

PROYECTO DE INVESTIGACIÓN

Título	Detectores de Fallos en Sistemas Distribuidos
Código UTN	FESIEC855
Dependencia	Dpto. Ingeniería en Sistemas de Información
Director/a	PESSOLANI, Pablo
Codirector/a	GUTIERREZ, María de los Milagros
Período de ejecución	01/04/2026 hasta 31/03/2028 - 2 años
Resumen	<p>Los sistemas distribuidos (DS - Distributed Systems) tienen como principal objetivo incrementar la capacidad de procesamiento, mejorar la escalabilidad y facilitar la programación, el despliegue y la gestión de aplicaciones distribuidas que se ejecutan sobre ellos. Es sabido que detectar un fallo de manera certera reduce los costos de operación y aumenta la confianza en el sistema distribuido. Sin embargo, esta tarea no es trivial.</p> <p>En la literatura suele mencionarse que la detección de fallos puede resolverse fácilmente mediante timeouts y reintentos. No obstante, estas estrategias no contemplan la complejidad propia de los sistemas distribuidos, como el problema del consenso distribuido [FLP].</p> <p>En este contexto, Chandra y Toueg [FD] proponen una abstracción denominada Failure Detector (FD), que funciona como un oráculo consultable por distintas aplicaciones para conocer el estado de los nodos y procesos miembros de un grupo o clúster. Esta propuesta demuestra ser una solución práctica al problema del consenso distribuido en forma eventual.</p> <p>En particular, este PID propone investigar diferentes mecanismos de detección de fallos basados en métodos clásicos como heartbeats, pings, mensajes Hello, solicitudes ARP, conexiones TCP, traceroute, entre otros. Asimismo, se analizarán mecanismos complementarios, tales como el análisis del emisor de los paquetes recibidos, el estado de componentes de red (switches, routers, etc.), los emisores del tráfico que circula por ellos y la información reportada por otros miembros del clúster. De esta manera, se busca brindar información no solo sobre el estado de los nodos, sino también sobre diversas capas de software hasta llegar a las aplicaciones.</p> <p>Además, con el surgimiento de los modelos de aprendizaje automático, capaces de predecir y aprender patrones de</p>

comportamiento, se indagará su aplicación en el contexto de la detección de fallos en sistemas distribuidos. El objetivo será comparar su desempeño frente a los mecanismos tradicionales ya establecidos en la comunidad tecnológica. En los últimos años han surgido investigaciones que exploran el uso de aprendizaje automático para la detección de fallos [MLFD][MLFD2].

Una de las hipótesis del proyecto es que las técnicas de aprendizaje automático pueden mejorar el desempeño de los algoritmos de detección de fallos, aumentando la precisión con la que cada miembro de una aplicación distribuida evalúa el estado de los demás. Esto permitiría tomar decisiones más acertadas y ejecutar acciones que mejoren la disponibilidad del sistema, manteniendo la integridad y consistencia de los datos.

Uno de los principales desafíos del proyecto es la construcción y/o búsqueda de un conjunto de datos adecuado para entrenar algoritmos de aprendizaje automático. Los tipos de fallos y los patrones de comportamiento varían entre distintos sistemas distribuidos, y es probable que los datos relacionados con fallos contengan información sensible o confidencial, lo que dificulta su recolección. Además, en muchos casos, estos datos son administrados por proveedores de servicios, lo que limita su acceso. Estas dificultades hacen que los datos reales sean escasos y difíciles de obtener.

Algunas propuestas sugieren el uso de datos sintéticos generados mediante simulaciones. Por otro lado, los registros de logs en sistemas distribuidos pueden constituir una fuente relevante de información. Para su aprovechamiento, podría ser necesario aplicar técnicas de procesamiento de lenguaje natural y análisis de series temporales, con el fin de extraer patrones y características que indiquen la presencia de fallos.