críticos.
2. Investigar y analizar arquitecturas tolerantes a fallas que aseguren la confiabilidad del sistema y que sean aplicables en la industria satelital.
3. Investigar y analizar protocolos de comunicación, para las capas superiores del modelo de OSI (modelo de interconexión de sistemas abiertos - ISO/IEC 7498-1), orientados a la tolerancia a fallas y confiabilidad de los sistemas. Realizar un estudio comparativo de los diferentes protocolos estudiados.
4. Investigar una metodología para lograr una medición de la tolerancia a fallas en arquitecturas de aviónica.

5. Desarrollar un estudio comparativo de arquitecturas tolerantes a fallas con el fin de obtener ventajas y desventajas de cada una de ellas.
6. Diseñar modelos alternativos de arquitecturas tolerantes a fallas, que tenga un grado de confiabilidad tal que permita la aplicación de componentes COTS.
7. Evaluar la confiabilidad de los modelos de arquitecturas (mediante métrica desarrollada en este trabajo o siguiendo otras estrategias).
8. Proponer el diseño de una nueva arquitectura tolerante a fallas, con un grado de confiabilidad suficiente para la aplicación de componentes COTS en aviónicas de vehículos satelitales.
9. Simular la arquitectura planteada para medir su grado de tolerancia a fallas y performance.

1.3.3. Preguntas de investigación

Para este trabajo de tesis se plantearon las siguientes preguntas de investigación, que se fueron respondiendo a lo largo de este trabajo

- ¿Es posible la realización de un método de medición del grado de tolerancia a fallas de una arquitectura de aviónica?
- ¿Cuál es la estrategia más indicada de tolerancia a fallas que permita brindar un alto grado de confiabilidad en la utilización de componentes COTS en sistemas críticos?
- ¿Cuál es la arquitectura más indicada que permita desarrollar tolerancia a fallas en sistemas críticos basados en componentes COTS?
- ¿Es factible la utilización de componentes COTS en sistemas espaciales?

Marco Teórico

Solo sé que no sé nada

Platón

2.1. Terminología

Existe una importante diferencia entre los significados de las palabras falla, error y avería¹, que es importante destacar antes de comenzar con el desarrollo de este trabajo.

Un **avería** de sistema ocurre cuando el servicio prestado por el sistema ya no coincide con las especificaciones del mismo (Hanmer, 2007). Esto quiere decir que existe un problema que tiene una consecuencia negativa en el sistema completo, logrando que este ya no logre cumplir con sus especificaciones. Cuando el sistema no se comporta de la manera que es especificada, este ha fracasado. Esto significa que lo que se espera de un sistema se encuentra descripto, comúnmente en especificaciones o requerimientos (Pullum, 2001).

Para la IEEE (1990) avería es “la inhabilitación de una sistema o componente a llevar a cabo las funciones requeridas en los requerimientos específicos de performance del mismo”.

Hanmer (2007) ejemplifica averías de sistemas cuando: el sistema se bloquea y se detiene cuando no debería hacerlo, el sistema calcula un resultado incorrecto, el sistema no está disponible, el sistema es incapaz de responder a la interacción con el usuario. Cuando el sistema no hace lo que debe hacer, el sistema ha fracasado. Las averías son detectados por los usuarios mientras usan el sistema.

Las averías son causados por los errores. Un **error** es una parte del estado del sistema que es susceptible de provocar un avería en el sistema, un error que afecta al servicio es una indicación de que un avería se ha producido (Hanmer, 2007). Un error se puede propagar, es decir dar a lugar otros errores (Pullum, 2001).

IEEE (1990) define error como “la diferencia entre un valor computado, observado o medido, con el valor verdadero, especificado o el teóricamente verdadero”.

Los errores se pueden clasificar en dos tipos: errores de tiempo y valores (Hanmer, 2007). Los errores de valores son aquellos que se manifiestan como valores discretos incorrectos o estados del sistema incorrecto. En cambio, los errores de tiempo pueden incluir aquellos que no cumplen con el total de las tareas.

¹En inglés: fault, error y failure.

Hanmer (2007) especifica los siguientes casos más comunes de errores:

- Timing: existe una falta de sincronización en la comunicación de los procesos.
- Bucles infinitos: ejecución de un bucle sin detenerse, esto consume memoria, y la avería del sistema.
- Error de protocolo: errores en el flujo de comunicación ya que no coinciden los protocolos. Mensajes enviados en formato diferente, en tiempos diferentes, a lugares de sistemas incorrectos.
- Inconsistencia de datos: errores son diferentes en diferentes lugares.
- Sobrecarga de sistema: el sistema es incapaz de hacer frente a la sobrecarga de actividades a la que es expuesta.

La causa adjudicada o la hipótesis de un error es una **falla**, también llamado “bugs”. Una **falla activa** es aquella que produce un error (Pullum, 2001). Una falla es un defecto que está presente en el sistema y que puede causar un error (Hanmer, 2007). Es la desviación actual de lo correcto Hanmer (2007).

Según IEEE (1990) una falla es “un defecto en un dispositivo de hardware o componente; como por ejemplo un corto circuito o un cable cortado”. También realiza una segunda definición diciendo que falla es “un paso incorrecto, proceso, o definición de dato en un programa de computadora” IEEE (1990). Esta última afirmación es la que se usa en el ámbito de este trabajo.

Algunas fallas introducidas en el Software (SW) se detallan en Hanmer (2007), lo cual señala que pueden incluir:

- Especificaciones incorrecta de requerimientos
- Diseño incorrecto
- Errores de programación

Entonces, como lo indica Pullum (2001) con la tolerancia a fallas, lo que se busca es prevenir la avería mediante la “tolerancia” de fallas, las cuales son detectables cuando un error aparece. Las fallas son el motivo de errores y los errores son motivos de avería (Dubrova, 2013).

También se suele utilizar el término anomalía en las operaciones de vehículos espaciales para referirse a comportamientos anómalos o no esperados del sistema (David M. Harland, 2005)

En Dubrova (2013) se describe un ejemplo para diferenciar correctamente estos conceptos. Se considera el SW de una planta nuclear, en la cual existe una computadora que es responsable de controlar la temperatura, la presión y demás variables de interés para la seguridad del sistema. Se da el caso de que uno de los sensores detecta que la turbina principal se encuentra girando a una velocidad menor a la correcta. Esta falla hace que el sistema envíe una señal para aumentar su velocidad (error). Esto produce un exceso de velocidad en la turbina, lo cual tiene como consecuencia que la seguridad mecánica apague la turbina. En esta situación el sistema no está generando energía. Esto se considera un avería, porque el sistema no está entregando el servicio según lo establecido por los requerimientos. Pero es un avería salvable.

Otro concepto es el de **mantenibilidad**, este es la capacidad de un sistema, bajo condiciones normales, de ser restaurado a un estado en el cual puede realizar sus funciones requeridas, cuando se realiza el mantenimiento (Rausand y Hoyland, 2004).

En secciones posteriores se ven los conceptos de confiabilidad, disponibilidad y seguridad (Sección 2.5.1, 2.5.2, 2.5.3, respectivamente).

2.2. La fiabilidad en el software

El objetivo final de la Tolerancia a Fallas (FT), es el desarrollo de un sistema fiable (Dubrova, 2013). Teniendo en cuenta que SW que se encuentran dentro de las naves espaciales, como satélites, lanzadores, y sobre todo vehículos tripulados son críticos, ya que de ellos dependen el éxito o fracaso de una misión, se debe llevar a cabo un sistema fiable. La fiabilidad de un sistema es la capacidad del mismo de entregar a los usuarios un nivel deseado de servicio (Dubrova, 2013).

La fiabilidad también se la puede considerar como una propiedad global que permite justificar la confianza de los servicios de un sistema (Hitt y Mulcare, 2070). Por lo tanto, como lo indica Hitt y Mulcare (2070) la fiabilidad es un término amplio y cualitativo que está relacionado con atributos no funcionales (o “-ilities”), que buscan generar un sistema “ideal”, especialmente cuando su funcionamiento es crítico.

Como se muestra en la Figura 2.1 la consecuencia de la fiabilidad es la relación entre la evitación de fallos y la reducción de fallos, así como también la FT.

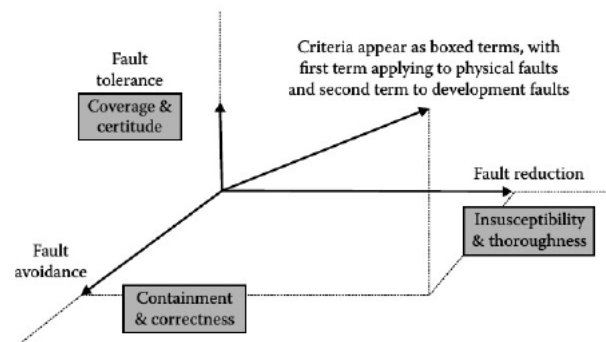


Figura 2.1: Fiabilidad (Hitt y Mulcare, 2070)

Según Pullum (2001) la fiabilidad puede ser clasificada en:

- Impedimentos: son aquellas cosas que se interponen en el camino de la fiabilidad. Son las fallas, errores y fracasos.
- Medios: los medios para lograr la fiabilidad, según el autor, se pueden dividir en dos grupos:
 1. Aquellos que son utilizados durante la construcción del SW (Evitación de Fallas (FA)² y FT).
 2. Aquellos que contribuyen con la validación del SW una vez desarrollado (Eliminación de Fallas (FR)³ y Predicción de Fallas (FF)⁴).
- Atributos: describen las propiedades de la fiabilidad y proporcionan una forma de evaluar el logro de esas propiedades.

2.3. Impedimentos de la confiabilidad

El impedimento de la confiabilidad o deterioro de la confiabilidad es definido en términos de fallas, errores, y fracaso (Dubrova, 2013). Los mismos fueron desarrollados en la sección 2.1. Lo que tienen en común estos

²En inglés, Fault Avoidance

³En inglés, Fault Removal

⁴En inglés, Fault Forecasting

tres conceptos, es que avisan, o dan alerta cuando algo está mal (Dubrova, 2013). La diferencia radica, en que las fallas son a nivel físico; los errores se dan a nivel computacional; mientras que los fracasos se dan a nivel de sistema Dubrova (2013).

2.3.1. Orígenes de la falla

Existen diversos orígenes de fallas. Estas pueden provenir desde terceros, en el caso de productos comprados, pueden deberse a una falta del conocimiento del problema, falta de tiempo, etc. Dubrova (2013) clasifica el origen de las fallas en cuatro grupos: *especificación incorrecta*, *implementación incorrecta*, *defectos de fabricación* y *factores externos*

Las *especificaciones incorrectas* son aquellas que surgen debidas a una incorrecta especificación de requerimiento o un mal diseño de una arquitectura o de un algoritmo (Dubrova, 2013). Estos orígenes de fallas son bastante comunes en el desarrollo de sistemas. Un ejemplo típico citado por Dubrova (2013), es el caso de requerimientos que ignoran aspectos del medio ambiente en el que opera el sistema. Una mala redacción de un requerimiento o el olvido de uno de ellos, puede traer graves problemas, atrasos y pérdida de dinero, en el diseño y producción de un sistema espacial.

Las *implementaciones incorrectas*, se refieren a las *fallas de diseño*, surgen cuando el sistema implementado no cumple con los requerimientos (Dubrova, 2013).

Otro origen de falla son los *defectos de los componentes* (Dubrova, 2013). Estos pueden incluir defectos de fabricación, defectos aleatorios dados en los componentes, etc.

Y por último se tienen las fallas que son causadas por *factores externos*, los cuales provienen del medio ambiente, usuarios u operadores (Dubrova, 2013). Ejemplos de estos factores externos pueden ser, vibraciones, cargas electrostáticas, temperatura, radiación electromagnética, envío incorrecto de comandos, etc.

2.3.2. Modos comunes de fallas

Un *Modo Común de fallas (CMF)*⁵ es una falla que ocurre simultáneamente en dos o más componentes redundantes (Dubrova, 2013).

Gangloff (1975) define los CMF como múltiples unidades de fracaso debido a una sola causa.

CMF son causados por fenómenos que crean dependencias entre unidades redundadas, lo que causa la falla de estas unidades simultáneamente (Dubrova, 2013).

Según como lo indica Dubrova (2013) el único enfoque para combatir los CMF, es mediante el diseño en diversidad. Diseño en diversidad es la implementación de más de una variante de la función en cuestión (Dubrova, 2013). Esto se puede lograr variando los algoritmos que se utilizan, diferentes equipos de trabajo realicen las mismas partes del sistema, de manera tal de tener redundancia en código, etc.

2.3.3. Fallas en el Software

El SW difiere en gran medida con el Hardware (HW). En primer lugar el SW no envejece, no se deforma, tampoco se puede quebrar ni ser afectado por el medio ambiente. El SW es determinístico, siempre responde de la misma manera en el mismo ambiente, al menos que falle.

⁵En inglés, common-mode faults

Por otro lado el SW se lo puede actualizar varias veces a lo largo del su ciclo de vida.

En tercer lugar, arreglar bugs de SW **no significa que el mismo sea más confiable**, al contrario pueden ocurrir nuevos errores (Dubrova, 2013).

Por último el SW es mucho más complejo y menos regular que el HW. Tests tradicionales y métodos de debug pueden ser inadecuados para los sistemas de SW

2.4. Medios de fiabilidad

Los medios de confiabilidad son métodos y técnicas que permiten el desarrollo de un sistema confiable (Dubrova, 2013). Los medios se pueden dividir en dos grandes grupos (Pullum, 2001):

1. Aquellos que son empleados durante el proceso de construcción del SW (Pullum, 2001),
2. y a aquellos que ayudan en la validación del SW después que fue desarrollado (Pullum, 2001).

Dentro del primer grupo se tiene:

- Evitación de Fallas
- Tolerancia a Fallas

Por otro lado, en el segundo grupo se puede mencionar los siguientes:

- Eliminación de Fallas
- Predicción de Fallas

2.4.1. Evitación de Fallas

FA son técnicas de mejoramiento de la fiabilidad utilizadas durante el desarrollo de SW para reducir el número de fallas introducidas durante la etapa mencionada (Pullum, 2001). Estas técnicas pueden estar presentes en las especificaciones y requerimientos del sistema, métodos de diseño de SW (Pullum, 2001).

Dubrova (2013) la denomina *prevención de fallas*⁶, y coincide con el autor anterior, definiendo FA como técnicas de control de calidad durante la especificación y fabricación de los procesos de diseño.

2.4.2. Tolerancia a Fallas

En la sección 2.6 (página 13) se discutirá con mayor detalle la FT. Esto es así ya que en este trabajo de tesis se tiene como principal punto de estudio la Tolerancia a Fallas.

⁶En inglés, Fault prevention

2.4.3. Eliminación de Fallas

La FR hace referencia a las técnicas utilizadas para mejorar la fiabilidad empleadas durante la validación y verificación del SW (Pullum, 2001). Estas técnicas mejoran la fiabilidad del SW mediante la detección de fallas, usando métodos de verificación y la validación, y eliminando las fallas que se van detectando (Pullum, 2001).

Por otro lado Dubrova (2013) indica que el FR se lleva a cabo durante fases de desarrollo de SW tanto como durante el ciclo de vida de un sistema. Durante la fase de desarrollo, FR consiste en tres pasos: *verificación, diagnóstico y corrección* (Dubrova, 2013). FR durante la vida operacional de un sistema, consiste en el mantenimiento preventivo y correctivo del mismo (Dubrova, 2013).

2.4.4. Predicción de Fallas

La FF se realiza mediante la realización de una evaluación del comportamiento del sistema con respecto a la ocurrencia o la activación de una falla (Dubrova, 2013). Esta evaluación puede ser:

- Cualitativa: que tiene como objetivo clasificar los modos de fallas o combinaciones de eventos que llevan al sistema al fracaso (Dubrova, 2013).
- Cuantitativa: que tiene como objetivo evaluar en término de probabilidad, el grado en el cual los atributos de fiabilidad son satisfechos (Dubrova, 2013).

FF incluye técnicas para aumentar la fiabilidad del sistema que son usados durante la validación del SW, con el objetivo de estimar la presencia de fallas y la ocurrencia o consecuencia de fracasos (Pullum, 2001)

2.5. Atributos de la fiabilidad

El objetivo final de la FT es desarrollar un sistema que sea fiable. Fiabilidad tiene muchas definiciones, pero comúnmente es expresado como la probabilidad de **no fallar** (Hitt y Mulcare, 2070). “La fiabilidad es la probabilidad de que un sistema continúe funcionando correctamente durante un intervalo de tiempo particular” (Torres-Pomales, 2000).

La fiabilidad de un sistema SW puede ser descrita por una serie de atributos, los cuales son mencionados a continuación.

2.5.1. Confiabilidad

La confiabilidad es la probabilidad de que un sistema continua operando correctamente durante un intervalo de tiempo dado (Torres-Pomales, 2000).

Dubrova (2013) coincide que la confiabilidad $R(t)$ ⁷ de un sistema es la probabilidad de que el sistema opere sin fracasos en el intervalo de tiempo $[0, t]$.

La confiabilidad es una medida de la entrega correcta del servicio que brinda un sistema (Dubrova, 2013).

⁷En inglés, Reliability

En sistemas críticos como el SW de vehículo espacial, es sumamente necesario que tenga una alta tasa de confiabilidad, ya que por ejemplo, perder el contacto con la nave, podría representar la pérdida de la misión, o una gran cantidad de datos. Otro ejemplo que se puede mencionar es de un satélite geoestacionario de comunicación, la pérdida de este servicio debe ser baja, casi nula (idealmente).

Coincidiendo Pressman (2001), define la confiabilidad como la “probabilidad de tener operaciones libre de fallas de un programa de computadora, en un ambiente específico para un tiempo específico”. El mismo autor también indica que la confiabilidad es la Tiempo Medio Entre Fallas (MTBF)⁸. Donde:

$$MTBF = MTTF + MTTR$$

Tiempo Medio de Fallas (MTTF) es el promedio de tiempo desde que empieza la operación del sistema hasta el tiempo que se produce la primera falla. Tiempo Medio de Reperación (MTTR) es el promedio de tiempo que se requiere para recuperarse, después de un fracaso, al correcto funcionamiento (Hanmer, 2007). MTBF es similar a MTTF, lo único que los diferencia es que MTBF es la suma de MTTF y MTTR. Según Hanmer (2007) MTBF es utilizado para aquellos sistemas que son reperables. Para el caso contrario se utiliza MTTF.

La IEEE (1990) define confiabilidad como “La capacidad del sistema o componente de realizar sus funciones requeridas bajo las condiciones establecidas durante un período de tiempo especificado”.

2.5.2. Disponibilidad

Es la probabilidad de que el sistema esté operando correctamente en un determinado instante de tiempo (Torres-Pomales, 2000). La disponibilidad $A(t)$ de un sistema en el instante de tiempo t es la probabilidad que el sistema esté funcionando correctamente en el instante t (Dubrova, 2013).

Dubrova (2013) realiza una definición matemática de $A(t)$, llamandola también como, *punto de disponibilidad* o *disponibilidad instantánea*. Y la define como:

$$A(T) = \frac{1}{T} \int_0^T A(t) dt$$

Para Hanmer (2007) la disponibilidad del sistema es el porcentaje de tiempo en el que es capaz de llevar a cabo una función determinada.

Para el caso de los sistemas que no pueden ser reparados el punto de disponibilidad es igual a la confiabilidad del sistema (Dubrova, 2013).

Los estados de disponibilidad pueden ser representados en términos de fuera de servicio por año. En la Tabla 2.1 que expone Dubrova (2013) se puede observar esta relación.

La IEEE (1990) define la disponibilidad como “El grado en el cual un sistema o componente se encuentra operativo y accesible cuando se requiere su uso. También es expresado en términos de probabilidad”

En satélites de órbita baja (LEO⁹), es necesario que el satélite se encuentre disponible al momento de su pasada por las estaciones terrestres para poder descargar los datos que se fueron almacenando. Del mismo modo, el satélite debe estar disponible para poder realizar las funciones necesarias para poder cumplir con su misión (como por ejemplo la registración de imágenes en una determinada zona terrestre).

⁸Del inglés, mean-time-between-failure

⁹Del inglés, Low Earth Orbit

Disponibilidad	Fuera de servicio
90 %	36.5 días/año
99 %	3.65 días/año
99.9 %	8.76 horas/año
99.99 %	52 minutos/año
99.999 %	5 minutos/año
99.9999 %	31 segundos/año

Tabla 2.1: Disponibilidad en relación con su baja de servicio por año. Tabla modificada de Dubrova (2013)

Diferente es el caso para los satélites geoestacionarios, ya que estos deberían estar disponible la mayor parte del tiempo, ya que en la mayoría de los casos son de comunicación.

2.5.3. Seguridad

La seguridad se considera como una extensión de la confiabilidad (Dubrova, 2013). Seguridad $S(t)$ es definida como la probabilidad que el sistema sea capaz de realizar su función correctamente o discontinuar su función en una manera a prueba de fallas (Dubrova, 2013).

Según Torres-Pomales (2000) la seguridad es la probabilidad de que el sistema llevará a cabo sus tareas de una manera no peligrosa. Un peligro se lo puede definir como “un estado o condición de un sistema, que juntos con otras condiciones ambientales de el sistema, conducirá inevitablemente a un accidente” (Torres-Pomales, 2000).

La seguridad es requerida para aquellas aplicaciones de seguridad crítica donde un fracaso puede resultar en lesiones humanas, pérdidas de vida o desastres ambientales (Dubrova, 2013).

Para satélites es importante que se tenga una alta seguridad, ya que una pérdida de una misión representa grandes cantidades de dinero perdido.

2.6. Tolerancia a falla

En sistemas críticos, como el de una planta nuclear, sistema médico, el sistema de vuelo de un avión, o el de un satélite, el SW (ni el hardware) deben fallar, ya que esto daría como resultado la pérdida de muchas vidas. Para el caso particular, del vehículo espacial (satélite, transbordador, lanzador), la falla del SW podría tener como consecuencia la pérdida de una misión, y/o una gran cantidad de dinero, y hasta vidas en algunos casos (vuelos tripulados). La principal diferencia entre el SW de una misión satelital, con la de un avión o una planta nuclear, o un sistema médico, es que ante alguna falla o error, se torna complicado llegar hasta el satélite para realizar una actualización o cargar un parche de SW.

La IEEE (1990) define como SW crítico a “aquel cuyo fracaso puede tener un impacto en la seguridad, o puede causar grandes pérdidas financieras o sociales”. El SW de estos sistemas críticos deben tener la capacidad de seguir funcionando, aún en la presencia de fallas, o errores. Imagínese el caso, de un avión comercial, con pasajeros a bordo, y de repente ocurre un problema debido al mal diseño del SW (por ejemplo un overflow de memoria). En esta situación es impensable que el SW se congele y que el piloto reinicie el sistema, esperar que se reestablezca al estado en el cual se encontraba antes del problema, para seguir funcionando. Lo mismo ocurre con el SW de naves espaciales, hay situaciones en la que no se puede esperar y es preferible que el sistema siga funcionando aún en la presencia de fallas.

Tal lo como indica Pressman (2001) las fallas de SW implica problemas cualitativos que son descubiertos después de que el SW es llevado a los usuarios y probados por ellos. Una gran cantidad de estudios indican que en las actividades de diseño se introducen entre un 50 y 65 por ciento de errores del total de errores que se dan durante el proceso del SW (Pressman, 2001). Esto no debe ocurrir en el ámbito espacial, ya que una vez que el sistema es utilizado, es muy difícil corregir los errores que surgen

Cabe aclarar que el SW al no ser un componente físico, no puede ser tratado de la misma manera que un componente hardware. Como ejemplifica Torres-Pomales (2000), las fallas que surgen a nivel de bit, como por ejemplo en un disco duro, son fallas del dispositivo de almacenamiento y pueden ser mitigadas con la aplicación de técnicas de redundancias. Esto no es así para el SW. Por lo tanto evitar los errores en el a nivel de SW no es tan trivial como en el hardware.

A nivel de SW las fallas son llamadas “bugs” (tal como se indica en la sección 2.1 en la página 6), y existe un solo tipo de fallas que es introducido durante el desarrollo del SW (Torres-Pomales, 2000). Las fallas en el SW son el principal motivo de que todo un sistema fracase.

La FT, puede ser utilizada como una capa más de protección (Torres-Pomales, 2000). Esta aplicada al SW se refiere al uso de técnicas que permiten seguir brindando el servicio en un nivel aceptable de performance y seguridad después que una falla de diseño ocurra.

Debe hacerse una diferencia entre FT y calidad. Hanmer (2007) lo define de la siguiente manera: “FT es la capacidad del sistema a ejecutarse apropiadamente a pesar de la presencia de fallas. FT ocurre en tiempo de ejecución”. Cuando se habla que un sistema es tolerante a fallas, significa que fue diseñado de tal manera, que puede seguir funcionando correctamente aún en la presencia de errores de sistemas (Hanmer, 2007).

En cambio calidad, tal como lo define Hanmer (2007), “se refiere a cuán libre de fallas está el sistema. Técnicas de calidad que indican cómo el SW es creado. Si el sistema fue testeado.”

Un sistema de alta calidad tendrá menor número de fallas, que esto representa menor número de fallas en tiempo de ejecución. La reducción del número de fallas no implica que los resultados de los defectos son menos severos (Hanmer, 2007). El sistema debe tomar medidas para reducir el impacto de los errores y fallas, y es allí donde surge la FT.

Un sistema tolerante a fallas provee una continua y segura operación, aún durante la presencia de fallas. Un sistema tolerante a fallas, es un elemento crítico para una arquitectura de vuelo, lo cual incluye hardware, SW, timing, sensores y sus interfaces, actuadores, elementos y datos de comunicación con los diferentes elementos (Hitt y Mulcare, 2070).

Este tipo de sistemas debería detectar los errores causados por fallas, evaluar los daños producidos por la falla, aislar a la misma y por último recuperarse, en ese caso se habla de arquitectura o sistemas FDIR¹⁰.

FT es la capacidad de un sistema a continuar funcionando a pesar de la ocurrencia de fallas (Dubrova, 2013). Un sistema tolerante a fallas debe ser capaz de manejar fallas tanto de hardware como de SW. La FT es necesaria debido a que es imposible construir un sistema perfecto.

El objetivo de la FT es el desarrollo de sistemas los cuales funcionen correctamente en presencia de fallas (Dubrova, 2013). La FT es alcanzada mediante la utilización de algunos tipos de redundancias (Dubrova, 2013). *Redundancia* es la provisión de capacidades funcionales que sería innecesario para entornos libres de fallos (Dubrova, 2013). Esto significa tener hardware adicionales, check bits en una cadena de datos, o algunas líneas de código que verifica el correcto resultado del SW. La redundancia permite enmascarar una falla, o detectarla, para luego localizarla, contenerla y recuperarse de esta (Dubrova, 2013). Las técnicas de tolerancia de fallas se emplean durante la adquisición, o desarrollo del SW. Permite al SW tolerar fallas después que este haya sido desarrollado (Pullum, 2001). Cuando una falla se da, las técnicas de FT proveen mecanismos al sistema de SW

¹⁰FDIR, del inglés: Failure detect, isolate and recover

para prevenir el fracaso del sistema (Pullum, 2001).

2.7. Clasificación de un sistema de control tolerante a fallas

2.8. Redundancia en el software

A pesar de lo comentado anteriormente, sobre la importancia del SW, todavía existe una creencia, de que el SW aparece por arte de magia, y que los programadores no son nunca, lo suficientemente capaces, de hacer un SW libre de errores. Salvo aquellas empresas u organizaciones que tienen un proceso maduro de desarrollo de SW, el resto cae en el error de pensamiento mencionado anteriormente.

Dubrova (2013) explica que la FT aplicado en el SW no está tan entendido, ni maduro, como es en el caso de la FT aplicada en hardware. Si una falla existiera en el SW, esta se haría “visible”, solo cuando las condiciones relevantes ocurran (Dubrova, 2013). Y muchas veces por tiempo o costo, no se realizan los tests cubriendo todos los posibles ambientes reales, lo cual tiene consecuencias desastrosas, tal como se expone en la sección 1.1 (página 3).

Para sistemas complejos o grandes, donde existe una gran cantidad de estados, implica que solo una pequeña porción del SW puede ser verificada correctamente (Dubrova, 2013). Los tests tradicionales y métodos de depuración actuales, no alcanzan para grandes sistemas (Dubrova, 2013). La utilización de métodos formales para describir las características requeridas por el comportamiento del SW, exigen gran complejidad computacional, y solo son aplicables en ciertas situaciones (Dubrova, 2013).

Las técnicas de FT pueden dividirse en dos grupos:

- técnicas de una sola versión, se utilizan cuando existe una sola versión del SW en el sistema.
- técnicas multi-versión, se utilizan cuando se desarrollan varias versiones de una misma función.

Estas se explican en las siguientes secciones.

2.8.1. Técnicas single version

Estas técnicas son utilizadas para tolerar parcialmente las fallas del diseño de SW (Pullum, 2001). Técnicas single-version de FT se basa en el uso de redundancia aplicada a una única versión de una pieza de SW para detectar y recuperarse de fallas (Torres-Pomales, 2000).

Estas técnicas a los software que cuentan con una sola versión, un número capacidades funcionales que no serían necesarias dentro de un ambiente libre de fallas (Dubrova, 2013).

2.8.1.1. Estructuras de software

En Torres-Pomales (2000) se mencionan dos técnicas de estructuración del SW que son muy buenas a la hora de mantener FT en el SW.

La definición de una arquitectura en el software es de suma importancia ya que proveen las bases para la implementación de FT (Torres-Pomales, 2000). Una de las técnicas utilizadas en el desarrollo del software es la modularización. Esta consiste en descomponer el problema en componentes manejables. Esto tiene como resultado que sea más eficiente la aplicación de la FT en el diseño de un sistema (Torres-Pomales, 2000).

El particionado es otra técnica mencionada en Torres-Pomales (2000), lo cual provee aislación entre módulos independientes del sistema. Esta técnica permite descomponer al problema en partes separadas (Pressman, 2001). El particionado puede ser horizontal u vertical. En el primero se descomponen el problema moviéndose en forma horizontal en la jerarquía, mientras que el segundo se parte de lo más general hasta llegar a lo detallado, moviéndose verticalmente en la jerarquía (Pressman, 2001).

Sistema de cierre es un principio de FT, en el cual ninguna acción es permitible sin una autorización expresa (Torres-Pomales, 2000). Siguiendo este principio ninguna de las funciones que componen al sistema deberían tener más capacidad de la necesaria (Torres-Pomales, 2000). Las ventajas de desarrollar un sistema bajo este principio, es que es sencillo el manejo de errores, y evitar la propagación de fallas si ocurriesen.

2.8.2. Técnicas multi-version

Las técnicas de multi-version utilizan dos o más versiones diferentes del mismo módulo de SW (Dubrova, 2013) (Torres-Pomales, 2000), lo cual satisface el requerimiento de diversidad.

El objetivo de utilizar diferentes versiones de SW es que es construido de diferentes maneras, por lo tanto fallarían de diferente maneras (Torres-Pomales, 2000).

2.8.3. Técnicas de detección de fallas

Para los SW tolerantes a fallas, de una sola versión, se suelen utilizar varios tests de “aceptación” para detectar fallas (Dubrova, 2013). Es necesario que estos SW cuenten con dos propiedades: auto protección¹¹ y auto check¹² (Torres-Pomales, 2000). La auto protección significa que los componentes de sistema tienen la capacidad de protegerse así mismo mediante la detección de errores (Torres-Pomales, 2000). La propiedad de auto check significa que los componente son capaces de detectar fallas internas y tomar las acciones necesarias para evitar la propagación del error.

El resultado del sistema depende del resultado de los tests. Si el resultado pasa exitosamente el test, este es el correcto, caso contrario significa la presencia de fallas (Dubrova, 2013). Un test es más efectivo si se puede calcular de una manera simple (Dubrova, 2013).

Las técnicas utilizadas son las siguientes:

- *Timing checks*: se agrega a los sistemas una restricción de tiempo. Basado en esa restricción se puede deducir si el comportamiento del sistema se desvió (Dubrova, 2013). Los más utilizado es el *watchdog timer*, este es un contador, actualizado con un *timer* que detecta si un módulo de SW se bloqueó o congeló, entonces se reinicia ese módulo o el sistema.
- *Coding checks*: se utiliza en los sistemas donde los datos se codifican usando técnicas de redundancia de datos (Dubrova, 2013).
- *Reversal checks*: son aquellos donde se toma los valores de salida, y con ellos se busca encontrar cuáles fueron los datos de entrada. Si los datos de entrada reales coinciden con los calculados (para una misma salida), este se encuentra libre de fallas (Dubrova, 2013).
- *Reasonableness checks*: usa propiedades semánticas en los datos para detectar fallas (Dubrova, 2013).

¹¹En inglés, self-protection

¹²En inglés, self-checking

- *Structural checks*: se basa en el conocimiento de las propiedades de la estructura de datos (Dubrova, 2013).
- *Replication checks*: se basa en la comparación de resultados de varios componentes (Torres-Pomales, 2000).

Se suelen utilizar árboles de fallas, como una técnica auxiliar en el desarrollo de sistemas para la detección de fallas (Torres-Pomales, 2000). El árbol de falla permite obtener un enfoque top-down de las diferentes fallas que se pueden dar. El árbol no cubre todas las fallas que puedan darse, pero si ayudan en un alto grado en el desarrollo de SW tolerante a fallas (Torres-Pomales, 2000).

2.8.4. Técnicas de recuperación de fallas

Una vez que la falla es detectada, el sistema debe proceder a recuperarse de aquella, y volver a un estado operacional normal (Dubrova, 2013). Si los mecanismos de detección y contención de fallas fueron desarrollados correctamente, esta es contenida dentro de un set de módulos en el momento de la detección (Dubrova, 2013).

2.8.4.1. Manejo de excepciones

En muchos SW y lenguajes de programación, se logra recuperarse mediante el manejo de excepciones. El manejo de excepciones es la interrupción del funcionamiento normal para responder a un funcionamiento anormal del sistema (Torres-Pomales, 2000). Los posibles eventos que pueden lanzar una excepción son:

1. Excepciones de interfaces, son lanzadas por un módulo cuando se da una solicitud inválida de algún servicio (Dubrova, 2013).
2. Excepciones locales, son lanzadas por algún módulo cuando sus propios mecanismos de detección de fallas encuentran un problema interno (Dubrova, 2013).
3. Excepciones de fracaso, son lanzadas cuando un mecanismos de detección encuentra una falla, pero es imposible recuperarse de esa falta (Dubrova, 2013).

2.8.4.2. Checkpoint y Restart

Para los software de una sola versión existen pocos mecanismos de recuperación. Checkpoint y restart es uno de ellos. También es conocido como *backward error recovery* (Dubrova, 2013). La mayoría de las fallas que se dan en los SW son debido a fallas que provienen del diseño, tal como se mencionó anteriormente. Estas fallas son activadas por entradas al sistema (Dubrova, 2013).

Este mecanismo cuenta con el módulo principal que se encuentra en ejecución combinado con un bloque que realiza tests de aceptación. Si se detecta una falla, en el bloque de testeo, se envía una señal de “reinicio”, para que el módulo principal vuelva al estado anterior, es decir, antes de producirse el error. Este estado anterior se encuentra almacenado en una memoria checkpoint (Dubrova, 2013). En la figura ??¹³, se muestra la representación de este mecanismo.

[INLINE]Agregar imagen

[INLINE]Agregar imagen

¹³Basado en Dubrova (2013) y Torres-Pomales (2000)

2.8.4.3. Diversidad de datos

La diversidad es una técnica utilizada para mejorar la eficiencia en los checkpoint y restart, usando diferentes entradas por cada reinicio (Dubrova, 2013). Esto se basa en que las fallas en el SW son dependientes de las entradas (Dubrova, 2013). Es poco probable que la misma falla se de con la misma secuencia de entrada (Dubrova, 2013).

2.8.4.4. Bloques de recuperación

Esta técnica combina las bases de la técnica de checkpoints y restart enfocada con múltiples versiones de un componente de SW en el sentido de que una versión de SW diferente es lanzada cada vez que se encuentra una falla (Torres-Pomales, 2000). Los checkpoints son creados antes de que una versión de SW se ejecuta (Torres-Pomales, 2000). La ejecución de las múltiples versiones pueden ser secuencial o paralelas dependiendo de la disponibilidad de la capacidad de procesamiento y performance requerida (Torres-Pomales, 2000).

La representación de esta técnica se puede observar en el figura [AGREGAR IMAGEN]. Las versiones son diferentes implementaciones de un mismo programa. Solo uno de estas versiones provee la salida del sistema. Si un error es detectado por el test de aceptación, se vuelve hacia atrás, se retoma el último checkpoint, y se vuelve a ejecutar el módulo de SW pero con una versión diferente a la que se ejecutó anteriormente (Dubrova, 2013).

Los checks del test de aceptación deben mantenerse simples para mantener la velocidad de la ejecución (Dubrova, 2013).

[INLINE]Agregar imagen

2.8.4.5. Programación N-version

2.8.4.6. Programación N-Auto Checking

2.9. Técnica de evaluación de fiabilidad

La evaluación de la fiabilidad es de suma importancia para el desarrollo de sistemas críticos, ya que permite identificar que aspectos del comportamiento del sistema juega un papel importante (Dubrova, 2013).

1. Modelado de un sistema en la fase de diseño.
2. Aseguramiento del sistema en la fases finales de desarrollo (testing).

El análisis confiabilidad tiene tres enfoques importantes:

- Confiabilidad de HW
- Confiabilidad de SW
- Confiabilidad humana

En este trabajo de tesis se pondrá énfasis en el estudio en la confiabilidad del SW a nivel de sistema.

La evaluación de la fiabilidad tiene dos aspectos. En primer lugar se tiene una *evaluación cualitativa* que permite identificar, clasificar y medir modos de fallas, o eventos combinacionales que puedan provocar una falla. El otro aspecto es la *evaluación cuantitativa*, la cual permite evaluar en términos de probabilidad los atributos de la fiabilidad (Sección 2.5), disponibilidad, seguridad.

El análisis de confiabilidad es de gran importancia ya que provee información que es la base de la toma de decisión. Esto es aplicado a diferentes áreas, tales como análisis de riesgos, protección ambiental, calidad, optimización de mantenimientos y operaciones y diseño de ingeniería (Rausand y Hoyland, 2004).

2.10. Medidas comunes de fiabilidad

Las medidas de fiabilidad más comunes son las siguientes: failure rate, tiempo medio a la falla, tiempo medio de reparación y tiempo medio entre fallas.

2.10.1. Failure rate

Failure rate λ es el número esperado de fallas por unidad de tiempo (Dubrova, 2013). Es usual utilizar la dimensión *fallas/horas*.

Generalmente, λ se encuentra a nivel de componente. Para conocer el failure rate del sistema completo, se puede realizar (a groso modo) una sumatoria de los λ de los componentes que integran el sistema.

$$\lambda = \sum_{i=1}^n \lambda_i$$

La evolución de λ a través del tiempo, no tiene el mismo comportamiento tanto para HW como para SW. Si se divide el ciclo de vida de un sistema en las siguientes fases: mortalidad prematura (I), vida útil (II), desgaste (III) (Dubrova, 2013) se aprecia, para el caso del HW, lo que se denomina *curva de la bañera* la cual puede observarse en la Figura 2.2. En una primera fase, λ decrece, ya que a través de los procesos de testing se van descubriendo y resolviendo los errores. Luego se da un periodo de estabilización. Y al final, el HW sufre el paso del tiempo, y se desgasta, aumentando la tasa de fallas.

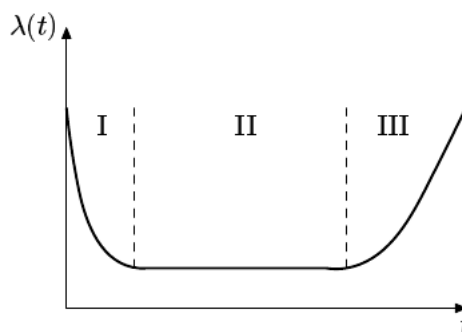


Figura 2.2: Failure rate de HW vs tiempo

Para el SW es totalmente diferente. En primer lugar cuando se realiza una actualización, se aumenta la complejidad, como así también la probabilidad de fallas, con ello el failure rate. Otra diferencia sustancial con el HW es que el SW no se desgasta con el tiempo. En la Figura 2.3 se aprecia a través de tiempo. Esta curva

suele llamarse curva serrucho. El failure rate del SW decrece en función del tiempo. En estos tipo de sistemas la tasa de falla depende de varios factores como pueden ser el proceso utilizado en el diseño y codificación, complejidad del SW, tamaño del SW, etc (Dubrova, 2013).

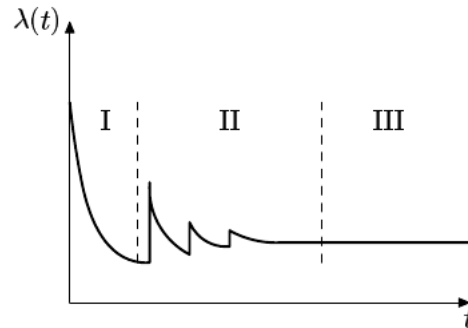


Figura 2.3: Failure rate SW vs tiempo

A lo largo de la vida de sistema se supone el failure rate λ como constante. Por lo tanto la confiabilidad del sistema varía exponencialmente con respecto al tiempo (Dubrova, 2013):

$$R(t) = e^{-\lambda t}$$

Esto se conoce como *ley de la falla exponencial* (Dubrova, 2013). El gráfico de confiabilidad $R(t)$ vs tiempo se muestra en la Figura 2.4.

Figura 2.4: Confiabilidad vs tiempo

2.10.2. Tiempo medio medio de falla

El *tiempo medio de falla* (MTTF¹⁴) de un sistema es el tiempo esperado que transcurra hasta la primera falla que se detecte en el sistema. En terminos de confiabilidad, MTTF se define de la siguiente manera (Dubrova, 2013) (Rausand y Hoyland, 2004)

$$\int_0^{\infty} R(t) dt$$

2.10.3. Tiempo medio de reparación

El *tiempo medio de reparación* (MTTR¹⁵) de un sistema, es el promedio de tiempo que se requiere para reparar al sistema. MTTR se especifica en términos de la tasa de reparación μ (Dubrova, 2013) (Rausand y Hoyland, 2004), el cual es el número esperado de reparaciones por unidad de tiempo:

$$MTTR = \frac{1}{\mu}$$

El MTTR depende de los mecanismos de recuperación ante fallas que se utilicen en el sistema, localización del sistema, scheduler de mantenimiento (Dubrova, 2013). Con esto se puede definir la disponibilidad como

¹⁴Del inglés, Mean Time To Failure

¹⁵Del inglés, Mean Time To Repair

sigue:

$$A(\infty) = \frac{MTTF}{MTTF + MTTR}$$

Muchas veces se utiliza MDT (¹⁶) en vez de MTTR, para denotar más claro que es el tiempo medio que el sistema se encuentra fuera de servicio.

2.10.4. Tiempo medio entre fallas

El tiempo medio entre fallas (MTBF¹⁷) de un sistema es el tiempo promedio entre dos fallas del sistema.

$$MTBF = MTTF + MTTR$$

2.10.5. Cobertura de fallas

La cobertura de fallas es la probabilidad de que el sistema no interrumpirá su actividad cuando una falla se presente. En términos matemáticos la cobertura de fallas la probabilidad condicional $P(A|B)$. Existen diferentes coberturas de fallas, dependiendo de si se está tratando con detección de fallas, localización de fallas, contención de fallas o recuperación de fallas (Dubrova, 2013). Siendo A detección, localización, contención o recuperación de fallas, y B la existencia de fallas.

2.11. Métodos de cálculos de fiabilidad

Para evaluar la fiabilidad de sistemas se pueden utilizar diagramas de bloque de confiabilidad y procesos de Markov (Dubrova, 2013)

2.11.1. Diagramas de bloques de confiabilidad

2.11.1.1. Cálculo de confiabilidad

Para medir la confiabilidad de un sistema mediante diagrama de bloques, se debe dividir el sistema objetivo en partes paralelas y en serie. Se computa la confiabilidad de las partes. La confiabilidad del sistema estará compuesta por la confiabilidad de ambas partes (Dubrova, 2013). Entonces

$$R(t) = \begin{cases} \prod_{i=1}^n R_i(t) & \text{para estructuras en serie} \\ 1 - \prod_{i=1}^n (1 - R_i(t)) & \text{para estructuras en paralelo} \end{cases}$$

Esto nos indica que un sistema paralelo, es más confiable que uno en serie, aún así si sus componentes son menos confiables. Tal como ejemplifica Dubrova (2013), si se diseña un sistema en serie de 100 componentes, con una confiabilidad de 0.999, el sistema completo tendrá una confiabilidad de: $0,999^{100} = 0,905$. Mientras que, para un sistema paralelo, con solo cuatro componentes, con una confiabilidad menor (0.95) la confiabilidad del sistema será $1 - (0,95)^4 = 0,99999375$. El punto en contra de los sistemas paralelos, es que representan un costo mayor que las estructuras en serie (Dubrova, 2013).

¹⁶Del inglés, Meann Downtime

¹⁷Del inglés, Mean Time Between Failure

2.11.1.2. Cálculo de disponibilidad

Si se asume que el tiempo de falla y de recuperación son independientes, entonces se puede utilizar diagramas de bloques para calcular la disponibilidad del sistema (Dubrova, 2013). Se puede observar que el cálculo es similar al cálculo de confiabilidad.

$$A(t) = \begin{cases} \prod_{i=1}^n A_i(t) & \text{para estructuras en serie} \\ 1 - \prod_{i=1}^n (1 - A_i(t)) & \text{para estructuras en paralelo} \end{cases}$$

2.11.2. Utilización de procesos de Markov

El principal objetivo del análisis de los procesos de Markov es calcular $P_i(t)$ la probabilidad de que el sistema se encuentre en el estado i en el tiempo t . Con esto se puede calcular fácilmente la confiabilidad, disponibilidad o seguridad del sistema (Dubrova, 2013).

Para determinar $P_i(t)$ se debe derivar una serie de ecuaciones diferenciales, una por cada estado del sistema. Estas ecuaciones se denominan ecuaciones de estado de transición. Las ecuaciones de los estados de transición se representan en una matriz M denominada *matriz de transición*. Cada elemento m_{ij} de la matriz M es un rate de transición entre los estados i y j

$$M = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{k1} \\ m_{12} & m_{22} & \dots & m_{k2} \\ & & \dots & \\ m_{1k} & m_{2k} & \dots & m_{kk} \end{bmatrix}$$

Haciendo $P(t)$ un vector en el cual el i -ésimo elemento es la probabilidad $P_i(t)$ de que el sistema se encuentra en el estado i en el tiempo t . Con ello se tiene:

$$\frac{d}{dt}P(t) = MP(t)$$

(Dubrova, 2013).

Para calcular confiabilidad, disponibilidad y seguridad solo es necesario reemplazar en la matriz M los rate correspondientes.

2.12. Modelos de falla

En esta sección se explica con un poco más de detalle algunos conceptos que se presentaron en secciones anteriores. Los conceptos que se detallarán son los siguientes:

- Función de confiabilidad $R(t)$
- Función de tasa de falla λ
- Tiempo medio hasta la falla (MTTF)
- Vida residual media (MRL)

También existen varias distribuciones de probabilidad que son utilizados para modelar el tiempo de vida de componentes de sistemas. Algunos de estos modelos de distribución del tiempo de vida son los siguientes:

- Distribución binomial
- Distribución exponencial
- Distribución gamma
- Distribución Weibull
- Distribución normal
- Distribución Birnbaum-Saunders
- Distribución inversa de Gauss

2.12.1. Tiempo hasta la falla

El tiempo hasta la falla (T) de un componente del sistema, es el lapso de tiempo desde que el componente empieza a operar hasta la primera falla que se produzca (Rausand y Hoyland, 2004).

Nota: Se asume que el componente que falla, no puede ser recuperado. Por otro lado, si un sistema falla, puede recuperarse (a través de técnicas de reconfiguración, componentes back-up).

T no significa que es solo una medida de tiempo. T puede significar (Rausand y Hoyland, 2004):

- Cantidad de kilómetros de un automóvil
- Número de rotación de un motor
- Número de ciclos de trabajo

Debe destacarse que esta no solo es una variable discreta. Una variable discreta puede ser aproximada a una variable continua. Suponiendo que T es una distribución continua con densidad de probabilidad $f(t)$ y una función de distribución:

$$F(t) = P(T) = \int_0^t f(u) du \text{ para } t > 0$$

$F(t)$ demuestra la probabilidad de que un componente falle dentro de un intervalo de tiempo $(0, t]$ (Rausand y Hoyland, 2004).

La función de densidad de probabilidad $f(t)$ se define como sigue (Rausand y Hoyland, 2004):

$$f(t) = \frac{d}{dt}F(t) = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t}$$

Esto indica que para un Δt pequeño,

$$P(t < T \leq t + \Delta t) \approx f(t)\Delta t$$

2.12.2. Función de confiabilidad

La función de confiabilidad se define como sigue:

$$R(t) = 1 - F(t) = P(T > t)$$

para $t > 0$.

2.12.3. Tasa de falla

La probabilidad de que un componente falle en un intervalo de tiempo $(t, t + \Delta t]$ dado que el componente ha funcionado hasta t , se tiene:

$$P(t < T \leq t + \Delta t | T > t) = \frac{P(t < T \leq t + \Delta t)}{P(T > t)} = \frac{F(t + \Delta t) - F(t)}{R(t)}$$

Si se divide esta probabilidad por un intervalo de tiempo Δt , y haciendo $\Delta t \rightarrow 0$, se obtiene la función de tasa de falla.

2.12.4. Tiempo medio hasta la falla

El MTTF de un componente está definido por (Dubrova, 2013) Rausand y Hoyland (2004):

$$MTTF = E(T) = \int_0^{\infty} t f(t) dt$$

Se demuestra en Rausand y Hoyland (2004) que

$$MTTF = \int_0^{\infty} R(t) dt$$

2.12.5. Vida restante media

Considerando un componente con tiempo hasta la falla T , que es colocado en operación en el tiempo $t = 0$ y continúa funcionando en el instante t . La probabilidad de que el componente en de edad t sobrevive en un intervalo x , entonces se tiene (Rausand y Hoyland, 2004):

$$R(x|t) = P(T > x + t | T > t) = \frac{P(T > x + t)}{P(T > t)} = \frac{R(x + t)}{R(t)}$$

Entonces:

$$MRL(t) = \frac{1}{R(t)} \int_t^{\infty} R(x) dx$$

Nótese que cuando $t = 0$, el componente es nuevo, y entonces $\mu(0) = \mu = MTTF$.

2.12.6. Distribución binomial

La distribución binomial es uno de lo más utilizados (Rausand y Hoyland, 2004). Esta distribución es utilizada en los siguientes casos:

- Cuando se tienen n ensayos independientes
- Cada ensayo tiene dos posibles resultados
- La $P(A) = p$ es la misma en todos los ensayos

Entonces

$$P(X = x) = \binom{n}{x} \cdot p^x (1 - p)^{n-x}$$

Donde $\binom{n}{x}$ es el coeficiente binomial y X es el número de n ensayos para alcanzar un resultado A (Rausand y Hoyland, 2004).

2.12.7. Distribución exponencial

Considerando un componente que entra en operación en el instante $t = 0$. El tiempo hasta la falla T del componente tiene una función de densidad de probabilidad de la siguiente forma Rausand y Hoyland (2004):

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{para } t > 0, \lambda > 0 \\ 0 & \text{para cualquier otro caso} \end{cases}$$

Esta distribución es la que se denomina distribución exponencial con parámetro λ .

La confiabilidad de esta función es:

$$R(t) = P(T > t) = \int_t^\infty f(u) du = e^{-\lambda t} \text{ para } t > 0$$

El MTTF es:

$$MTTF = \int_0^\infty R(t) dt = \int_0^\infty e^{-\lambda t} dt = \frac{1}{\lambda}$$

La tasa de falla de esta distribución es λ

La distribución exponencial es lo que se utiliza más comúnmente en los análisis de confiabilidad.

El MTTF es:

$$MTTF = \int_0^\infty R(t) dt = \int_0^\infty e^{-t} dt = \frac{1}{\lambda}$$

Si hacemos:

$$R(MTTF) = R\left(\frac{1}{\lambda}\right) = e^{-1} \approx 0,3679$$

se puede calcular la probabilidad de que un componente sobreviva durante el tiempo medio hasta la falla (Rausand y Hoyland, 2004).

Mientras que la tasa de falla es:

$$z(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

NOTA IMPORTANTE: Como se mencionó anteriormente, existen numerosas distribuciones de probabilidad para medir el tiempo de vida de los componentes. En este trabajo de tesis, se trabaja únicamente con la distribución exponencial, debido a que es una de las más sencillas, y la más utilizada en la bibliografía.

2.13. Protocolos de comunicación de tiempo real

2.13.1. Sistemas de tiempo real

Un sistema de tiempo real es un sistema de computadora que depende del correcto comportamiento funcional con respecto al dominio del tiempo, esto significa que el sistema debe llevar a cabo sus funciones y obtener

como resultados (correctos) dentro de las restricciones de tiempo (Lisner, 2007).

Los sistemas de tiempo real se dividen en *sistemas de tiempo real "soft"* y *sistemas de tiempo real "hard"*. En los soft real-time systems es importante cumplir con los tiempos del planificador, pero el no cumplimiento no tiene un impacto en la seguridad del sistema. Por otro lado, en los hard real-time systems, el no cumplimiento de las restricciones del tiempo, tiene como consecuencia un impacto catastrófico en el sistema.

Existen numerosos protocolos de comunicación, la mayoría abocados en la implementación de la capa 1 y 2 del modelo de ISO/OSI. Estos protocolos especifican las restricciones de hardware, la topología de red, la arquitectura del nodo, acceso al medio, los mecanismos de detección de error, etc. (Lisner, 2007). En los sistemas críticos como en el caso de los vehículos satelitales, es necesario la aplicación de sistemas de tiempo real, que garantice una máxima latencia y un comportamiento determinístico.

La tolerancia a fallas aplicadas en protocolos de comunicación de tiempo real, permiten que la red continúe funcionando aún cuando algunos de sus componentes fallaron. Los sistemas tolerantes a fallas son diseñados a partir de un modelo de falla dado (Lisner, 2007). El modelo de falla describe la estructura del sistema y los tipos de fallas que pueden ocurrir (Lisner, 2007).

Otro aspecto importante de los protocolos de comunicación tolerantes a fallas de tiempo real, es conocer el momento en el que un nodo está disponible para enviar mensajes. Esto es llevado a cabo por las estrategias de acceso al medio. Los métodos de acceso al medio pueden ser clasificados como event-triggered y time-triggered. En el primero, un mensaje es enviado, si algún nodo requiere ese mensaje. Este no es una buena opción para los sistemas de tiempo real, ya que no asegura los límites de tiempo (Lisner, 2007).

Los métodos time-triggered es hecho en ciclos. Los nodos tienen acceso al medio a través de intervalos periódicos de tiempo (Lisner, 2007). Una de las principales ventajas de este enfoque, es que es predecible, los intervalos son pre configurados. Como desventaja se puede mencionar que se hace un uso deficiente del medio, cuando algún nodo no tiene nada para enviar en su ventana de tiempo.

2.14. Estrategia de acceso al medio

2.14.1. CSMA

El esquema CSMA (Acceso Múltiple por Detección de Portadora¹⁸) proviene de ideas del Ethernet alámbrico. En esta técnica los nodos esperan durante un intervalo corto de tiempo y aleatorio antes de transmitir (Tanenbaum, 2003). Este método es utilizado comúnmente en redes inalámbricas. Al protocolo CSMA se la puede dividir en:

- CSMA con Detección de Portadora
- CSMA con Detección de Colisiones

En la primera, el nodo intenta enviar un mensaje cuando el canal está inactivo. Si otro nodo lo está usando espera hasta que se desocupa y luego transmite. En el caso de una colisión de mensajes, espera un tiempo aleatorio antes de enviar el mensaje nuevamente. Esta tiene algunas desventajas, que no la hacen apropiadas para su aplicación en sistemas críticos. Para hacer frente a las desventajas del CSMA con detección de portadora, se utiliza CSMA/CD (CSMA con Detección).

Por otro lado, también existe otro método denominado CSMA/CA (CSMA con Evitación de Colisiones), utilizado en el 802.11. Este protocolo que es similar al CSMA/CD de Ethernet, con detección de canal antes

¹⁸Del inglés, Carrier Sense Multiple Access

de enviar y retroceso exponencial desopués de las colisiones (Tanenbaum, 2003). Esta estrategia además de ser utilizada en el Wireless, se utiliza en CAN.

2.14.2. TDMA

TDMA (time division multiple access) es una técnica de multiplexación del tiempo que es utilizada en redes inalámbricas, comunicación de satélites y en diferentes protocolos de tiempo real (Lisner, 2007).

Los accesos al medio se realiza mediante ciclos, en el cual se subdivide dentro de slots de tiempo de ancho estático. Solo un nodo está permitido enviar en un solo slot.

2.14.3. Minislotting

El uso de slots de tiempo con un ancho fijo y estático del tiempo, como es el que se utiliza en TDMA, puede convertirse en un problema. En el TDMA los nodos tiene que esperar todo un ciclo para poder enviar un mensaje. Además, el nodo debe enviar mensajes vacios (*null*) durante su slots, cuando no tiene nada que enviar. Solución de este problema es el minislotting, ya que permite la utilización de slots con ancho variable. Esto logra recortar los tiempos de espera cuando no se está utilizando el medio (Lisner, 2007). Este método permite un uso más eficiente de los ciclos. La transmisión puede tener diferntes anchos, y no hay necesidad de enviar mensajes null. Este método se basa en la prioridad y para evitar la monopolización del medio, algunos protocolos como el ARIN 629, utiliza timeouts para denegar el acceso al medio después de un determinado tiempo (Lisner, 2007).

2.15. Revisión de protocolos

2.15.1. CAN

CAN es un protocolo de comunicación desarrollado por Bosch para ser utilizado en automóviles. Existe un estándar de ISO para CAN. CAN es un protocolo del tipo event-triggered y utiliza CSMA/CA.

CAN utiliza diferentes mecanismos de detección de errores. Utiliza un CRC de 15 bits. Cada nodo puede enviar un mensaje de diagnóstico de errores, además de que lleva un contador de errores propios. Si este número de errores de grande, el mismo nodo entra en modo *error activo* (envía flags de “error activo”), *error pasivo*(envía flags de “error pasivo” y espera 8 bit times antes de repetir el envío del mensaje) y *bus off*(el nodo se apaga) (Lisner, 2007)

2.15.2. byteflight

Este se basa en el protocolo Minislotting y utilizado por BMW en automóviles. Este utiliza un clock master de alta presición para la sincronización.

Tanto CAN como byteflight, no poseen un mecanismo o técnica para proteger el canal de alguna falla de los nodos (Lisner, 2007).

2.15.3. ARINC 659 o SAFEbus

SAFEbus es una implementación del estandar ARINC 659 para su utilización en aviones comerciales. También se suele utilizar en vehículos espaciales. SAFEbus es una arquitectura que es altamente redundante.

2.15.4. TTP/C

TTP/C proviene de la familia TTP. Este protocolo utiliza el esquema TDMA en una arquitectura de doble canal. La configuración se encuentra pre definida. Esto permite al sistema ser predecible, lo cual facilita la tolerancia a fallas. Su alto grado de tolerancia a fallas, permite desarrollar aplicaciones más confiables que otros protocolos (Lisner, 2007).

2.16. Posibles fallas en una red

Las principales fallas que se pueden producir en una red son debidas al canal de comunicación y los nodos/controladores, esto se debe a que son los componentes más importantes de toda red (Lisner, 2007).

Las fallas en el nodo pueden ser múltiples tipos. El nodo puede codificar erróneamente un mensaje; se pueden dar errores en el timing, es decir actúan en tiempo equivocado; el nodo puede tener un comportamiento inusual y se bloquea el servicio que está brindando.

Asimismo, las fallas en el canal de comunicación pueden ser de diferentes tipos. El canal puede generar ruido, este ruido puede modificar un mensaje (hasta el momento un mensaje correcto); el mensaje no llega a los destinatarios.

También se pueden observar que tanto errores en el nodo como en el canal se produzcan simultáneamente (Lisner, 2007).

2.17. Protocolo CAN

2.17.1. Consideraciones previas

En este trabajo de tesis se decidió continuar el desarrollo de la arquitectura con BUS-CAN ya que este es de interés para los proyectos desarrollados en INVAP. Cabe aclarar que el objetivo principal de esta tesis no es llevar a cabo una comparación exhaustiva de protocolos de comunicación tolerantes a fallas, sino elegir la más indicada y que se adapte a las necesidades de INVAP. Por tal motivo, se expone, en esta sección, las características importantes de CAN.

2.17.2. Introducción

El bus CAN (Controller Area Network) fue desarrollado por Bosch. Comenzó su desarrollo en 1983 y tuvo su primer release en 1986. Fue estandarizada por la Organización Internacional de Estandarización (ISO) bajo el nombre de ISO 11898. El Bus CAN surgió como respuesta al rápido crecimiento de la electrónica en la industria automotriz (ESD Electronics, 2017). Este protocolo se definió con el objetivo de proveer comunicación determinística de sistemas distribuidos, y que necesiten un alto grado de confiabilidad. CAN permite la conec-

tividad vía bus serial. El bus está compuesto por 2 cables que pueden estar blindados o no (ESD Electronics, 2017).

Las características de CAN que la convierten en una tentadora opción para la aplicación en el sector espacial, son:

- Bajo costo
- Operabilidad en ambientes eléctricos complicados
- capacidades de tiempo real.
- facilidad en el uso

La especificaciones de CAN, originalmente desarrollados por Bosch cubre solamente las capas físicas y de enlaces de datos del modelo de referencia de ISO/OSI. Luego de la estandarización ISO provee detalles adicionales de la capa física de CAN.

A diferencia de otros protocolos como USB o Ethernet, CAN no envía grandes bloques de datos de un punto a otro. CAN envía mensajes cortos como temperatura, o RPM (Revoluciones por minutos), y son enviados como broadcast a todo el sistema (Corrigan, 2002).

El protocolo de comunicación CAN, en su estandar ISO 11898, explica como la información viaja entre los dispositivos conectados a una arquitectura siguiendo el modelo de OSI. La arquitectura planteada por la ISO 11898 se puede observar en la FIGURA (Corrigan, 2002)

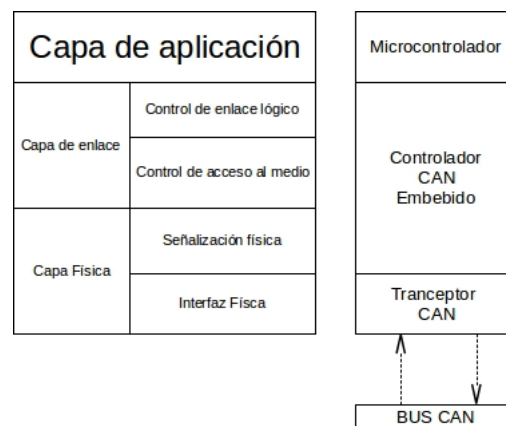


Figura 2.5: Arquitectura estándar propuesta por ISO898

En resumen, la capa física de CAN describe la definición del bit timing, bit encoding y la sincronización, los niveles aceptables de la señal (corriente y voltaje), los conectores y características físicas de los cables (Corrigan, 2002). Por otro lado la capa de enlaces de datos, provee todos los servicios necesarios para la transmisión de un stream de bits desde un nodo a otro.

El bus CAN es un bus de tipo broadcast, es decir que todos los nodos escuchan todas las transmisiones. No hay manera de enviar un mensaje a un determinado nodo, por lo tanto no es necesario el direccionamiento de nodos. (Kvaser, 2017).

Actualmente la industria aeronáutica la utiliza. También el área agrícola en sus maquinarias, control de tráfico, sistemas de control industrial, sistemas domóticos y además, en equipamientos médicos.

2.17.3. Elementos necesarios

Como se puede observar en la Figura 2.5

Cada nodo conectado a la red necesita de

- Central Processing Unit (CPU), o microprocesador, que decida que mensajes debe recibir y que debe enviar
- Controlador CAN, muchas veces integrado en el controlador, el cual recibe la serie de bits del bus. También es el encargado de transmitir mensajes (cuando se requiera), esto significa que al mensaje lo convierte en una serie de bits.
- Transceiver, definido en el ISO 11898.

2.17.4. Capa física

El estándar de CAN define el bit encoding, el timing, y la sincronización. La capa física es la encargada de convertir 1 y 0 en pulsos eléctricos para enviar mensajes por el canal, y también el proceso inverso cuando recibe mensajes. La capa física es implementada enteramente en HW (Corrigan, 2008).

2.17.4.1. Bus de comunicación

Para iniciar una transmisión de mensaje es necesario como mínimo dos nodos conectados al bus CAN. Esto se debe a que el dispositivo que envía un mensaje está también recibiendo su propio mensajes, con esto puede chequear cada bit que ha enviado. De esta manera, un segundo nodo responde con un ACK mientras el bit todavía está siendo transmitido por el primer nodo. Esto demuestra por qué se necesitan de dos nodos para completar la transmisión de mensajes (Corrigan, 2008).

En la Figura 2.6 se observa un ejemplo extraído de (Corrigan, 2008). En esta se muestra un nodo A que envía un mensaje. Luego se ve que B y C contestan con un ACK indicando que el mensaje fue recibido sin errores. Luego B y C comienzan a transmitir hasta que C gana el bus debido a que tiene un bit dominante, y termina este completando su mensaje.

S O F	11-bit Identifier	S R R	I D E	18-bit Identifier	R T R	r1	r0	DLC	0 ... 8 Bytes Data	CRC	ACK	E O F	I F S
-------------	----------------------	-------------	-------------	----------------------	-------------	----	----	-----	--------------------	-----	-----	-------------	-------------

Figura 2.6: Tráfico en el BUS CAN

El medio físico es una línea de bus con dos cables, terminada por una resistencia. Los cables pueden ser paralelos, trenzados y/o blindados. Los segmentos de cable para la conexión de los nodos deben ser lo más cortos posibles.

Para mayores detalle técnico referentes al bus será necesario estudiar el estándar ISO 11898 o en Corrigan (2008).

El largo del bus va a determinar el bitrate de la comunicación. Para alcanzar un bitrate de 1Mbps el máximo del canal de comunicación es de 40m. Si se tiene un bus de 1km el bitrate es de 0.05Mbps. Con esto se puede observar que el bitrate decae cuando la distancia se incrementa. Para el CAN el bitrate también está determinado

por el total del delay del sistema (Corrigan, 2008). Esto es la suma de los delays del nodo, y el delay propio del cable.

Debe destacarse qde que existen diferentes capas físicas para el protocolo CAN:

- El tipo más común es el establecido en el estandar ISO 11898-2. Este protocolo cuenta con dos cables, cada uno identificado como CAN_H y CAN_L. También es llamado *CAN de alta velocidad* (1 Mbit/s). Este requiere que el largo del cable sea como máximo 40m.
- También en el estándar de ISO se establece el ISO 11898-3, el cual define otro esquema de doble cable balanceado pero para bajas velocidades. Este es tolerante a fallas si alguno de los cables entra en corto circuito. Este es llamado también como *CAN de baja velocidad* (125 kbit/s)
- Se define el SAE J2411 que utiliza un solo cable para la capa física. Este es utilizado en autos (velocidad por encima de los 50 kbit/s). No tiene demasiados requerimientos en cuanto al bitrate ni el largo del canal de comunicación. EL estandar define 32 nodos por red.
- Existen modificaciones del estandar RS485 para su funcionamiento con CAN

2.17.4.2. Bit timing

El tipo de señal es codificado con Non Return Zero (NRZ). Los bits son codificados en dos estados llamados “recesivo” y “dominante”(el bit 0 es asociado al bit dominante). El protocolo permite acceso al bus multi-master con una resolución determinística ante colisiones. Para lograr la sincronización, el protocolo sincroniza durante las transiciones. Por tal motivo, deben evitarse el envío de largas cadenas de bits en un mismo estado. CAN utiliza una técnica que se denomina “bit stuffing” o “bit padding” en la cual luego de enviar 5 bits con el mismo estado, se insertan bits de relleno.

2.17.4.3. Asignación eficiente del bus

La asignación del bus depende de su aplicación. Generalmente existen 2 tipos de clases de asignación.

- **Asignación en un tiempo fijo:** La asignación se hace secuencialmente a cada participante sin importar si este necesita el bus en ese momento. Esta técnica, asigna tiempo del bus a cada nodo, y si no tiene nada que enviar, el bus se encuentra ocioso.
- **Asignación en base a sus necesidades:** el bus se asigna según las necesidades del nodo (utilizando CSMA o CSMA/CD). En CAN la asignación del bus es negociado entre los mensajes que esperan ser transmitido. CAN utiliza esta técnica.

2.17.5. Capa de Enlace

CAN utiliza el control de acceso al medio tipo *CSMA/CS+CR* (Acceso múltiple con detección de portadora, detección de colisión más resolución de colisión). CAN resuelve el problema de la colisión con la supervivencia de una de las tramas que chocan en el bus. La trama "ganadora.^{es} aquella que tiene mayor prioridad. Por lo tanto se puede indicar que CAN por naturaleza tiene en cuenta la prioridad.

Como ya se mencionó anteriormente el bit *dominante* es el 0 y el bit *recesivo* es el 1, la resolución se realiza con una operación lógica AND de todos los bits transmitidos simultaneamente. Cada transmisor se encuentra

continuamente observando y comprobando que el bit enviado se corresponda con el bit que envía. Cuando no coincide, el controlador retira el mensaje del bus y se convierte automáticamente en receptor. Como puede observarse la capa de enlace se comporta de manera similar a la capa física.

La única diferencia que presenta la capa de enlace de CAN es que todos los errores a nivel de un solo bits son detectados. Los errores de múltiples bits son detectados con una alta probabilidad (CAN-CIA, 11 de abril del 2017) s

2.17.6. Formato del mensaje

CAN utiliza un formato de mensajes cortos (94 bits) En el mensaje no está explícito ninguna dirección, por este motivo el mensaje puede ser escuchado por todos los nodos de la red (Kvaser, 2017).

Los tipos de mensajes son los siguientes:

- Frame de datos (Data Frame)
- Frame remoto (Remote Framte)
- Frame de error (Error Frame)
- Frame de sobrecarga (Overload Frame)

2.17.6.1. Data Frame

Este es el frame más común. Las partes más importantes son:

- Campo de arbitraje, el cual determina la prioridad del mensaje
- El campo de datos, que contiene desde cero hasta ocho bytes de datos
- El CRC, que está conformado por 15 bits utilizados para calcular el checksum del mensaje
- Un campo de ACK. Cualquier controlador que haya recibido el mensaje envía un bit de acuse de recibo al final de cada mensaje. El transmisor comprueba la presencia del bit ACK. En caso de no detectar este bit reenvía el mensaje. Al no poder conocer la dirección de los nodos, no se sabe si el mensaje fue recibido correctamente por el node receptor, solo se sabe que el mensaje fue recibido por uno o más nodos.

2.17.6.2. Remote Frame

El frame remoto es un Data Frame con dos diferencias, es marcado como Remote Frame, esto es el bit RTR es recesivo; y por otro lado no hay un campo de datos. Este frame es utilizado para pedir la transmisión de un determinado frame de datos. Por ejemplo si el nodo A transmite un Remote Frame con el campo de arbitraje en 234, entonces el nodo B, reponderá con un frame de datos con el campo de arbitraje seteado a 234 (Kvaser, 2017). Este frame es poco utilizado.

2.17.6.3. Error Frame

Este frame se envía cuando un node detecta alguna falla en el mensaje. El envío de este frame provoca la retransmisión inmediata del mensaje. El Error Frame consiste en una bandera, el cual está compuesto por 6 bits del mismo valor, y un Error Delimiter que está compuesto por 8 bits recesivos.

2.17.6.4. Overload Frame

Este frame es similar al frame de error con el mismo formato. Es enviado cuando el nodo está ocupado. Este Frame es muy poco usado. El único controlador que generaba Overload Frame está obsoleto (Kvaser, 2017)

2.17.6.5. CAN estándar y CAN extendido

El protocolo de comunicación CAN, es un protocolo de multiple acceso con detección de colisión y arbitraje según la prioridad de los mensajes (CSMA/CD+AMP) (Corrigan, 2002). CSMA ya fue estudiado en 2.14.1. CD+AMP significa que las colisiones son resueltas mediante arbitración por corrimiento de bits. El identificador con mayor prioridad es el que siempre gana el bus.

El CAN estandar (Figura 2.7) tiene los siguientes campos:

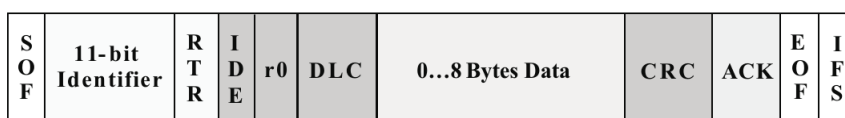


Figura 2.7: Frame de mensaje del CAN estándar

- SOF: es el bit dominante de inicio de frame, este marca el comienzo de un mensaje.
- Identificador: este es un identificador de 11 bits, en el cual el valor binario más bajo es el de mayor prioridad.
- RTR: El bit denominado Remote Transmission Request (RTR) es dominante cuando la información es requerida desde otro nodo. Todos los nodos reciben el pedido, pero el identificador determina node específico.
- IDE: Es una extensión del bit identificadorque significa que se está por transmitir un un identificador estandar CAN sin extensión.
- r0: Bit reservado.
- DLC: 4 bits que contiene el número de bytes que van a transmitirse.
- Data: Hasta 64 bits, es el mensaje que se transmite.
- CRC: Son 16 bits que conforman el Cyclic Redundancy Check (CRC) para chequear la existencia de algún error.
- ACK: Todos los nodos que reciben el mensaje correctamente sobreescribe este bit recesivo, indicando que han recibido el mensaje sin errores. Si un nodo detecta un error y detecta la persistencia de este bit recesivo. Con esto, se descarta el mensaje enviado y el nodo transmisor vuelve a enviar el mensaje.

- IFS: Son 7 bits que representan el espacio entre frames. Es el tiempo que requiere el controlador para mover el mensaje hasta el buffer de mensajes recibidos.

El CAN extendido (Figura 2.8) es idéntico al CAN estándar, salvo por algunos detalles:

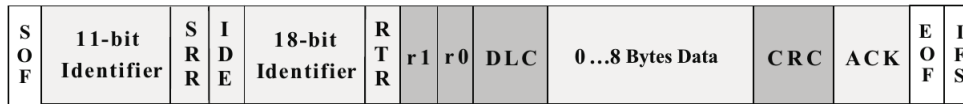


Figura 2.8: Frame del mensaje del CAN extendido

- SRR: Es el bit de petición remota substituta que reemplaza al bit RTR del mensaje estándar.
- IDE: Un bit recesivo en el identificador de extensión (IDE) indica que siguen más bits de identificación. A este le siguen 18 bits.
- r1: Este bit viene después del RTR, es un bit adicional reservado.

2.17.6.6. Arbitraje

Cualquier controlador CAN, cuando detecta que el bus está desocupado puede comenzar a transmitir. Esto puede ocasionar que al mismo tiempo, dos o más controladores comiencen a transmitir al mismo tiempo. Este conflicto se resuelve de una manera ingeniosa. El nodo transmisor a medida que transmite los bits, se encuentra monitoreando el bus. Si por ejemplo, el nodo detecta un bit dominante cuando este envía un bit recesivo, el nodo inmediatamente deja de transmitir y se convierte en un receptor. Cuando el bus se encuentra desocupado, recién vuelve a transmitir (Kvaser, 2017).

Una condición importante que se debe cumplir es que dos o más nodos no pueden transmitir el mismo campo de arbitraje.

2.17.7. CANAerospace

CANAerospace es una definición de protocolos y datos que fue diseñada para ser utilizada en sistemas de comunicación altamente confiables, sobre todo aeroespaciales, que estén basadas en el Controller Area Network (CAN) (Stock Flight Systems, 2017). Este, fue desarrollado por Stock Flight Systems, creado en 1997 y tiene su primer Release en 1998, y fue utilizado en varias aeronaves, como por ejemplo SOFIA Boeing 747SP¹⁹. También es utilizado en las interfaces de simuladores de vuelo. CANAerospace es un proyecto open source, y se continúa desarrollando. Fue publicado por NASA como el Advanced General Aviation Transport Experiments Database Standard (AGATE).

Este protocolo nace con el propósito de crear un estándar para la creación de aplicaciones que requieran un constante monitoreo del flujo de datos y una sincronización de tiempo en sistemas con redundancias Stock Flight Systems (2017).

El protocolo CAN no por sí solo no cubre problemas como la representación de datos, direccionamiento, protocolo orientados a la conexión. Además, la utilización de CAN en aplicaciones espaciales requieren de la definición de un estándar que esté orientado a los requerimientos específicos de la misión. CANAerospace es

¹⁹Ver: <https://www.sofia.usra.edu/public/about-sofia/sofia-aircraft>

creada para enfrentar estos problemas. Este protocolo es una pequeña capa de software que permite un manejo fácil del bus de datos que cumple con los requerimientos específicos de sistemas de aviónicas. CANAerospace fue instalado en una gran cantidad de aeronaves desde 1998 y ha demostrado una excelente confiabilidad en ambientes hostiles.

Algunas de las características más destacable de CANAerospace es que, no existen esclavos y maestros, por ello se dice que es una red democrática. Los mensajes se encuentran bien identificados. Existen mecanismos para informar eventos de emergencias. Es sencillo de aplicar. Y es un proyecto open source, por lo tanto no tiene costo alguno, y existen tutoriales libres y gratuitos.

2.17.7.1. Formato del mensaje CANAerospace

En la Figura 2.9 se puede observar el formato básico del mensaje de CANAerospace. Este está compuesto principalmente por bytes, los cuales se describen a continuación:

- **NODE-ID:** Este byte es utilizado cuando se produce un error en algún dispositivo y comienza a funcionar su redundancia. De esta manera, para aquellas arquitecturas que lo permiten, el protocolo puede identificar estos cambios.
- **Data-Type:** CANAerospace permite múltiples tipos de datos para cada mensaje. Este byte permite identificar cada tipo de datos de cada mensaje.
- **Service Code:** Para datos que se utilizan en las operaciones normales. Este dato debe reflejar el estado de los datos, esto ayuda a la implementación de un monitoreo integral de datos.
- **Message Code:** El número de mensajes permite detectar si el mensaje se perdió y verificar si la unidad de transmisión trabaja correctamente.

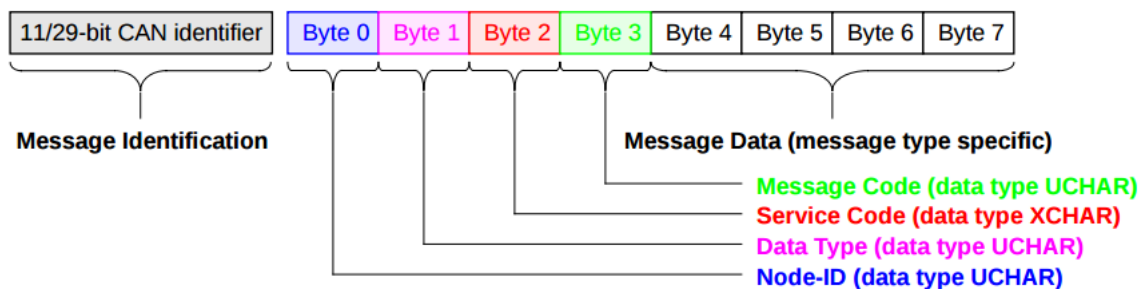


Figura 2.9: Formato del mensaje de CANAerospace

Estado del arte

Locura es hacer lo mismo una y otra vez
esperando obtener resultados diferentes

Albert Einstein

3.1. Árboles binarios

El concepto de arquitecturas de árboles binarios es aplicable en el desarrollo de sistemas de computadoras jerárquicas, y sobre todo en computadoras de alta performance (Raghavendra et al., 1984). Existen dos diferentes mecanismos de tolerancia a fallas (Raghavendra et al., 1984):

1. Esquemas con back up.
2. Esquemas con degradación de performance.

Teniendo en cuenta que estas arquitecturas son aplicadas principalmente en la construcción de circuitos VLSI¹ (Singh y Youn, 1991), se asume su aplicabilidad a arquitecturas de aviónica.

La FT y la performance de los sistemas dependen de las capacidades de las redes que se utilizan para la comunicación entre unidades de procesamiento (Raghavendra et al., 1984).

Un árbol binario está compuesto por nodos y enlaces (links). Existe un nodo central dónde se desprenden dos nodos hijos, estos se encuentran enlazados al nodo padre. Así recursivamente, se van generando dos nuevos hijos, por cada uno de los nodos. Los árboles binarios están divididos en niveles, que representa cada una de las generaciones de nodos.

Este tipo de topología tiene algunas problemas que la FT debe hacer frente. En las arquitecturas basadas en árboles binario, existe una cierta probabilidad de que un nodo o un link falle (Raghavendra et al., 1984). Las arquitecturas de árbol binario son en general físicamente estáticas. Por lo tanto, cualquier falla en uno de sus nodos (o links) demandaría una avería a nivel sistema, lo cual daría lugar a una pérdida de misión. Para ello se debe dotar a la arquitectura de un mecanismo de reconfiguración.

¹Del inglés, Very Large Scale Integration

La tolerancia a fallas en arquitecturas binarias ya fueron estudiadas en profundidad en Hayes (1976), Raghavendra et al. (1984), Singh y Youn (1991). Para lograr FT en estas arquitecturas se las deben diseñar con un número mínimo de nodos de backup y links redundantes, de modo tal de hacer frente cualquier punto de falla simple en la arquitectura.

3.1.1. Esquema de árbol binario con backups

El esquema planteado por Raghavendra et al. (1984) es similar al que se muestra en la Figura 3.1. En este esquema se agrega nodos y links redundantes como técnica de FT. Existe un nodo de backup por cada nivel del árbol.

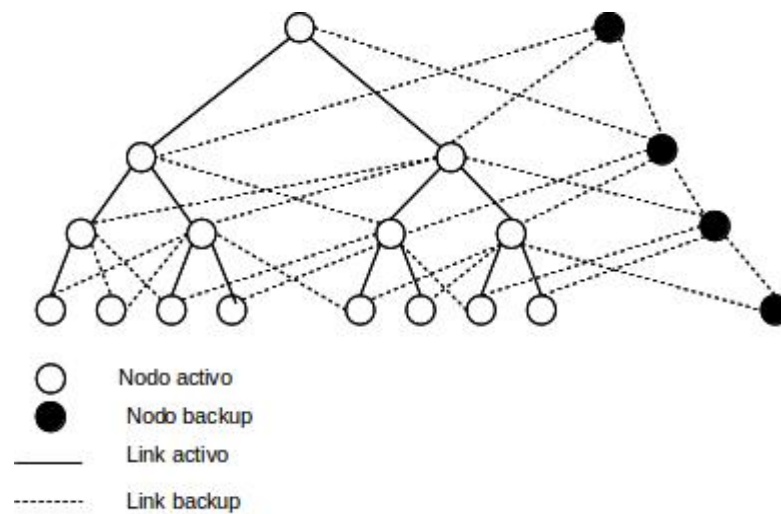


Figura 3.1: Árbol binario de 4 niveles

Esta arquitectura cuenta con una restricción, la cual indica que solo se puede tolerar una falla singular, por cada nivel de la arquitectura. Arquitecturas de este tipo, podrían tolerar más de una falla, sólo si se dan en diferentes niveles del árbol (Raghavendra et al., 1984). Para agregar más tolerancia, se deberían agregar más redundancias.

Notese en la Figura 3.1 que cuando una unidad de procesamiento (nodo) falla, todos los links se deben reajustar hacia el nodo de la derecha. En este punto es importante mencionar, que ante esta situación se requiere una reconfiguración a nivel de HW, además de una reconfiguración a nivel de SW. Es necesario algoritmos de ruteo dinámicos, de modo tal de conocer los nuevos caminos que intercomunican nodos, para mantener la operabilidad del sistema.

3.1.1.1. Estimación de la confiabilidad de un árbol binario con backup

Se asume que la probabilidad de fallas de los links es muy baja en comparación con la de los nodos. Teniendo que el rate de falla es de λ , la confiabilidad de un nodo es $R = e^{-\lambda t}$. También se sabe que un árbol binario con n niveles, se tiene $2^n - 1$ nodos en total. Entonces la confiabilidad de todo el sistema es:

$$R_{nr} = R^{2^n - 1}$$

Raghavendra et al. (1984) incluye en sus cálculos un factor de cobertura c , el cual es la probabilidad condicional de que se lleve a cabo una recuperación exitosa, luego de que una falla se haya detectado. Entonces la

confiabilidad del sistema para una arquitectura de árbol binario con redundancias (un nodo backup por nivel) es el siguiente:

$$R_{sys} = \prod_{k=0}^{n-1} [R^{2^{k+1}} + R^{2^k}(1-R) + 2^k c R^{2^k}(1-R)]$$

Simplificando:

$$R_{sys} = R^{2^{n+1}} \prod_{k=0}^{n-1} [(2^k c + 1) - 2^k c R]$$

En la Figura 3.2, se observa la confiabilidad de una arquitectura de árbol binario de 4 niveles, con una cantidad de $2^4 - 1 = 15$ nodos. En color negro se grafica una arquitectura sin redundancia, mientras que en color rojo, azul y cian, se muestra arquitecturas redundadas con diferentes c (0.98, 0.99, 1, respectivamente).

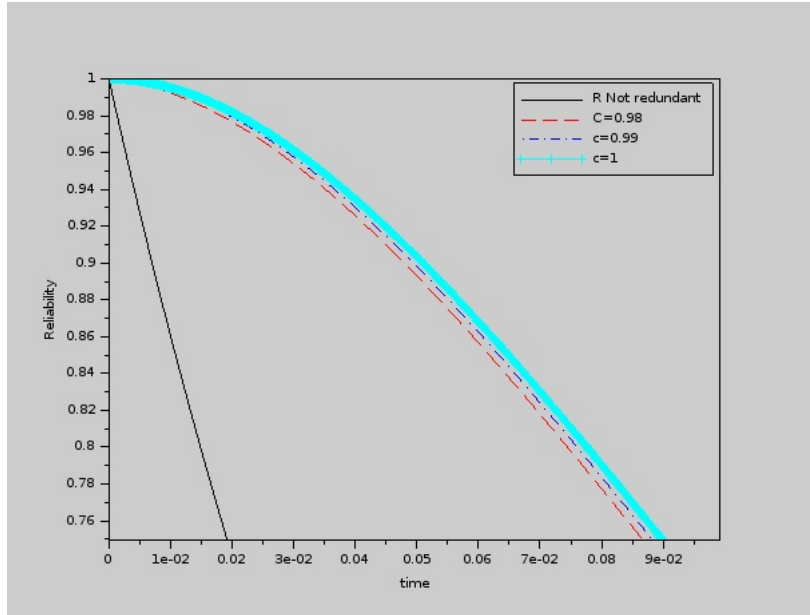


Figura 3.2: Confiabilidad con respecto al tiempo de una arquitectura de árbol binario de 4 niveles

Con esto podemos indicar, como es de esperarse, una arquitectura de árbol binario redundanda permite mantener un alto nivel deseable de confiabilidad, durante un mayor lapso de tiempo, a diferencia de un sistema no redundando. También se puede concluir que con un factor de cobertura c más próximo a uno, maximiza los niveles de confiabilidad con respecto al tiempo.

3.1.1.2. Extensión de la arquitectura con backup

3.1.2. Esquema de árbol binario con degradación de perfomance

3.2. Sistemas Hypercube

Escribir algo

3.3. Redes distribuídas

Una red de computadoras hace referencia a un conjunto de computadoras autónomas que se encuentran interconectadas, esto quiere decir que pueden intercambiar información (Tanenbaum y Wetherall, 2012). Una red distribuída tiene como principal característica que no existe un nodo central que "gestione" toda la red. Todas las cargas de las tareas y/o actividades son distribuídas entre los nodos que forman parte de la red.

Una red distribuída es tolerante a fallas si los nodos pueden formar subredes (Stivaros, 1992). Es decir, la red se debe mantener activa y conectada, con diferentes topologías e interconexiones, que permitan tolerar posibles fallas producidas en algunos nodos y permitir mantener la performance (Stivaros, 1992). Debido a que el procesamiento de la red se encuentra distribuída en todo el sistema, esto brinda una ventaja por encima a los sistemas centralizados desde el punto de vista de la confiabilidad (Pradhan y Reddy, 1982). Un componente importante de las redes distribuídas tolerantes a fallas, es la topología del sistema (Pradhan y Reddy, 1982).

Siguiendo la notación de Pradhan y Reddy (1982) para describir la topología del sistema se utiliza, un gráfico sin direccionamiento $G = \langle V, E \rangle$, donde V representa un set de nodos y E representa un set de relaciones. Stivaros (1992) agrega que $V(G) = \{v_1, v_2, v_3, \dots, v_n\}$ representa un vector de nodos, los cuales tiene probabilidades de operación $P = (p_1, p_2, p_3, \dots, p_n)$; y define una función de asignación π , la cual es una función que asigna V con una probabilidad $P_{\pi(v)}$.

La FT de la red G , dado el vector de probabilidades y una función de asignación π , tal como se viene discutiendo anteriormente, es la probabilidad de que la red continúe funcionando, es decir continúe conectada, aún en la falla (aleatoria) de alguno de sus nodos, esto se denota $FT(G; \vec{P}, \pi)$.

Se dice que un subset de nodos S , es un estado tolerante del sistema G , cuando estos nodos se mantienen conectados y funcionales. Se utiliza θ para indicar el conjunto de todos los S posibles. Un estado tolerante S contribuye $\prod_{v \in S} P_{\pi(v)} \prod_{v \notin S} (1 - p_{\pi(v)})$ a la probabilidad de FT.

Para calcular el total de la FT del sistema, incluyendo todo los S , se hace:

$$FT(G; \vec{P}; \pi) = \sum_{S \in \theta} Pr(S) = \sum_{S \in \theta} \prod_{v \in S} P_{\pi(v)} \prod_{v \notin S} (1 - p_{\pi(v)})$$

Una relación entre nodos es representado como ij , lo cual representa un enlace bidireccional entre nodos. El grado del nodo i representa el número de relaciones que inciden en ese nodo, el cual se escribe d_i . Así, d_i está limitado por el número de puertos de entrada y salida disponibles por cada nodo (Pradhan y Reddy, 1982). k_{ij} representa el número mínimo de *hop* (*hop* representa la transmisión a través de un link de datos) (Pradhan y Reddy, 1982).

Stivaros (1992) y Pradhan y Reddy (1982) mencionan la existencia de varias topologías que pueden ser aplicadas en una red distribuída tolerante a fallas. Una topología estrella posee una baja distancia entre nodos, pero una pobre tolerancia a fallas (Pradhan y Reddy, 1982) (Stivaros, 1992). La topología anillo, permite un simple ruteo, pero existen grandes distancias internodo. Un sistema completamente interconectado, presenta buenas características tolerantes a fallas, pero tiene un alto costo (Pradhan y Reddy, 1982). Pradhan y Reddy (1982) propone una topología para una arquitectura distribuída de comunicación. Esta es una topología robusta, y puede llegar a ser compleja a medida que aumentan los nodos.

3.3.1. Algoritmo de ruteo

Los algoritmos de ruteos son necesarios en el desarrollo de una arquitectura tolerante a fallas y reconfigurable. Los algoritmos de ruteo se los pueden dividir en dos: *algoritmos primarios* y *algoritmo alternativo*. El primero es utilizado cuando no hay fallas de nodos. Se deben mantener los caminos para llegar, correctamente,

al nodo destino. El segundo se utiliza en la presencia de alguna falla, este requiere que sea capaz de detectar la presencia de fallas, para luego reconfigurar el sistema. Estos algoritmos deben ser simples y requerir una mínima cantidad de HW y SW.

Agregado a lo mencionado en el párrafo anterior, deben existir algoritmos de diagnóstico distribuido de falas, basados en los algoritmos de ruteo (Pradhan y Reddy, 1982).

3.4. Redes Ethernet en aviónica

TTEthernet es una tecnología de red de computadoras comercializada por TTTech Computertechnik AG para el desarrollo de aplicaciones seguras. SAE International² estandarizó esta red como SAE AS6802. TTEthernet se basa en el Ethernet clásico, en el cual se pone énfasis en las características principales que deben respetarse en sistemas críticos, tales como latencias de mensajes determinísticos, precisión de tiempo real, tolerancia a fallas (Loveless, Andrew T., 2015). Tiene la capacidad de transmitir datos 100 veces más rápido que la que lo hacen las tecnologías tradicionales tales como el MIL-STD-1553.

El Ethernet clásico presenta ventajas, tales como su alta velocidad de transmisión de datos, flexibilidad, y su disponibilidad y bajo costo (ya que se trata de un componente COTS)(Loveless, Andrew T., 2015), hacen deseable su aplicación en el área espacial. fue utilizado en diferentes proyectos aeroespaciales y en misiones importantes tales como el Space Shuttle y la Estación Espacial Internacional (ISS) (Loveless, Andrew T., 2015). A pesar de esto, el Ethernet no cumple con el determinismo requerido por las aplicaciones de tiempo real de un vehículo espacial. Por tal motivo, se desarrolla el sistema TTEthernet, el cual introduce un reloj de sincronización descentralizado, permitiendo la transmisión mensajes Time-Triggered (TT). En este tipo de red, existe una herramienta de planning que asigna a cada dispositivo un intervalo de tiempo, en el cual puede utilizar para transmitir frames. Estos sistemas utilizan Links Virtuales (VL) para permitir el envío de mensajes time-triggered. Cada VC es asociado con un time-triggered frame através de un identificador de tráfico crítico (CTID), este reemplaza el control de acceso al medio (MAC) (Loveless, Andrew T., 2015). Una red simple se muestra en la Figura 3.3.

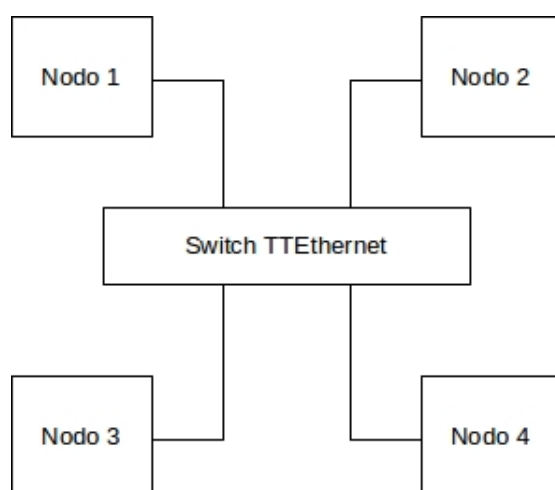


Figura 3.3: Arquitectura básica TTEthernet

TTEtheret puede actuar en dos clases de tráfico, con el obojtivo de soportar diferentes niveles de cricticidad de mensajes. Estas clases de tráficos son las siguientes (Loveless, Andrew T., 2015) (Steiner, Willfried, 2013):

²<http://www.sae.org>

- Time-Triggered, permite enviar mensajes de acuerdo a una planificación predefinida (scheduling),
- Rate-Constrained (RC), en el cual se llevan algunas restricciones de tamaño y rate de transmisión de frames,
- Best-Effort (BE), el cual se comporta de manera similar que el Ethernet

El paquete TTEthernet tiene una gran similitud con el frame del estandar IEEE 802.3 (Ethernet). Se denota en la Figura 3.4 que en lugar de la MAC del 802.3 frame se divide en el CT Marker y el CTID. El primero es un identificador estático utilizado para distinguir paquetes TT de otros tipos de tráfico.

7 bytes	1 byte	4 bytes	2 bytes	6 bytes	2 bytes	46-1500 bytes	4 bytes	12 bytes
Preamble	SFD	CT Marker	CTID	Source Address	Length	Data Payload	FCS	IPG

Figura 3.4: Frame de mensaje TTEthernet

El estandar SAE AS6802 define el protocolo de sincronización del TTEthernet. El protocolo TTP fue diseñado para reducir la complejidad de las arquitecturas distribuídas tolerantes a fallas (TTTech, 20 de Enero del 2017). Existe un grupo de dispositivos que relojes locales que permite la sincronización requerida por la comunicación TT, a esto se lo denomina dominio de sincronización (Loveless, Andrew T., 2015). Cada dominio de sincronización es asignado a uno de los siguientes roles:

- Compression Master (CM)
- Synchronization Master (SM)
- Synchronization Cliente (SC)

3.4.1. Experiencia de vuelo

Loveless, Andrew T. (2015) demuestra la aplicación de esta tecnología para computadoras de vuelo redundantes en misiones de naves simuladas. Esta tecnología toma tanto interés que el Sistema de Exploración Avanzada (AES)³ de la NASA, lleva a cabo un proyecto denominado Avionics and Software (AS). TTEthernet también fue utilizada en una arquitectura tolerante a fallas en El Integrated Power, Avionics and Software (IPAS) del Johnson Space Center (JSC) y en el Core Flight Software (CFS) del Asteroid Redirect Mission (ARM) simulado⁴ (Loveless, Andrew T., 2015). También fue utilizado exitosamente durante la misión Orion (TTTech).

TTEthernet simplifica el diseño de sistemas espaciales que deben contar con tolerancia a falla y una alta disponibilidad. La seguridad y la redundancia se mantiene sin ningún tipo de aplicación extra.

3.5. Arquitectura de red basada en BUS

En Tai et al. (1999) se presenta una arquitectura tolerante a fallas basada en BUS. Esta topología forma parte de un programa denominado X2000, el cual tiene como objetivo desarrollar una arquitectura tolerante a

³Del ingles, Advanced Exploration Systems

⁴Los mencionados anteriormente son ejemplos de la utilización de TTEthernet

fallas y basada en componentes COTS. la cual pertenece a NASA. Esta este programa se desarrolla bajo una filosofía de misiones espaciales la cual dice: "Más rápido, mejor, más económica"⁵.

Esta arquitectura utiliza una topología novedosa denominada *stack-tree topology* (Chau et al., 1999) (Tai et al., 1999). Una stack-tree es un árbol donde cada nodo rama se encuentra conectado como mucho a tres otros nodos de los cuales, como máximo, dos son nodos ramas (Tai et al., 1999). Una Complete Stack-Tree (CST) es aquella en donde cada nodo rama está conectado al menos 1 nodo rama (Tai et al., 1999).

En las Figura 3.5 se pueden observar ejemplos de stack-tree. Mientras que en esta otra Figura 3.6 no es una stack-tree. Las imagenes fueron extraídas de Tai et al. (1999)

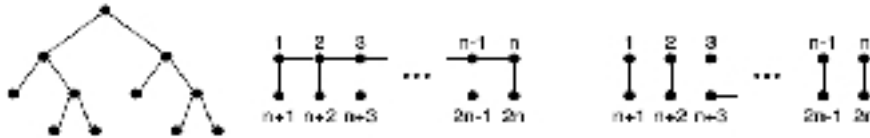


Figura 3.5: Arquitecturas stack-trees

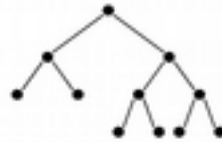


Figura 3.6: Arquitecturas que no responden al modelo stack-trees

Como el objetivo principal es desarrollar una arquitectura tolerante a fallas, esto es, que aún cuando un nodo o un link entre nodos falle, el sistema completo debe continuar funcionando, sin ningún tipo de degradación en su servicio. Para cumplir con este objetivo Tai et al. (1999) trabajan con un esquema de denominado CST de esquema dual (CST_D)⁶. Esta puede observarse en la Figura 3.7 extraída de citeTai99.

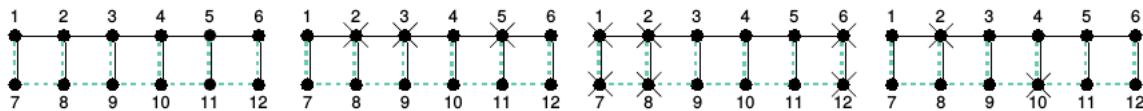


Figura 3.7: Esquema CST_D

Para aumentar la confiabilidad de la arquitectura ante la acurrencia de fallas, se agrega links de backups que unen los nodos iniciales con el final, dando un efecto 3D de anillo (Tai et al., 1999).

3.5.1. Evaluación de la confiabilidad de arquitecturas basadas en BUS

La confiabilidad de esta arquitectura (como ya se viene mencionando) es la probabilidad de que, a lo largo del tiempo de vida de la misión t , la arquitectura se encuentra funcional, en la cual todos los nodos (aquellos que no hayn fallado) se encuentren conectados (Tai et al., 1999). (Tai et al., 1999) y (Chau et al., 1999) asumen que la probabilidad de que un nodo falle es mucho mayor a que un enlace (el bus físico) falle.

En Tai et al. (1999) indica que la confiabilidad de la red depende del tamaño k (k es el número de nodos ramas). La confiabilidad de una red basada en CST simplex es la probabilidad $U(k)$ de que los nodos no fallen,

⁵En ingles, faster, better, cheaper

⁶En ingles,CST dual scheme

o que ante una falla se detecte y se lleve a cabo correctamente la reconfiguración del sistema.

$$U(k) = (1-q) \sum_{j=0}^k \binom{k}{j} (1-q)^{k-j} (cq)^j$$

donde $q = 1 - e^{-\lambda t}$ es la probabilidad de que un nodo falle durante el tiempo de vida de la misión f . Entonces

$$R_s^{CST} = \sum_{k=1}^n (n-k+1) U(k) (cq)^{2(n-k)}$$

donde $(cq)^{2(n-k)}$ es la probabilidad que los $2(n-k)$ nodos que conforman un cluster fallan, y esta falla es detectada y se lleva a cabo una correcta reconfiguración.

Para un CST dual ocurre de manera similar. Se define $V(k)$ de la siguiente manera:

$$V(k) = 2(1-q)^k \sum_{j=1}^k \binom{k}{j} (1-q)^{k-j} (cq)^j + (1-q)^{2k}$$

Entonces la medida de $R_D^{CST} = \sum_{k=1}^n (n-k+1) V(k) (cq)^{2(n-k)}$ p

3.6. Métrica y modelado de la confiabilidad de sistemas

Es de suma importancia llevar a cabo el análisis de la confiabilidad, disponibilidad y mantenibilidad (RAM⁷) de sistemas satelitales, durante la fase de diseño (Hoque et al., 2015), a fin de lograr la mínima cantidad de fallas o incrementar el MTBF (Peng et al., 2013). Llevar a cabo esto, es de gran importancia, ya que permite el desarrollo de estrategias que permitan altos grados de confiabilidad, disponibilidad y mantenibilidad (Hoque et al., 2015).

Existen dos categorías de medición de la confiabilidad y predicción (Schneidewind, 1997), estas son utilizadas para asegurar la seguridad del software de sistemas críticos (Schneidewind, 1997), las cuales son:

- medición y predicción que están asociadas con las fallas y errores residuales.
- medición y predicción que están asociadas con la disponibilidad del sistema a sobrevivir durante la misión sin experimentar fallas en el fallas (o pérdidas) en el sistema.

Las dos categorías mencionadas anteriormente son explicadas en Schneidewind (1997).

Según Liu et al. (2014) las severidades de las fallas son clasificadas como críticas, peligrosas o triviales, teniendo en cuenta la contribución de esa falla a la pérdida de la misión. Es importante, además, conocer el riesgo de una falla. El riesgo, se define como la posibilidad de que una falla produzca una lesión (por ejemplo, un astronauta en vuelos tripulados), algún daño material (por ejemplo, la destrucción del satélite), o una pérdida (por ejemplo, la pérdida de la misión).

Dependiendo de la misión, un criterio para definir si un sistema es seguro o no, es reduciendo las fallas que pueden provocar pérdidas de vida, pérdida de la misión o la obligación de abortar una misión (Schneidewind, 1997). Schneidewind (1997) define dos criterios que deben satisfacerse:

- $r(t_i) < r_c$,

⁷Del ingles, Realibility, Availability, Maintainability

$$\blacksquare T_F(t_i) > t_m$$

dónde t_i es el Tiempo total de testing (observado o predicho); $r(t_i)$ son las fallas restantes hasta t_i ; r_c es un valor crítico de fallas restantes; $T_F(t_i)$ es la métrica para medir el riesgo; y t_m es la misión de la duración.

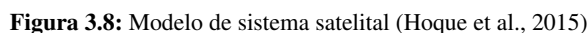
Lo anterior significa que un sistema crítico será seguro si: las fallas restantes en el tiempo de prueba son menores a un valor crítico de cantidad de fallas, o la duración de una misión es mayor al tiempo que se de la siguiente falla.

En la literatura se utilizan modelos matemáticos para modelar los sistemas críticos y calcular así su confiabilidad. La mayoría de ellos asumen, que todas las fallas tienen igual tasa de detección de fallas, como así también la misma severidad, lo cual no es correcto Liu et al. (2014). Las técnicas de verificación formal tiene un fuerte enfoque para la verificación de sistemas complejos, y permitir verificar las propiedades deseadas de los sistemas (Peng et al., 2013). Es importante llevar a cabo una verificación de modelo⁸, esta incluye técnicas de verificación automática para sistemas de estados concurrentes finitos. En estos modelos, las especificaciones de los sistemas son escritos en una lógica temporal y proposicional, y el procedimiento de verificación es una búsqueda del espacio de estado del diseño (Hoque et al., 2015).

En Hoque et al. (2015) se utiliza la *verificación probabilística de modelos*, la cual es utilizada para verificar sistemas, de los cuales los comportamientos son estocástico por naturaleza. Este se basa en la construcción y análisis de modelos probabilísticos como cadenas de Markov. En Hoque et al. (2015) se indica que estas técnicas fueron aplicadas en misiones de NASA. Este modelo se basa en la construcción y en el análisis de sistema de un modelo probabilístico de el sistema, típicamente una cadena de Markov (Hoque et al., 2015). Los modelos de Markov son muy utilizados para los análisis de confiabilidad y disponibilidad de sistemas complejos (Hoque et al., 2015).

En Hoque et al. (2015) se lleva a cabo un modelo simplificado de un sistema satelital este se muestra en la Figura 3.8. La descripción del modelo se llevó a cabo usando el lenguaje PRISM. El modelo que se describe en Hoque et al. (2015) y Peng et al. (2013) se refiere a un modelo a nivel de sistema/misión, muy diferente a lo que se pretende realizar en este trabajo de tesis, pero es factible basarse en los conceptos que describe la literatura, ya que lograron aumentar la confiabilidad del modelo propuesto, en contraposición del modelos clásicos"

⁸Model checking


$$\lambda = \frac{-\ln R_e}{MTBF}$$

El estado de la cuestión sobre la temática tratada en esta sección, indica que debería utilizarse modelos formales para el análisis de los sistemas. Se llega a la conclusión de que no existe un método (o modelo) completo y exclusivo para lo que se pretende desarrollar en este trabajo de tesis. Por ello se deberá utilizar y moldear las metodologías vistas en esta sección, para lograr cumplir con los objetivos del presente trabajo.

agregar info

45

Análisis y desarrollo de arquitectura tolerante a fallas

La raza humana necesita un desafío intelectual.
Debe ser aburrido ser Dios y no tener nada que descubrir

Stephen Hawking

4.1. Introducción

En este capítulo, en primer lugar, se plantea una serie de requerimientos (no estrictos) que ayudan en el desarrollo de una arquitectura para una misión satelital ficticia basada en componentes COTS. Luego, en base a estos requerimientos, se presentan modelos para la medición de confiabilidad de tres tipos diferentes de topologías de comunicación de subsistemas satelital. Con ello, se pudo elegir cual es la topología que asegura una mayor tolerancia a fallas.

Como siguiente paso

4.2. Requerimientos para el análisis

En esta sección no se pretende realizar una lista de requerimientos formales para el desarrollo de una arquitectura, esto se hace el capítulo siguiente. El objetivo de esta sección, es llevar a cabo una guía sencilla de las partes principales de un sistema satelital. Con esto en mente se podrá desarrollar diferentes topologías de comunicación, y luego analizar su tolerancia a falla.

Se supondrá una misión de 15 años (sin carga útil para simplificar el análisis) cuyo sistema estará compuesto por los siguientes subsistemas (en inglés para mantener correspondencia con la literatura) basado en Fortescue et al. (2003):

- Power Subsystem
- Attitude and Orbit control Subsystem

- Telemetry and Subsystem
- Thermal Subsystem
- Propulsion Subsystem
- Data Handling Subsystem

El sistema, entonces, está compuesto por 6 nodos. Los nodos se suponen computadoras con capacidad de procesamiento suficiente para cada subsistema. Cada una de estas computadoras/nodos es un componente COTS con un cierto grado de confiabilidad (se suponen lo suficientemente bajo como para no ser utilizado en forma directa en el desarrollo de satélites). A nivel de HW estos nodos cuenta con tolerancia a fallas, lo que aumenta su confiabilidad. El subsistema de Data Handling tiene que tener comunicación con todos los subsistemas, ya que será la encargada de controlar el correcto funcionamiento del sistema, enviar comandos, y empaquetar telemetría de los sensores.

4.3. Nomenclatura

Durante el análisis se utilizará la siguiente nomenclatura:

T_m	Tiempo de la misión
$TTNF$	Tiempo hasta la siguiente falla
λ	Tasa de falla
$MTBF$	Tiempo medio entre fallas
$MTTR$	Tiempo medio de reparación
$A(t)$	Disponibilidad
$R(t)$	Confiabilidad

Como se definió anteriormente el T_m es de 15 años. Para simplificar los trabajos de cálculos y ejecución del presente trabajo, se supone que la arquitectura puede fallar solo una vez durante la misión satelital. Por lo tanto, el $TTNF$ será de 15 años. La tasa de falla se define como:

$$\lambda = \frac{1}{15}$$

. Suponiendo que la arquitectura no debería fallar durante los 15 años de misiones se da la siguiente situación:

$$MTBF = MTTR = 15$$

Por otro lado, La disponibilidad es $A(t) = 99\%$. La confiabilidad, como se estudió en secciones anteriores, es $R(t) = e^{-\lambda t}$

4.4. Estudio de topologías de arquitecturas

Luego de un estudio exhaustivo del estado de la cuestión (Capítulo 3) se llegó a la conclusión de que las topologías más estudiadas, por ende más maduras y sencillas de aplicar a las actividades que se pretenden realizar en la presente tesis son:

- Árbol binario
- Red distribuída
- Arquitectura hypercube

En esta sección, se definirá modelos para medir la confiabilidad de cada una de las topologías de arquitectura que aseguren una mayor tolerancia a fallas a nivel de sistema, de modo tal que si se llega a producir una falla en cualquiera de los nodos (sistemas de procesamiento) de la arquitectura, esta puede reconfigurarse, permitiendo que esta continúe funcionando, sin sufrir ningún tipo de degradación, aún en la presencia de fallas.

4.4.1. Árbol binario

En primer lugar se planteó un árbol binario de cuatro niveles con back up, basandose en el diseño de Raghavendra et al. (1984) 3.2. Este diseño cuenta con $2^n - 1 = 15$ nodos. De lo estudiado en Sección 3.1 la confiabilidad puede ser calculada de la siguiente manera:

$$R_{sys} = R^{2^{n+1}} \prod_{k=0}^{n-1} [(2^k c + 1) - 2^k c R]$$

Con esto se puede observar la confiabilidad de la red con respecto al tiempo 4.1.

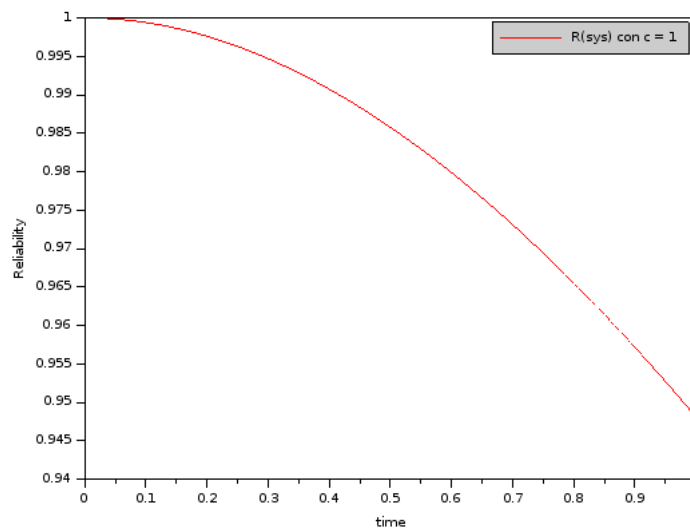


Figura 4.1: Confiabilidad con respecto al tiempo de árbol binario de 4 niveles

La cantidad de niveles que se eligió para esta topología depende del número de subsistemas que se requieren.

Los nodos back up se mantienen inactivos, es decir no participan en el procesamiento durante la vida normal del sistema. En caso de producirse una falla, se supone que estos nodos de redundancia, comienzan a funcionar automáticamente, sin ningún tipo de retraso. Esto, que no corresponde con la realidad, permite simplificar los cálculos para este trabajo de tesis.

4.4.2. Red distribuida

El estudio de una topología de red distribuida no es tan sencilla como la que se plantea para un árbol binario. Para el desarrollo de esta red, además de los 6 nodos que representa cada uno de los subsistemas requeridos, se agrega 2 nodos de redundancia. Por lo tanto se tiene una red de 8 nodos. Siguiendo la metodología de desarrollo presentado por Pradhan y Reddy (1982) se llevó a cabo la red que se presenta en la Figura 4.2.

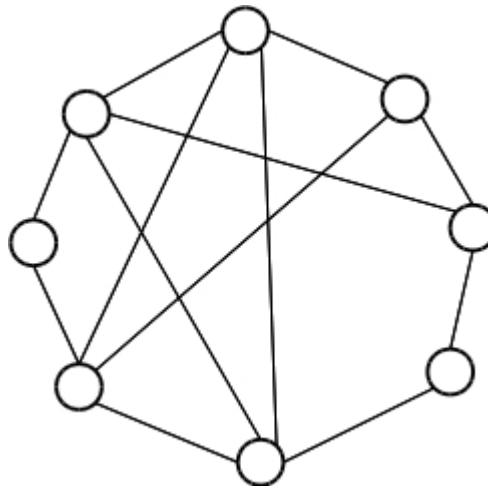


Figura 4.2: Red distribuida

Esta cuenta con 4 nodos de grado 4, 2 nodos de grado 3, y 2 nodos de grado 2. Ante cualquier falla de alguno de los nodos de la red, esta topología tiene la capacidad formar subredes, que mantienen todos los nodos conectados (a excepción del fallado), asegurándose la funcionalidad y reconfiguración del sistema completo. Estratégicamente y para lograr la condición mencionada anteriormente, se crea la **condición** de que pueden fallar hasta 4 nodos simultáneamente. Esto aseguraría de que la red continuará funcionando aún en la presencia de fallas (tolerante a fallas).

Se modificó la fórmula desarrollada por Stivaros (1992), para lograr una coherencia de los modelos que se vienen planteando en el presente trabajo. Teniendo en cuenta que la confiabilidad del sistema completa es:

$$R(t) = \prod_{v \in S} e^{-\lambda t}$$

donde v representa el nodo y S es el subsistema funcional. Es decir, que este modelo recorre todos los nodos funcionales. Cuando existen nodos con fallas, y que dejan de ser funcionales, el modelo es el siguiente:

$$R_{sys} = \sum_{i=0}^k ((\prod_{v \in S} R(t)) - (\prod_{v \notin S} (1 - R(t)c)))$$

donde c se definió como una *constante de degradación del sistema* para modelar una degradación del sistema.

En la Figura 4.3 se puede observar la *curva de confiabilidad* del sistema, para el caso en el que todos los nodos se encuentran funcionales.

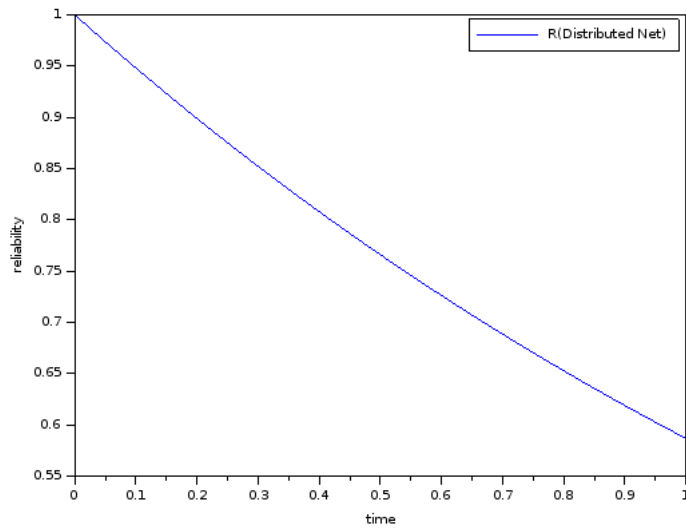


Figura 4.3: Confiabilidad de red distribuida

Para el caso de la falla de todos los nodos la *curva de confiabilidad* (Figura 4.4) muestra correctamente la degradación esperada de la confiabilidad, con respecto al sistema funcionando correctamente sin ninguna falla.

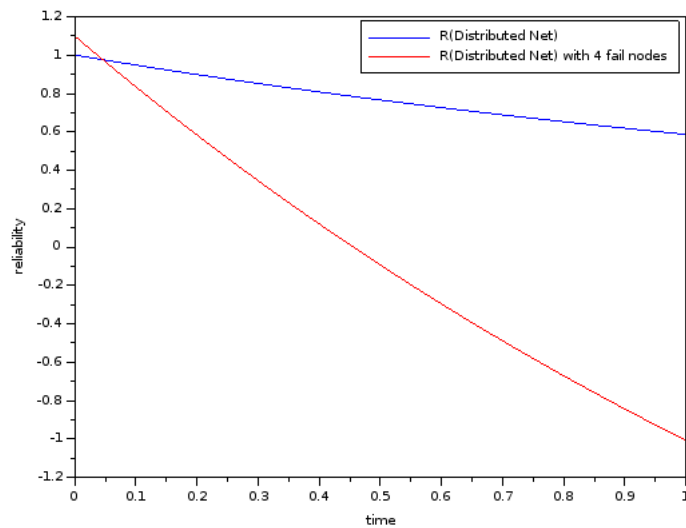


Figura 4.4: Confiabilidad de red distribuida con 4 nodos fallando

4.4.3. Red hypercube

Para el caso de la red hypercube se llevaron a cabo modificaciones al modelo planteado por Abd-El-Barr y Gebali (2014), con el propósito de mantener coherencia en los modelos y cálculos que se realizan en este trabajo. Se diseñó una red 3-dimensional, con 8 nodos, de los cuales 6 nodos corresponden a los diferentes subsistemas y 2 nodos son utilizados de redundancia (Figura 4.6).

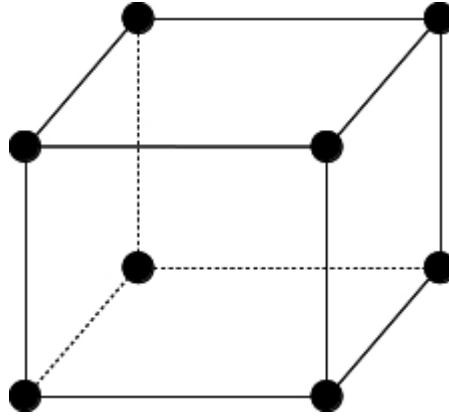


Figura 4.5: Red Hypercube

En este trabajo de tesis, la confiabilidad se calculó por medio del siguiente modelo:

$$R_{sys} = 1 - [N(1 - e^{-\lambda t})]$$

Como se puede observar el modelo es similar al modelo de árboles binario. Como se pudo estudiar en la bibliografía, árboles binarios y redes hypercube presentan varias similitudes, incluso se emebebe árboles binarios en este tipo de red. La *curva de confiabilidad* de esta red se la puede observar en la figura

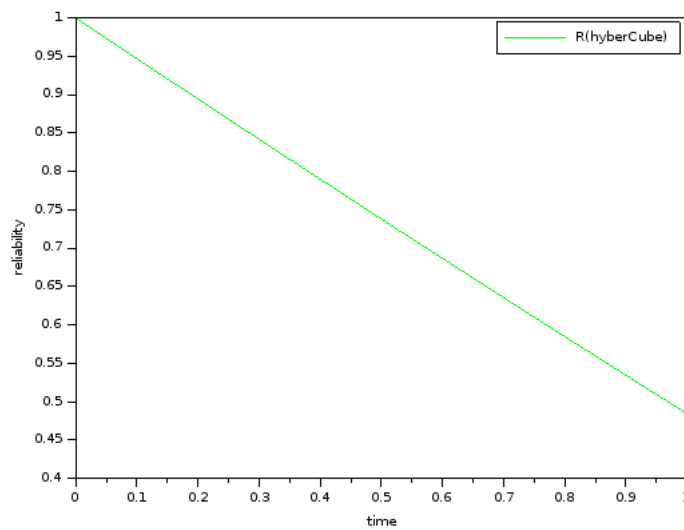


Figura 4.6: Confiabilidad de red hypercube

4.5. Topología utilizada en la arquitectura a diseñar

Teniendo en cuenta los modelos presentados anteriormente, se procede a realizar una comparación de las diferentes curvas. El resultado de esto, permitirá conocer de manera analítica qué topología presenta una mayor tolerancia a fallas a travez del tiempo. Además, teniendo en cuenta lo estudiado en el estado de la cuestión, se puede realizar una comparación conceptual de las tres topologías.

La aplicación de árboles binario podría resultar una buena opción, ya que representa un desarrollo sencillo. En contraposición se puede indicar que existe un alto grado riesgo de que se produzca una falla en el nodo raíz y de su redundancia, lo cual pondría en peligro la misión. Así mismo, presenta otro punto negativo que se puede mencionar y es la gran cantidad de enlaces que esta necesita para mantener a todos los nodos de la red conectados.

Por otra parte, la red distribuída cuenta con la capacidad de distribuir el trabajo en todos sus nodos. Esto quiere decir, que si se produce una falla irrecuperable en cualquiera de sus nodos, la arquitectura podría continuar funcionando sin verse afectada por la ausencia de dicho nodo. Esto demanda un procesamiento computacional extra, y la necesidad de desarrollar algoritmos de ruteo especiales. Además, como punto negativo se puede mencionar que también, al igual que los árboles binarios, necesitan una gran cantidad de enlaces.

Por último, las topologías hypercube tiene un excelente respaldo teórico, exigen menor cantidad de enlaces, y pueden tolerar la falla de una gran cantidad de nodos (hasta el 50 % de los nodos). Su complejidad aumenta en gran medida, cuando se desarrollan arquitecturas de más dimensiones, sin embargo, esto aumenta su confiabilidad.

Como se mencionó en el primer párrafo, se realizó una comparación de modelos de confiabilidad de las tres topologías estudiadas. Se asumió que la distribución de la confiabilidad es exponencial, con una tasa de falla fija de $1/15$, y se estudió su evolución en un rango $[0, 1]t$. El resultado de esta comparación se la puede observar en la Figura 4.7 y en la Tabla 4.1.

Se observa que tanto las redes distribuidas como la hypercube presenta una mayor confiabilidad a través del tiempo que los árboles binarios. Si bien, las redes distribuidas y la hypercube tienen una curva similar, la primera presenta una mayor confiabilidad sostenida en el tiempo.

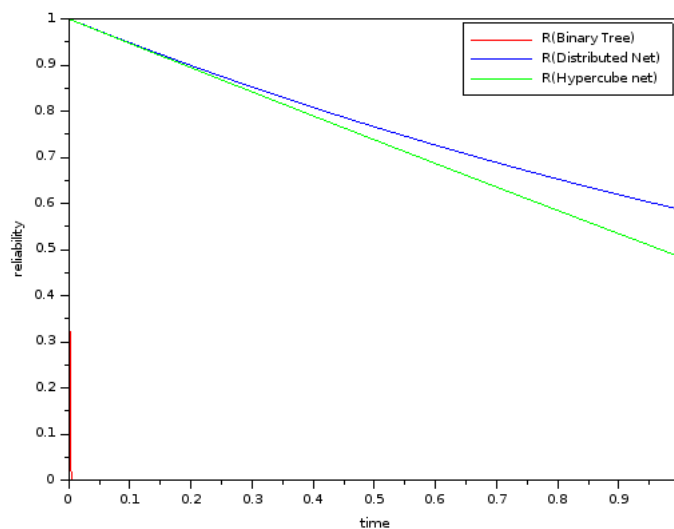


Figura 4.7: Comparación de confiabilidad

Tabla 4.1: Comparación de confiabilidad de topologías

T	Topologías de red		
	Tree Net	Distr Net	Hyper Net
0	1	1	1
0.001	0.803766	0.999947	0.999947
0.002	0.64604	0.999893	0.999893
0.003	0.519265	0.99984	0.99984
0.004	0.417368	0.999787	0.999787
0.005	0.335466	0.999733	0.999733
...
0.995	0	0.948317	0.947109
0.996	0	0.948266	0.947056
0.997	0	0.948216	0.947003
0.998	0	0.948165	0.94695
0.999	0	0.948115	0.946897

Con esto se puede concluir que la topología que presenta un mayor grado de confiabilidad es la que responde a una filosofía distribuida (bajo las condiciones en las que fueron estudiadas). Por lo tanto, la arquitectura satelital, tolerante a fallas y basada en componentes COTS que se desarrolla en la presente tesis se basa en una **topología distribuida** para interconectar los diferentes subsistemas.

4.6. Topología propuesta

Sobre la base de los resultados presentados en (Arias y Wiman, 2017) se puede establecer que la topología propuesta es la más adecuada para el desarrollo de una arquitectura tolerante a fallas como se presenta en la Figura 4.8. En esta se puede observar que cada subsistema (térmico, power, telemetría, etc.) tiene su propia CPU controladora. Estas CPU se conectan a los nodos.

El modelo exige como requerimiento que cada nodo debe estar compuesto por una computadora (componente COTS) que es la encargada de realizar el procesamiento de las tareas. También, debe existir un puente de comunicación entre la red y la CPU del subsistema. De este modo se hace frente a posibles fallas en la computadora del nodo. Esta conexión se observa en la Figura 5.7

4.7. Protocolo de comunicación

Como se estudió en el marco teórico, en una arquitectura de comunicación, el protocolo CAN trabaja en las capas inferiores del modelo de OSI. CAN propone cómo debe ser el medio físico, y la manera de transmisión de las señales. Además, preve cuál es la estrategia a utilizar para sincronizar la transmisión de datos sin ningún tipo de latencia, ni tampoco colisiones de mensajes. Así, este protocolo permite que estos sean enviados teniendo en cuenta sus prioridades, eliminando las colisiones, y por lo tanto también, el tiempo de espera en la que los nodos se quedan relegados cuando han tratado de enviar un mensaje al mismo tiempo.

Este protocolo no establece nada sobre ruteos de mensajes, tipos de mensajes y tipos de datos, que a la hora de desarrollar una arquitectura de aviónica es importante tener en cuenta esos detalles. Es necesario, que exista una capa de servicios que facilite la comunicación entre aplicaciones de usuarios de diferentes nodos, y la comunicación entre las aplicaciones de usuario y el protocolo CAN (de bajo nivel) en sí. Además, teniendo

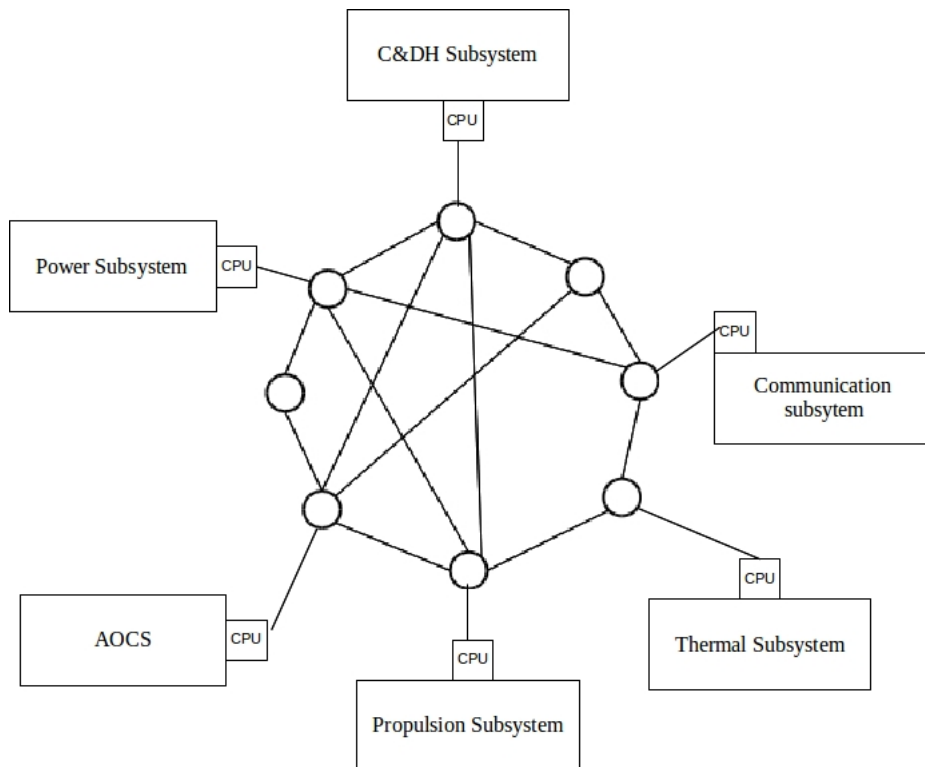


Figura 4.8: Arquitectura propuesta utilizando topología de red distribuida

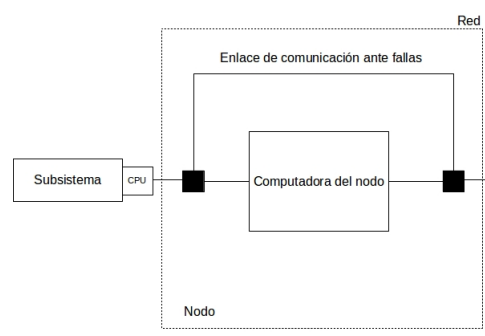


Figura 4.9: Conexión entre la red y el subsistema

en cuenta los resultados obtenidos en las secciones anteriores, se necesita además, diseñar un protocolo de comunicación basado en redes distribuidos, y que a su vez, esté basado en CAN.

Por tales motivos, surge la necesidad de diseñar un protocolo de comunicación basado en CAN, que se “monte” sobre las capas superiores del modelo de referencia de OSI; y que permita la distribución de las tareas y el procesamiento llevado a cabo por los nodos.

El protocolo CANae se encuentra en su primera versión 0.1 Alpha. Esto hace referencia a que en esta instancia de trabajo se tratará de comprender el problema y llevar a cabo un diseño preliminar de protocolo. De esta manera se podrá extraer, tanto puntos positivos, como puntos negativos; o más bien fortalezas y debilidades del stack de servicios que brinda CANae.

CANae actúa en la capa de aplicación del modelo de referencia de OSI. CANae divide esta capa en dos subcapas:

- CANae Application Layer
- CANae High Application layer

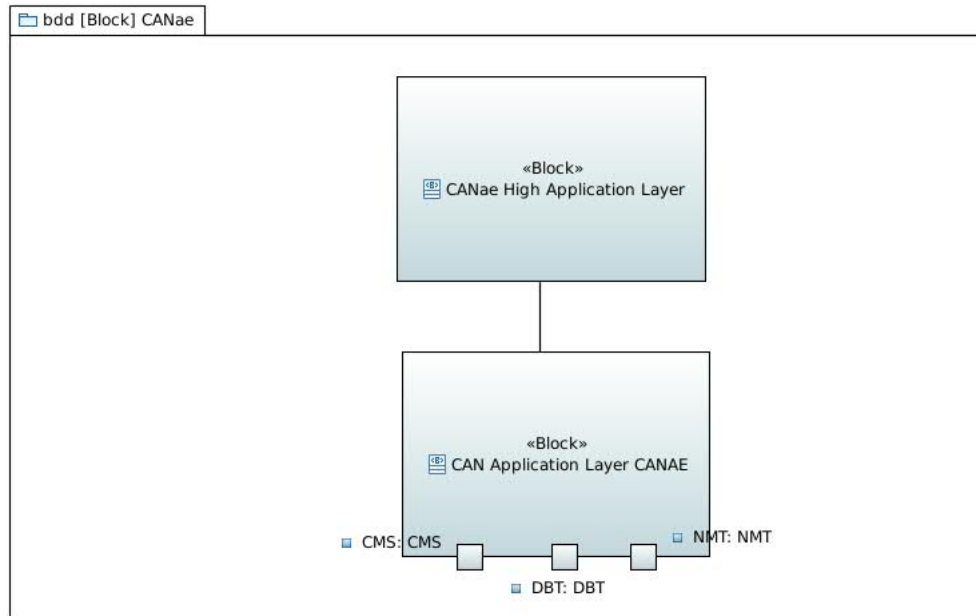


Figura 4.10: Estructura de la capa de aplicación de CANae en alto nivel

Esto puede observarse en la Figura ?? . En el gráfico se observa que la CANae application Layer, cuenta con 3 interfaces:

- **CMS:** ofrece un ambiente orientado a objeto para diseñar aplicaciones de usuario. Esta entidad ofrece variables y eventos, y especifica cómo un módulo puede acceder a las interfaces de CAN.
- **NMT:** ofrece un ambiente orientado a objetos para permitir que un módulo (el NMT Master) se ocupe de la inicialización y posibles fallas de otros módulos (NMT Slaves).
- **DBT:** ofrece el servicio para distribuir dinámicamente el identificador de para los diferentes nodos.

Debe destacarse que, el servicio CMS ofrece un ambiente orientado a objeto para diseñar aplicaciones de usuario, es decir, que ofrece la posibilidad de modelar el comportamiento del sistema en forma de objeto. Este servicio ofrece la posibilidad de crear variables y eventos, según las necesidades de los usuarios que son utilizadas para diseñar y especificar cómo la funcionalidad de un módulo puede acceder a las interfaces de bajo nivel de CAN. Esto propone un grado de innovación al tratar al protocolo, a los datos, mensajes y eventos como objetos, permitiendo así un modelado orientado a objetos. Por tal motivo, este protocolo fue modelado con SysML (System Modeling Language). Para mayor detalle de estas interfaces se debe consultar en el apéndice A.

Agregando, dentro de esta capa existen dos entidades que juegan un papel importante (consultar apéndice A) las cuales son:

- **Gestor de mensajes:** este se encarga de gestionar los mensajes que son enviados y recibidos desde la red. Esta entidad debe ser capaz de determinar si el mensaje contiene datos o eventos. Trabaja en conjunto con el CMS.

- **Gestor de nodo:** esta entidad se encarga de llevar el control de los nodos existentes en la red. En este nodo se encuentra la tabla de ruteo primario y secundario necesarios para la correcta comunicación.

La subcapa denominada CAN High Application Layer tiene una función más del lado de la gestión de nodos y tareas. En esta capa se desarrollan los algoritmos necesarios para realizar el ruteo y la reconfiguración de la red ante fallas, por lo tanto en la High Application Layer se debe implementar la FDIR. El protocolo no define ningún algoritmo de ruteo, por lo que queda para el usuario la definición de los algoritmos. El protocolo recomienda desarrollar dos algoritmos, el primario y el secundario. Para aumentar la tolerancia a fallas se pueden desarrollar más algoritmos secundarios, y el switch entre algoritmos, puede depender de medidas de performance.

Así, queda conformada el protocolo de comunicación CANae como uno de los principales productos de este trabajo de tesis.

Arquitectura propuesta

Dadme un punto de apoyo y moveré el mundo

Arquímedes

5.1. Introducción

La arquitectura propuesta se centra en la comunicación de la aviónica de un vehículo espacial, con la característica de que sus componentes (nodos) son de baja confiabilidad (se piensa en la utilización de componentes COTS) por lo tanto, es necesario que esta arquitectura sea tolerante a fallas. En este trabajo el protocolo de comunicación (tolerante a fallas) propuesto, se encuentra en las capas superiores del modelo de referencia OSI¹

Aquí hago la propuesta de la arquitectura.

El capítulo estará dividido en

- Arbol de requerimientos
- Diseño estructural (Diagramas de bloques, bloques internos)
- Diseño dinámico (Diagramas de secuencia, interacción, maquina de estados,)

5.2. Árbol de requerimientos

A continuación se detallan los requerimientos que van a guiar el diseño y desarrollo de la propuesta de arquitectura de aviónica tolerantes a fallas basada en componentes COTS para satélites. En la Figura 5.1 y en la Tabla 5.1 se pueden observar los requerimientos.

A continuación se explica cada uno de los requerimientos:

- **L1_000:** Este es el objetivo principal de la presente tesis: desarrollar una arquitectura de aviónica tolerante a fallas basada en componentes COTS para vehículos espaciales.

¹ISO/IEC 7498-1 - Open System Interconnection

Tabla 5.1: Tabla de Requerimientos

	ID	Name	Detail	Kind of Requirements
0	L1_001	L1_001	The architecture components shall be Components Off-The-Shelf category	Constraint
1	L2_000	L2_000	The architecture shall be reconfigure when a node fail.	Funcional
2	L2_001	L2_001	Each subsystem shall represented for a Components Off-The-Shelf	Constraint
3	L1_000	L1_000	Shall be develop an avionics architecture for spacecraft using Components Off-The-Shelf	Constraint
4	L1_002	L1_002	The architecture shall have at least 6 subsystems	Constraint
5	L1_003	L1_003	The architecture shall have fault tolerance techniques to assurance the the mission life	No funcional
6	L1_004	L1_004	The main bus to interconnect components shall be the Controller Area Network Bus developed for Bosch	Constraint
7	L1_005	L1_005	The architecture shall be sufficient to make a master degree thesis	No funcional
8	L2_002	L2_002	The architecture shall implements a based Controller Area Network protocol to intercommunicate the components	Funcional
9	L3_000	L3_000	The intercommunication between nodes shall used a Controller Area Network protocol based developed in the master degree theis.	Constraint
10	L1_006	L1_006	The architecture shall ssurance at least 10 years mission life	No funcional
11	L3_001	L3_001	The architecture shall use the distributed network philosophy	Constraint
12	L1_007	L1_007	The thesis shall be finished in 1 year	Constraint
13	L2_003	L2_003	The nodes shall send and receive message from any nodes connected to the architecture.	Funcional

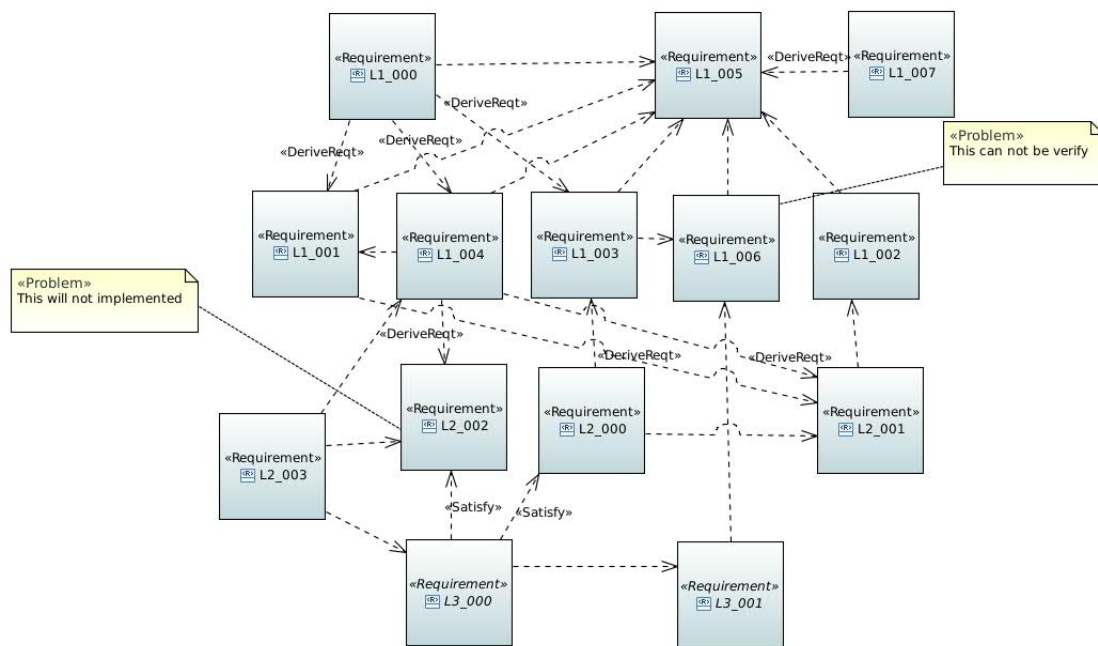


Figura 5.1: Diagrama de requerimientos de la arquitectura propuesta

- **L1_001:** El principal requerimiento de este trabajo de tesis, es desarrollar una arquitectura basada en componentes COTS, lo cual agrega un grado importante de innovación tecnológica y complejidad, siendo de especial interés para INVAP trabajar con estos tipos de componentes.
- **L1_002:** Este requerimiento hace referencia que se debe asegurar el correcto funcionamiento de la arquitectura con una cantidad de al menos 6 subsistemas.
- **L1_003:** Este requerimiento exige que la arquitectura sea tolerante a fallas para así poder satisfacer el requerimiento **L1_006**.
- **L1_004:** Este requerimiento (constraint) indica que se debe pensar la arquitectura para ser utilizada con el bus CAN de Bosch. Este requerimiento es de interés par INVAP.
- **L1_005:** Este requerimiento indica que la arquitectura que se desarrolle debe ser suficiente para lograr terminar una tesis de maestría. Por el diseño y desarrollo debe ser solo el suficiente y necesario para cumplir con el presente trabajo.
- **L1_006:** Este requerimiento indica que debe asegurarse que la arquitectura sea lo suficientemente robusta como para asegurar el tiempo de vida de la misión de 10 años como mínimo. Existe un **problema** con este requerimiento, y es que para los alcances de esta tesis, será imposible verificar si se cumple con este requerimiento.
- **L1_007:** Este requerimiento indica que el trabajo de tesis tiene que ser finalizado en menos de un año. Esto tiene relación con el requerimiento **L1_005** debido a que no se logrará alcanzar el detalle necesario para lograr la correcta implementación, verificación y validación de la arquitectura.
- **L2_000:** Este requerimiento exige que la arquitectura pueda reconfigurarse cuando un nodo falle.
- **L2_001:** Este requerimiento indica que para esta instancia de trabajo cada subsistema será tratado como un nodo dentro de la arquitectura.

- **L2_002:** Este requerimiento indica que se debe utilizar dentro de la arquitectura un protocolo de comunicación basado en el Bus CAN de Bosch.
- **L2_003:** Este requerimiento asegura que los nodos deben poder enviar y recibir mensajes de cualquier otro nodo conectado a la red CANae.
- **L3_000:** Este requerimiento indica que el protocolo de comunicación utilizado en esta arquitectura debe estar basado en el el protocolo CAN. Este protocolo fue desarrollado en el presente trabajo de tesis bajo el nombre de CANae 0.1 Alpha (Vease: Apéndice A).
- **L3_001:** Este requerimiento exige la utilización de una filosofía de red distribuída para su diseño y desarrollo.

5.3. Casos de Uso

A continuación se muestra el diagrama de Casos de Uso para la arquitectura propuesta. El diagrama de Casos de Uso tiene como propósito explicar de manera general el comportamiento a nivel de sistema de la arquitectura propuesta. Los actores identificadores para esta arquitectura son los diferentes subsistemas que se conectan a la red a través de los nodos. Los subsistemas interactúan con el nodo a través de la aplicación de usuario, es decir se encuentra asociados. Se puede observar en la Figura 5.2 que de forma general existen tres casos de usos que comprenden el comportamiento general de la arquitectura. Se tiene *Enviar mensaje* y *Recibir Mensaje* los cuales son básicos para cualquier arquitectura de comunicación. A esto se le agrega el caso de uso de *FDIR* que es el agregado que se le hace en este trabajo de tesis a través del protocolo de comunicación desarrollado denominado CANae (Ver apéndice A). Debe aclararse que en la Figura 5.2 aparece un resumen de los Actores interactuando con el sistema.

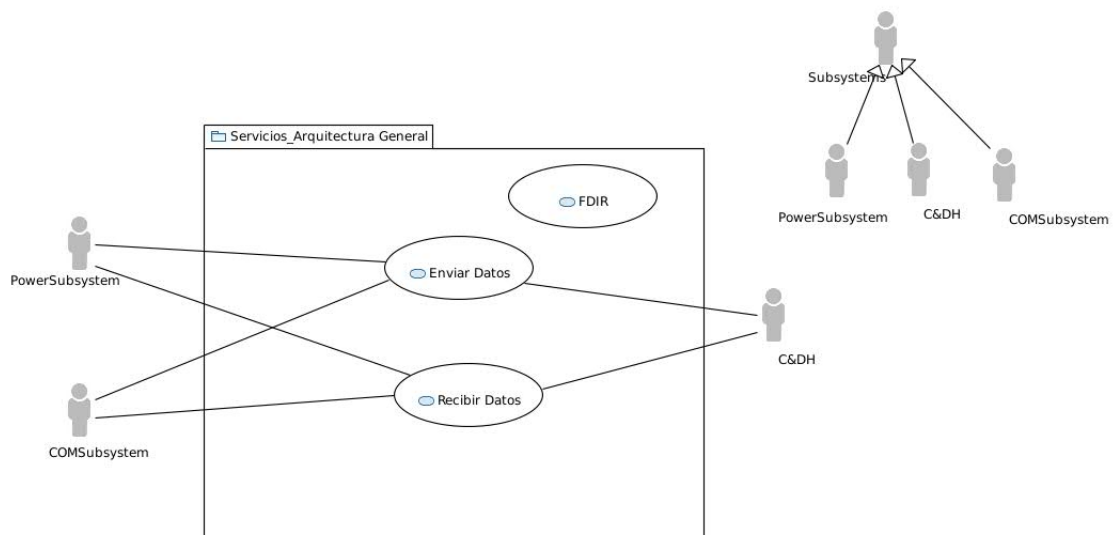


Figura 5.2: Diagrama de Casos de Uso General de la arquitectura Propuesta

De manera más detallada en la Figura 5.3 se observa un diagrama de casos de uso exployado. En este se pueden observar su comportamiento.

Le especificación de los casos de usos se describen en B. Como se puede observar en la Figura 5.3 el sistema tiene 9 casos de uso.

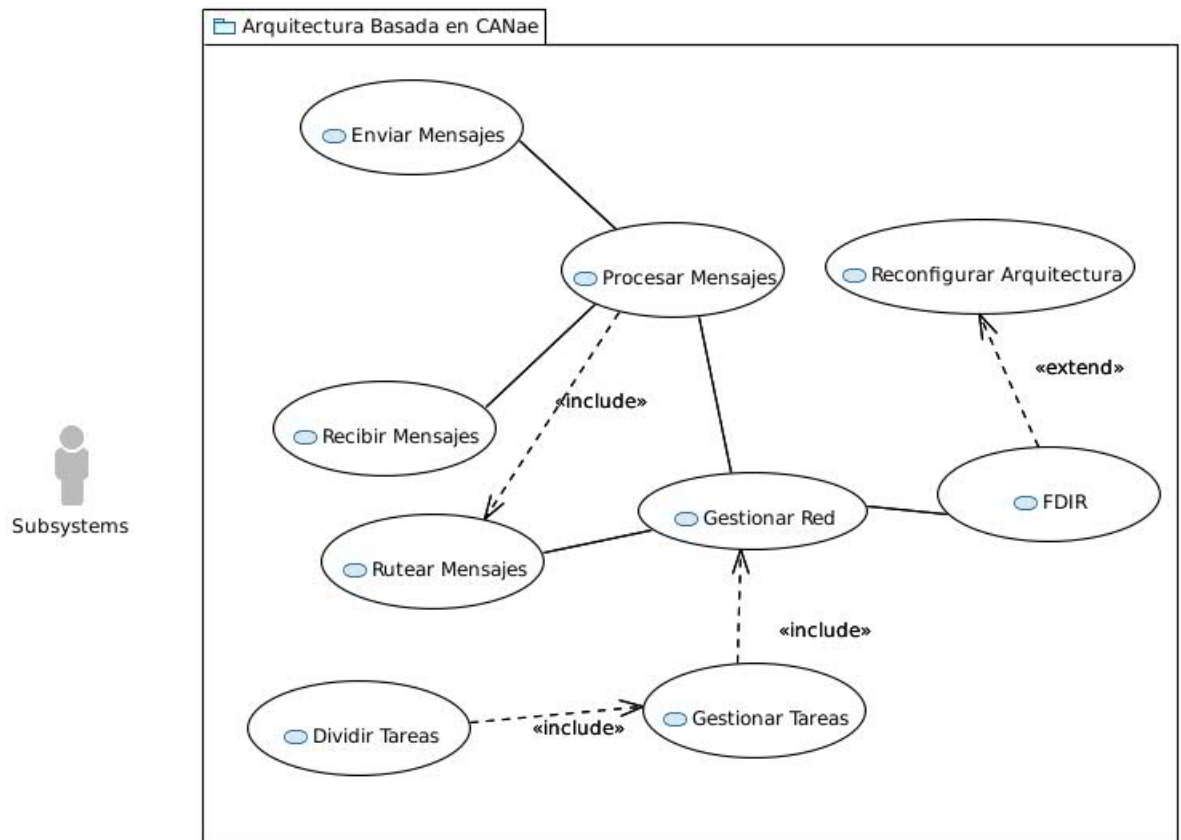


Figura 5.3: Diagrama de Casos de Uso de arquitectura propuesta

Los casos de uso de *Enviar Mensaje* y *Recibir Mensaje* denota la capacidad de la arquitectura de enviar y recibir mensajes (Data Frames de CANae) desde y hacia las aplicaciones de usuario. Cuando una aplicación necesita enviar un mensaje, estos deben primer pasar por el caso de uso *Procesar Mensaje* este es el encargado de empaquetar los mensajes. También en este caso de uso se encuentra el caso de uso *Rutear Mensaje*.

El caso de uso *Rutear Mensaje* se encarga de aplicar los algoritmos necesarios para lograr entregar el mensaje de una manera eficiente, siguiendo la tabla de ruteo del CANae. Para esto se basa del caso de uso *Gestionar Red*.

El caso de uso *Gestionar Tareas* y *Dividir Tareas* se encarga de de conocer (periódicamente) el estado de la red y de sus nodos. Con esto puede decidir cómo se reparten las tareas de una manera eficiente, siguiendo la filosofía de arquitectura distribuida.

Por último tenemos el caso de uso *FDIR* que hace referencia a la Detección, Aislación, y Recuperación del sistema ante fallas. Debido a la complejidad de esto, no se lo desarrolla a esta caso de uso en este trabajo. El cual queda para **trabajos futuros??**. En este caso de uso se encuentra el *Reconfigurar Arquitectura*. Este es una algoritmo del Detección, Aislación y Recuperación de Fallas (FDIR) que cuando se produce una falla en el algunos de los nodos, la red se tiene que reconfigurar, para lograr continuar funcionando, aún cuando haya nodos inactivos. Esto le brindaría a la arquitectura una grado más de tolerancia a fallas.

5.4. Diseño Estructural

Este trabajo de tesis plantea una arquitectura a nivel de sistema. Esta arquitectura se encuentra basada en la red CAN debido a un interés especial de INVAP. Además se sigue una filosofía de red distribuida por las razones estudiadas en la sección 4.4. Esto le da la característica de innovador a este proyecto, ya que no se encontró en estado de la cuestión la utilización de esta filosofía en el desarrollo de vehículos espaciales, mucho menos en Argentina, que se sigue una filosofía más tradicional y conservadora a la hora de desarrollar satélites.

La arquitectura propuesta, como se mencionó anteriormente tiene como base el bus CAN. A la red se pueden conectar un número N de nodos ($N < 128$), siendo estos componentes COTS de baja confiabilidad. Esta arquitectura trabaja en las capas superiores de los protocolos de comunicación, por lo tanto, en esta instancia de trabajo, el nodo se conecta a la red mediante un microcontrolador CAN, el cual a su vez está compuesto por el CAN transceiver.

La red a bajo nivel, es decir desde el cable CAN, el CAN Transceiver y el microcontrolador CAN, deben trabajar bajo las normas preestablecidas por algún protocolo existente, que cuya selección de este, cae bajo la responsabilidad del diseñador y/o ingeniero en sistema que vaya a implementar la arquitectura. Para esta arquitectura se aconseja el uso del protocolo CANOpen (CAN-CIA, 11 de abril del 2017). Esto se puede observar en la Figura 5.4 que el Node (nodo de la red) tiene como *inout flowport* el microcontrolador CAN y el CAN Transceiver. El CAN Transceiver simplemente emite y recibe señales eléctricas. Las especificaciones (por ejemplo resistencias y voltaje, que son mostrados en el diagrama solo con propósito) se pueden encontrar en los estándares (CAN-CIA, 11 de abril del 2017).

Dentro de cada nodo se encuentra la aplicación de usuario que será de responsabilidad del desarrollador implementar esta aplicación haciendo uso del protocolo CANae desarrollado en este trabajo.

Para agregar tolerancia a fallas a la arquitectura se plantea la necesidad de llevar a cabo un Bridge tolerante a fallas. Este Bridge permite que la comunicación de la red con los subsistemas no se vean afectadas ante la falla de un nodo. La utilización de esta técnica se explica en la siguiente sección.

De esa manera queda definida la estructura a nivel de bloques de la arquitectura planteada.

5.4.1. Bridge tolerante a fallas

En esta sección se explica la necesidad de la utilización de un Bridge Tolerante a Fallas. No se pretende llevar a cabo una especificación técnica del Bridge, ya que escapan a los propósitos de este trabajo de tesis. En primer lugar, como la arquitectura persigue la filosofía de red distribuida, y la posibilidad de distribuir las tareas entre todos los nodos, surge la dificultad de que muchas veces no existe compatibilidad entre las computadoras de los diferentes subsistemas. Es decir, una computadora del subsistema de control térmico, por ejemplo, no puede prender y/o pagar componentes del subsistema de propulsión. Tal como se muestra en la Figura 5.5 la ejecución de tareas de un subsistema en otro está prohibido.

Esto lleva a una segunda propuesta en la cual el nodo conectado a la red distribuida, debe estar separado de la computadora que se comunican directamente con los componentes del subsistema. Con esto surge una segunda propuesta de desarrollo de la estructura de arquitectura. En esta, el subsistema se encuentra separado del nodo, con su propia CPU, y se conecta a la red mediante el nodo. Suponiendo que un nodo falle, debido a su baja confiabilidad (es un componente COTS), ese subsistema queda totalmente aislado de la red. En la Figura 5.6 se observa que, cuando falla el Nodo del *Subsistema 1* cualquier mensaje que envíe el *Subsistema 2* al primero, nunca se entregará. Esto provoca que la confiabilidad del sistema completo, disminuya linealmente con la confiabilidad del componente COTS.

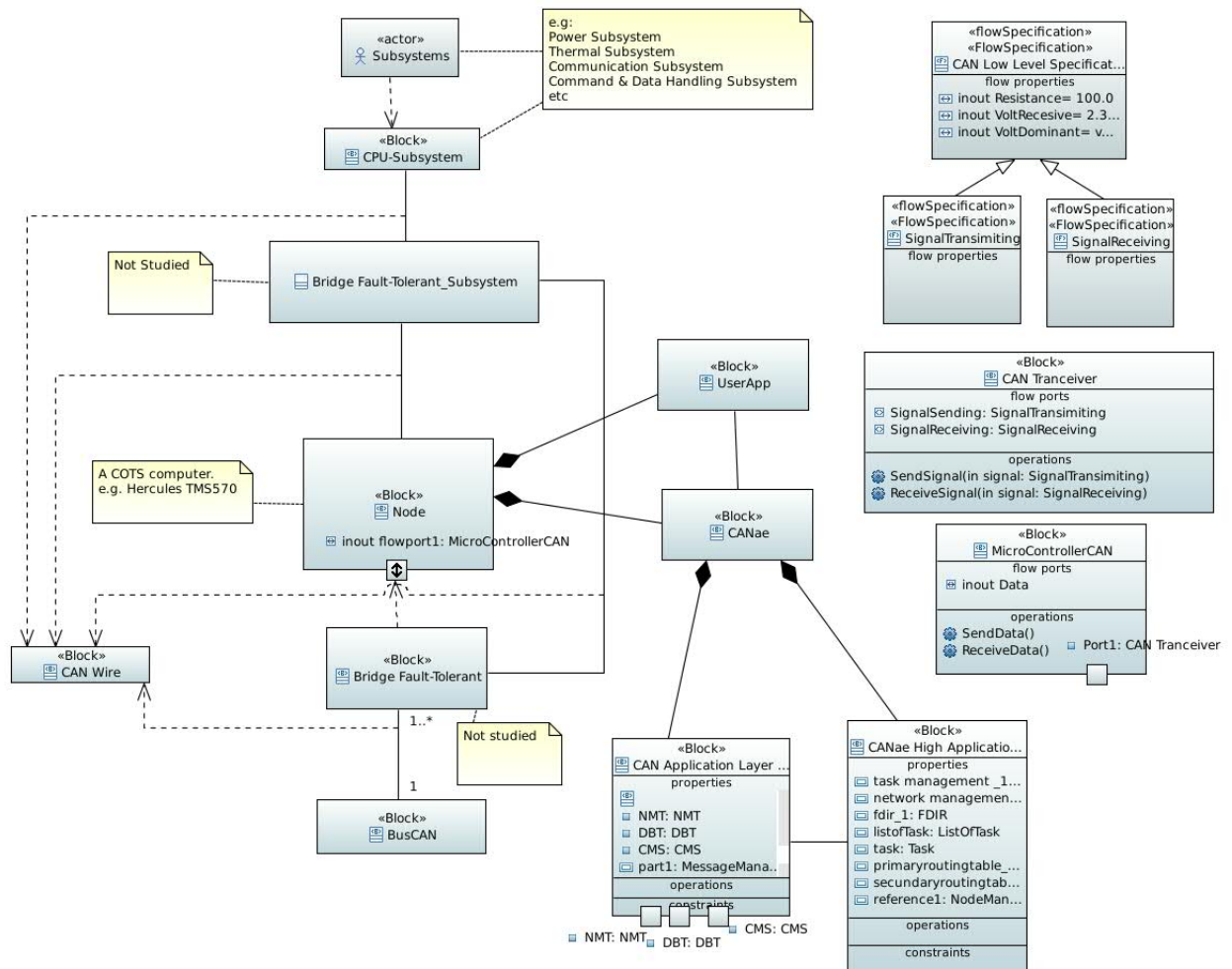


Figura 5.4: Diagrama de bloques de la arquitectura completa

La tercera opción y la ideal es la existencia de un Brdige, que se lo llama: *Bridge Tolerantes a Fallas*. Esto crea un puente entre la red y la computadora del subsistema permitiendo aumentar la tolerancia a fallas de la arquitectura. En caso falla en el nodo, el Bridge se activa ignorando al *nodo fallado* y conectando a la computadora del subsistema directamente en la red. Esto se puede observar en la Figura 5.7, siendo el Bridge el cuadrado negro.

Los detalles técnicos, electrónicos y de implementación no se desarrollan en este trabajo de tesis.

Como puede observarse en al Figura 5.4 el protocolo CANae se encuentra dentro del Nodo. El desarrollador se debe encargar de implementar los protocolos, como un stack de servicios más del sistema operativo (o sistema embebido) que se encuentre en el nodo.

5.4.2. Bloques Internos

En esta sección se muestra el diagrama de bloques internos que conforman el nodo que es aquel que implementa el protocolo de comunicación CANae. Como puede observarse en la Figura ?? el nodo, representa la parte más compleja de la red. En este dispositivo se encuentra toda la complejidad de la arquitectura, y es el encargado tanto de llevar a cabo las tareas, como así también comunicarse con los demás nodos, a través del protocolo CANae y coordinar las tareas a desarrollarse. Además, debe notarse que estos nodos se suponen

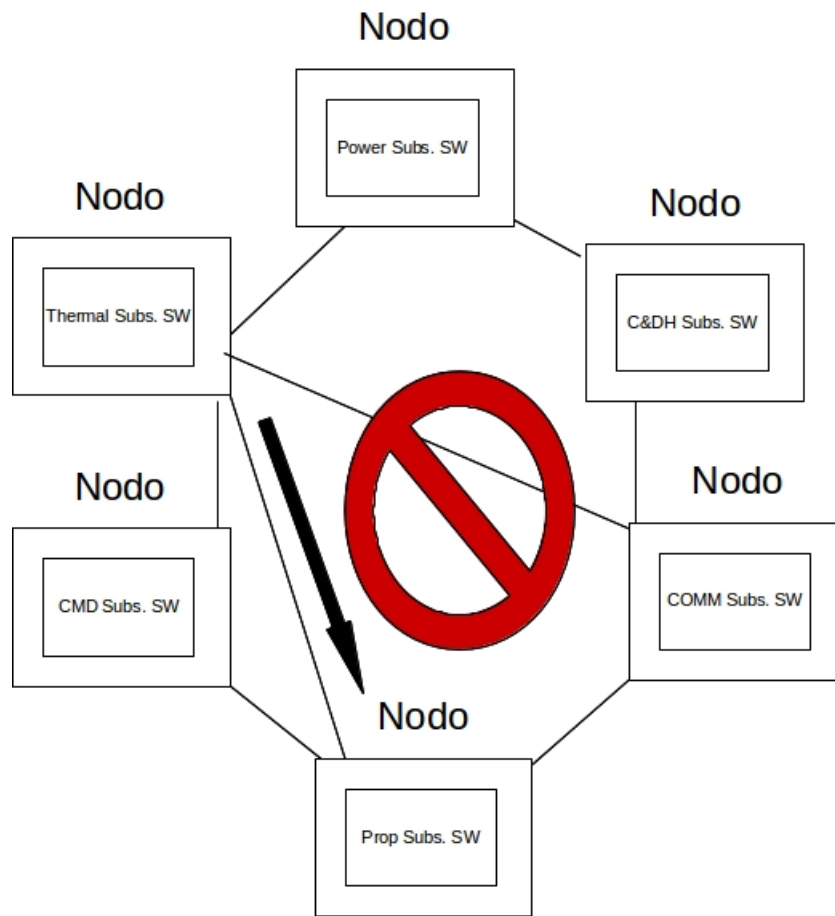


Figura 5.5: Posibilidad de interconexión de nodos.

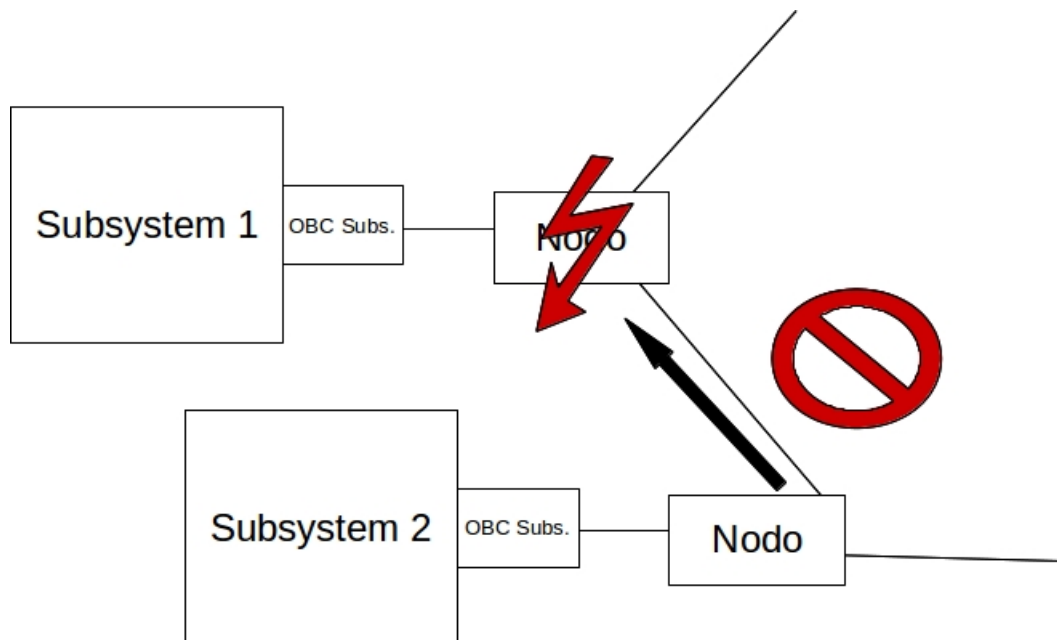


Figura 5.6: Posibilidad de interconexión de nodos.

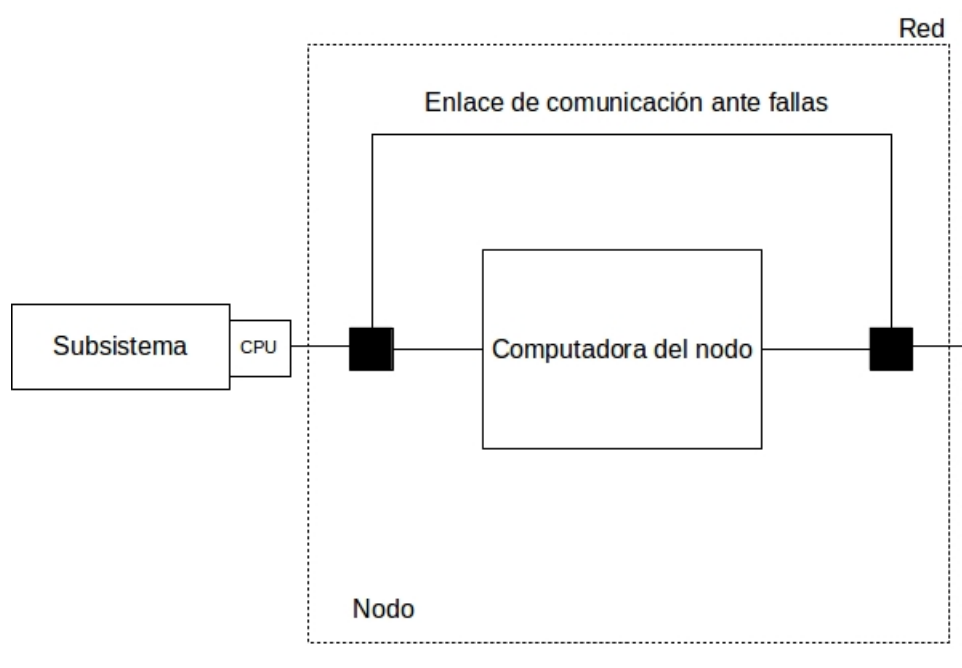


Figura 5.7: Conexión entre la red y el subsistema

que son componentes de estantería o componentes COTS los cuales tiene una muy baja confiabilidad pero un alta performance. Por lo tanto, es de esperarse que el software de los nodos esté basado o desarrollado sobre un Sistema Operativo de Tiempo Real (RTOS) y estos deben aplicar técnicas de tolerancia a fallas que escapen de los alcances de este trabajo de tesis. Todos los bloques internos que se observa en la Figura ?? deben desarrollarse como módulos pertenecientes al stack del servicio del sistema operativo, o bien, módulos de servicio del sistema embebido de los nodos.

Como se estudió en el marco teórico (Véase sección 2.17) el estándar del protocolo CAN (CAN-CIA, 11 de abril del 2017) trabaja en las capas inferiores del modelo de OSI. Cada Nodo que se deba conectar a una red basada en CAN debe, en primer lugar, contar con un CAN Tranceiver y un Microcontrolador CAN, estos son modelados en el diagrama de bloques internos como Ports. CANae (Véase apéndice A) por su parte, trabaja en la capa de aplicación del modelo de OSI. Esta es modelada como una Parte (Part) del Nodo. Como se puede observar en el modelo ??, CANae es la capa de comunicación de la red que se comunica directamente (como es de esperarse) con la aplicación del usuario.

5.4.3. Nodo Monitor

En la versión actual del protocolo CANae (0.1 Alpha) preve la existencia de un nodo monitor. El objetivo principal de este nodo es lograr la coordinación de todos los nodos en las fases iniciación del protocolo. Una vez que la red CANae ya fue creada el nodo monitor se puede ser retirada de esta.

5.5. Modelo dinámico

En esta sección se modelará el comportamiento de la arquitectura. A diferencia de lo modelos que se estudió en la sección anterior (diagramas de bloques, diagrama de bloques internos) que hacen referencia al *qué* del sistema, en el modelado del comportamiento se hace rereferencia al *cómo* del sistema. Este *cómo* es descripto

por el modelado del comportamiento dentro del sistema (Holt y Perry, 2008).

En esta sección se verán diferentes 3 niveles de abstracción diferentes de los modelos de comportamiento del sistema. En primer lugar, el nivel más alto de abstracción es la máquina de estado, que considera la interacción entre bloques. El siguiente diagrama que se verá es el diagrama de secuencia, el cual pone en manifiesto los diferentes mensajes que son enviadas entre las entidades. Y por último, se desarrolla el diagrama de actividades, siendo este el nivel más bajo de abstracción del modelado del comportamiento del sistema.

5.5.1. Máquina de estado

En esta subsección se describe el ciclo de vida de los bloques que conforman la arquitectura. En resumen, lo que se pretende demostrar, es el compartamiento de los objetos de la arquitectura.

En la Figura 5.9 se presenta la máquina de estado de la arquitectura completa. En esta se puede observar todos los estados por los cuales la arquitectura debe recorrer.

A continuación se procede a explicar la máquina de estado. Todo inicia en el momento en el que la red es alimentada (eléctricamente) siguiendo los protocolos de CAN (CAN-CIA, 11 de abril del 2017). En primer lugar, la red pasa a un estado de inicialización de los nodos, en los cuales, estos son alimentados, se encienden, y comienzan con todos los procedimientos normales de cualquier sistema embebido. Los cuales son Booteo, chequeo del funcionamiento y del estado de salud del propio HW y los HW conectados a él. Esto incluye la iniciación, del protocolo CANae como un stack más del servicio del sistema operativo o sistema embebido que gobierna a cada uno de los nodos. Esto se observa en la Figura 5.10.

Una vez que todos los nodos se encuentran inicializados, y la red se encuentra estable², la arquitectura pasa a un estado de Stand-By. En este estado, la arquitectura espera por la orden del nodo monitor para comenzar a configurar la red. Aquí se sigue lo indicado en el protocolo CANae (Véase apéndice A). En primer lugar el se crean las instancias de red localmente. Luego el nodo monitor coordinar la creación de las instancias de red remota. Luego se crean los objetos nodos tanto locales como remotos. Estas actividades son necesarias para lograr entablar una correcta comunicación entre nodos utilizando el protocolo CANae (Versión 0.1 Alpha).

Una vez finalizada la creación de la red, la arquitectura completa pasa a un estado Ready, donde el nodo monitor ya puede ser retirado. Luego, la arquitectura ignora el nodo monitor (Si todavía se encuentra conectado) y ya puede comenzar a trabajar normalmente.

5.5.2. Diagrama de secuencia

A continuación se muestra los diagramas de secuencias, para entender cual es el comportamiento de la arquitectura, en un grado de abstracción menor.

En la Figura 5.11 se puede observar como en el momento de alimentar (eléctricamente) a los nodos, este comienza en primer lugar se ejecuta el bootstrap. Este bootea al sistema operativo (o sistema embebido según sea la tecnología que se utilice). Una vez que el sistema operativo es cargado se deben ejecutar (preferentemente) dos tipos diferentes de chequeos:

- Hardware Check
- Software Check

²Defínase como estable, la situación en la que el protocolo CANae se inicializó correctamente y está listo para pasar a la segunda etapa.

Estos son dos piezas de código. El primero tiene como tarea principal es chequear el estado de salud de los equipos de hardware, tanto del propio nodo, como así también del subsistema que tiene más proximo. Como se expuso en la subsección 5.4.1 cada nodo tiene próximo a él (físicamente hablando) un subsistema (con su computadora de vuelo). Por lo tanto el nodo se debe encargar de chequear el estado de salud de los diferentes componetes de ese subsistema. El segundo chequeo, se ocupa de verificar el estado de salud del software tanto del nodo, como del subsistema más próximo.

Luego de que los chequeos se hayan resuelto exitosamente se crea una instancia del protocolo CANae y da inicio a su servicio NMT. Este servicio se puede observar en la Figura ???. En primer lugar se debe esperar que el nodo monitor esté inicializado, y este le envía un mensaje de *crete_remote_network()* a todos los nodos conectados a la arquitectura. Este mensaje llega al NMT del protocolo de CANae de todos los nodos.

Luego, siguiendo el protocolo de CANae, cada nodo crea su instancia de nodo. Esto puede observarse en la Figura 5.13. El objeto de *CANAppLayerController* envía el mensaje *create_node()* al *NodeManagement* de *create_newtork_object()* con el objetivo de crear una instancia local de la red.

Luego el nodo monitor envía el mensaje *prepare_remote_node*. El protocolo CANae preve los pasos que se muestran en la Figura 5.15

Por último, se conectan todo los nodos a la red, para lo cual deben actualizar el *Node Table Configuration*, y finalmente ya ya pueden comenzar a comunicarse. Esto se observan en las Figuras 5.15 y 5.16

5.5.3. Diagrama de actividades

En esta subsección se verá algunos diagramas de actividad para modelar el comportamiento de la arquitectura propuesta. Los diagramas de actividades comúnmente son utilizados generalmente para describir el comportamiento a muy bajo nivel (a nivel de componentes), en otras palabras para modelar procesos, mientras que los diagramas de secuencia muestran el comportamiento entre elementos y las máquinas de estados muestra el comportamiento dentro de los elementos (Holt y Perry, 2008).

En la Figura 5.17 se puede observar el proceso, en detalle, del inicio de los nodos. Todo comienza cuando se alimenta (eléctricamente) el nodo. En primer lugar, se ejecuta (como ocurre normalmente) el bootstrap, este bootea al sistema operativo o al sistema embebido (dependiendo de como se implemente). Los siguientes pasos son los esperados para cualquier sistema espacial. Se chequea el HW propio del nodo, se chequea memoria y componentes. Luego, se chequea el estado de salud del subsistema. Para ello, se puede pedir el chequeo a la CPU del subsistema, y este a vez, chequea los componentes propios del subsistema.

Con estos pasos se puede obtener el Houskeeping (HK) tanto de HW propio del nodo como del subsistema. Este HK es recibido por el sistema operativo (o sistema embebido), y se comienza a hacer un chequeo del lado del software. Esto significa que se pueden realizar cualquier tipo de actividad para chequear que las tareas (o procesos) se ejecutan correctamente.

Para finalizar, el sistema operativo crea una instancia del stack de servicio de comunicación, dónde se encuentra el protocolo de comunicación CANae.

Debe observarse que el chequeo de HW tanto propio como el del subsistema se encuentra dentro de lo que se denomina *zona interruptible*, esto significa que si se da alguna *Falla* (Failure en la Figura), debe activarse el sistema FDIR del nodo. Este lo que busca en primer lugar es detectar la falla, dónde se produjo y por qué. Luego, El sistema trata de aislar la falla para que no afecte a ningún otro componente. Por último el sistema hará el intento de recuperarse de la falla, en el caso de que se recupere correctamente, se regresa hacia las tareas de chequea para llevar a cabo estas actividades nuevamente; en caso contrario se termina la actividad.

Otra parte importante a modelar de la arquitectura propuesta, es el momento de la creación de la red CANae.

Como se mencionó anteriormente el protocolo CANae (Véase apéndice A) requiere de la existencia de un nodo monitor para coordinar la creación de la red CANae. Este modelo se puede observar en la Figura 5.18.

En nodo monitor inicia la creación y configuración del nodo enviando la señal *create_remote_network()* a todos los nodos conectados a la red. Cada nodo que recibe la señal, lleva a cabo todas las actividades prevista por el protocolo CANae para la creación y configuración de la red. Crea las instancias de la red remota, los objetos red, crea la tabla de configuración de nodos y crea por último las instancias de los nodos.

En segundno paso el nodo monitor envía la señal de *Prepare Remote Node*, la cual es recibida por todos los nodos y se preparan internamente para comenzar a comunicarse.

Por último, el nodo nodo monitor envía la señal para que los nodos se conecte a la red. Cada nodo se conecta, y actualiza su tabla de configuración de nodo. A partir de este momento los nodos ya pueden comenzar a enviar mensajes y el nodo monitor ya puede ser extraído de la red.

Por último también se modela el “cómo” se comporta la arquitectura al momento de tratar de enviar un mensaje de uno nodo a otro nodo. Esto se puede observar en la Figura 5.19. En primer lugar el nodo emisor, debe “preparar” el mensaje. Esto significa que debe empaquetarlo para, cumplir tanto con los protocolos CANae, como así también con el procolo CAN.

En segundo paso se debe consultar a la tabla de configuración de nodo, para conocer cuál es el camino óptimo para llegar a entregar el mensaje. Para ello se le aplica algún algoritmo de ruteo.

Por último, se envía el mensaje através del microcontrolador CAN. Este es recibido por el nodo destino y luego es procesado. Esto puede observarse con mayor detalle en el apéndice A dónde se desarrolla el protocolo CANae 0.1 Alpha.

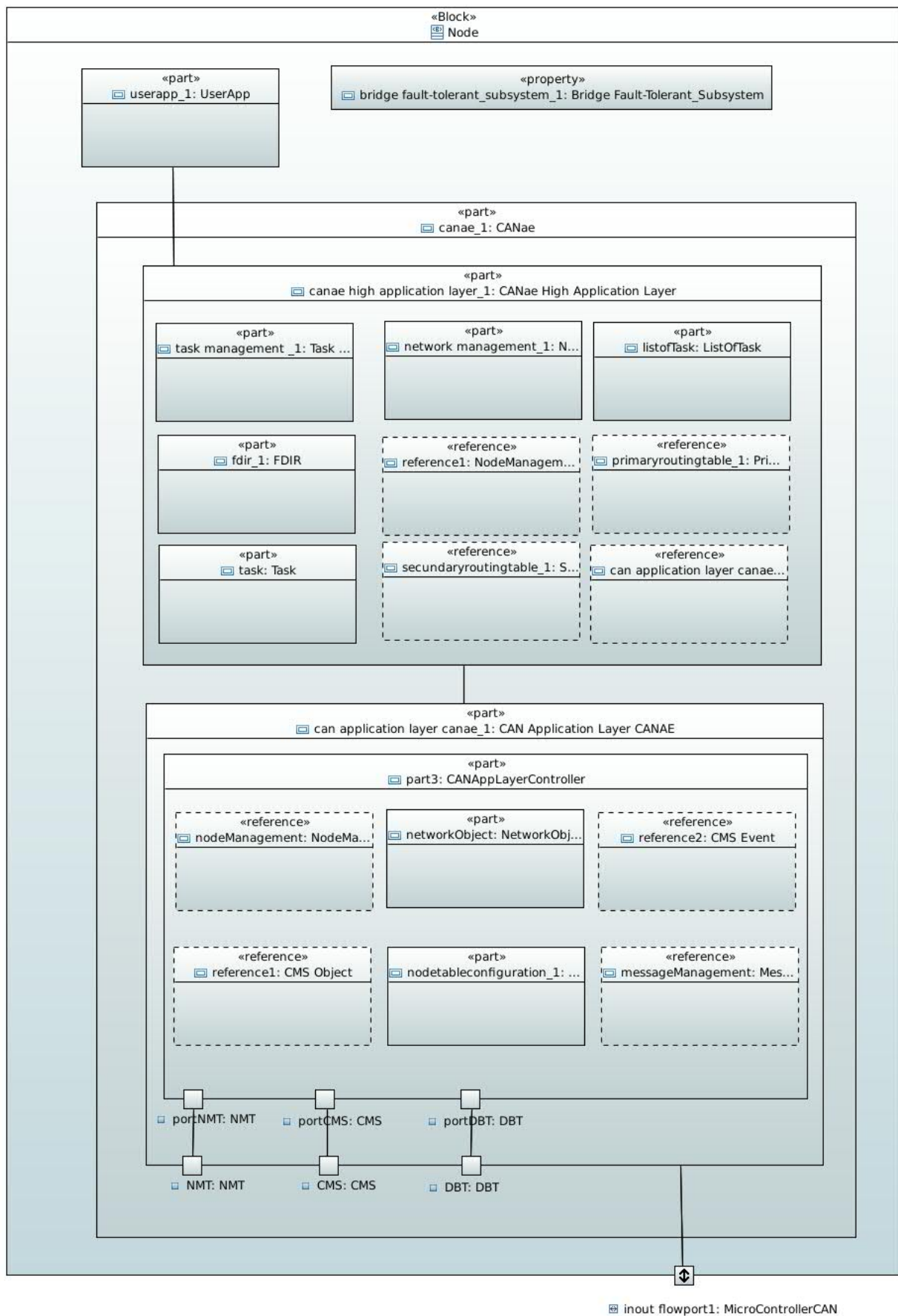


Figura 5.8: Diagrama de bloques interno del nodo.

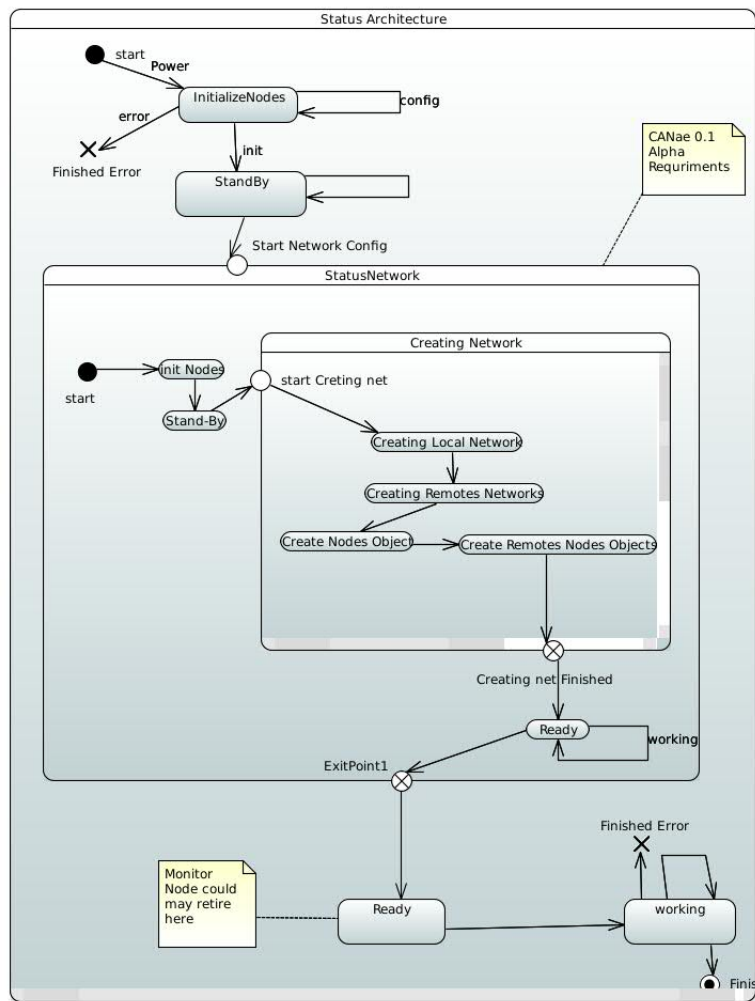


Figura 5.9: Máquina de estado de la arquitectura completa

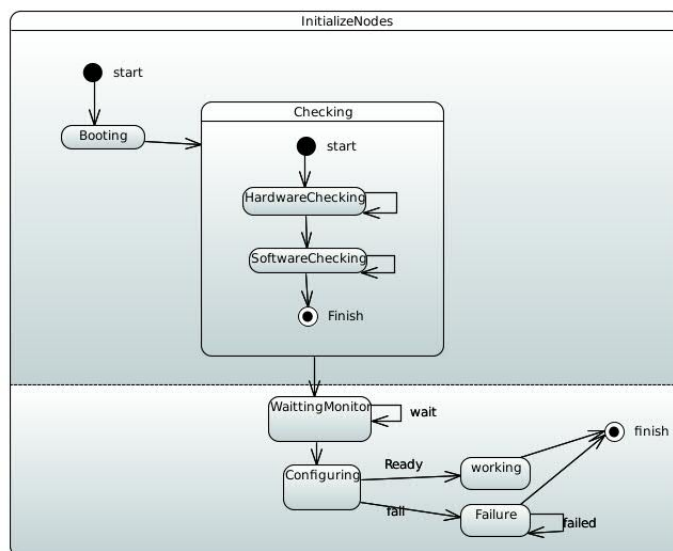


Figura 5.10: Máquina de estado de los nodos.

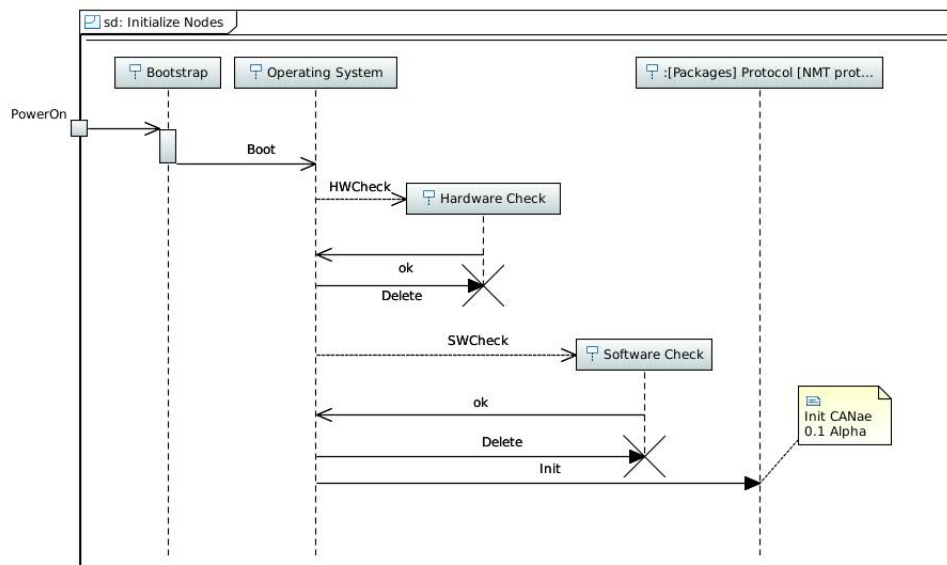


Figura 5.11: Diagrama de secuencia del inicio de la arquitectura.

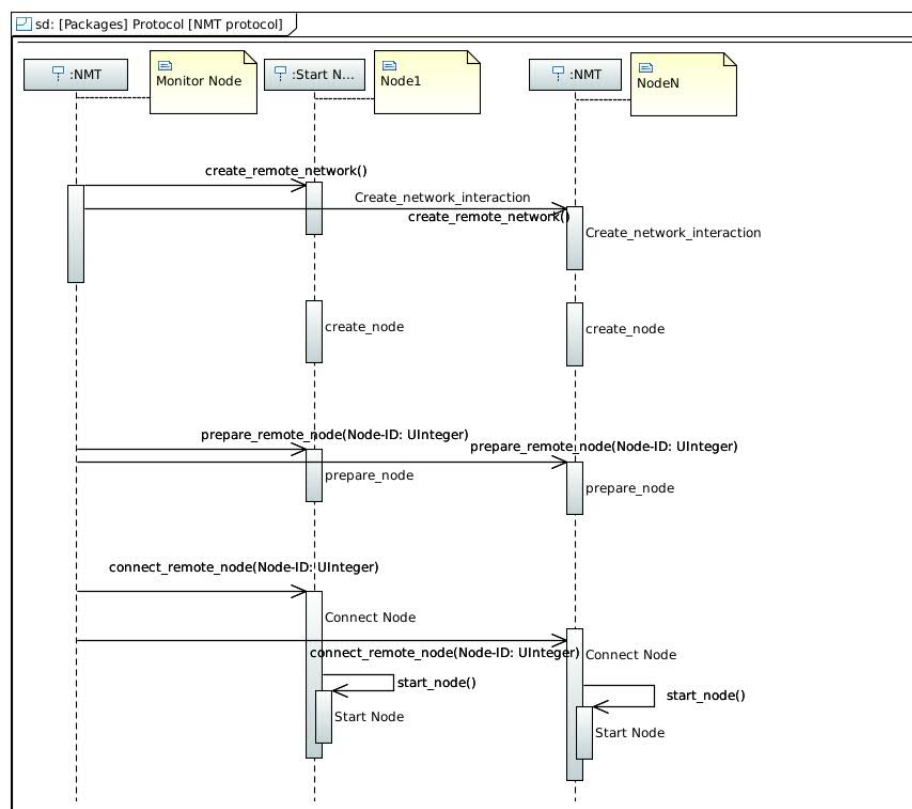


Figura 5.12: Diagrama de secuencia de CA

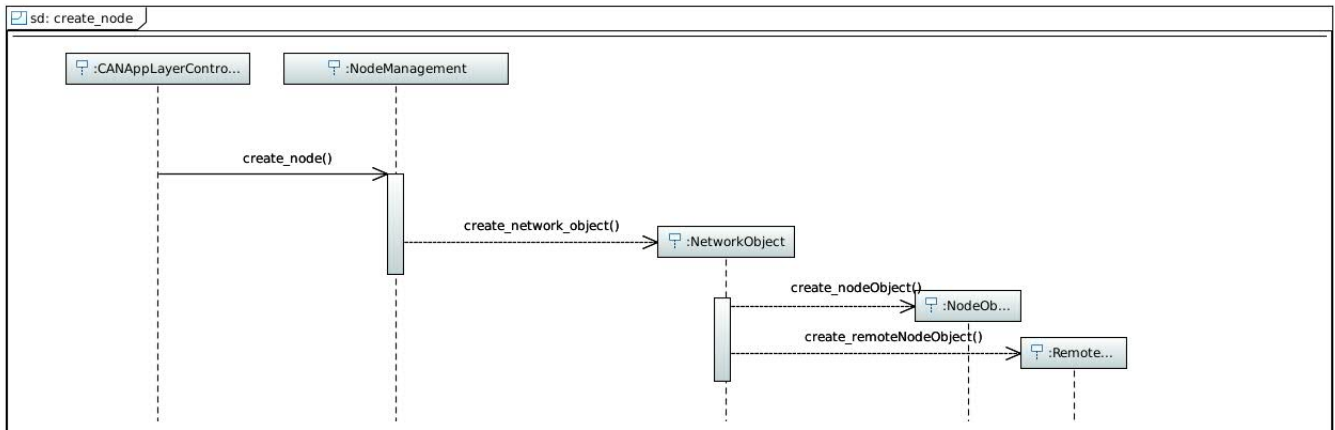


Figura 5.13: Crear Objeto Red, Objeto Nodo y Objeto Nodo Remoto

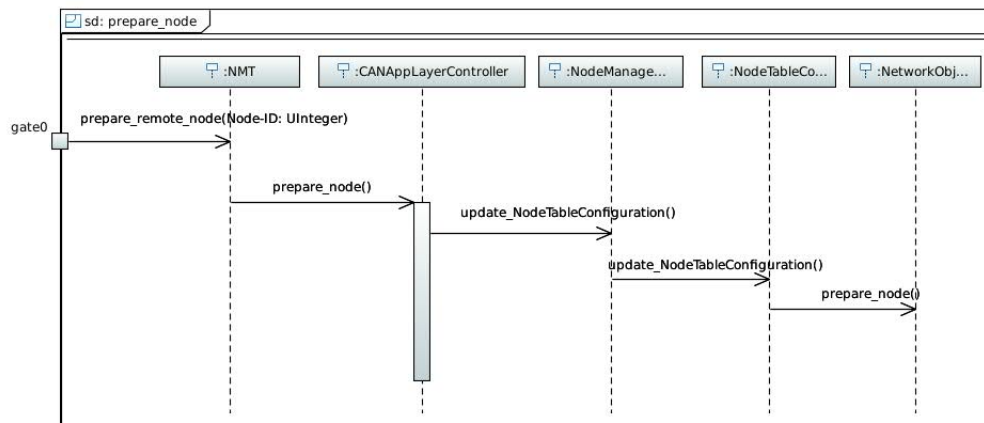


Figura 5.14: Preparar nodo para la conexión a la red CAN

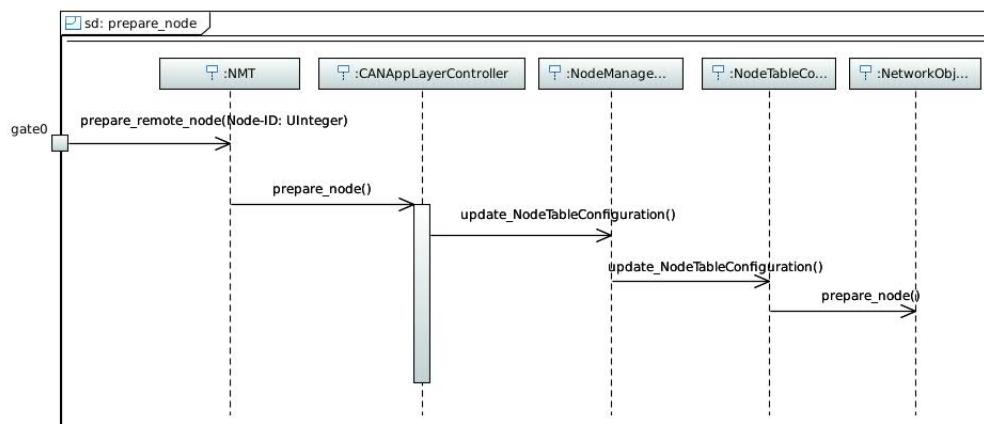


Figura 5.15: Preparar nodo para la conexión a la red CAN

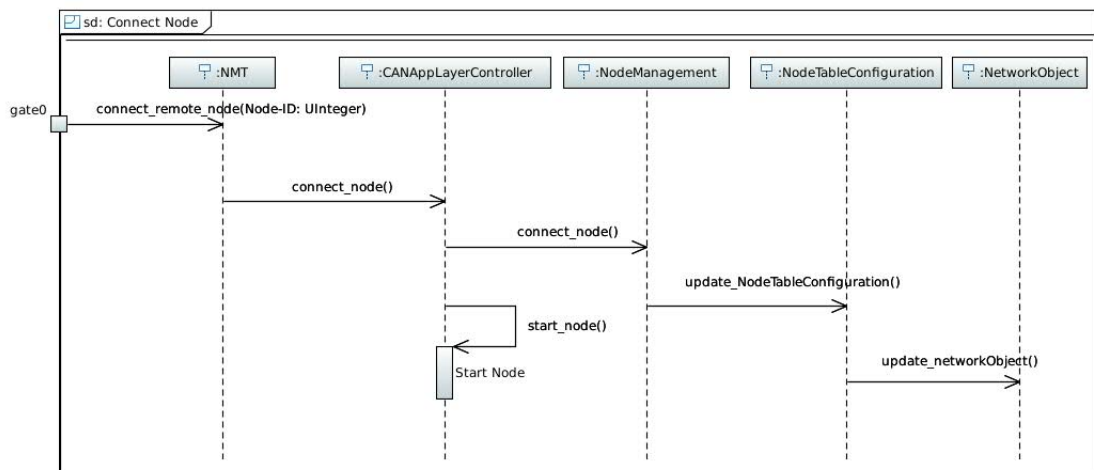


Figura 5.16: Conectar el nodo a la red CAN

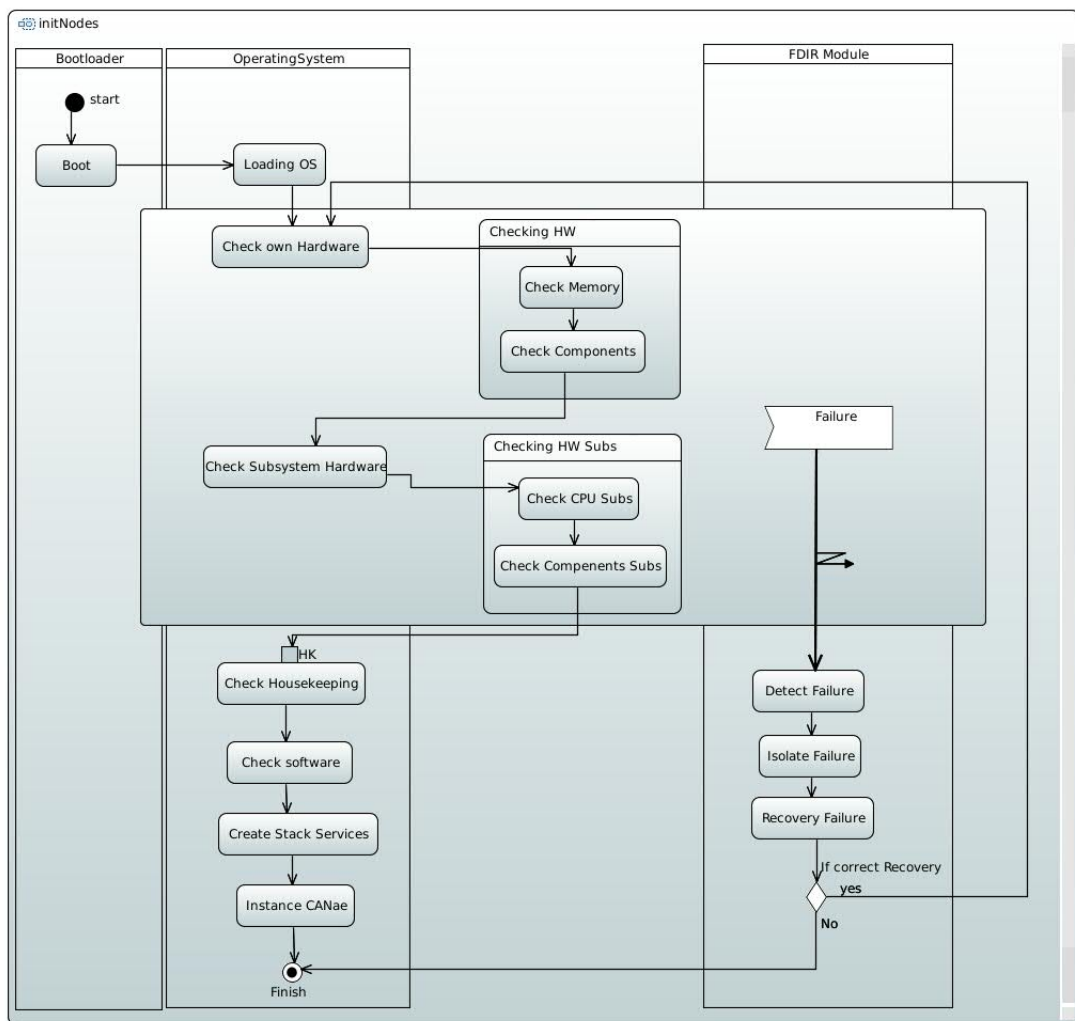


Figura 5.17: Diagrama de actividades del inicio de nodos

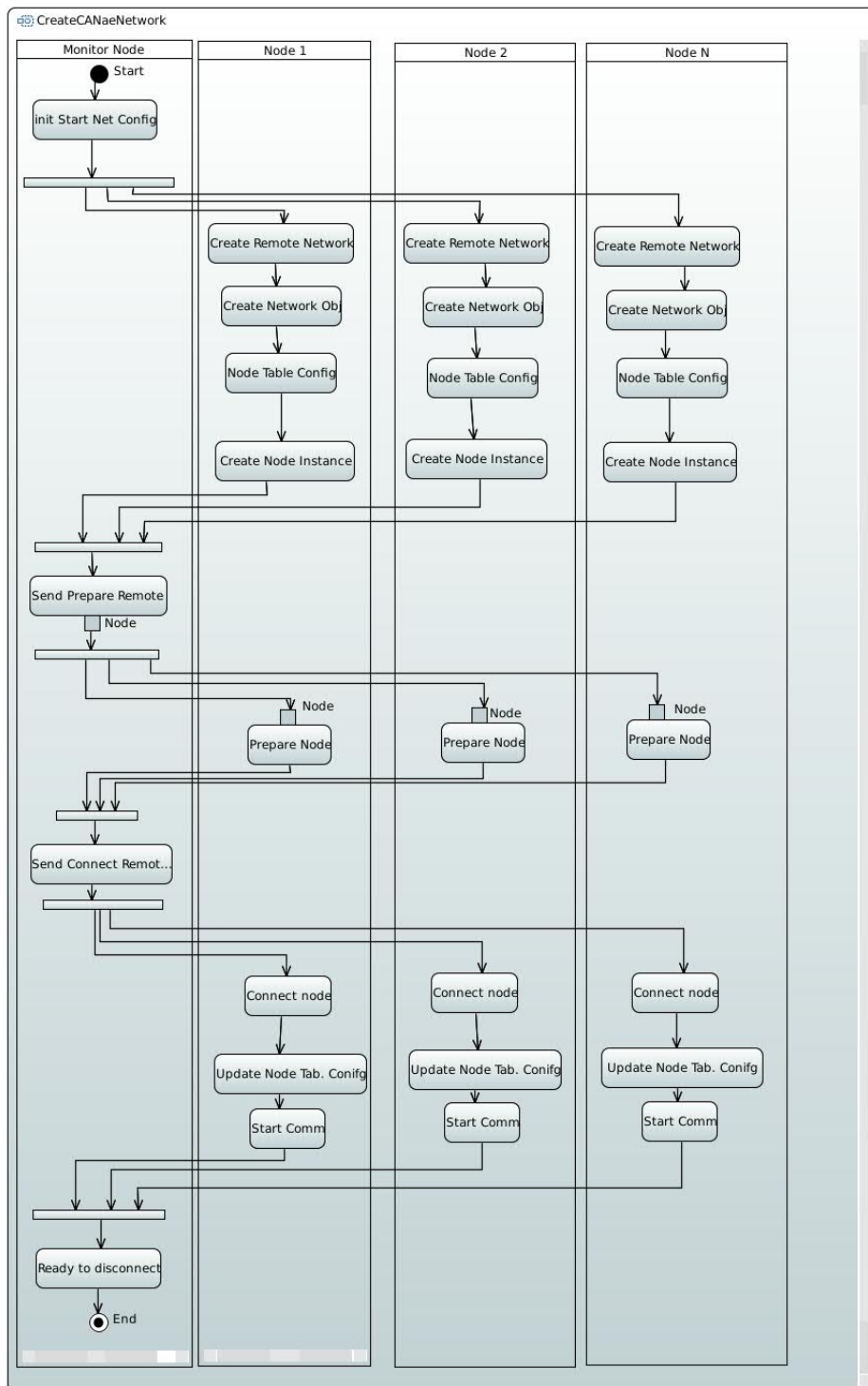


Figura 5.18: Diagrama de actividades del inicio de nodos

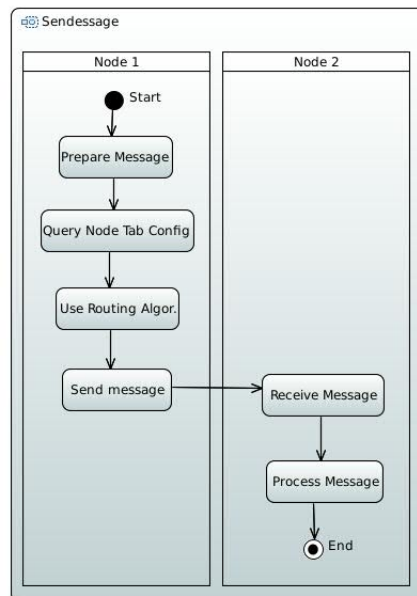


Figura 5.19: Diagrama de actividades del inicio de nodos

Conclusión

Las personas que están lo suficientemente locos como para pensar que pueden cambiar el mundo, son los que lo hacen.

Steve Jobs

No es un secreto de que llevar a cabo una misión satelital representa grandes costos y largos períodos de tiempo de desarrollo de la misión. La NASA también padeció estos problemas hace algunas décadas. Pero aquella, al ser una organización consciente de sus habilidades (y sobre todo con recursos), se puede dar el lujo de innovar, de ir en contra de las corrientes y filosofías normales de desarrollo. Así, es como surge su lema *Faster, Better and Cheaper* (FBC). Este lema fue introducido por el administrador de la NASA Dan Goldin a principios de la década de los 90's. Dan Goldin tenía una vasta experiencia en el desarrollo de pequeños satélites. Él estaba convencido de que el desarrollo "clásico" de misiones satelitales de NASA necesitaba un cambio de enfoque, para lograr reducir el tiempo de desarrollo, y pasar de escalas de tiempo de décadas a años y sobre todo lograr reducir costos (Paxton, 2007).

Es sabido que este lema produce una gran cantidad de fallas y carencias en las misiones satelitales (Paxton, 2007). Sin embargo, esto es lo que impulsa la innovación, y el desarrollo y entrenamiento de nuevos managers, ingenieros y científicos, los cuales serán los líderes de las nuevas generaciones. (Paxton, 2007). Por tales motivos, es importante, que empresas y organizaciones del ámbito espacial en Argentina, dejen de lado el "clasicismo", y se deben animar, y por su puesto, alentar a las futuras generaciones, a transformar el tradicional *caminar con los manuales de NASA bajo el brazo*, por la innovación y el desarrollo de nuevas tecnologías y técnicas para la realización de satélites (que en este trabajo se denominan *vehículos de nueva generación*). Esta visión la tienen empresas tales como INVAP, y otras tanto nacionales como extranjeras, que están creciendo a pasos agigantados llevando en su interior el cambio de filosofía que se exige, para reducir costos, y con la capacidad necesaria para producir a gran escala.

Este trabajo tiene como objetivo demostrar que es posible desarrollar vehículos espaciales utilizando componentes COTS de baja confiabilidad, trayendo consigo una reducción considerable del costo y del tiempo de desarrollo. De más está decir, que esta tesis no busca desarrollar en profundidad una arquitectura altamente confiable llegando a niveles de detalles, que son esperables en una misión satelital, en la que participan varios equipos de decenas de ingenieros, managers y científicos. Si no, más bien, ser el puntapié inicial de diversas líneas de investigación y desarrollo, ya que como se ha demostrado en este trabajo, es factible el desarrollo de estas arquitecturas "económicas", con la aplicación de técnicas y filosofías que escapen de lo tradicional.

Luego de llevar a cabo un análisis de los sistemas satelitales y de aviónica que se han publicado en diferentes trabajos de investigación, y como resultado de diversas reuniones con varios equipos dentro de la Unidad de Desarrollo INVAP, se llegó a la conclusión, de que se debe trabajar sobre: el desarrollo de aviónicas confiables aún cuando sus componentes sean de baja confiabilidad, la comunicación entre componentes, y que existe una tendencia a implementar redes de componentes dentro de los sistemas.

Así, una vez realizado este análisis previo, se estudiaron diferentes redes que tengan la característica de ser tolerante a fallas y que sean factibles de ser aplicadas en el ambiente espacial, estas se estudiaron en el Marco Teórico (Capítulo 2 Página 6) y en el Estado del Arte (Capítulo 3 Página 36). Las cuales fueron las siguientes:

- Árboles binarios
- Redes hypercube
- Redes distribuídas

Los resultados del análisis de la confiabilidad se encuentra detallada en la sección 4.4 (Página 48) y en Arias y Wiman (2017). Resumiendo, como se puede observar en la Figura 6.1, una vez realizada una comparativa entre redes tolerantes a fallas, la que presenta un mayor grado de confiabilidad sostenible en el tiempo es la que corresponde a la red distribuida. Este resultado representa un importante aporte para el conocimiento de tipos de topologías confiables tolerantes a fallas, y el desarrollo de modelos para lograr medir la confiabilidad de esas topologías. Los resultados de este estudio fueron presentados en el *Congreso Argentino de Tecnologías Espaciales 2017 (Córdoba, Argentina)*

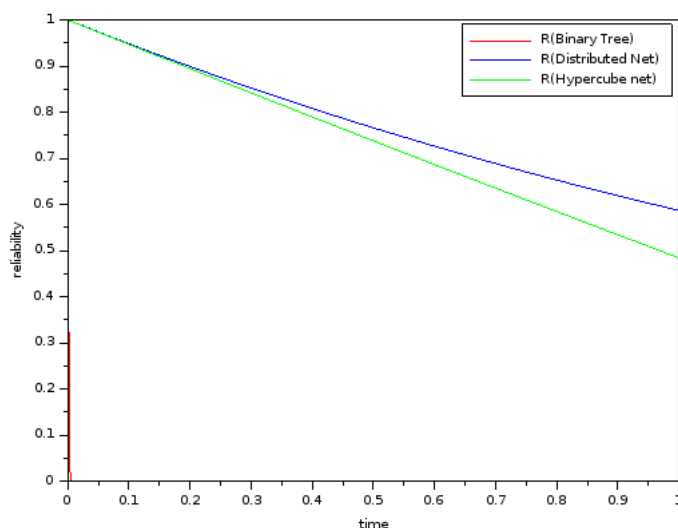


Figura 6.1: Comparación de confiabilidad

Un requerimiento que fue planteado por la Unidad de Desarrollo INVAP fue trabajar con el bus de comunicación CAN. CAN, es un protocolo de comunicación de bajo nivel, que demuestra ser altamente confiable por lo que es utilizado (y es de donde nació) mayormente en la industria automotriz. Esto, nos da una clara razón de que CAN, es un protocolo confiable, debido a que en la fabricación de automóviles, el producto final (el automóvil) puede poner en riesgo la vida de personas. Debido a su confiabilidad, y a su excelente respuesta en sistemas de tiempo real, CAN es utilizado en también en el ambiente espacial (entre otras industrias).

En base a estos resultados y restricciones se plantea el desafío de implementar, crear, diseñar, una arquitectura que sea capaz de trabajar, con una filosofía de red distribuida, esta puede estudiarse en 5 (Página 57). El grado de innovación que coloca esta nueva arquitectura se ve reflejada en que no existe una computadora única (la connotación “única” hace referencia a la computadora propiamente dicha más sus redundancias, o bien, cualquier (micro)controlador más sus redundancias) por cada subsistema como se trabaja “tradicionalmente” sino que, existen computadoras híbridas capaz de llevar a cabo tareas de todos los sistemas en forma distribuida.

Para ello se planteó un protocolo basado en CAN y CANOpen. Este protocolo, hasta el momento de desarrollar este trabajo, se encuentra en su versión 0.1 Alpha. El protocolo presenta la innovación de estar orientado a objetos, y para ello se utilizó el lenguaje SysML (System Modeling Language) para llevar a cabo su modelado y la documentación del mismo. Esto facilita la interpretación del stack de servicio que deben ser implementados a la hora del desarrollo. Este protocolo se encuentra documentado en el apéndice A (Página 85).

En base a los resultados obtenidos y a la arquitectura propuesta basada en CANae (apéndice A) se concluye de que es factible desarrollar una *arquitectura de aviónica tolerantes a fallas basada en componentes COTS para vehículos satelitales* utilizando el bus CAN, como bus principal de comunicación. Si bien esto exige un cambio de paradigma importante, este trabajo concluye con una clara demostración de que es posible diseñar arquitecturas fiables basadas en componentes baratos y de baja confiabilidad.

6.1. Cumplimiento de los objetivos

En la Tabla 6.1 se describe si los objetivos planteados fueron cumplimentados o no.

En cuanto a las preguntas de investigación que se plantearon para este trabajo de tesis, las respuestas de estas se describen a continuación:

- **¿Es posible la realización de un método de medición del grado de tolerancia a fallas de una arquitectura de aviónica?** *No fue posible responder. Se logró desarrollar una manera de comparar la confiabilidad de topologías de arquitecturas, basándose en la tolerancia a fallas.*
- **¿Cuál es la estrategia más indicada de tolerancia a fallas que permita brindar un alto grado de confiabilidad en la utilización de componentes COTS en sistemas críticos?** *En este trabajo se demostró que la estrategia más indicada es desarrollar una arquitectura basada en red distribuida.*
- **¿Cuál es la arquitectura más indicada que permita desarrollar tolerancia a fallas en sistemas críticos basados en componentes COTS?** *La arquitectura basada en CANae desarrollada en este trabajo de tesis, cumple con las necesidades necesarias para sentar las bases de trabajos futuros en la implementación de arquitecturas confiables basadas en componentes COTS.*
- **¿Es factible la utilización de componentes COTS en sistemas espaciales?** *Este trabajo de tesis demostró que es factible la utilización de componentes COTS.*

6.2. Trabajos a futuros

Luego de desarrollado este trabajo de tesis, se logra ver el gran potencia que existe en el estudio de arquitecturas de aviónicas tolerantes a fallas, y el estudio de los componentes COTS introducidos en sistemas, donde el sistema esperado debe ser más confiable que los componentes que la conforman.

Para ello se recomienda proseguir con los siguientes líneas de trabajo futuros:

Tabla 6.1: Tabla de objetivos cumplidos

Objetivo planteado	Resultado
El objetivo de este trabajo es investigar y analizar arquitecturas de comunicación de los subsistemas de aviónica tolerante a fallas basada en componentes COTS para vehículos satelitales de nueva generación.	Cumplido
Realizar un estudio del estado de la cuestión sobre arquitecturas tolerantes a fallas para sistemas críticos.	Cumplido
Investigar y analizar arquitecturas tolerantes a fallas que aseguren la confiabilidad del sistema y que sean aplicables en la industria satelital.	Cumplido
Investigar y analizar protocolos de comunicación, para las capas superiores del modelo de OSI (modelo de interconexión de sistemas abiertos - ISO/IEC 7498-1), orientados a la tolerancia a fallas y confiabilidad de los sistemas. Realizar un estudio comparativo de los diferentes protocolos estudiados.	No cumplido. Debido a la complejidad del objetivo demandaría una cantidad de tiempo extra, y no alcanzaría para ser cumplido en el tiempo preestablecido. Se decidió trabajar directamente con CAN.
Investigar una metodología para lograr una medición de la tolerancia a fallas en arquitecturas de aviónica	Cumplido. Se investigó metodologías para la medición de topologías.
Desarrollar un estudio comparativo de arquitecturas tolerantes a fallas con el fin de obtener ventajas y desventajas de cada una de ellas.	No cumplido. No se compararon arquitecturas completas debido a que no alcanzaría a ser cumplido en el tiempo preestablecido. En contraposición se compararon topologías.
Diseñar modelos alternativos de arquitecturas tolerantes a fallas, que tenga un grado de confiabilidad tal que permita la aplicación de componentes COTS.	No cumplido. Solo se desarrolló una arquitectura debido a cuestiones de tiempo.
Evaluar la confiabilidad de los modelos de arquitecturas (mediante métrica desarrollada en este trabajo o siguiendo otras estrategias).	No cumplido.
Proponer el diseño de una nueva arquitectura tolerante a fallas, con un grado de confiabilidad suficiente para la aplicación de componentes COTS en aviónicas de vehículos satelitales.	Cumplido.
Simular la arquitectura planteada para medir su grado de tolerancia a fallas y performance.	No cumplido.

- Diseño detallado, desarrollo e implementación del protocolo CANae.
- Diseño detallado de la arquitectura propuesta.
- Estudio de nuevas técnicas de tolerancia a fallas aplicadas a los diferentes niveles de detalle de las arquitecturas de aviónica.
- Desarrollo de algoritmos de ruteo para la distribución de tareas en redes distribuida.
- Estudio de tecnología Wireless como medio de comunicación alternativo al cableado.

Bibliografía

- Mostafa Abd-El-Barr y Fayez Gebali. Reliability analysis and fault tolerance for hypercube multi-computer networks. *Information Sciences*, 276:295 – 318, 2014.
- R. Alena, R. Gilstrap, J. Baldwin, T. Stone, y P. Wilson. Fault tolerance in ZigBee wireless sensor networks. En *Aerospace Conference, 2011 IEEE*, páginas 1–15, 2011.
- Thomas Anderson y John C. Knight. A Framework for Software Fault Tolerance in Real-Time Systems. *IEEE Trans. Software Eng.*, (3):355–364, 1983.
- Emmanuel Arias y Gustavo Wiman. Estudio de la confiabilidad de arquitecturas tolerantes a fallas basada en componentes cots para aviónicas de vehículos espaciales. Asociación Argentina de Tecnología Espacial - Instituto Universitario Aeronautico, Apr 2017.
- Klaus Becker y Sebastian Voss. *A Formal Model and Analysis of Feature Degradation in Fault-Tolerant Systems*. Springer International Publishing, 2016.
- CAN-CIA. CAN-CIA, 11 de abril del 2017. URL <https://www.can-cia.org>.
- Savio. N. Chau, L. Alkalai, Ann T. Tai, y J. B. Burt. Design of a fault-tolerant COTS-based bus architecture. *IEEE Transactions on Reliability*, 48(4):351–359, 1999. ISSN 0018-9529.
- Chau, Savio N. and Smith, Joseph and Tai, Ann T. A design-diversity based fault-tolerant cots avionics bus network. En *Dependable Computing, 2001. Proceedings. 2001 Pacific Rim International Symposium on*, páginas 35–42, 2001.
- Steve Corrigan. Introduction to the Controller Area Network (CAN). Technical Report SLOA101A, Texas Instruments, 2002.
- Steve Corrigan. Controller Area Network Physicalnn Layer Requeriments - ICP, Industrial Interface. Technical Report SLLA270, Texas Instruments, 2008.
- Edward Crawley, Olivier de Weck, Steven Eppinger, Christopher Magee, Joel Moses, Warren Seering, Joel Schindall, David Wallace, y Daniel Whitney. Engineering Systems Monograph, The influence of architecture in engineering systems. 2004.
- Dr Ralph D. Lorenz (auth.) David M. Harland. *Space Systems Failures: Disasters and Rescues of Satellites, Rockets and Space Probes*. Springer Praxis Books. Praxis, 1 edition, 2005. ISBN 978-0-387-21519-8,978-0-387-27961-9.
- Douglas Isbell and Don Savage. Mars Climate Orbiter Failure Report - NASA, 1999.

- Elena Dubrova. *Fault-Tolerant Design: an introduce*. Springer-Verlag New York, New York, U.S.A., 1 edition, 2013.
- Guillaume Jacques Joseph Ducard. *Fault-Tolerant Flight Control and Guidance Systems for a Small Unmanned Aerial Vehicle*. Tesis de Doctorado, Swiss Federal Institute of Technology Zurich, 2007.
- Cristopher Edwards, Thomas Lobaerts, y Hafid Smaili. *Fault Tolerant Flight Control. A benchmarck challenge*. Springer-Verlag, 2010. ISBN 978-3-642-11689-6.
- Jens Eickhoff. *Onboard Computers, Onboard Software and Satellite Operations*. Springer-Verlag Berlin Heidelberg, Alemania, 2012. ISBN 978-3-642-25169-6.
- ESD Electronics. ESD Electronics, 2017. URL <http://www.esd-electronics-usa.com/Controller-Area-Network-CAN-Introduction.html>.
- S. Esposito, C. Albanese, M. Alderighi, F. Casini, Giganti L., M. L. Esposti, C. Monteleone, y M. Violante. Cots-based high-performance computing for space applications. *IEEE Transactions on Nuclear Science*, 62 (6):2687–2694, 2015.
- Marie-Aude. Esteve, Joost-Pieter Katoen, VienY. Nguyen, Bart Postma, y Yuri Yushtein. Formal correctness, safety, dependability, and performance analysis of a satellite. páginas 1022–1031, 2012. doi: 10.1109/ICSE.2012.6227118.
- J. S. Eterno, J. L. Weiss, D. P. Looze, y A. Willsky. Design issues for fault tolerant-restructurable aircraft control. En *Decision and Control, 1985 24th IEEE Conference on*, páginas 900–905, 1985.
- Kevin Forsberg y Mooz Harold. 4 System Engineering for Faster, Cheaper, Better. *INCOSE International Symposium*, 9(1):924–932, 1999. ISSN 2334-5837.
- Peter Fortescue, John Stark, y Graham Swinerd. *Space Systems Engineering*. Wiley, West Sussex, England, 3 edition, 2003. ISBN 0171619515.
- Peter Fortescue, John Stark, y Graham Swinerd. *Space Systems Engineering*. Springer International Publishing, 2016.
- Sanford Friedenthal, Alan Moore, y Rick Steiner. *A Practical Guide to SysML: Systems Modeling Language*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2008.
- W. C. Gangloff. Common mode failure analysis. *IEEE Transactions on Power Apparatus and Systems*, 94(1): 27–30, 1975. ISSN 0018-9510.
- Hamilton, Deirdre L. and Walker, Ian D. and Bennett, John K. Fault tolerance versus performance metrics for robot systems. *Reliability Engineering & System Safety*, 1999.
- Robert S. Hanmer. *Patterns for Fault Tolerant Software*. John Wiley and Sons Ltd, England, 2007. ISBN 978-0-470-31979-6.
- John P. Hayes. A graph model for fault-tolerant computer systems. *IEEE Transactions on Computers*, 1976.
- Hess, Ronald and Vetter, T. K. and Wells, S. R. Design and evaluation of a damage-tolerant flight control system. En *Institution of Mechanical Engineers Part G Journal of Aerospace Engineering*, 2005.
- Ellis F. Hitt y Dennis Mulcare. Fault-Tolerant Avionics. En Cary R. Spitzer, editor, *Avionics Development and Implementation Second Edition*, capítulo 8. CRC Press, Williamsburg, Virginia, U.S.A., 2070.

- Jon Holt y Simon Perry. *SysML for systems engineering*. The Institution of Engineering and Technology, London, United Kingdom, 2008. ISBN 978086341825.
- K.A. Hoque, O.A. Mohamed, y Y. Savaria. Towards an accurate reliability, availability and maintainability analysis approach for satellite systems based on probabilistic model checking. volumen 2015-April, páginas 1635–1640, 2015.
- IEEE. IEEE Standard Glossary of Software Engineering Terminology. Std. 610.12-1990, IEEE, 1990.
- INVAP. INVAP Sociedad del Estado, 2016. URL <http://www.invap.com.ar/es/>.
- Baback A. Izadi y Füsün Özgüner. An augmented k-ary tree multiprocessor with real-time fault-tolerant capability. *The Journal of Supercomputing*, 27(1):5–17, 2004.
- Jim Krodel y George Romanski. Handbook for real-time operating systems integration and component integration considerations in integrated modular avionics systems. Techn. Rep. DOT/FAA/AR-07/48, U.S. Department of Transportation, 2008.
- Kvaser. Kvaser, 2017. URL <https://www.kvaser.com/can-protocol-tutorial/>.
- Jens C. Lisner. *A Fault-tolerant Dynamic Time-triggered Protocol*. Tesis de Doctorado, Dependability of Computing Systems - University of Duisburg-Essen, 4 2007.
- Yu Liu, Duo Li, y Chao Guo. Software reliability modeling with fault detection data when knowing fault severity. páginas 558–562, 2014. doi: 10.1109/ICRMS.2014.7107257.
- Loveless, Andrew T. capítulo On TTEthernet for Integrated Fault-Tolerant Spacecraft Network. AIAA SPACE Forum. American Institute of Aeronautics and Astronautics, Aug 2015. doi: 10.2514/6.2015-4526. 0.
- Michael R. Lyu. *Software Fault Tolerance*. John Wiley and Sons Ltd, Chichester, New York, USA, 1995. ISBN 9780471950684.
- Pignol M. Dmt and dt2: two fault-tolerant architectures developed by cnes for cots-based spacecraft supercomputers. En *12th IEEE International On-Line Testing Symposium (IOLTS'06)*, 2006.
- MARTE. <http://www.omgmarTE.org/>, 3 de Junio de 2016.
- Jackson R. Mayo, Robert C. Armstrong, y Geoffrey C. Hulet. *Leveraging Abstraction to Establish Out-of-Nominal Safety Properties*. Springer International Publishing, 2016.
- Victor P. Nelson. Fault-tolerant computing: fundamental concepts. *Computer*, 23(7):19–25, 1990.
- Papyrus. <http://eclipse.org/papyrus/>, 3 de Junio de 2016.
- Larry J. Paxton. “faster, better, and cheaper” at nasa: Lessons learned in managing and accepting risk. *Acta Astronautica*, 61(10):954 – 963, 2007. ISSN 0094-5765.
- Zhaoguang Peng, Yu Lu, Alice Miller, Chris Johnson, y Tingdi Zhao. A probabilistic model checking approach to analysing reliability, availability, and maintainability of a single satellite system. páginas 611–616, 2013. doi: 10.1109/EMS.2013.102.
- Dhiraj K. Pradhan y Sudhakar M. Reddy. A Fault-Tolerant Communication Architecture for Distributed Systems. *IEEE Transactions on Computers*, C-31(9):863–870, Sept 1982. ISSN 0018-9340.
- Roger S. Pressman. *Software Engineering, A Practioner's Approach*. McGraw-Hill, fifth edition edition, 2001.

- Laura Pullum. *Software fault tolerance techniques and implementation*. Artech House, England, 2001. ISBN 1-58053-137-7.
- C. S. Raghavendra, AVIvizenis A., y M. D. Ercegovic. Fault tolerance in binary tree architectures. *IEEE Transactions on Computers*, (6):568–572, 1984.
- M. Rausand y A. Hoyland. *System Reliability Theory. Models, statistical methods, and applications*. John Wiley Sons, Inc., Hoboken, New Jersey, USA, 2 edition, 2004.
- RTI. *RTI Purchase Order Terms and Conditions v1.12. Spanish Translation*. RTI International, 3040 East Cornwallis Road, Research Triangle Park, USA, 2015.
- Norman F. Schneidewind. Reliability modeling for safety-critical software. *IEEE Transactions on Reliability*, 46(1):88–98, 1997.
- Norman F. Schneidewind y Allen P. Nikora. Issues and methods for assessing cots reliability, maintainability, and availability. 1998.
- Adit D. Singh y Hee Y. Youn. A modular fault-tolerant binary tree architecture with short links. *IEEE Transactions on Computers*, 1991.
- Steiner, Willfried. An Introduction to TTEthernet. TTTech Computertechnik AG, Apr 2013.
- Constantine Stivaros. A measure of fault-tolerance for distributed networks. En *Computing and Information, 1992. Proceedings. ICCI '92., Fourth International Conference on*, páginas 426–429, 1992.
- Stock Flight Systems. CANAerospace, 2017. URL <http://www.stockflightsystems.com/canaerospace.html>.
- T. Stone, R. Alena, J. Baldwin, y P. Wilson. A viable cots based wireless architecture for spacecraft avionics. En *Aerospace Conference, 2012 IEEE*, páginas 1–11, 2012.
- SysML. <http://sysml.org/>, 2 de Junio de 2016.
- A. T. Tai, S. N. Chau, y L. Alkalai. Cots-based fault tolerance in deep space: Qualitative and quantitative analyses of a bus network architecture. En *High-Assurance Systems Engineering, 1999. Proceedings. 4th IEEE International Symposium on*, páginas 97–104, 1999.
- Andrew S. Tanenbaum y David J. Wetherall. *Redes de computadora*. Pearson Educación, Mexico, quinta edition, 2012.
- A.S. Tanenbaum. *Redes de computadoras*. Editorial Alhambra S. A. (SP), 2003. ISBN 9789702601623.
- Wilfredo Torres-Pomales. Software Fault Tolerance: A tutorial. Technical Report NASA/TM-2000-210616, National Aeronautics and Space Administration Langley Research Center, Hampton, Virginia, 2000.
- TTTech. Technical report, TTTech Computertechnik AG, Vienna, Austria.
- TTTech. <http://tttech.com/>, 20 de Enero del 2017.
- Christopher B. Watkins y Randy Walter. Transitioning from federated avionics architectures to integrated modular avionics. En *2007 IEEE/AIAA 26th Digital Avionics Systems Conference*, 2007.
- Dave Woerner, Les Deutsch, y Chris Salvo. The X2000 Program: An Institutional Approach to Enabling Smaller Spacecraft. 2000.

BIBLIOGRAFÍA

Cong Zhang, I. M. Jaimoukha, y F. R. S. Sevilla. Fault-tolerant observer design with a tolerance measure for systems with sensor failures. En *2016 American Control Conference (ACC)*, páginas 7523–7528, 2016.

Youmin Zhang y Jin Jiang. Bibliographical review on reconfigurable fault-tolerant control systems. *Annual Reviews in Control*, 32(2):229–252, 2008.

Protocolo de comunicación: CANae 0.1 Alpha Version

A.1. Introducción

En este apéndice se explica el diseño de la versión Alpha del protocolo CANae. Este protocolo se basa en el la capa de aplicación del protocolo CAN y CANopen. El nombre CANae proviene de la unión de CAN y CONAE.

A.1.1. Revisiones

Tabla A.1: Revisiones de CANae

Ver.	Who.	When	What
0,1 Alpha	Emmanuel Arias	25/06/2017	Initial Version

A.2. Capa de aplicación

La capa de aplicación de CANae está dividida entre la denominada *CANae Application Layer* y la *CANae High Application Layer* Figura A.1. La funcionalidad que brinda la capa de aplicación está dividida en diferentes elementos de servicios dentro de la capa de aplicación. Un elemento de servicios brinda un determinado servicio.

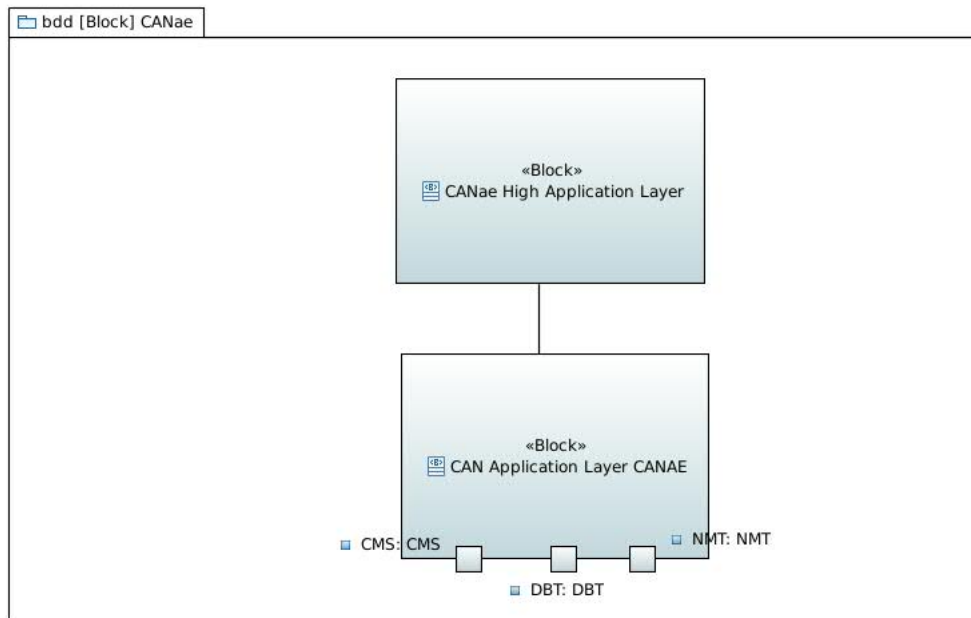


Figura A.1: Estructura de la capa de aplicación de CANae en alto nivel

Una capa de aplicación tiene las siguientes 4 funciones primitivas:

- Request: Una aplicación emite una petición a la capa de aplicación para solicitar un servicio.
- Indication: Emitido por la capa de aplicación con destino una aplicación, para reportar de algún evento interno detectado por la capa de aplicación, o indicar que un servicio fue solicitado.
- Response: Emitido por la aplicación con destino la capa de aplicación para responder a una Indication previa.
- Confirm: Emitido por la capa de aplicación para reportar el resultado de una petición previa.

A.3. Tipos de servicios de la capa de aplicación

Un tipo de servicio define las primitivas que intercambian las aplicaciones de usuario y los diferentes elementos de servicio que se encuentran dentro de la capa de aplicación. Los tipos de servicios posibles en CANae son los siguientes:

- Local service: Solo envuelve un elemento de servicio local.
- Confirmed service: Implica que uno o más elementos de servicio pares. La aplicación de usuario emite un request a su local service element. Este se transmite a sus elementos de servicios pares, la cual es

enviada a la aplicación como una indicación. La otra aplicación emite un response que se transmite a la aplicación originante para confirmar la recepción del request.

- Provide initiated service: Implica solo el elemento de servicio local. El elemento de servicio detecta un evento no solicitado por la aplicación y emite una Indication.

A.4. CANae Application Layer

En CANae se divide la tradicional capa de aplicación del modelo OSI en dos capas. En esta sección se estudia la *CANae Application Layer*. Esta capa está conformada por 3 elementos de servicio:

- CAN based Message Specification (CMS): ofrece un ambiente orientado a objeto para diseñar aplicaciones de usuario. Esta entidad ofrece variables y eventos, y especifica como un módulo puede acceder a las interfaces de CAN.
- Network Managment (NMT): ofrece un ambiente orientado a objetos para permitir que un módulo (el NMT Master) se ocupe de la inicialización y posibles fallas de otros módulos (NMT Slaves).
- Distributor: ofrece el servicio para distribuir dinámicamente el identificador de para los diferentes nodos.

Estos están basados en la capa de aplicación de CAN para la industria (CAN in Automation [CiA] International Users and Manufacturers Group e.V. CAN Application Layer for Industrial Applications). Estos elementos de servicios determinan el 'qué' puede hacer la capa. Estos son representados como interfaces (Figura A.2).

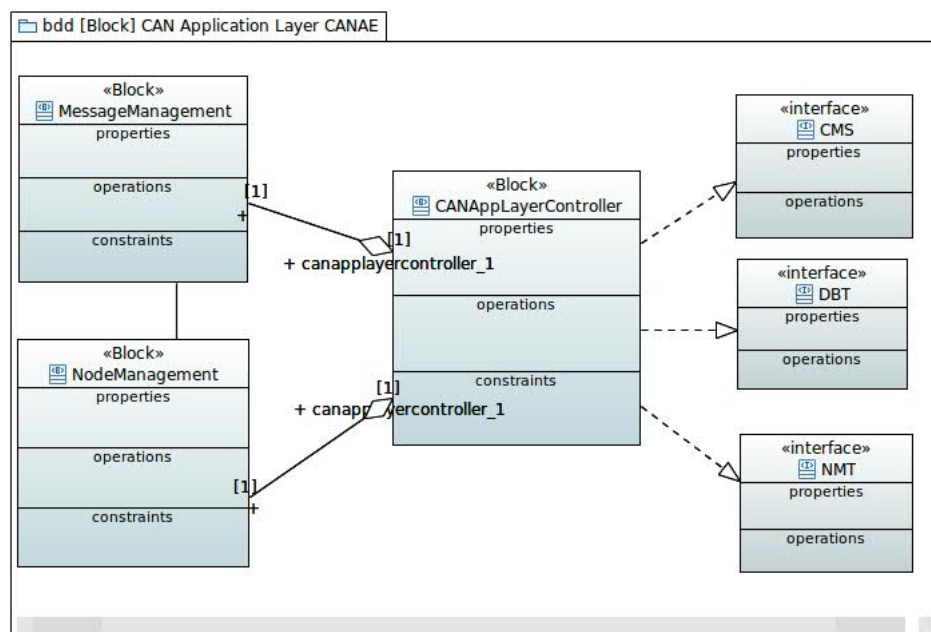


Figura A.2: CANae Application Layer

Dentro de esta capa existen dos entidades importantes:

- Gestor de mensajes: este se encarga de gestionar los mensajes que son enviados y recibidos desde la red. Esta entidad debe ser capaz de determinar si el mensajes contiene datos o eventos. Trabaja en conjunto con el CMS.

- Gestor de nodo: esta entidad se encarga de llevar el control de los nodos existentes en la red. En este nodo se encuentra la tabla de ruteo primario y secundario necesarios para la correcta comunicación.

A.5. CMS

CMS ofrece un ambiente orientado a objeto para diseñar aplicaciones de usuario. CMS ofrece a la aplicación la posibilidad de modelar su comportamiento en la forma de objeto. Este elemento de servicio ofrece variables, eventos que se utilizan para diseñar y especificar como la funcionalidad de un módulo puede acceder a las interfaces CAN.

El servicio asume que no hay fallas provenientes de la capa de enlace de datos y la capa física de la red CAN. Las fallas, si ocurren, son resueltas por el NMSE (Network Management Service Element).

A.5.1. Prioridades de los objetos

La arbitración del protocolo CAN se realiza teniendo en cuenta la prioridad de los identificadores de los frames que se envía. La arbitración es no destructiva, debido a la necesidad de tener un procesamiento de mensajes de tiempo real. La prioridad se determina teniendo en cuenta cada uno de los bits. El indentificador de los objetos y eventos de CMS está compuesto por 8 bits más 3 bits que los diferenciará entre objetos y eventos. Los eventos tiene mayor prioridad.

000	(ID)	Tipo de evento
111	(ID)	Prioridad objeto

La prioridad de los objetos es un UInteger de 1 byte. La prioridad más baja es 0 y la prioridad más alta de 1. Por lo tanto la prioridad va desde 00000000b (mayor prioridad) hasta 11111111b (menor prioridad).

A.5.2. Objeto CMS

Un objeto CMS define una estructura de dato que se debe respetar para poder enviar información (variables) a través de la red. El objeto está compuesto por los siguientes campos:

- Name: string de 6 caracteres (6 bytes)
- User_type: client, server (1 bit)
- Priority: UInteger (1 bytes)
- Datatype: UInteger identificador del tipo de datos (1 byte)
- Data: Indefinido.

A.5.3. Servicios de CMS

Con respecto a las variables de CMS existen una serie de servicios brindados por CMS, que pueden ser tanto *local services* como *remote services*. Su clasificación depende de si se está accediendo a variables almacenadas dentro del nodo, o bien se está intentando acceder (escribir o leer) una variable remota. Un evento está definido de la siguiente manera:

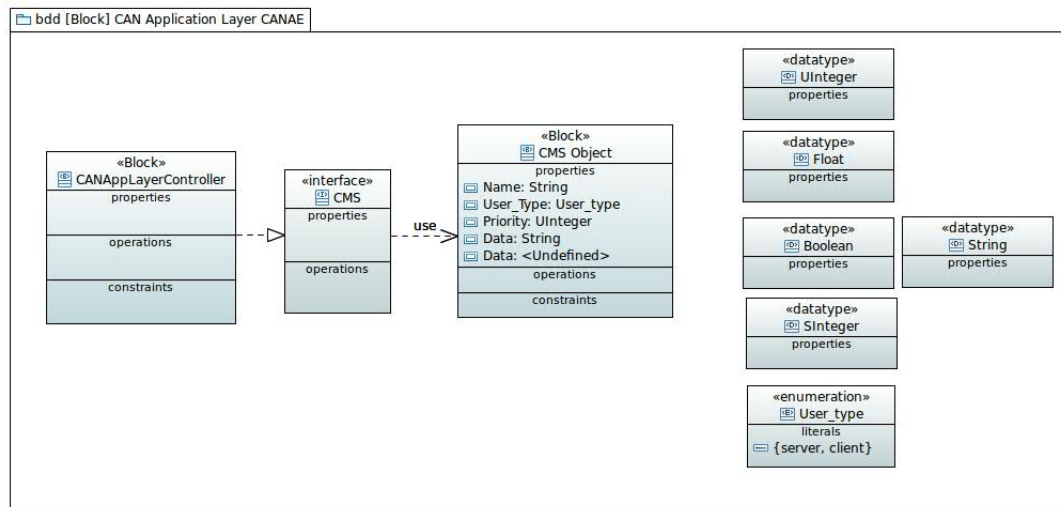


Figura A.3: Definición del Objeto CMS

- Name: string de 6 caracteres (6 bytes)
- User_type: client, server (1 bit)
- Priority: UInteger (1 bytes)
- Event_Type: UInteger identificador del tipo de datos (1 byte)
- Data: Indefinido.

A.5.3.1. Servicios locales

- Boolean define_variable(String name, DataType data_type)
 - **String name:** Nombre de la variable.
 - **DataType data_type:** Tipo de dato de la variable.
- Boolean set_variable(String name, UInteger destination, Data data, UInteger priority)
 - **String name:** Nombre de la variable.
 - **UInteger destination:** Dirección del destinatario. Esta dirección debe ser provista por el NMT.
 - **Data data:** Datos a enviar.
 - **UInteger priority:** Prioridad de la variable.
- Boolean update_variable(String name, data data)
 - **String name:** Nombre de la variable
 - **Data data:** Datos a enviar

A.5.3.2. Servicios remotos

- Boolean write_variable(String name, Data data)
 - **String name:** Nombre de la variable a escribir
 - **Data data:** Dato a escribir
- Data read_variable(String name)
 - **String name:** Nombre de la variable a leer

A.5.4. Eventos

Los eventos se utilizan para modelar comportamientos asíncronos, tales como temperatura excedida en un tiempo determinado. La ocurrencia de los eventos es detectada por el nodo, que en ese momento actuará como server y se notifica a todos los clientes. Los servicios al igual que las variables se dividen los eventos brindados por servicio local o servicio remoto. Los servicios se describen a continuación.

A.5.4.1. Servicios locales

- define_event(String name, UInteger priority, UInteger event_type)
 - **String name:** Nombre del evento
 - **UInteger priority:** Prioridad del evento
 - **UInteger event_type:** Tipo de evento según tabla A.2

A.5.4.2. Setvicios remotos

- Event notify_event()
- Boolean store_event(String name)
 - **String name:** Nombre de evento a almacenar
- Event read_event(String name)
 - **String name:** Nombre de la variable a leer

A.5.4.3. Prioridad de eventos

La prioridad de los eventos se define como una tabla que se define al momento de implementar el protocolo. Al desarrollar la tabla de prioridades se debe cuidar la relevancia de los eventos. La prioridad más baja es 0. El tamaño de la prioridad del evento es de 1 byte. Por lo tanto la prioridad va desde 00000000b (mayor prioridad) hasta 11111111b (menor prioridad).

Esta tabla de prioridad de eventos, debe ir acompañada de un diccionario de eventos, donde se almacena los datos relevantes al evento. En esta se hace una descripción del evento. Sirve de guía para la implementación y desarrollo de aplicaciones que identifican eventos.

El diccionario debería incluir los siguientes campos:

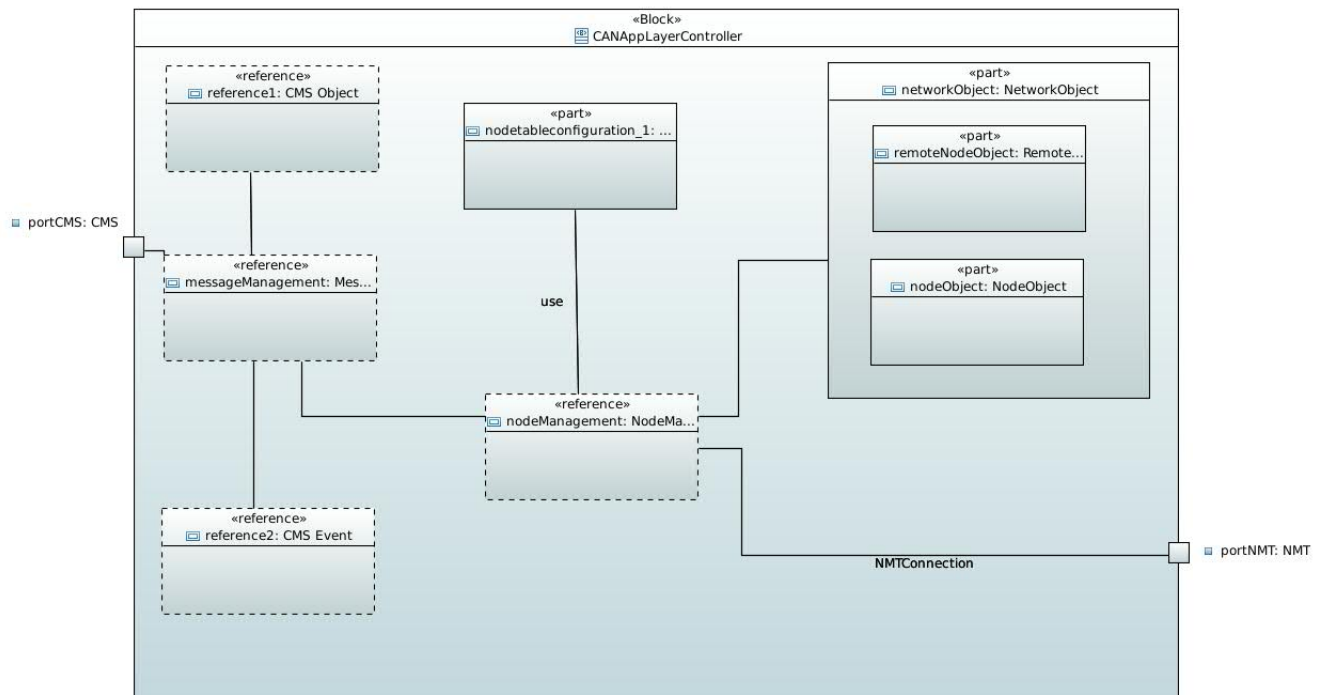
Tabla A.2: Prioridad eventos

00000000	Nombre evento	Mayor prioridad
00000001	Nombre evento	
00000010	Nombre evento	
...
11111111	Nombre evento	Menor prioridad

- Nombre del evento
- Descripción del evento
- Prioridad
- Umbral límite inferior (En el caso de que existiera)
- Umbral límite superior (En el caso de que existiera)

A.6. NMT

Esta es una entidad de la capa de aplicación que permite el correcto funcionamiento de la red. Esta entidad tiene los servicios necesarios para que cada nodo tenga un conocimiento de todos los nodos conectados en la red. Esto permite el armado de una tabla de nodos, que es mantenida en todo los nodos. Esta entidad exige la existencia de un nodo monitor, que es el encargado de monitorear y configurar la red. Luego, una vez en funcionamiento este nodo monitor no será vital para la red. En el NMT se deben implementar los algoritmos de ruteo. El entorno NMT se puede observar en la Figura A.5. Se puede observar también el diagrama de bloques internos en la Figura A.4

**Figura A.4:** Definición de la entidad NMT (bloques internos)

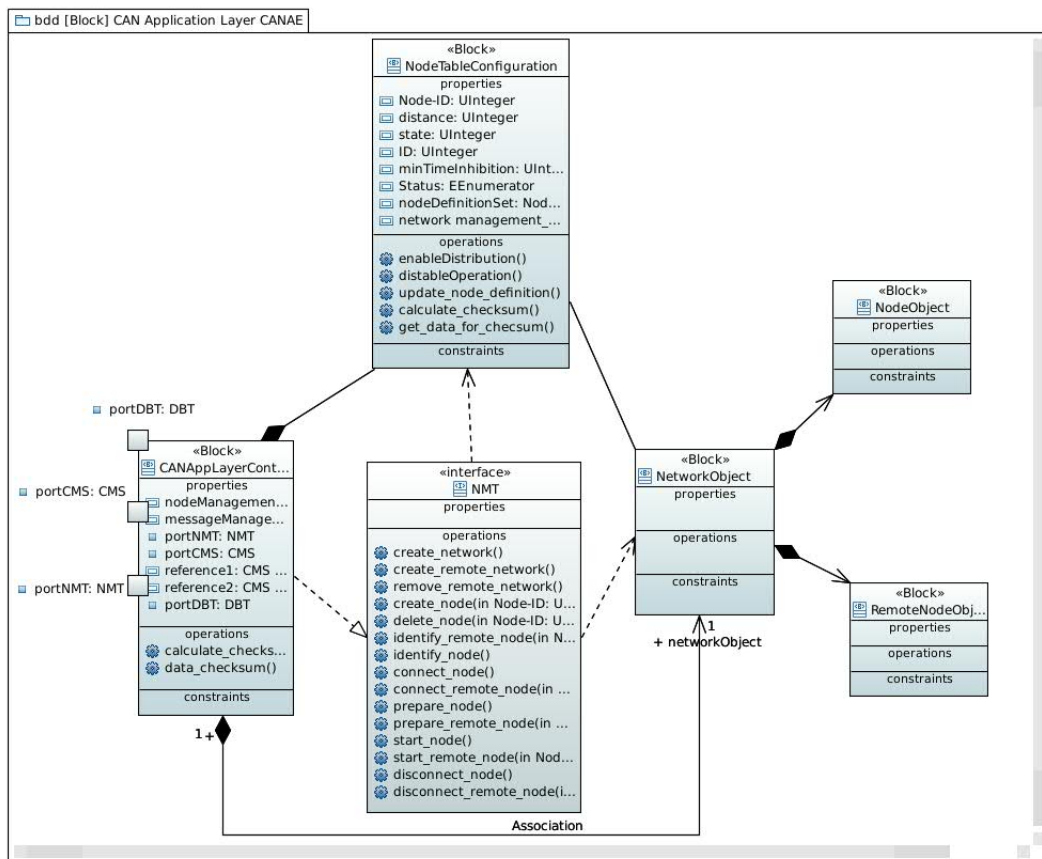


Figura A.5: Definición de la entidad NMT

Por medio de este servicio es posible conocer el estado de la red CAN. Los mensajes de NMT CAN son enviados con máxima prioridad.

A.6.1. Objetos NMT

El Network Management utiliza tres objetos diferentes para modelar una red CAN.

- **Network Object**: Representa todos los modulos conectados en la red CAN. El objeto de red existe en todos los nodos.
- **Remote Node Object**: Cada nodo conectado a la red CAN tiene representado todos los demás nodos.
- **Node Object**: Cada nodo que es gestionado por el servicio NMT está representado por un node object. Este objeto se encuentra modelado en cada uno de los nodos.

Cada nodo y su objeto (remoto y local) es unívocamente identificado en la red por su NMT Address. Esta dirección no se puede cambiar. Esto significa que por cada nodo existe un objeto nodo y un objeto nodo remoto con el mismo NMT Address, replicada en todos los nodos de la red.

A.6.2. Servicios NMT

La entidad NMT ofrece los siguientes servicios:

A.6.2.1. Module Control Service

EL NMT monitor inicializa los nodos NMT que forman parte de la red CAN distribuida, a través de este servicio se asegura que todos los nodos se encuentran configurados y funcionando correctamente.

A.6.2.2. Error Control Services

Através de este servicio, el NMT detecta fallas en la red CAN, ya sea fallas producidas en la capa de enlace de datos (fallas locales) y/o fallas producidos en otros nodos.

A.6.2.3. Configuration Control Services

Através de este servicio se realiza la configuración de la red.

A.6.2.4. Servicios a implementar

La funciones típicas que se deben implementar en el NMT son los siguientes:

- Boolean create_network()
- Boolean create_remote_network()
- Boolean remove_remote_network()
- Boolean create_node(UInteger Node-ID, UInteger address, String description)
 - **UInteger Node-ID:** ID del nodo a crear.
 - **UInteger address:** Dirección del nodo creado. UInteger entre [0,255]
 - **String description:** Descripción del nodo.
- Boolean delete_node(UInteger Node-ID)
 - **UInteger Node-ID:** Dirección del nodo a eliminar. UInteger entre [0,255]
- UInteger identify_remote_node(UInteger Node-ID)
 - **UInteger Node-ID:** Dirección del nodo a identificar
- UInteger identify_node()
- Boolean connect_node()
- Boolean connect_remote_node(UInteger Node-ID)
 - **UInteger Node-ID:** ID del nodo remoto a conectar.
- Boolean prepare_node()
- Boolean prepare_remote_node(UInteger Node-ID)
 - **UInteger Node-ID:** ID del nodo a pasar a modo preparado

- Boolean `start_node()`
- Boolean `start_remote_node(UInteger Node-ID)`
 - **UInteger Node-ID:** ID del nodo que comenzará a participar en la red.
- Boolean `disconnect_node()`
- Boolean `disconnect_remote_node(UInteger Node-ID)`
 - **UInteger Node-ID:** ID del nodo a desconectar.

A.6.3. Protocolos NMT

A.6.3.1. Crear red CANae

Para conectar correctamente la red CANae se debe respetar el protocolo NMT para crear la red. En primer lugar el nodo monitor debe enviar el mensaje `create_remote_network()` para avisar a todos los nodos que deben crear sus propias instancias de *NetworkObject* a través de `create_network()`. Automáticamente, los nodos deben crear una propia instancia del nodo mediante `create_node()`. Luego el nodo monitor envía el mensaje de `prepare_remote_node()` para preparar todos los nodos a conectarse a la red CANae. Luego el nodo monitor envía la señal de `connect_remote_node(UInteger Node-ID)` con la dirección de todos los nodos que figuran en su *NodeTableConfiguration* preprogramada. Luego cada nodo, correctamente preparado, se conecta a la red a través de `connect_node()`.

Este protocolo puede ser observado en la Figura A.6

A.6.3.2. Crear Red

Para lograr la correcta conexión de los nodos a la red, estos deben configurarse. Cuando se alimenta la red, los nodos deben iniciar en estado de *escucha* esperando la orden del nodo monitor de crear la red mediante `create_remote_network()`. Esta se envía en broadcast a todos los nodos conectados. Cuando el controlador de la capa de aplicación recibe el mensaje a través del servicio NMT, este manda un mensaje al gesto del nodo interno, el cual crea la *tabla de configuración del nodo* y actualiza con la información necesaria.

A.6.3.3. Crear Objeto Red, Objeto Nodo y Objeto Nodo Remoto

Antes de comenzar a funcionar la red, cada nodo debe crear un objeto de red exigidos por el protocolo NMT. Una vez que el proceso Crear Red termina, el Gestor de la capa de Aplicación envía un mensaje `create_node()` al Gestor del Nodo. El Gestor de Nodo crea un *Network Object*. Este a su vez crea un *Node Object* y un *Remote Node Object*. Estos objetos pertenecen a la entidad NMT (Figura A.8).

A.6.3.4. Preparar nodos para conexión

Para lograr la conexión de los nodos a la red CAN, es necesario que el nodo monitor envía la señal `_prepare_remote_node(UInteger Node-ID)` a cada uno de los nodos conectados a la red siguiendo la propia tabla de configuración de nodos preprogramada. Cada nodo recibe la señal procesa aquella que le pertenece, y comienza a preparar el nodo para lograr la correcta conexión. Para ello debe actualizar su propia tabla de configuración de nodos, como así también actualizar su *Objeto Red* (Figura A.9).

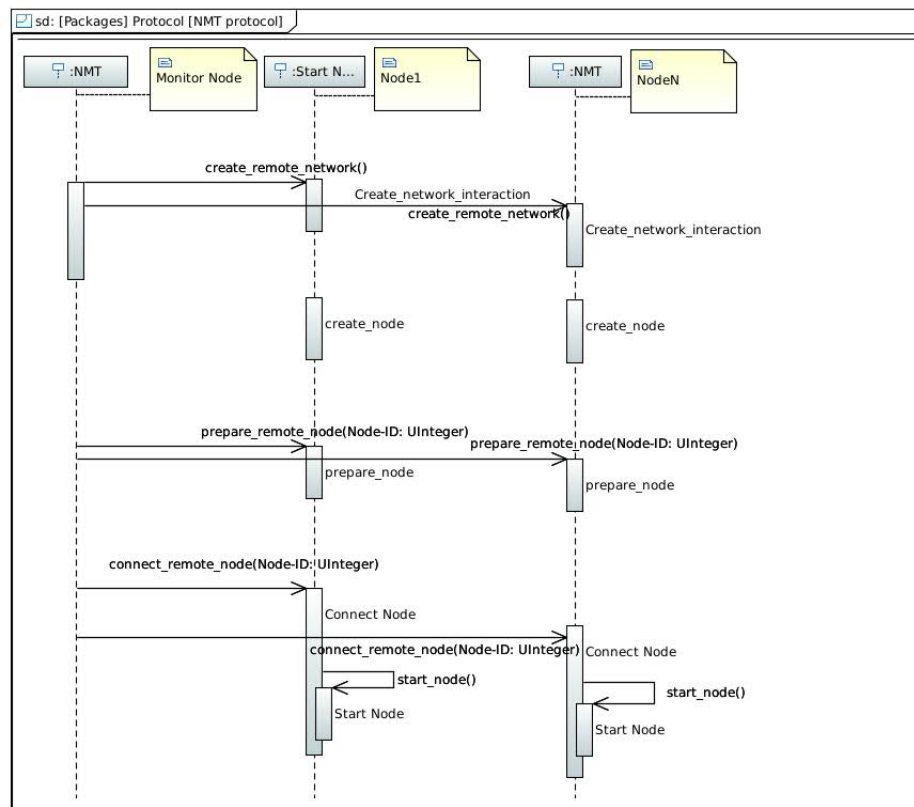


Figura A.6: Protocolo NMT

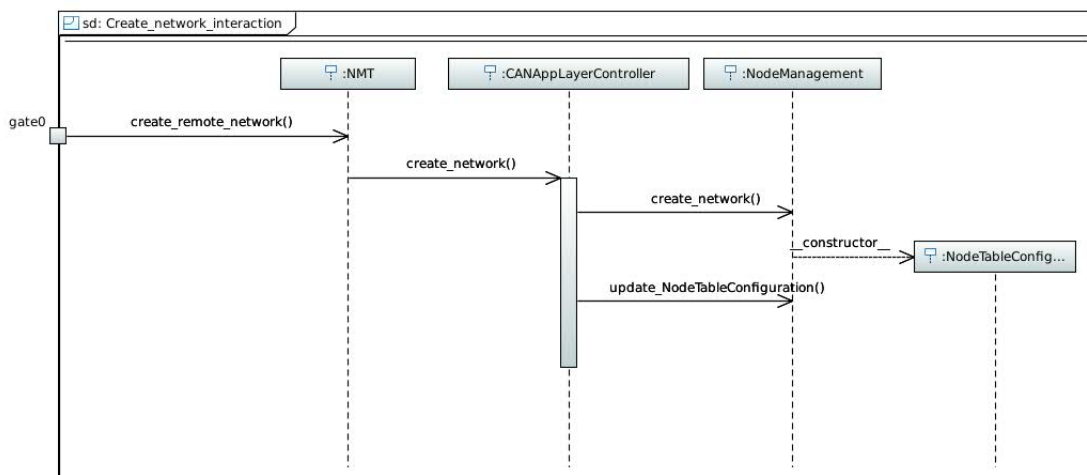


Figura A.7: Crear Red

A.6.3.5. Conectar nodos en la red

Para conectar los nodos a la red se tiene que modificar la información de la *Node Table Configuration* y en el *Objeto de Red*. Una vez realizado esto se puede comenzar con la comunicación de los nodos (Figura A.10).

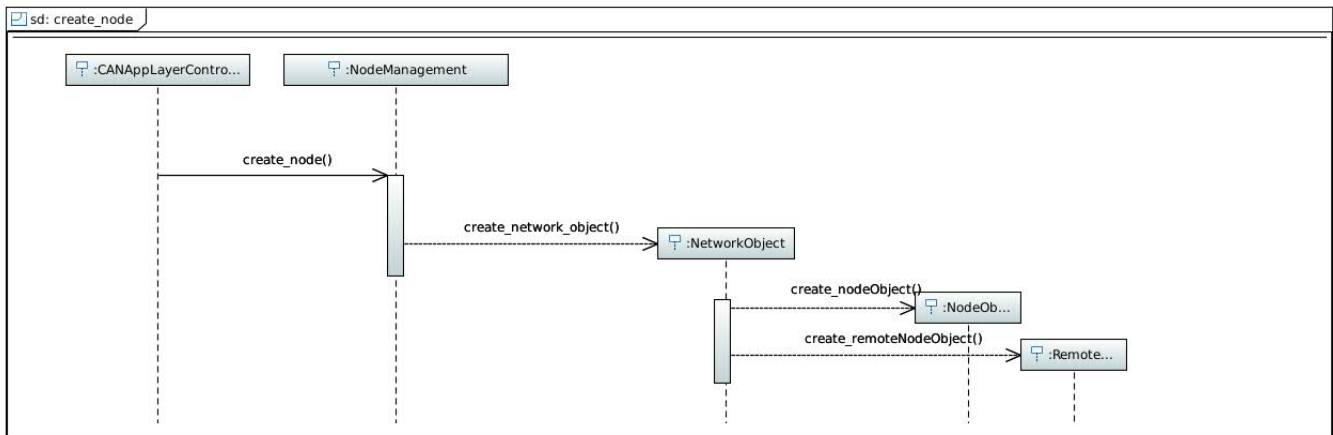


Figura A.8: Crear Objeto Red, Objeto Nodo y Objeto Nodo Remoto

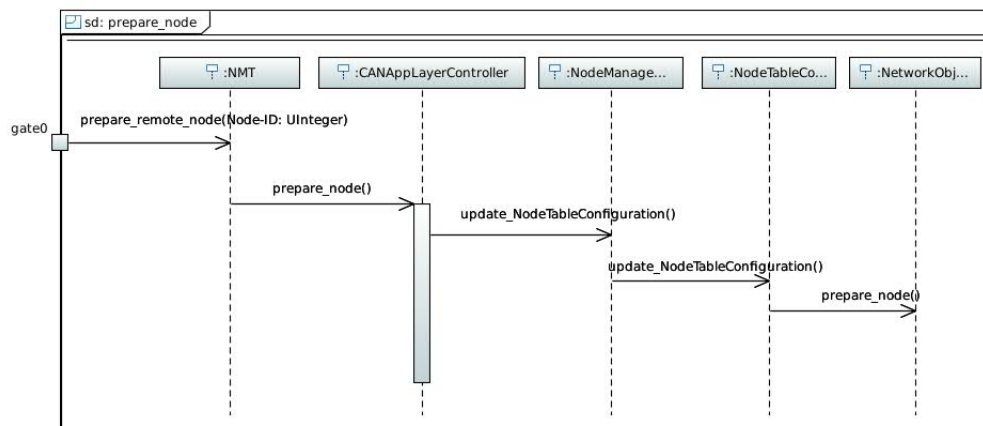


Figura A.9: Preparar nodo para la conexión a la red CAN

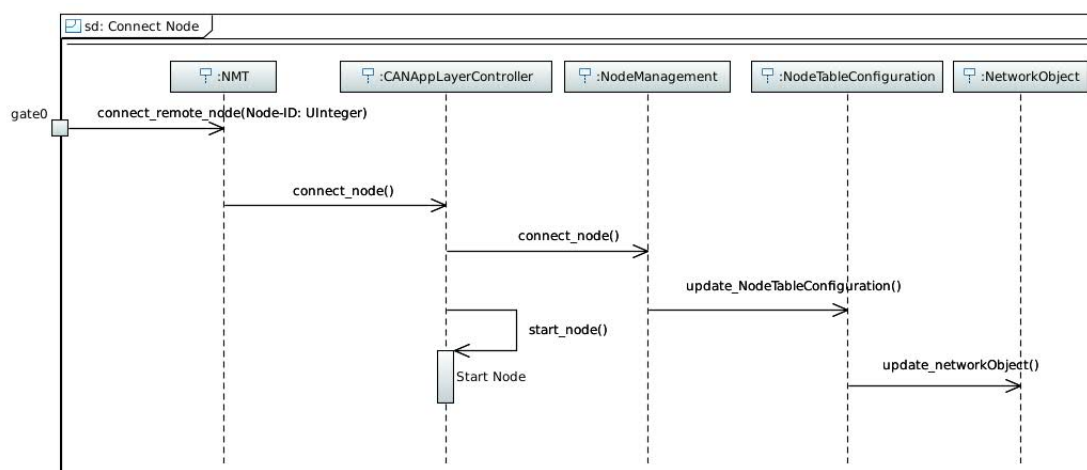


Figura A.10: Conectar el nodo a la red CAN

A.6.3.6. Comenzar la comunicación

Para que el nodo pueda comenzar a comunicar debe encontrarse en un estado *Start*. Esto representa una actualización de la *Node Table Configuration* y del *Objeto de Red* (Figura A.11).

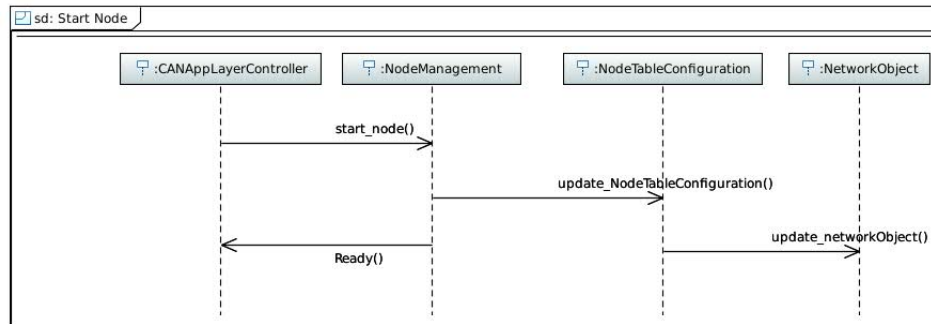


Figura A.11: Comenzar la comunicación de nodos

A.7. DBT

La principal complicación del desarrollo de una red distribuída basada en el BUS CAN, es que puedan comunicarse correctamente entre sí, cómo se deben asignar los identificadores (Node-ID) correctamente a cada uno de los nodos y cómo se asignan los tiempos de inhibición. Los Node-ID y los tiempos de inhibición se deben distribuir entre los nodos de tal manera de asegurar que:

- Se prevean los conflictos entre nodos. Por ejemplo diferentes funciones que usan el mismo identificador.
- Se prevean la falta de coincidencia. Por ejemplo que existan diferentes indentificadores para el mismo nodo.
- ofrecer el control integrado para un comportamiento dinámico del sistema.

El protocolo CAN (CAN Application Layer for Industrial Applications CiA/DS204-1) dispone la existencia de 3 métodos para distribuir identificadores y tiempo de inhibición a un módulo.

- **Distribución estándar:** en este método los identificadores y tiempos de inhibición son estandarizados por el módulo proveedor (nodo monitor. Una distribución estándar requiere la estandarización de todas las funciones y su correspondiente identificador.
- **Distribución estática:** En este método los identificadores y los tiempos de inhibición son fijos. Todos los identificadores y los tiempos de inhibición son fijados en el momento del desarrollo.
- **Distribución dinámica:** En este método los identificadores y los tiempos de inhibición son distribuidos vía red CAN a través del servicio estándar y un protocolo definido.

En esta instancia de trabajo, para el desarrollo del protocolo CANae, se utilizó el *método de distribución estática*. Esto exige que los nodos ya cuenten con un identificador y tiempos de inhibición por defecto. Esto se asigna en el momento del desarrollo del diseño e ingeniería de la arquitectura de aviónica. Esto reduce considerablemente la complejidad de la entidad DBT. Si bien originalmente, el DBT es un elemento de servicio

de la capa de aplicación CAN que ofrece una distribución dinámica de identificadores y tiempos de inhibición, en CANae es utilizado para asegurar la correcta comunicación y consistencia de los nodos y la red en su conjunto.

La arquitectura del DBT se puede observar en la Figura A.12

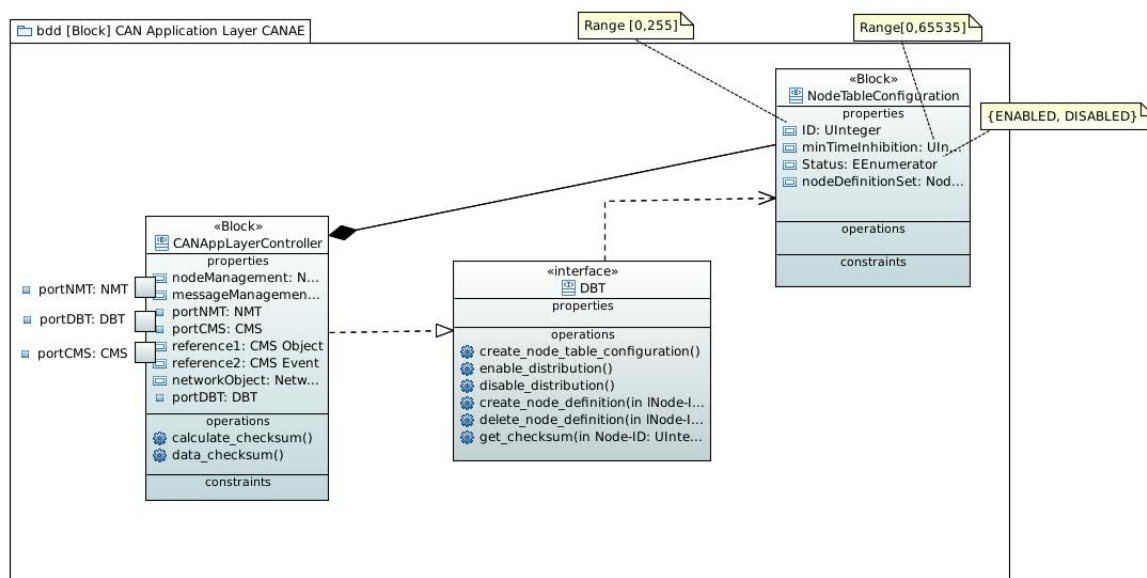


Figura A.12: Arquitectura de la entidad DBT

A.7.1. Objetos y servicios de DBT

El DBT de CANae utiliza 2 objetos para modelar su funcionamiento:

- **Node-ID Database (NodeTableConfiguration)**: Esta tabla contiene la definición de todos los nodos conectados a la red. Esta tabla está compuesta por Node-ID Definitions
- **Node-ID Definition**: define todos los atributos de un Node-ID. Contiene una definición de usuario creado por el usuario.

El DBT de CANae ofrece las siguientes categorías de servicio:

- **Distribution Control Services**: esta es tarea del nodo monitor de enviar el identificador y los tiempos de inhibición a todos los nodos conectados a la red. Los datos que son enviados por el nodo monitor son corroborados por los nodos, comprobando que el ID enviado coincida con su propio Node-ID.
- **Consistency Control Services**: através de este servicio cada nodo puede detectar inconsistencia en el NodeTableConfiguration e inconsistencia entre nodos.

A.7.2. Descripción de los servicios DBT

Los servicios se describen en forma de tabla que contiene los parámetros de cada función que se define para ese servicio. Los parámetros determinan el tipo de servicio. Todos los servicios asumen que no ocurrieron ningún tipo de error en la capa de física ni en la capa de enlace de datos.

A.7.3. Objetos DBT

Los objetos que utiliza la entidad DBT para modelar el servicio son los siguientes:

- **Node Table Configuration (Node-ID Database)**

- **Atributos**

- Estado: Uno de los valores ENABLED, DISABLED. Este atributo indica si el DBT Monitor es capaz de distribuir los Node-ID y tiempos de inhibición. En esta instancia de trabajo el valor del estado en el DBT del Nodo Monitor será DSIABLED
 - Node Definition Set set de todas las definciones.

- **Node-ID Defintion**: Las definiciones contiene los siguientes datos:

- ID: Valor en el rango de [0,255]
 - Mínimo tiempo de inhibición: en el rango [0,65535] indicando el valor mínimo en unidades de $100\mu\text{sec}$ para el tiempo de inhibición.

A.7.4. Servicios DBT

A.7.4.1. Distribution Control Services

- `create_node_table_configuration()`
- `enable_distribution()`
- `disable_distribution()`
- `create_node_definition(UInteger lNode-ID, UInteger hNode-ID, UInteger minInhibitTime)`
 - **UInteger lNode-ID**: Límite inferior del rango de IDs.
 - **UInteger hNode-ID**: Límite superior del rango de IDs.
 - **UInteger minInhibitTime**: Tiempo mínimo de inhibición de cada nodo.
- `delete_node_definition(UInteger lNode-ID, UInteger hNode-ID)`

A.7.4.2. Consistency Control Service

- `get_checksum(UInteger Node-ID, Float checksum)`
 - **UInteger Node-ID**: ID del nodo que desea controlar la consistencia.
 - **UInteger checksum**: checksum de la DBT Definition.

A.7.5. Protocolos DBT

A.7.5.1. Crear NodeTableConfiguration

El La tabla de configuración del nodo juega un papel similar al que lo haría el Node-ID DATABASE del estándar CAN. La principal diferencia, es que esta tabla se encuentra presente en todos los nodo y no en el

“Master”, ya que este protocolo está pensado para ser utilizado en redes distribuidas, donde no existe un nodo central o master.

El nodo monitor envía la señal a los nodos de crear la tabla de configuración del nodo a través de *create_node_table_configuration()*. Este es recibido a través de la interfaz DBT del nodo, y crea el objeto *NodeTableConfiguration*. Esto se observa en la Figura A.13.

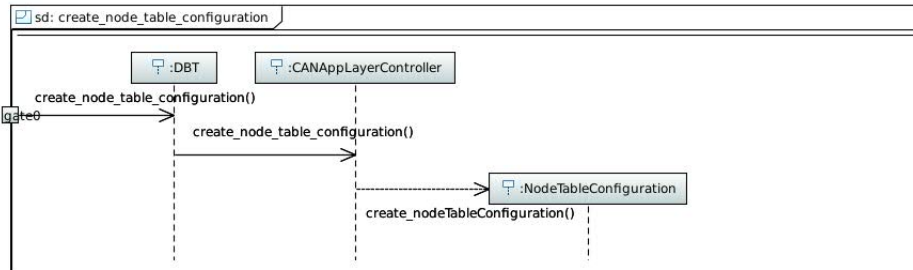


Figura A.13: Protocolo create_node_table_configuration

A.7.5.2. Habilitar distribución

La distribución de Node-ID es utilizada cuando se utiliza el método de asignación dinámica. En esta versión del protocolo la distribución del ID se encuentra desactivada. En la Figura A.14 se puede observar el protocolo.

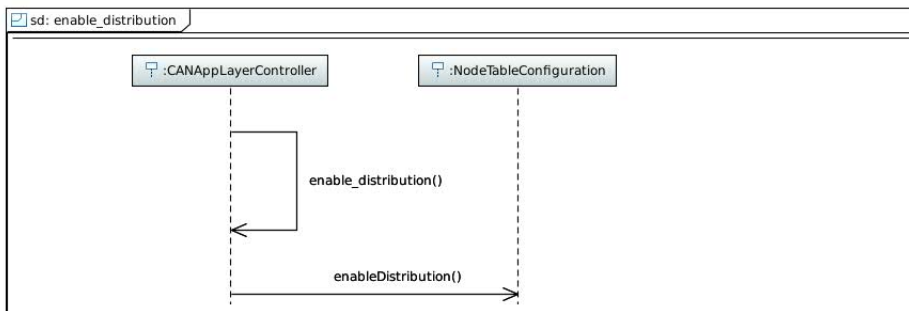


Figura A.14: Protocolo enable_distribution

A.7.5.3. Deshabilitar distribución

Esta opción se encuentra por defecto en esta versión del protocolo CANae. El ID se distribuye estáticamente. Son preestablecidos durante el desarrollo. En la Figura A.15

A.7.5.4. Crear definición del nodo

El nodo monitor envía la señal para que sea creen las definiciones de nodo. Para ello es necesario que el nodo monitor envíe los rangos permitidos para la definición de los nodos. Esto se define a través de las variables *lNode-ID* y *hNode-ID*. En esta versión del protocolo estas variables no son utilizadas ya que los Node-ID ya se encuentran definidas estáticamente. También el nodo monitor envía el tiempo mínimo de inhibición explicado anteriormente. El nodo monitor hace este proceso con todos los nodos de la red a través del mensaje *create_node_definition*. Este se lo puede observar en la Figura A.16.

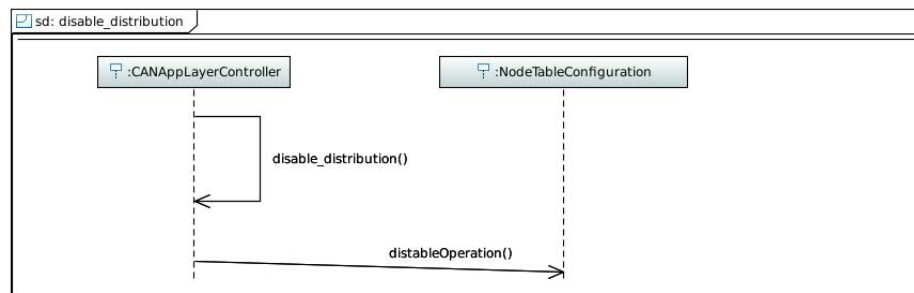


Figura A.15: Protocolo disable_distribution

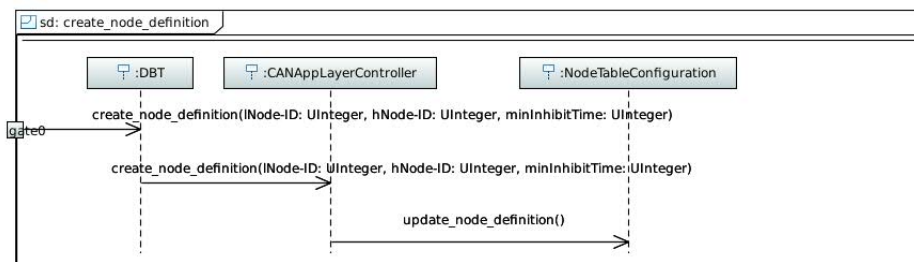


Figura A.16: Protocolo create_node_definition

A.7.5.5. Eliminar la definición del nodo

Esta función es utilizada para eliminar la definición del nodo. Se lleva a cabo para eliminar un nodo de la red. Se hace a través de la función *delete_node_definition()*. Esta se la puede observar en la Figura A.17

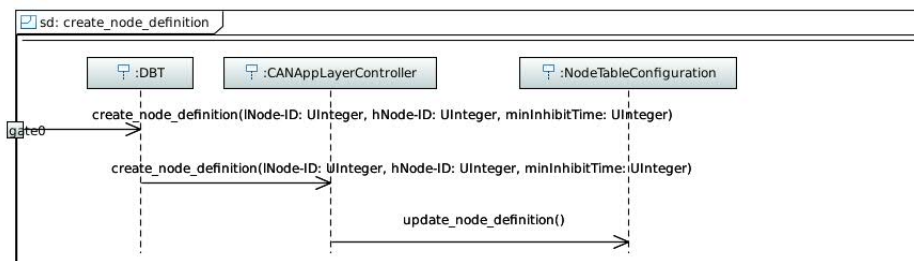


Figura A.17: Protocolo delete_node_definition

A.7.5.6. Checksum

Para poder asegurar la integridad de los nodos conectados a la red, se puede utilizar este mecanismo para comprobar de que la información, referente a la definición de los nodos, se encuentra correctamente en todos los nodos. Para ello, cada nodo lleva a cabo un checksum de su información y la envía por la red. Como esta está basada en el Bus-CAN, el mensaje llega correctamente a todos los nodos. Los nodos realizan sus comprobaciones y responden. Si la información es correcta deben responder con un mensaje que contegan bits NO dominantes. En el caso de que se produzca alguna diferencia en los checksum deberán responder con bits dominantes. Cuando esto se produzca deberán llevarse a cabo medidas de recuperación o reconfiguración de la red. Este protocolo se lo puede observar en la Figura A.18

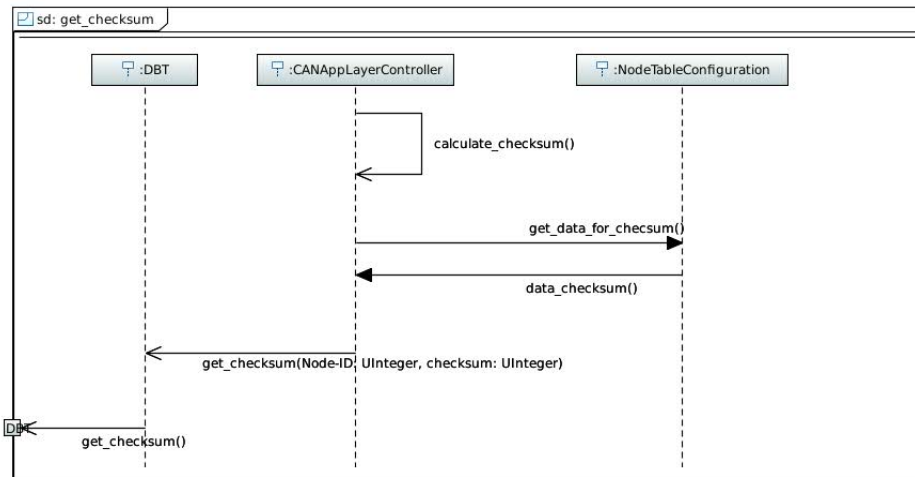


Figura A.18: Protocolo get_checksum

A.8. Gestor de mensajes

El **Message Management** se encarga de gestionar los mensajes que son enviados y recibidos desde la red CAN. Esta entidad debe ser capaz de determinar si el mensaje (enviado o recibido) contiene datos o eventos. El gestor de mensajes tiene cuatro entidades:

- *Receiver Manager*: es el encargado de recibir los mensajes de la capa inferior a través de la interfaz NMT. Esta entidad recibe y despaquetiza los mensajes.
- *Prepared Message*: es el encargado de recibir los mensajes desde las capas superiores para ser enviado a la red.
- *Sorter Message*: es el encargado de clasificar los mensajes, ya sea aquellos que envía a la red CAN como los que recibe de esta. Su principal función es verificar el tipo de mensaje que está recibiendo o enviando (datos o eventos). Los eventos tienen mayor prioridad que los datos y deben atenderse con urgencia.
- *Buffer Message*: este sirve de buffer para almacenar mensajes tanto que se envían como lo que se reciben. Este buffer es utilizado por el *Sorter Message*.

En la Figura A.19 se muestra la composición del gestor de mensajes. Por otra parte en la Figura A.20 muestra el diagrama interno del gestor de mensajes.

A.8.1. Receiver Manager

Esta entidad perteneciente al *Message Management* se encarga de recibir los mensajes de la capa inmediatamente inferior a través de la interfaz NMT (y sus protocolos). Esta entidad recibe y despaquetiza los paquetes. Se supone que los mensajes llegan sin ningún tipo de error.

Las principales funciones de esta entidad son las siguientes:

- `receive_message(Message message)`: esta es la operación que se debe llamar para recibir el mensaje proveniente de la capa inferior.
 - **Data message**: Es el mensaje que recibe de las capas inferiores de la red CAN.

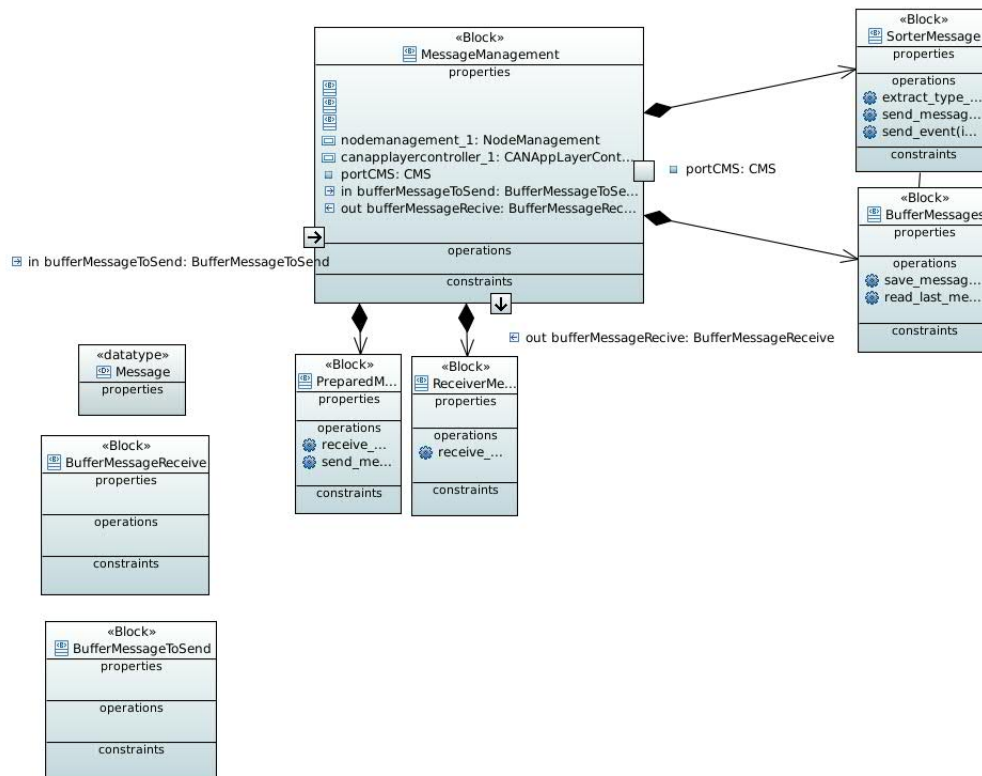


Figura A.19: Arquitectura de *Message Management*

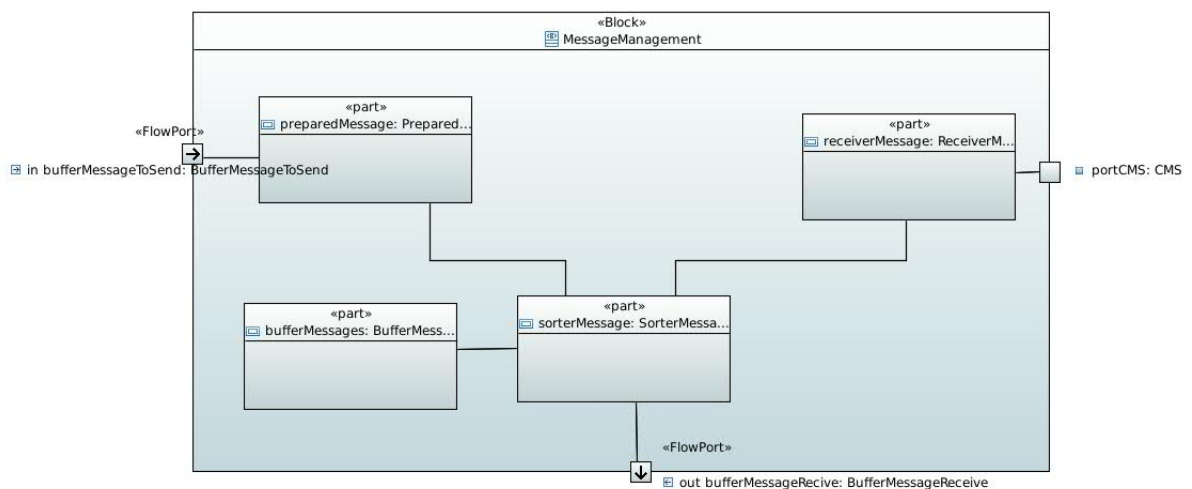


Figura A.20: Diagrama interno del *Message Management*

A.8.2. Prepared Message

Esta entidad es la que recibe los mensajes desde las capas superiores y lo envía al *Sorter Message*.

- **receive_message(Message message):** esta es la operación que se debe llamar para recibir el mensaje proveniente de la capa superior.
 - **Message message:** Es el mensaje que recibe de las capas inferiores de la red CAN.

- `send_message(Message message)`: esta es la operación que se debe llamar para enviar el mensaje al *Sorter Message*.
 - **Message message**: Es el mensaje que se envía al *Sorter Message*

A.8.3. Sorter Message

Esta entidad es el encargado de clasificar los mensajes que provienen tanto del *Receiver Message* y *Prepared Message*. Este clasifica los mensajes recibidos o por enviar, de modo tal de conocer aquellos que sean datos o eventos. Debe tenerse en cuenta que los eventos tiene mayor prioridad que los datos, debido a que (en su mayoría) informan de algún problema a la red.

Las principales funciones con las que cuenta esta entidad son las siguiente:

- `Boolean extract_type_message(Message message)`: esta operación permite extraer el tipo de información que contiene el mensaje: datos o evento.
 - **Message message**: es el mensaje que recibe de las capas inferiores de la red CAN.
 - **Return Boolean**: retorna 1 si se trata de un evento, 0 si se trata de un dato.
- `Boolean send_message(Message message)`: esta función envía los mensajes a las capas inferiores a través del protocolo NMT.
 - **Message message**: es el mensaje formateado listo para enviar.
 - **Return Boolean**: retorna 1 si la operación se llevó correctamente.
- `Boolean send_event(Message message)`: esta función envía el evento en forma de mensajes formateado.
 - **Message message**: es el mensaje formateado listo para enviar.
 - **Return Boolean**: retorna 1 si la operación se llevó correctamente.

A.8.4. Buffer Messages

Esta es la entidad que permite almacenar los mensajes que son recibidos y enviados. Esta entidad debe controlar la integridad de los mensajes que se encuentran almacenados. Además se encarga de escribir y leer los mensajes almacenados en el buffer. El buffer es FIFO¹.

Las principales funciones de esta entidad son las siguientes:

- `save_message(Message message)`: esta función almacena el mensaje en el buffer de mensajes
 - **Message message**: es el mensaje formateado listo para enviar.
- `Message read_last_message()`: esta función lee el último mensaje.
 - **Return Message**: retorna el mensaje extraído del buffer.

¹First In First Out

A.8.5. Buffer Message Receive y Buffer Message To Send

La comunicación entre el Gestor de Mensajes y las capas, tanto superior como inferior, se realizan por medio de buffers. Se pueden dar las siguientes situaciones:

- Llegada de un mensaje de la capa superior para ser enviada a través de la red. En este caso la aplicación de usuario necesita enviar un mensaje, para ello debe colocar el mensaje en el *Buffer Message To Send*. El *Gestor de Mensaje* debe detectar (por medio de algún mecanismo) que se ha escrito en el buffer, para tomar el mensaje.
- Enviar un mensaje a la red. Cuando el gestor necesita enviar un mensaje lo tiene que almacenar en el *Buffer Message To Send*. Esto “alertaría” por medio de algún mecanismo a las capas inferiores, de que existe un nuevo mensaje a enviar

A.9. Gestor de Nodo

El *Node Management* es una entidad que se encuentra en el NMT, y es la encargada de llevar a cabo el control de los nodos existentes en la red. En esta entidad se encuentra almacenada la tabla de ruteo primario y secundario, utilizado para lograr la correcta comunicación y configuración de la red.

El objetivo del gestor del nodo, es conocer todos los nodos que se encuentran conectados en la red, y la distancia que existe entre el nodo actual y los demás.

Por medio de las tablas de ruteo (primaria o secundaria en caso de fallas) el nodo conoce a qué nodos puede enviar mensajes y a cuáles no. Debe aclararse que nada impide enviar un mensaje a un nodo “desconectado”, pero esta desición es ineficiente y altamente peligrosa, ya que se puede estar enviando mensajes críticos a un nodo “caído”.

El gestor de nodo de cada dispositivo conectado a al red, es el encargado de llevar a cabo el protocolo *heartbeat*, el cual le indica a todos los nodos conectados a la red que este está “vivo”

En la Figura A.21 se puede observar la arquitectura del Gestor de Nodos.

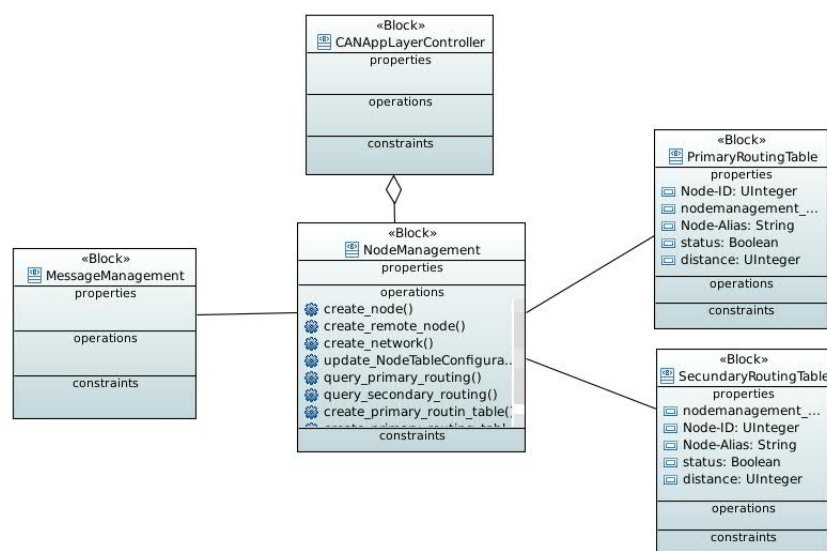


Figura A.21: Arquitectura del *Node Management*

A.9.1. Tabla primaria y secundaria de ruteo

Los principales objetos del *Node Management* son las tablas de ruteo. Estas se las puede representar como una lista, en la cual se incluye información de importancia para mantener actualizada la estructura de la red. Los nodos se basarán sobre esta lista para tener conocimiento de los nodos conectados a la red y con esto, llevar a cabo una correcta distribución de las tareas.

Los algoritmos de ruteos se deben basar en estas tablas, ya que contienen toda la información necesarias para cumplir con sus objetivos.

A continuación se detalla la tabla primaria y tabla secundaria de ruteo.

A.9.1.1. Tabla primaria

La Tabla Primaria de Ruteo es la tabla principal estática utilizada por el *Node Management* para la gestión de los nodos de la red. Esta tabla mantiene la a los n-ísimos nodos conectados a la red, desde el punto de vista del nodo “actual”. Es decir, para cada nodo existe una y sola una tabla, y esa tabla es única en la red, ya que es creada desde el punto de vista del nodo “host”.

La estructura de la lista es simple. Cada renglón de la lista es un nodo, siendo el primer renglón obligatoriamente el nodo de la tabla, llamado nodo “host”. Los siguientes renglones de la lista representan los demás nodos conectados a la red. Por convención, los nodos se listan comenzando por los vecinos de derecha a izquierda. Luego se continúa la lista tratando de seguir la convención de derecha a izquierda.

La tabla tiene la siguiente estructura:

- **UIntegerNode-ID:** Es el ID único para cada nodo.
- **String Node-Alias:** Es el alias para cada nodo. Este alias puede ser utilizado para comodidad por parte del usuario para referenciar los nodos.
- **UInteger Address:** Es la dirección del nodo.
- **Boolean Status:** Es el estado del nodo. Pude ser { CONNECTED, DISCONNECTED }.
- **UInteger Distance:** Es la distancia (cantidad de nodos) que hay entre el nodo “host” y los demás nodos de la red.

A.9.1.2. Tabla secundaria

La Tabla Secundaria de Ruteo es una tabla estática similar que la Tabla Primara de Ruteo, con la diferencia sustancial que es utilizada cuand hay fallas en ruteo. Esta tabla se genera cuando se producen fallas en uno o más nodos, y se necesita conocer la estructura de la red, para luego tomar desiciones sobre la configuración de la misma.

La estructura de la Tabla Secundaria de Ruteo es similar a la Tabla Primara de Ruteo. La diferencia sutancial es que no se encuentran todos los nodos de la red, sino aquellos que se encuentran “funcionales”.

La estructura de la tabla se detalla a continuación:

- **UIntegerNode-ID:** Es el ID único para cada nodo.

- **String Node-Alias:** Es el alias para cada nodo. Este alias puede ser utilizado para comodidad por parte del usuario para referenciar los nodos.
- **UInteger Address:** Es la dirección del nodo.
- **Boolean Status:** Es el estado del nodo. Puede ser { CONNECTED, DISCONNECTED }.
- **UInteger Distance:** Es la distancia (cantidad de nodos) que hay entre el nodo “host” y los demás nodos de la red.

A.9.2. Servicios

Los servicios brindados por el *Node Management* son los siguientes:

- **Boolean create_node(UInteger Node-ID, UInteger address, String description):** esta función crea el objeto nodo. Este servicio ya fue descrito en A.6.2.4. Entre otras cosas, este servicio crea la primera entrada en la Tabla Primaria de Ruteo.
 - **UInteger Node-ID:** ID del nodo a crear.
 - **UInteger address:** Dirección del nodo creado. UInteger entre [0,255]
 - **String description:** Descripción del nodo.
- **Boolean connect_remote_node(UInteger Node-ID):** envía la señal a sus nodos vecinos para que creen una referencia de este nodo en los demás.
 - **UInteger Node-ID:** ID del nodo remoto a conectar.
- **Boolean create_network():** esta función fue descrita en A.6.2.4. Este servicio crea una instancia de la Tabla Primaria de Ruteo.
- **Boolean update_NodeTableConfiguration():** esta función actualiza la tabla de configuración del nodo (Vista en: A.6.2.4).
- **Data query_primary_routing():** esta función lleva a cabo una consulta a la Tabla Primaria de Ruteo. Se la utiliza cuando se necesita extraer datos de la configuración de la red.
- **Data query_secondary_routing():** esta función lleva a cabo una consulta a la Tabla Secundaria de Ruteo. Se la utiliza cuando se necesita extraer datos de la configuración de la red luego de fallas en los nodos.
- **Boolean add_node(UInteger Node-ID, String Node-Alias, Boolean status, UInteger distance, UInteger address):** este servicio se lleva a cabo cuando llega el mensaje `connect_remote_node(UInteger Node-ID)`. Este servicio agrega un nodo (renglón) en la Tabla Primaria de Ruteo.
 - **UInteger Node-ID:** ID del nodo a agregar.
 - **String Node-Alias:** Alias del nodo a agregar.
 - **Boolean status:** Estado del nodo. CONNECTED=1, DISCONNECTED=0. Por defecto los nodos están en estado 1 (CONNECTED).
 - **UInteger distance:** Distancia que existe entre el nodo “host” y el nodo a agregar.
 - **UInteger address:** Dirección del nodo a agregar.

A.10. Formato de mensajes

Los mensajes enviados entre los nodos deben tener un formato compatible con las redes CAN convencionales. Por ello el mensaje tiene un tamaño que varía desde los 45 bits hasta los 109 bits (5 bytes hasta los 13 bytes), dependiendo de la cantidad de datos en el *Data Field*.

En la Figura A.22 se observa el formato de Frame de datos de los mensajes CANae.

SOF	Priority	Node-ID	RTR	Type of Message	Reserved bit	Data length	Data field	CRC	CRC delimiter	ACK	ACK Delimiter	EOF
1	4	7	1	1	1	0-8	0-64	1	1	1	1	1

Figura A.22: Frame de datos CANae

Los campos del mensaje son los siguientes:

- Start of Frame (SOF) - 1 bit: Este bit marca el comienzo de un frame. Suele ser 0.
- Priority - 4 bits: Este campo declara la prioridad del frame.
- Node-ID - 7 bits: El ID del Nodo origen.
- Remote transmission request (RTR) - 1 bit: Bit dominante (0) para frames de datos o eventos, y bit recesivo (1) para frames remotos. Un frame remoto es un mensaje que es enviado en forma automática por algun sensor. Los frames remotos se utilizan para enviar telemetría por parte de los sensores y componentes.
- Type of Message (TOM) - 1 bit: Bit dominante (0) si es un evento, y bit recesivo (1) para frames remotos.
- Bit reservado - 1 bit: Este bit está reservado para futuras aplicaciones. En esta versión debe ser un bit dominante (0).
- Data length - 6 bits: Cantidad de bytes del campo de datos.
- Data field - 0-64 bits (0-8 bytes): Datos.
- CRC - 15 bits: Cyclic redundancy check.
- CRC delimier - 1 bit: Debe ser recesivo (1).
- ACK - 1 bit: El transmisor debe poner este bit en recesivo (1), y el receptor puede responder este bit con un bit dominante (0).
- ACK delimiter - 1 bit: Debe ser recesivo (1).
- End-of-Frame (EOF) - 7 bits: Final del frame de datos.

El *Data Field* tiene el siguiente formato: el primer byte siempre es el ID de la función, evento y/o comando. Esto significa que es posible definir un número de $2^8 = 256$ funciones, eventos y comandos diferentes.

Los siguientes bytes (del 1 al 63) son argumentos de las funciones o datos.

Debe entenderse que esto no es mandatorio, ya que se pueden definir ID de dos o más bytes reduciendo la cantidad de datos.

A.10.1. Start of Frame

Este bit marca el comienzo de un mensaje CANae. Este es simplemente un bit dominante (0). Para que un dispositivo pueda enviar mensajes, el bus debe estar en estado IDLE.

A.10.2. Priority

Este campo indica la prioridad del mensaje. La prioridad está compuesta por 4 bits, por lo tanto acepta una cantidad de 2^4 niveles de prioridad. La prioridad máxima es 0000, mientras que la mínima es 1111.

Tabla A.3: Prioridades de mensajes

Priority Field	Prioridad	
0000	0	Alta prioridad
0001	1	...
...
1111	15	Baja prioridad

A.10.3. Node-ID

Este indica el ID del nodo que envía el mensaje. Este campo está compuesto por 7 bits, lo cual posibilita la existencia de hasta $2^7 = 128$ nodos conectados a la red.

A.10.4. Remote transmission request (RTR)

Este permite diferenciar un frame que sea remoto o un mensaje “normal”. Un frame remoto es aquel que se utiliza para enviar telemetría. Comúnmente es enviado por sensores y componentes. El bit dominante (0) indica que es un frame “normal” (contiene un mensaje o un evento), un bit recesivo (1) indica que es un frame remoto.

A.10.5. Type of Message (TOM)

Este campo es de 1 bit. Diferencia frames de mensajes con frames de eventos. Si el bit es dominante (0) significa que se trata de un frame de evento. Si el bit es recesivo (1) significa que se trata de un frame de mensaje. Debe tenerse en cuenta que si el bit RTR es recesivo (1), el bit TOM obligatoriamente debe ser recesivo (1) indicando que el frame trae consigo un mensaje con los datos de telemetría.

A.10.6. Bit reservado

Bit reservado.

A.10.7. Data length

Indica el tamaño del campo de datos (*Data Field*). Este campo esta compuesto por 6 bits.

A.10.8. Data field

En este campo se agregan los datos del mensaje. El primer byte del *Data Field* es el código de la función y/o evento. Los bytes subsiguientes pueden ser los argumentos y/o datos de la función o evento que se está enviando en el frame. Se pueden enviar solo una función a la vez, y hasta 7 datos o argumentos.

A.10.9. CRC

En este campo se almacena el CRC del frame. Para calcular el CRC se utiliza el polinomio de CAN.

$$x^{15} + x^{14} + x^{10} + x^8 + x^7 + x^4 + x^3 + 1$$

El algoritmo para resolver el CRC se muestra en A.1.

Listing A.1: Función para resolver CRC

```
uint16_t crc_c(uint8_t data , uint16_t crc){
    uint8_t i;

    crc ^= (uint16_t) data << 7;
    for (i = 0; i < 8; i++){
        crc <<= 1;
        if (crc & 0x8000){
            crc ^= 0xc599;
        }
    }
    return crc & 0x7fff;
}
```

A.10.10. CRC delimiter

Este es un bit recesivo (1) para indicar que el CRC Field finalizó.

A.10.11. ACK

El transmisor debe poner este bit en recesivo (1), y el receptor puede responder el mensaje colocando este bit en dominante(0).

A.10.12. ACK delimiter

Este bit indica el final del ACK. Este es un bit recesivo (1).

A.10.13. End-of-Frame (EOF)

Indica el final del frame de datos. Debe ser recesivo, es decir, los 7 bits en 1.

A.11. High Application Layer CANae

La denominada *High Application Layer* tiene una función más del lado de la gestión de nodos y tareas. En esta capa se desarrollan los algoritmos necesarios para realizar el ruteo y la reconfiguración de la red ante fallas, por lo tanto en la *High Application Layer* se debe implementar la FDIR. El protocolo no define ningún algoritmo de ruteo, por lo que queda para el usuario la definición de los algoritmos. Se recomienda desarrollar 2 algoritmos, el primario y el secundario. Para aumentar la tolerancia a fallas se pueden desarrollar más algoritmos secundarios, y el switch entre algoritmos, puede depender de medidas de performance.

En las Figuras [A.23] y [A.24] se pueden observar los diagramas de bloques y diagramas de bloques internos para el *High Application Layer* de CANae.

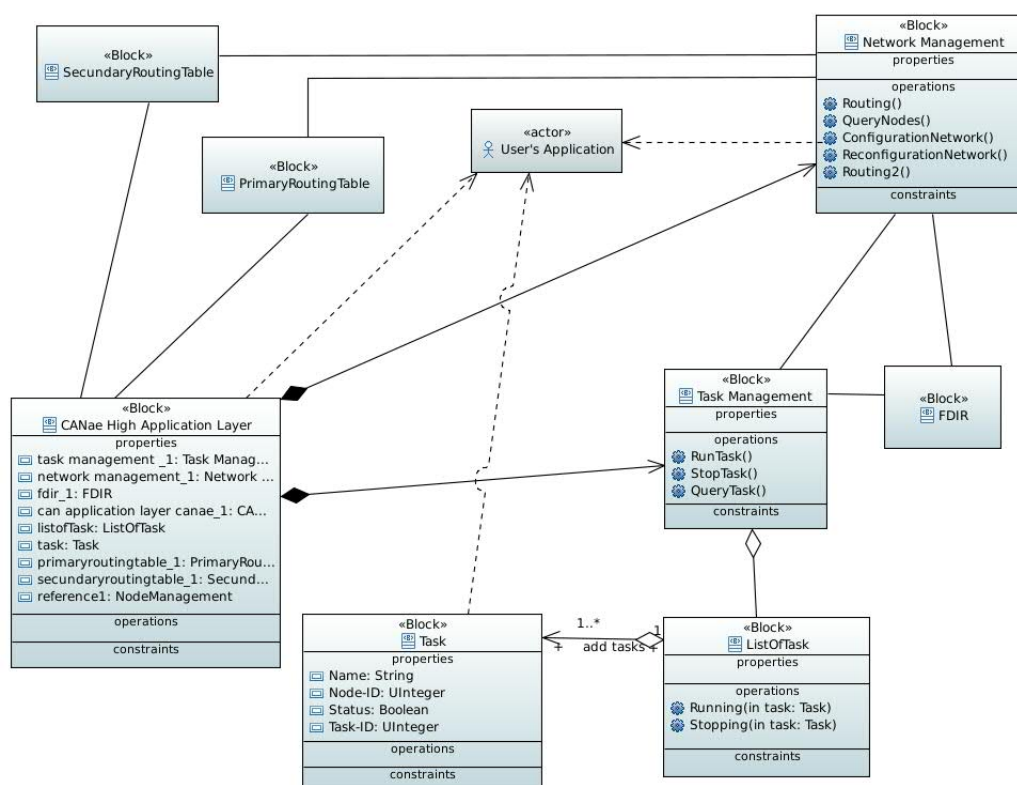


Figura A.23: Diagrama de bloques del High Application Layer de CANae

Esta capa está compuesta por varias entidades, las cuales tienen solo una instancia por entidad.

En la *High Application layer* se encuentra la entidad más importante, la cual es el FDIR. El desarrollo de esta entidad se encuentra fuera del alcance de esta versión (0.1 Alpha) del protocolo.

A.11.1. Network Management

Esta entidad se encarga de la gestión de la red desde el punto de vista del nodo. La entidad utiliza el *NodeTableConfiguration* del *CANae Application Layer*.

Los servicios brindados por el *Network Management* son los siguientes:

- **Routing()**: este servicio lleva a cabo el ruteo de los mensajes. Este servicio consulta la tabla *NodeTableConfiguration*.

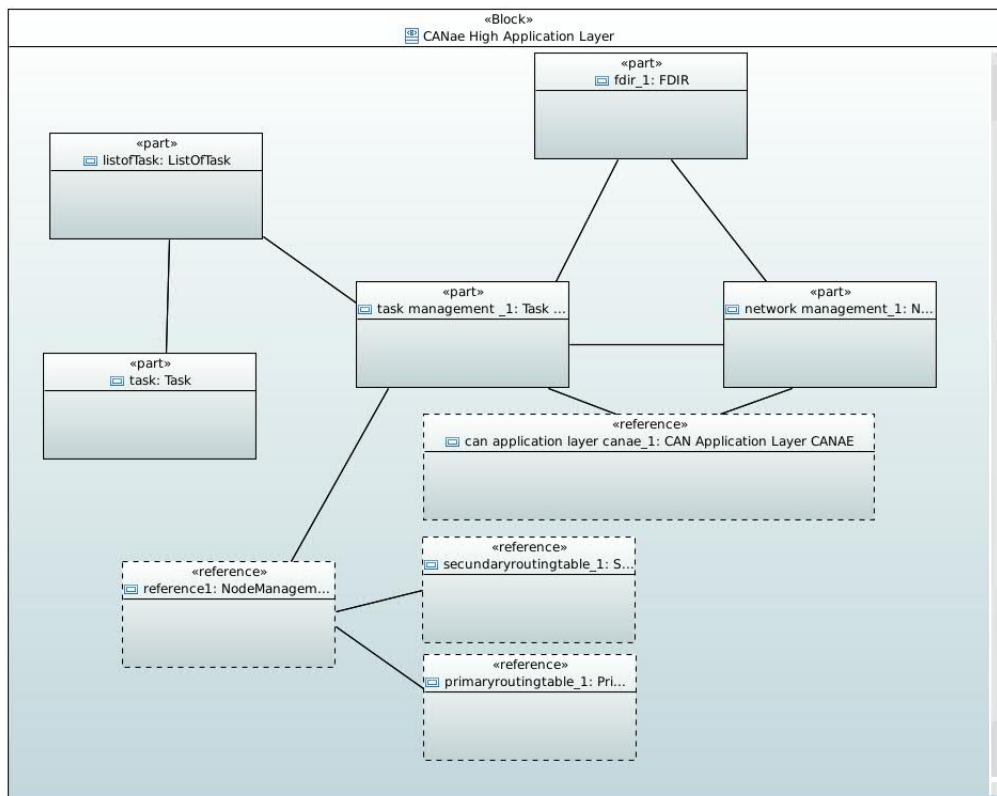


Figura A.24: Diagrama de bloques internos del High Application Layer de CANae

- **QueryNodes():** este servicio hace consulta de los nodos conectados. Este puede ser utilizado por el servicio Routing()
- **ConfigurationNetwork():** este servicio se lleva a cabo solo una vez, durante el encendido y configuración al nodo.
- **ReconfigurationNetwork():** este servicio lleva a cabo la reconfiguración de la red. Este servicio es activado por el FDIR una vez que detecta algún error.
- **Routing2():** Este servicio se lleva a cabo cuando se detecta un error en la red. Lo activa el FDIR.

El comportamiento del *Network Management* es sencillo, cuando se enciende por primera vez el nodo, de manera automática, la aplicación de usuario debe llamar el servicio *ConfigurationNetwork()* al *Network Management* esta entidad se comunica con el *Node Management* para que se genere la “Tabla Primaria de Ruteo”. Una vez que esta tabla se encuentre generada (tal como se hace referencia en A.9.1), la aplicación de usuario puede enviar señales de *QueryNodes()* y *Routing()*. Esto se puede ver visualmente en la Figura A.25.

En caso de errores, esto debe ser captado por el FDIR del sistema. El comportamiento del mismo se muestra en la Figura A.26

A.11.2. Task Management

Esta entidad se encarga de la gestión de las tareas. Debe aclararse que esta es una abstracción, y no depende del Sistema Operativo que se utilice para la codificación del sistema. El objetivo de esta entidad es mantener la distribución de las tareas entre los nodos de la red. Esto permitiría que cuando deja de funcionar (detectado por

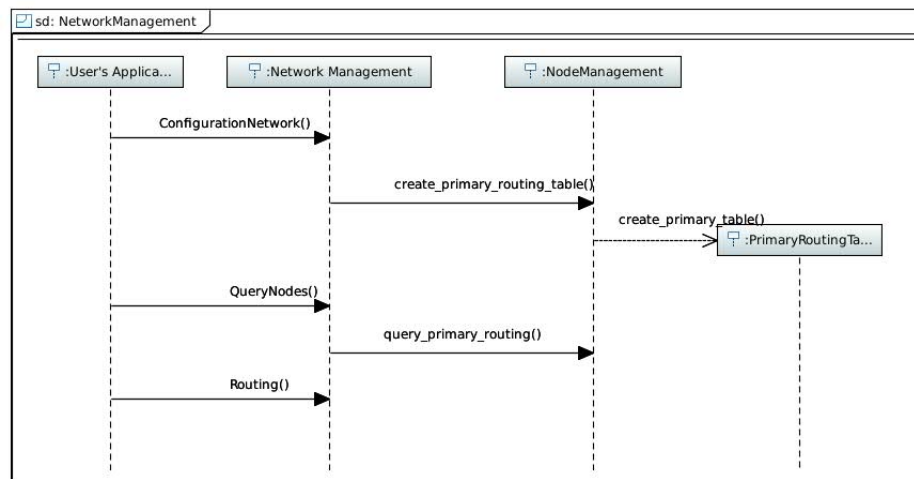


Figura A.25: Diagrama de interacción del comportamiento del *Network Management*

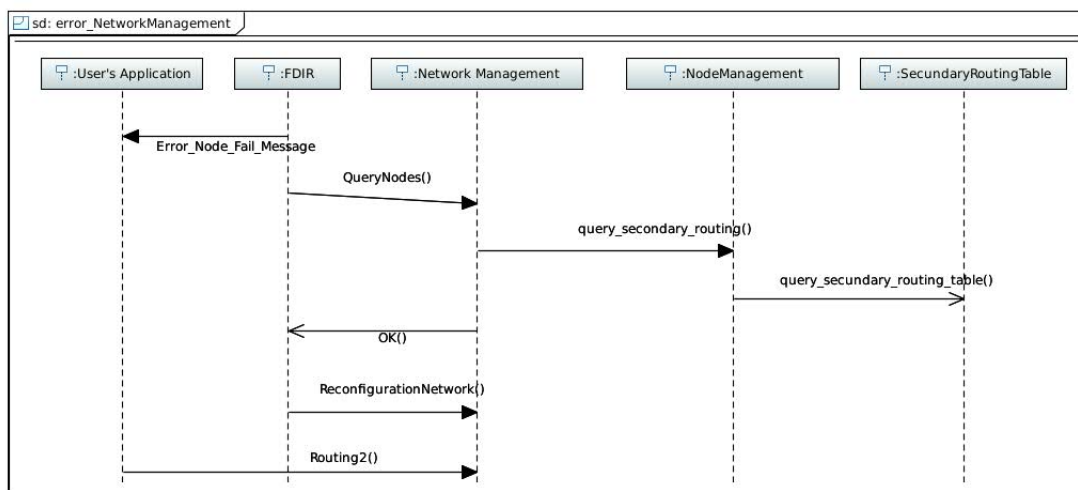


Figura A.26: Diagrama de interacción del comportamiento del *Network Management* en caso de errores

el FDIR) uno de los nodos, el *Task Management* gestiona las tareas que se desarrollan en los nodos. Para ellos se hace uso de la “Lista de Tareas” (*ListOfTask*).

Los servicios que ofrece el *Task Management* son los siguientes:

- **RunTask()**: este servicio activa las tareas para ejecutarse en el nodo.
- **StopTask()**: este servicio detiene las tareas que se ejecutan en el nodo.
- **QueryTask()**: este servicio hace consulta de las tareas que existene la “lista de Tareas”.

El comportamiento de esta entidad es sencilla. La aplicación de usuario le envía una señal de *QueryTasks()* con el motivo de conocer todas las tareas que se están ejecutando en la red. El *Task Management* lleva a cabo un *query_tasks_in_network()* al *Network Management* el cual le provee la información. La primera comprobación que debe realizar la aplicación de usuario es que la tarea no existe en otro nodo ejecutandose. Luego de esta comprobación ya tiene permitido llevar a cabo el *RunTask()*. Por último el *Task Management* manda el mensaje *Running(Task task)* al *ListOfTasks*(ver A.11.3). El comportamiento de la entidad se la puede observar en la Figura A.27.

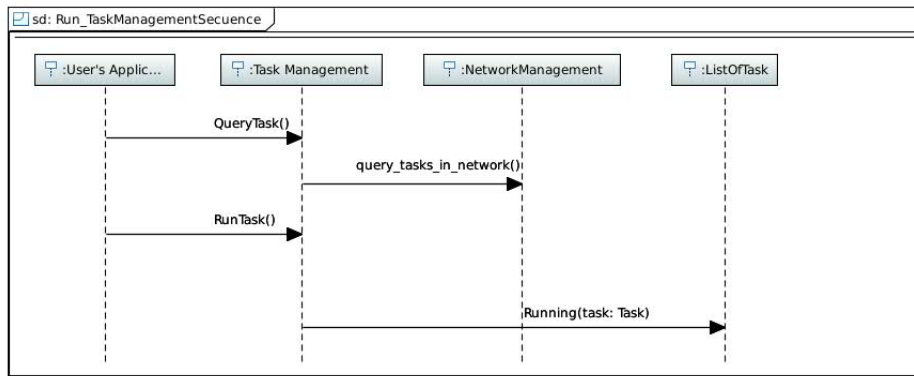


Figura A.27: Diagrama de secuencia de la ejecución de tareas del Task Management

Por otro lado cuando la aplicación de usuario necesita para una tarea, o debido a que el FDIR detecto una falla y necesita detener la ejecución de una tarea, el proceso que debe llevar el *Task Management* es el siguiente, en primer hace un *QueryTasks()* para conocer las tareas que se están ejecutando y sus estados. Luego manda un señal de *StopTask()* al *Task Management*. Esta, envía mensajes de *query_tasks_in_network()*, y luego el *Stopping(Task task)* para detener la ejecución del proceso. En la Figura A.28

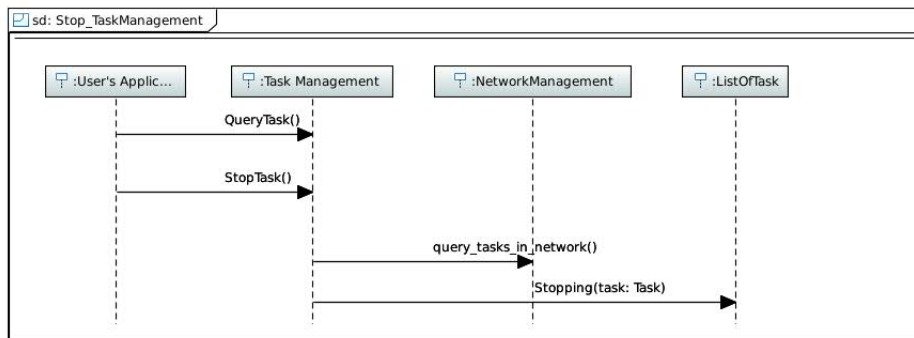


Figura A.28: Diagrama de secuencia para detener procesos por parte del Task Management

A.11.3. List of Tasks

Este es un objeto que mantiene una lista de las tareas que se están definidas en el sistema. Estas tareas pueden o no estar ejecutandose. Esta lista tiene un instancia a todas las tareas.

Los servicios que se ejecutan en esta entidad son los siguientes:

- **Running(Task task):** este servicio ejecuta una tarea.
 - **Task task:** tarea a ejecutar.
- **Stopping(Task task):** este servicio detiene una tarea.
 - **Task task:** tarea a ejecutar.

A.11.4. Task

Este es un objeto que representa la abstracción de las tareas. Las propiedades de esta entidad son las siguientes:

- **String Name:** es el nombre de la tarea.
- **UInteger Task-ID:** es el ID de la tarea.
- **UIntenger Node-ID:** es el ID del nodo donde se está ejecutando la tarea.
- **Boolean Status:** es el status de la tarea {RUNNING, STOPED}

A.12. Arquitectura completa del CANae

CANae nace en el principio de la ingeniería en sistemas basada en modelo utilizando el lenguaje de modelado SysML para su diseño y desarrollo. Debido a la complejidad de los sistemas espaciales, a la necesidad de lograr un entendimiento completo por parte de los desarrolladores, y para lograr una correcta interpretación del protocolo se decidió basarse en este principio. En esta sección se pretende mostrar solo con motivo informativo la arquitectura estática completa del protocolo, para demostrar la complejidad del mismo, pero a la vez lo sencillo que resulta su interpretación, con conocimientos básicos de SysML.

En esta se puede observar como interactúan las diferentes entidades, objetos y eventos que se desarrollaron a lo largo de este documento.

En la Figura A.29 se observa la arquitectura de la capa de aplicación de CANae denominada: **CAN Application Layer CANae**. La cual como se mencionó con anterioridad se encuentra basada en la capa de aplicación de diferentes protocolos que existen para las redes CAN.

A.13. Estados de nodos y red

El protocolo de comunicación CANae está diseñado para ser sencillo en su uso, por lo tanto, los nodos como la red pasan por estados de funcionamiento que son sencillos y atómicos. En esta versión se pretende desarrollar las bases para los futuros avances sobre el mismo.

En primer lugar el ciclo de vida de un nodo comienza cuando es alimentando. El nodo empieza en un estado *Stand-by* hasta que se inicializa internamente. Luego pasa a un estado de *Waiting Monitor*, donde espera las ordenes de un *Monitor Node*. Luego cuando recibe la señales del monitor comienza a configurarse internamente (ver: y). Luego de esto pasa a un estado *Working*. Aquí pueden ocurrir dos situaciones. En primer lugar, el nodo se apaga intencionalmente. La segunda situación es que el nodo falla y es detectado por el FDIR, entonces pasa a un estado *Failure*.

Por otro lado en el entorno de la red CANae, el ciclo de vida de la red comienza con la alimentación de los nodos. En primer lugar entra en un estado de *Init Nodes* donde todos los nodos de la red se encuentran desarrollando su ciclo de vida, tal como se indicó anteriormente. Luego la red pasa a un estado *Stand-By* para luego comenar el estado de *Creating Network*. Dentro de este estado, se dan diferentes estados, se necesita que todos los nodos se encuentren coordinados. Estos se detallan en A.13. Luego de su creación toda la red se encuentra en estado *Ready* donde se lleva a cabo su normal funcionamiento.

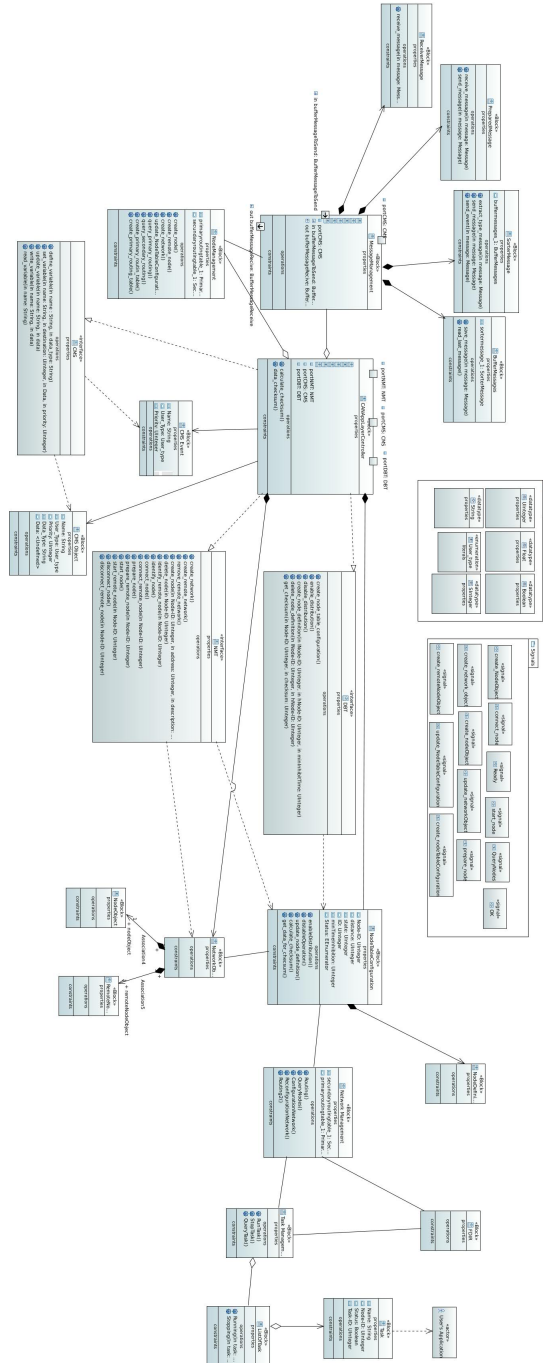


Figura A.29: Arquitectura completa del CANae

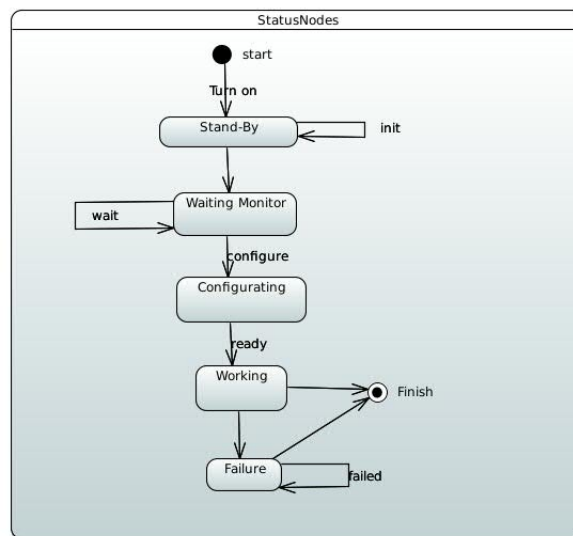


Figura A.30: Definición de la entidad NMT

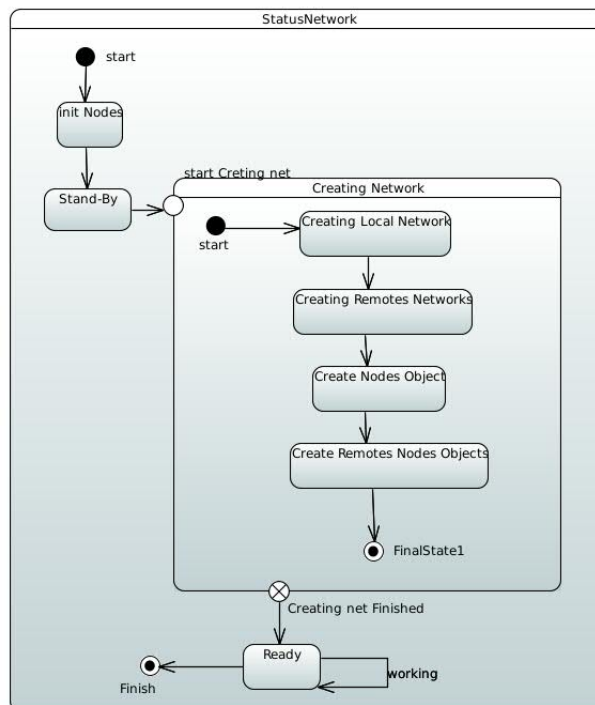


Figura A.31: Definición de la entidad NMT

Especificación de Casos de Uso

B.1. Enviar Mensaje

Use Case ID	El ID del Caso de uso es ARQ001
Historia del Caso de Uso	
Creado por	Arias Emmanuel
Fecha de Creación	12 de Junio del 2017
Fecha de la última actualización	N/A
Definición del Caso de Uso	
Actor	<ul style="list-style-type: none"> ▪ <i>UserApplication</i>.Este es la aplicación del usuario que se encuentra ejecutandose en el nodo.
Descripción	Este caso de uso recibe los mensajes que la aplicación de usuario necesita enviar por la red hacia otro nodo en el sistema.
Precondición	<ul style="list-style-type: none"> ▪ El nodo debe estar funcional y conectado a la red. ▪ El protocolo CANae debe haber finalizado su inicialiación. ▪ El nodo destino tiene que estar funcional.
Postcondición	<ul style="list-style-type: none"> ▪ Debe asegurarse el correcto envío del mensaje.
Prioridad	Prioridad media.

Curso normal de eventos	El caso de uso recibe el mensaje proveniente de la aplicación del usuario. En el siguiente paso prepara el mensaje siguiendo los lineamientos del protocolo CANae. Por último lo envía a las capas inferiores para ser enviado a través del protocolo CAN.
Cursos alternativos	N/A
Excepciones	En caso de producirse un error en el caso de uso, se lleva a cabo una excepción, para avisar a la aplicación de usuario que se produjo un error.
Includes	N/A
Notas y problemas	N/A

B.2. Recibir Mensaje

Use Case ID	El ID del Caso de uso es ARQ002
Historia del Caso de Uso	
Creado por	Arias Emmanuel
Fecha de Creación	12 de Junio del 2017
Fecha de la última actualización	N/A
Definición del Caso de Uso	
Actor	<ul style="list-style-type: none">▪ <i>UserApplication</i>. Este la aplicación del usuario que se encuentra ejecutandose en el nodo.
Descripción	Este caso de uso recibe los mensajes proveniente de las capas inferiores de CAN y luego es enviado a la aplicación del usuario.
Precondición	<ul style="list-style-type: none">▪ El mensaje tiene que llegar sin errores.▪ El nodo debe estar conectado y funcional en la red.▪ El nodo debe estar preparada (no ocupado) para recibir el mensaje.
Postcondición	N/A
Prioridad	Prioridad media.
Curso normal de eventos	N/A
Cursos alternativos	N/A
Excepciones	N/A
Includes	N/A
Notas y problemas	N/A

B.3. Procesar Mensaje

Use Case ID	El ID del Caso de uso es ARQ003
Historia del Caso de Uso	
Creado por	Arias Emmanuel
Fecha de Creación	12 de Junio del 2017
Fecha de la última actualización	N/A
Definición del Caso de Uso	
Actor	N/A
Descripción	Este caso de uso se encarga de recibir los mensajes desde la aplicación de usuario y de las capas inferiores de CAN, para realizar un procesamiento. Este procesamiento incluye empaquetado y desempaquetado, y verificación de la ocurrencia de errores, clasificación de mensajes que contienen datos o eventos.
Precondición	<ul style="list-style-type: none">■ El nodo debe estar conectado y funcional en la arquitectura.
Postcondición	N/A
Prioridad	Prioridad alta.
Curso normal de eventos	El caso de uso recibe el mensaje proveniente de las capas inferiores del CAN. Se suponen que llegan sin errores. Se desempaqueta, se clasifica (si es datos o evento) y luego se envía a la aplicación de usuario.
Cursos alternativos	N/A
Excepciones	N/A
Includes	Ver Rutear Mensajes. B.4
Notas y problemas	N/A

B.4. Rutear Mensaje

Use Case ID	El ID del Caso de uso es ARQ003.1
Historia del Caso de Uso	
Creado por	Arias Emmanuel
Fecha de Creación	12 de Junio del 2017
Fecha de la última actualización	N/A
Definición del Caso de Uso	
Actor	N/A
Descripción	Este es el encargado de rutear los mensajes. Debe indicar que nodos intermedios debe pasar el mensaje hasta llegar al nodo destino.

Precondición	<ul style="list-style-type: none">■ El nodo debe estar conectado y funcional en la arquitectura.■ El mensaje debe estar clasificado y empaquetado.■ Se debe conocer el nodo destino.
Postcondición	N/A
Prioridad	Prioridad alta.
Curso normal de eventos	Este caso de uso una vez que se clasifica y empaqueta el mensaje, debe consultar al Caso de Uso <i>Gestion de Red</i> para verificar que nodos intermedios debe pasar el mensaje hasta llegar al nodo destino.
Cursos alternativos	N/A
Excepciones	N/A
Includes	N/A
Notas y problemas	N/A

B.5. Gestionar Red

Use Case ID	El ID del Caso de uso es ARQ004
Historia del Caso de Uso	
Creado por	Arias Emmanuel
Fecha de Creación	12 de Junio del 2017
Fecha de la última actualización	N/A
Definición del Caso de Uso	
Actor	N/A
Descripción	Este caso de uso se encarga de gestionar la red. Se vale de las bondades ofrecidas por CANae para conocer cuales son los nodos conectados a la red y la distancia que existe entre el nodo objetivo y el destino.
Precondición	<ul style="list-style-type: none">■ El nodo debe estar conectado y funcional en la arquitectura.■ El protocolo CANae debe haber finalizado su inicialización.■ El protocolo CANae tiene que haber realizar la tabla de ruteo.
Postcondición	N/A
Prioridad	Prioridad alta.

Curso normal de eventos	El caso de uso utiliza las bondades del protocolo CANae para conocer cuales son los nodos conectados a la red y su distancia. Este caso de uso es la base para que se realice una correcta gestión de tareas a realizar en el sistema. Luego con esto se pueden dividir las tareas que se llevarán a cabo en los diferentes nodos de la red.
Cursos alternativos	N/A
Excepciones	N/A
Incluye	Ver Gestionar Tareas B.6
Notas y problemas	N/A

B.6. Gestionar Tareas

Use Case ID	El ID del Caso de uso es ARQ004.1
Historia del Caso de Uso	
Creado por	Arias Emmanuel
Fecha de Creación	12 de Junio del 2017
Fecha de la última actualización	N/A
Definición del Caso de Uso	
Actor	N/A
Descripción	Este caso de uso se encarga de gestionar las tareas que se deben ejecutar en la red. Para ello debe implementarse un algoritmo que asegure la correcta comunicación con los nodos para lograr la gestión de tareas del sistema.
Precondición	<ul style="list-style-type: none">■ El nodo debe estar conectado y funcional en la arquitectura.■ El protocolo CANae debe haber finalizado su inicialización.■ El protocolo CANae tiene que haber realizado la tabla de ruteo.
Postcondición	N/A
Prioridad	Prioridad alta.
Curso normal de eventos	En primer lugar debe consultar la tabla de ruteo para conocer la posición de los nodos. Con ello y lista de tareas del sistema, lleva a cabo un algoritmo de gestión de tareas, donde se negocia con los demás nodos que tareas se van a llevar a cabo en los diferentes nodos.
Cursos alternativos	N/A
Excepciones	N/A

Includes	Ver Dividir Tareas B.7
Notas y problemas	N/A

B.7. Dividir Tareas

Use Case ID	El ID del Caso de uso es ARQ004.1.1
Historia del Caso de Uso	
Creado por	Arias Emmanuel
Fecha de Creación	12 de Junio del 2017
Fecha de la última actualización	N/A
Definición del Caso de Uso	
Actor	N/A
Descripción	Este caso de uso se encarga de realizar al división de las tareas, una vez que se ha llevado a cabo la negociación.
Precondición	<ul style="list-style-type: none">■ El nodo debe estar conectado y funcional en la arquitectura.■ El protocolo CANae debe haber finalizado su inicialización.■ El protocolo CANae tiene que haber realizar la tabla de ruteo.
Postcondición	<ul style="list-style-type: none">■ Se actualiza la lista de tareas en los nodos.
Prioridad	Prioridad alta.
Curso normal de eventos	Luego de recibir el resultado de la negociación, actualiza la lista de tareas que se llevan a cabo en los nodos. E informa a los demás nodos.
Cursos alternativos	N/A
Excepciones	N/A
Includes	N/A
Notas y problemas	N/A

B.8. FDIR

Use Case ID	El ID del Caso de uso es ARQ005
Historia del Caso de Uso	
Creado por	Arias Emmanuel
Fecha de Creación	12 de Junio del 2017

Fecha de la última actualización	N/A
Definición del Caso de Uso	
Actor	N/A
Descripción	Este caso de uso se encarga de la detección, aislación y recuperación de errores que se produzcan tanto en el nodo como en la red.
Precondición	<ul style="list-style-type: none">■ El nodo debe estar conectado y funcional en la arquitectura.■ El protocolo CANae debe haber finalizado su inicialización.■ El protocolo CANae tiene que haber realizado la tabla de ruteo.
Postcondición	<ul style="list-style-type: none">■ Se toma una medida para lograr aislar y recuperar la arquitectura del error que se ha producido.
Prioridad	Prioridad alta.
Curso normal de eventos	Este caso de uso debe observar como se encuentra la red en cada momento. Para ello hace uso de las bondades que le puede brindar el caso de uso <i>Gestionar Red</i> . Cuando detecta que se ha producido una falla en el nodo o en la red, lanza una excepción para manejar el problema.
Cursos alternativos	Ver <i>Reconfigurar Arquitectura B.9</i>
Excepciones	N/A
Includes	N/A
Notas y problemas	No se llega a realizar un diseño y desarrollo de este caso de uso en esta instancia del trabajo de tesis.

B.9. Reconfigurar Arquitectura

Use Case ID	El ID del Caso de uso es ARQ006
Historia del Caso de Uso	
Creado por	Arias Emmanuel
Fecha de Creación	12 de Junio del 2017
Fecha de la última actualización	N/A
Definición del Caso de Uso	
Actor	N/A

Descripción	Este caso de uso se encarga de la detección, aislación y recuperación de errores que se produzcan tanto en el nodo como en la red.
Precondición	<ul style="list-style-type: none">■ El nodo debe estar conectado y funcional en la arquitectura.■ El protocolo CANae debe haber finalizado su inicialización.■ El protocolo CANae tiene que haber realizado la tabla de ruteo.
Postcondición	<ul style="list-style-type: none">■ Se toma una medida para lograr aislar y recuperar la arquitectura del error que se ha producido.
Prioridad	Prioridad alta.
Curso normal de eventos	Este caso de uso debe observar como se encuentra la red en cada momento. Para ello hace uso de las bondades que le puede brindar el caso de uso <i>Gestionar Red</i> . Cuando detecta que se ha producido una falla en el nodo o en la red, lanza una excepción para manejar el problema.
Cursos alternativos	Ver <i>Reconfigurar Arquitectura</i> B.9
Excepciones	N/A
Includes	N/A
Notas y problemas	No se llega a realizar un diseño y desarrollo de este caso de uso en esta instancia del trabajo de tesis.