



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

Paralelización de procesos de modelamiento de tráfico urbano por medio de la
contenerización del software Simulation of Urban MObility (SUMO) para
Supercomputadores

TESIS PARA OPTAR AL TÍTULO DE MAGÍSTER EN CIENCIAS, MENCIÓN
COMPUTACIÓN

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERE CIVIL EN COMPUTACIÓN

PABLO VILLAR MASCARÓ

PROFESOR GUÍA:
Javier Bustos Jiménez

PROFESOR CO-GUÍA:
Patricio Reyes

MIEMBROS DE LA COMISIÓN:
NOMBRE COMPLETO UNO
NOMBRE COMPLETO DOS
NOMBRE COMPLETO TRES

SANTIAGO DE CHILE
2025

RESUMEN DE LA TESIS PARA OPTAR
AL TÍTULO DE INGENIERE CIVIL
EN COMPUTACIÓN Y MAGÍSTER EN
CIENCIAS, MENCIÓN COMPUTACIÓN
POR: PABLO VILLAR MASCARÓ
FECHA: 2025
PROF. GUIA: Javier Bustos Jiménez

Paralelización de procesos de modelamiento de tráfico urbano por medio de la contenerización del software Simulation of Urban MObility (SUMO) para Supercomputadores

El presente trabajo de tesis aborda la problemática del escalamiento de simulaciones de tráfico vehicular urbano para su uso en el modelamiento de flujos de tráfico a nivel microscópico a través de grandes áreas metropolitanas. Se encuentra orientado a su implementación en ambientes de supercomputación, insertándose en un proyecto conjunto entre NIC Chile Research Labs y el *Barcelona Supercomputing Center* (BSC) para el desarrollo de gemelos digitales para las ciudades de Barcelona y Kobe (Japón).

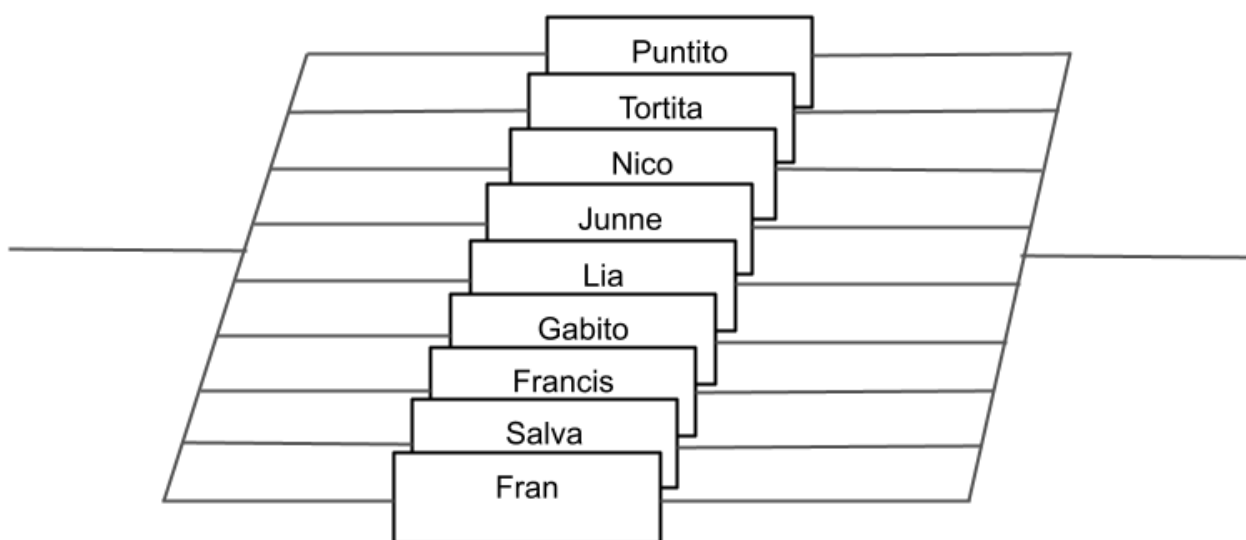
Dado el escalamiento de comportamiento exponencial de las simulaciones que se estudian dentro del trabajo preliminar de esta investigación, se propone una solución basada en la paralelización de estos procesos, buscando aumentar el *speedup* de la ejecución de las simulaciones, con el objetivo de analizar y comparar el crecimiento entre ambos modelos en cuanto al tiempo de ejecución.

Por medio del uso del *software* SUMO (*Simulation of Urban MObility*), el uso del algoritmo de partición de grafos *SPartSim* para la división de mapas y la implementación de una arquitectura de paralelización orientada a dispositivos de memoria compartida a través de *OpenMP*, se logra desarrollar una versión paralela de simulaciones generadas para las ciudades de Barcelona y Viladecans.

Para medir la escalabilidad de la solución implementada, se realiza un test de carga inicial para la versión secuencial de las simulaciones, midiendo sus tiempos de ejecución y el uso de CPU en cada caso, mientras que para la versión paralelizada se realiza un procedimiento similar.

A partir de estos experimentos, se obtiene que a medida que se aumenta el número de nodos de ejecución, se logra un escalamiento de comportamiento más lineal respecto al comportamiento de la versión secuencial, además de cotas más bajas para el tiempo máximo que ocupan las simulaciones en su ejecución.

Para todos los amores paralelos de mi vida



por todo su cariño y soporte a través de estos años inciertos y hostiles.

Agradecimientos

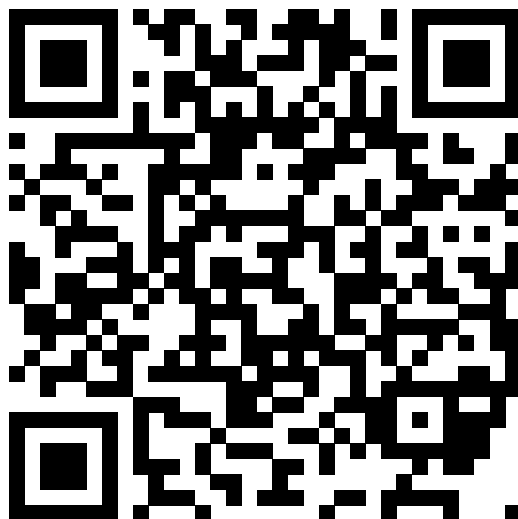


Tabla de Contenido

1. Introducción	1
2. Problema	4
2.1. Pruebas de escalabilidad en SUMO	4
3. Preguntas de investigación, hipótesis y objetivos	8
3.1. Preguntas de investigación	8
3.2. Hipótesis	8
3.3. Objetivos	8
4. Solución propuesta	10
4.1. Diseño de la solución	12
4.2. Herramientas utilizadas	12
4.3. Partición de mapas de Barcelona y Viladecans	13
4.4. Generación y partición de rutas vehiculares	14
4.5. Contenerización de simulaciones secuenciales	15
4.6. Paralelización de las simulaciones particionadas	16
4.7. Sincronización de las simulaciones particionadas	17
4.8. Diseño e implementación de test de carga	18
5. Marco teórico	19
5.1. Simuladores de tráfico urbano	19
5.2. Simulation of Urban MObility (SUMO)	22
5.3. Gemelos Digitales (<i>Digital Twins</i>) para ciudades	23
5.4. Simulaciones de tráfico con SUMO para el desarrollo de gemelos digitales	24
5.5. Simulaciones de tráfico vehicular urbano en High-Performance Computing	25
6. Resultados	27
7. Discusión	35

8. Conclusiones	37
Bibliografía	39
A. Ejemplos de archivos	42
1.1. Archivo XML para la descripción de redes de caminos en SUMO	42
1.2. Archivo XML para la descripción de rutas en SUMO	46
1.3. Archivo XML para la configuración de una simulación en SUMO	48
1.4. Resultados para tiempos de ejecución de simulaciones secuenciales	49
1.5. Resultados para tiempos de ejecución de simulaciones paralelizadas	50
1.6. Resultados para la eficiencia de paralelización de la solución propuesta	52

Índice de Tablas

Tabla 1: Comparación entre simuladores de tráfico vehicular (adaptada de [1] y [2]) ...	20
Tabla 2: Comparación entre <i>performance</i> de simuladores de tráfico vehicular (adpatada de [1] y [2])	21
Tabla 3: Comparación entre características de simuladores de tráfico vehicular (Adaptada de [3])	21
Tabla 5: Speedup para 4 particiones	30
Tabla 6: Speedup para 8 particiones	30
Tabla 7: Speedup para 16 particiones	31
Tabla 8: Speedup para 32 particiones	31
Tabla 9: Speedup para 64 particiones	32
Tabla 10: Tiempos para simulación secuencial	49
Tabla 11: 4 particiones	50
Tabla 12: 8 particiones	50
Tabla 13: 16 particiones	50
Tabla 14: 32 particiones	51
Tabla 15: 64 particiones	51
Tabla 16: Eficiencia calculada para 4 particiones	52
Tabla 17: Eficiencia calculada para 8 particiones	53
Tabla 18: Eficiencia calculada para 16 particiones	54
Tabla 19: Eficiencia calculada para 32 particiones	55
Tabla 20: Eficiencia calculada para 64 particiones	56

Índice de Ilustraciones

Figura 3: <i>Pipeline</i> para simulaciones basadas en el tamaño de la red.	5
Figura 4: <i>Pipeline</i> para simulaciones basadas en la carga de tráfico vehicular.	5
Figura 5: Resultados para la prueba de escalabilidad basada en el tamaño de la red de caminos.	6
Figura 6: Resultados para la prueba de escalabilidad basada en la carga de tráfico vehicular.	6
Figura 7: Diagrama de conversión de archivos XML a estructuras de tipo Graph	14
Figura 9: Entradas y salidas del simulador SUMO (adaptado de [4])	22
Figura 10: Ejemplo de simulación con SUMO de la ciudad de Barcelona.	23
Figura 11: Resultados para la prueba de escalabilidad basada en la carga de tráfico vehicular para la versión paralelizada de SUMO para 1 y 4 particiones.	27
Figura 12: Resultados para la prueba de escalabilidad basada en la carga de tráfico vehicular para la versión paralelizada de SUMO para 8 y 16 particiones.	28
Figura 13: Resultados para la prueba de escalabilidad basada en la carga de tráfico vehicular para la versión paralelizada de SUMO para 32 y 64 particiones.	28
Figura 14: Comparación de escalabilidad de simulaciones con su versión secuencial.	29
Figura 15: Comparación del <i>speedup</i> para distintos conjuntos de particiones.	33
Figura 16: Comparación de la eficiencia de paralelización para distintos conjuntos de particiones.	33
Figura 17: Comparación de escalabilidad de uso de CPU por simulaciones ejecutadas con diferentes números de particiones.	34

Capítulo 1

Introducción

El modelamiento de tráfico vehicular urbano por medio del uso de *software* de simulación resulta ser de gran relevancia al momento de proveer información necesaria para diseños y decisiones en torno a la movilidad de distintos agentes dentro de un área determinada. En particular, se puede considerar crucial para el desarrollo de *Gemelos Digitales* de ciudades y áreas metropolitanas. Un *Gemelo Digital* consiste en «una representación de un producto activo o un sistema de productos-servicios único que incluye una selección de características, propiedades, condiciones y comportamientos por medio de modelos, información y datos a través de una o múltiples fases de su ciclo de vida» [5]. Es decir, un *Gemelo Digital* consiste en una representación en tiempo real de una entidad física en un ambiente digital [6]. En este contexto, el presente trabajo se desarrolla bajo la colaboración entre NIC Research Labs y el equipo de Data Analytics and Visualization Group del Barcelona Supercomputing Center (BSC) para la creación de una plataforma generadora de *Gemelos Digitales* para las ciudades de Kobe (Japón) y Barcelona (España), los cuales tienen por objetivo contribuir a la toma de decisiones para la implementación de políticas de urbanismo en las ciudades, con el fin de mejorar la calidad de vida mediante la ejecución de cambios estructurales en éstas.

En dicho sentido, el uso de estas herramientas de simulación cobra especial importancia al momento de querer obtener cierto nivel de precisión en el modelamiento de diferentes escenarios, tales como puntos de congestión y flujo de tráfico en horas críticas para la movilidad de las personas. Sin embargo, el obtener simulaciones detalladas a grandes escalas implica un alto costo en cuanto a recursos computacionales, por lo que durante los últimos años se ha buscado desarrollar métodos de distribución y paralelización de estos procesos, con el objetivo de optimizar los recursos disponibles para simular, por ejemplo, áreas metropolitanas grandes y concurridas.

El problema propuesto aborda el estudio de la escalabilidad de estas simulaciones, con el objetivo de proponer una solución que mejore su *performance* y el uso de recursos computacionales, en especial para escenarios que contemplen una alta carga de tráfico

vehicular urbano y de esta manera poder implementar un sistema de simulaciones para dispositivos de memoria compartida tales como supercomputadores, facilitando su uso para el desarrollo de modelos tales como *Gemelos Digitales*. En un estado preliminar de este trabajo, se realiza un estudio de la escalabilidad de simulaciones sin paralelizar, el cual muestra que los tiempos de ejecución de estas simulaciones crecen de una manera exponencial. Adicionalmente, se muestra que el uso de CPU en cada simulación se mantiene constante en un 100%; es decir, cada una de las simulaciones ejecutadas utiliza la totalidad de los recursos asignados, lo cual, en adición al crecimiento de los tiempos de simulación, es inconveniente para la implementación de modelos a gran escala y con alta demanda de tráfico.

Dada esta problemática, la presente tesis aborda la implementación de una arquitectura de paralelización y sincronización para simulaciones de tráfico vehicular urbano orientada a su uso en supercomputadores, con el objetivo de estudiar la escalabilidad de simulaciones microscópicas (es decir, de alta granularidad), las cuales son realizadas mediante el *software* SUMO, para áreas urbanas extensas y con alta carga de tráfico vehicular. Para esto, se propone el uso de herramientas orientadas a High-Performance Computing, tales como *Singularity Containers* y *Message Passing Interface* (MPI), lo cual permite tanto la contenerización de procesos y aplicaciones complejas para su uso en supercomputadores, como la comunicación y sincronización de dichos procesos. En este caso, este trabajo se encuentra orientado a su implementación en el supercomputador *MareNostrum*, ubicado en el Centro Nacional de Supercomputación de Barcelona.

El proyecto se dividió en tres etapas, a ser: la partición y contenerización de las simulaciones, la paralelización de las mismas, y la determinación de los parámetros necesarios para la comunicación entre los procesos de simulación. En las tres etapas se hizo uso de *Singularity Containers*, mientras que en la segunda y tercera se sumó el uso de *OpenMP*.

El presente informe se divide en siete capítulos, a ser:

- **Problema:** donde se plantea el problema de investigación y se presentan los resultados de la primera prueba de escalabilidad de SUMO.
- **Preguntas de investigación, hipótesis y objetivos:** donde se postulan las preguntas de investigación, de las cuales se desprenden la hipótesis y los objetivos de este trabajo.
- **Solución propuesta:** donde se detalla la creación de la solución propuesta, su desarrollo, avances y los impedimentos presentados en el camino.
- **Marco teórico:** donde se profundizan y estudian materias relacionadas a la solución propuesta, tales como distintos simuladores de tráfico existentes y trabajos orientados al desarrollo de simulaciones en ambientes de *High-Performance Computing*.
- **Resultados:** donde se presentan los resultados obtenidos de la medición de los tiempos de simulación para la solución propuesta, su *speedup* y su eficiencia.
- **Discusión:** donde se analizan los resultados obtenidos, considerando detalles importantes del proceso de paralelización.

- **Conclusiones:** donde se concluye el trabajo y se da cuenta de los posibles trabajos futuros a desarrollar a partir de éste.

Capítulo 2

Problema

2.1. Pruebas de escalabilidad en SUMO

En un estudio preliminar al desarrollo de este trabajo de tesis, se implementó y ejecutó un conjunto de distintas pruebas de escalabilidad para observar el crecimiento de los tiempos de ejecución de SUMO en un computador personal. En particular, dos pruebas de escalabilidad fueron realizadas considerando dos aspectos importantes de los requerimientos para la creación de simulaciones de tráfico urbano: en primer lugar, se considera el tamaño de la red de caminos en número de aristas del grafo que representa el mapa, mientras que para el segundo se toma en consideración el número de agentes o vehículos por simulación.

Ambos experimentos se encuentran implementados en el lenguaje de programación C++ y fueron testeados en un computador personal con Intel® Core™ i7-106G7 CPU @1,30GHz. Para cada test se asignó únicamente una CPU, con el objetivo de mantener la simplicidad de los resultados. Adicionalmente, sólo fue considerado un tipo de agente para las simulaciones (automóviles) y cada simulación representa 1 hora de tráfico continuo en la ciudad de Barcelona, España, cuya área es de aproximadamente 101,9 km².

2.1.1. Prueba de escalabilidad basada en el tamaño de la red para computadores personales

Para este experimento, se implementa una forma automatizada para obtener los datos necesarios desde OpenStreetMap (OSM). Esta automatización toma una localización descrita en coordenadas, y consulta los datos requeridos a través de 30 iteraciones, donde el tamaño de la red es más grande con cada iteración.

Luego, se crean las simulaciones correspondientes mediante el *software* SUMO. Los archivos que describen los viajes, es decir, los puntos de partida y llegada de cada vehículo, fueron creados con *randomTrips.py* [7] por simplicidad, haciendo uso de una semilla de aleatoriedad definida; en tanto, los archivos correspondientes a la descripción de las rutas fueron creados mediante el uso de la herramienta *duarouter* de SUMO [8]. Finalmente, se midió el tiempo de ejecución de cada simulación versus el número de aristas del grafo correspondiente. El *pipeline* para la generación de estas simulaciones se muestra en la siguiente figura:

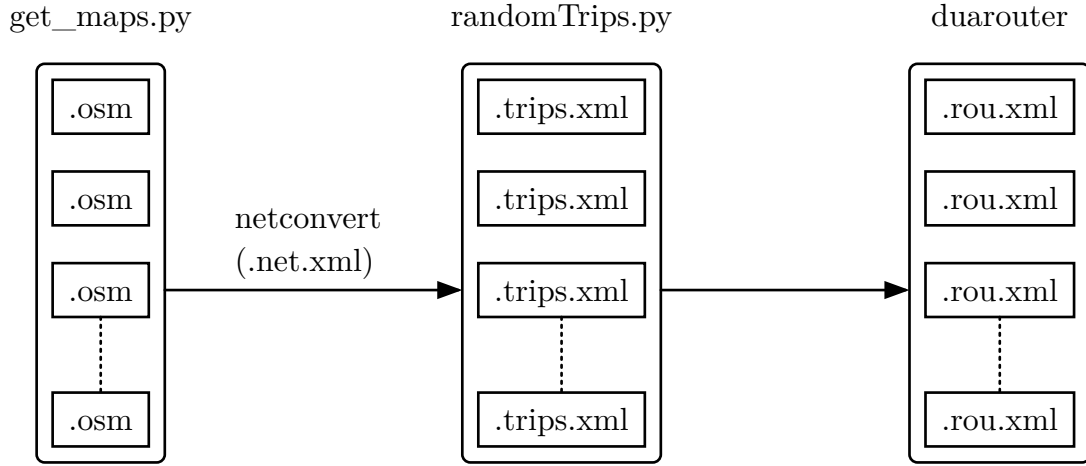


Figura 3: *Pipeline* para simulaciones basadas en el tamaño de la red.

2.1.2. Prueba de escalabilidad basada en el número de agentes por simulación para computadores personales

Por otro lado, en orden de testear la escalabilidad de las simulaciones respecto al número de agentes, se determina una red de caminos de tamaño fijo sobre la ciudad de Barcelona y se crean las simulaciones correspondientes con *randomTrips.py* y *duarouter*. El *pipeline* para este experimento se muestra en la Figura 4.

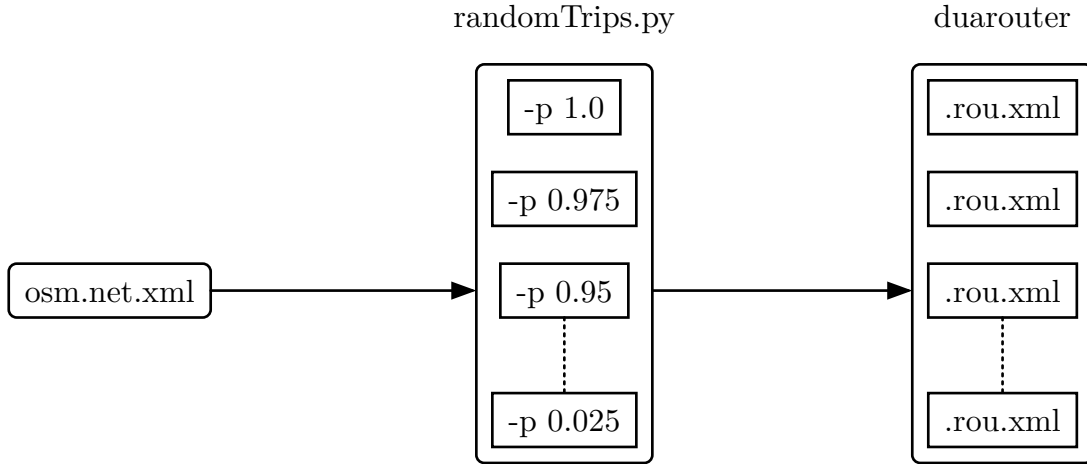


Figura 4: *Pipeline* para simulaciones basadas en la carga de tráfico vehicular.

Para poder incrementar de manera uniforme la cantidad de vehículos presentes en la simulación, se configuró la tasa de inserciones de agentes de *randomTrips.py* [9] para tomar valores decrecientes en el intervalo $[0.025, 1.0]$, mientras que la configuración de *duarouter* no fue alterada.

En total, 40 simulaciones con diferentes tasas de inserción fueron realizadas, midiendo el tiempo de ejecución de cada una. El tiempo total destinado a ejecutar todas estas simulaciones fue de aproximadamente 48 horas cronológicas.

2.1.3. Resultados

Los resultados obtenidos para ambas pruebas se muestran en los gráficos expuestos a continuación. A la izquierda, se muestran los tiempos de simulación en función del número de aristas de la red, mientras que a la derecha, se muestra el uso de CPU por cada simulación ejecutada:

Resultados para simulaciones basadas en el tamaño de la red

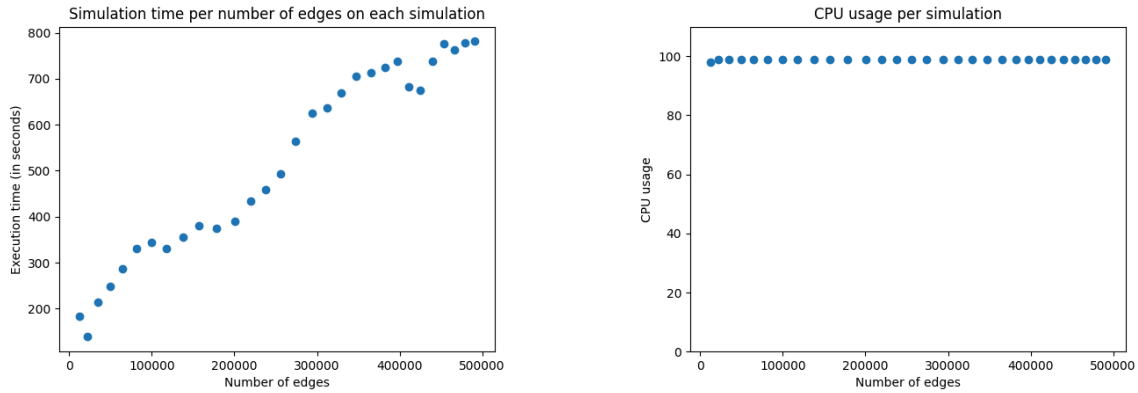


Figura 5: Resultados para la prueba de escalabilidad basada en el tamaño de la red de caminos.

Como es posible observar, en este caso existe un crecimiento lineal de los tiempos de simulación en relación al crecimiento de la red de caminos, lo cual nos muestra que dicho crecimiento no es realmente un problema al momento de ejecutar simulaciones sobre áreas extensas. Aún así, cabe destacar que el uso de CPU se mantiene constante en un valor máximo.

Resultados para simulaciones basadas en la carga de tráfico vehicular

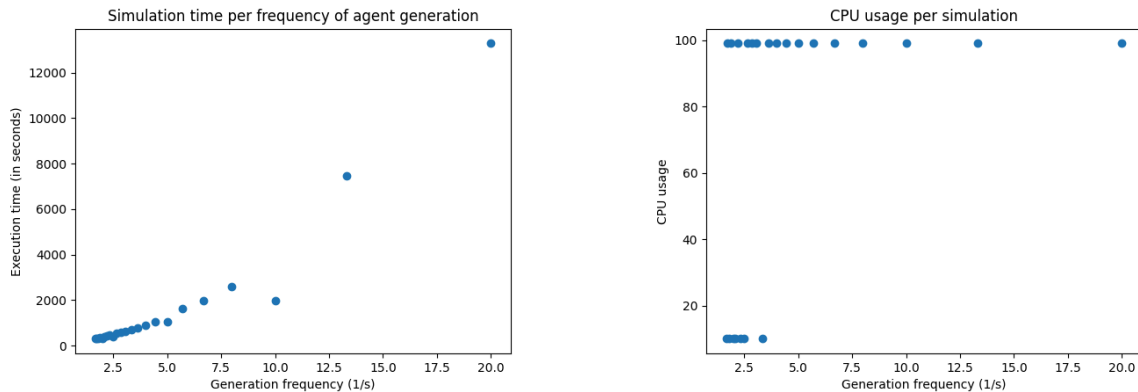


Figura 6: Resultados para la prueba de escalabilidad basada en la carga de tráfico vehicular.

En este caso, se puede observar un crecimiento exponencial de los tiempos de ejecución de cada simulación en relación a la frecuencia de inserción de vehículos en ésta. Es relevante mencionar que, a pesar de aumentar los tiempos de manera exponencial, las simulaciones no realizaron uso de la memoria *swap* del computador, lo que deja en evidencia que el problema principal para la ejecución de las simulaciones en términos de procesamiento refiere a la carga de vehículos o agentes dentro de la simulación más que el tamaño de la red, y sugiere la necesidad de paralelización de estos procesos. Por otro lado, se obtiene que el uso de CPU también se mantiene constante en un valor máximo para la mayoría de los casos.

Como es posible observar a partir de lo ya expuesto, existen principalmente dos dificultades al momento de buscar mejorar la *performance* de las simulaciones de tráfico urbano. La primera de ellas, y la más importante, refiere a la cantidad y variedad de agentes presentes en cada simulación. Dado que de la ejecución secuencial se tiene que los tiempos de ejecución aumentan de forma exponencial, al momento de querer simular redes de caminos con alta congestión de tránsito, el tiempo de ejecución de la simulación se alejará cada vez más de lo que se desea como una simulación en tiempo real.

La segunda dificultad a considerar tiene que ver con la paralelización misma y la sincronización de los procesos paralelos, dado que existe la posibilidad de generar cuellos de botella en las comunicaciones entre nodos y de esta forma perjudicar la *performance* de las simulaciones.

El problema que se plantea a partir de esto radica en la necesidad de paralelizar simulaciones que no sólo representen grandes áreas metropolitanas, sino que también presenten una alta carga de tráfico de agentes (ya sea vehicular o peatonal), manteniendo la granularidad y precisión de dichas simulaciones en un ambiente de supercomputación y mejorando el *speedup* de éstas de manera significativa, reduciendo también los posibles cuellos de botella en la comunicación de los procesos paralelizados.

Capítulo 3

Preguntas de investigación, hipótesis y objetivos

3.1. Preguntas de investigación

- ¿Cómo se diferencian en cuanto a *performance* la versión secuencial de simulaciones en SUMO respecto a una versión paralelizada en un ambiente de supercomputación?
- ¿Es posible mejorar la *performance* de simulaciones de tráfico urbano en orden de poder simular áreas metropolitanas completas a nivel microscópico?
- ¿Cómo se comporta asintóticamente la influencia del área a simular en la *performance* de las simulaciones de tráfico urbano con SUMO?
- ¿Cómo se comporta asintóticamente la influencia de la carga vehicular a simular en la *performance* de las simulaciones de tráfico urbano con SUMO?

3.2. Hipótesis

Es posible aumentar la escalabilidad de simulaciones de alta granularidad mediante la paralelización y sincronización de los procesos del *software* SUMO en un ambiente de supercomputación, incrementando el *speedup* de éstos en al menos un 5% de los tiempos de simulación secuenciales, y optimizando así el uso de recursos para simulaciones que contemplen áreas metropolitanas de gran extensión y alta carga de tráfico vehicular.

3.3. Objetivos

3.3.1. Objetivo General

Avanzar en el estado del arte acerca de la paralelización de procesos de simulación de tráfico urbano en supercomputadores, comprobando el aumento en la escalabilidad de dichos procesos luego de su correspondiente paralelización.

3.3.2. Objetivos específicos

1. Aplicar la contenerización de SUMO en *Singularity Containers* para su ejecución en supercomputadores.
2. Diseñar e implementar un modelo de paralelización de los procesos contenerizados, determinando los parámetros necesarios para la efectiva comunicación y sincronización entre los procesos.
3. Medir la escalabilidad de la solución implementada en un ambiente de supercomputación y comparar con las mediciones previamente realizadas.

Capítulo 4

Solución propuesta

La paralelización de procesos es una herramienta ampliamente utilizada para la optimización de operaciones en computación. Mediante la división de tareas en tareas más pequeñas que se ejecutan de manera simultánea, es posible reducir los tiempos de ejecución.

Durante los últimos años, han surgido distintas técnicas y herramientas que podrían facilitar las simulaciones de tráfico en paralelo aplicado a ambientes de High-Performance Computing, tales como supercomputadores. Algunas de estas son:

1. **Paralelización híbrida:** consiste en la combinación de múltiples técnicas de paralelización, incluyendo estándares de transmisión de mensajes para computación paralela tal como lo es MPI (*Message Passing Interface*) u *OpenMP* para dispositivos de memoria compartida, con el objetivo de alcanzar una mejor *performance* y escalabilidad. Esto permite una distribución más granular de tareas de simulación a través de múltiples nodos.
2. **Particionamiento de modelos:** refiere a la partición de modelos de simulación en submodelos que pueden ser ejecutados en paralelo, lo cual puede ayudar a reducir los costos de comunicación y mejorar la escalabilidad de las simulaciones.
3. **Simulaciones cloud-based:** implica la ejecución de modelos de simulación en Cloud Computing, lo que ofrece la oportunidad de tener simulaciones más flexibles y escalables. Sin embargo, la latencia de la red y las posibles amenazas en torno a la seguridad de los datos son desventajas que esta opción presenta.

Un estudio que aborda el problema de la paralelización de SUMO refiere al estudio llevado a cabo por Arroyave et. al. [10], en donde se presenta un modelo de particiones de mapa simplificadas para la simulación en paralelo de tráfico en zonas urbanas. Este modelo consiste en la división de las redes de caminos en forma de cuadrículas, lo cual incrementa el nivel de error de la simulación de manera proporcional con el número de particiones realizadas. Además de esto, dicho estudio señala que el método de paralelización presentado es más lento que su contraparte centralizada dado el proceso de sincronización

implementado. Esto plantea la necesidad de abordar el problema desde una perspectiva que considere un esquema de partición de grafos más eficiente, que permita distribuir mejor la carga vehicular en los nodos de simulación.

Por otro lado, el trabajo titulado «QarSUMO: A Parallel, Congestion-optimized Traffic Simulator» [11], presenta una versión paralelizada de SUMO considerando la complejidad de la distribución de las redes de caminos por medio de la implementación de un esquema de partición de grafos no-regulares, abreviado como METIS [12]. Si bien este estudio logra establecer un método de paralelización de alto nivel, incrementando la eficiencia de las simulaciones para situaciones de alta congestión de tráfico, su alcance en cuanto a áreas urbanas no alcanza una escalabilidad a nivel de simulaciones completas al no considerar su implementación en ambientes de *High-Performance Computing*. Sumado a esto, el *software* disminuye el nivel de granularidad de las simulaciones por medio de la simplificación del modelamiento de un gran número de vehículos a una sola entidad. Esto plantea, a su vez, la posibilidad de obtener un mejor escalamiento y una mejor tasa de granularidad en las simulaciones teniendo una mayor capacidad de cómputo, como es el caso de los supercomputadores.

En cuanto a la escalabilidad de las simulaciones en un contexto de paralelismo, si bien existen limitaciones respecto a cuánto se puede llegar a aumentar el *speedup* de una tarea mediante la paralelización, tal como dicta la Ley de Amdahl, la cual señala que la reducción en los tiempos de ejecución de un programa se verá limitado por aquellas porciones de programa que no son posibles de paralelizar, la paralelización de procesos resulta ser útil al considerar la disponibilidad de múltiples unidades de cómputo y tareas con alta posibilidad de concurrencia.

En particular, los procesos de simulación de tráfico vehicular urbano consumen una gran cantidad de recursos y tiempos de ejecución al correr de forma secuencial. Dado que la escalabilidad de dichas simulaciones depende de la carga de vehículos que ésta tenga, es posible pensar en una solución que distribuya esta carga a través de múltiples nodos que computen simultáneamente rutas divididas según zonas geográficas, de manera que cada nodo maneje una zona distinta, y se comuniquen con los nodos correspondientes mediante una entidad central que gestione la salida y entrada de vehículos a distintas particiones. Es importante, en este caso, considerar el *overhead* que dichos procesos de comunicación implican, dado que plantean casos de exclusión mutua que serializan parte de las tareas y, por lo tanto, limitan su *speedup*.

A raíz de lo ya expuesto, es posible extraer la posibilidad de implementar un modelo de paralelización orientado a su uso en dispositivos de alta disponibilidad, solucionando el problema de la escalabilidad, y haciendo uso de herramientas tales como algoritmos de partición de grafos no-regulares para distribuir apropiadamente la carga vehicular, de manera que se optimice el uso de recursos y no se arrastre un margen de error proporcional

al número de particiones realizadas para las simulaciones a ejecutar, conservando un mayor porcentaje de precisión para éstas.

4.1. Diseño de la solución

La solución propuesta al problema planteado contempla el desarrollo e implementación de un modelo de paralelización para simulaciones de tráfico vehicular urbano a partir de la contenerización de instancias del *software* SUMO. Para esto, se trabajó con el lenguaje C++ y las librerías *libtraci* y *OpenMP*.

El diseño considera el levantamiento de distintos nodos de ejecución, al cual se les asigna el mapa a simular y las rutas correspondientes a su partición, y, mediante una instancia contenerizada de SUMO, también asignada al nodo, se ejecuta la simulación a través de estas particiones que corren en paralelo. Es importante destacar la necesidad de la contenerización de SUMO, dado que cada nodo necesita su propia instancia del *software* para poder funcionar sin problemas de lectura/escritura en dispositivos de memoria compartida.

El particionamiento de los mapas a utilizar para las simulaciones contempla el uso del algoritmo de particionamiento de grafos *SPartSim* [13], mientras que el posterior corte de las rutas para cada partición se realiza a partir de la herramienta *cutRoutes.py*, provista por el *software* SUMO.

Para medir la eficiencia de la solución implementada, se propone el diseño e implementación de tests de carga que midan el tiempo de ejecución y uso de CPU para simulaciones paralelizadas con distintos *sets* de particiones, con el objetivo de comparar estos resultados con los previamente obtenidos y mostrados en la Sección 2.1 para simulaciones ejecutadas de manera secuencial.

Dado el contexto en el que este trabajo se inserta, los mapas utilizados para las pruebas de concepto y la posterior ejecución de los test de carga corresponden a las áreas de Barcelona y Viladecans, ambas ciudades de la región de Cataluña, España.

4.2. Herramientas utilizadas

4.2.1. Singularity Containers

Singularity es un proyecto *open-source* que permite la creación de contenedores de forma portable y reproducible [14]. A diferencia de *Docker*, *Singularity* fue creado para ejecutar aplicaciones complejas en *clusters* HPC, cobrando popularidad de manera significativa para su uso en supercomputadores; sin embargo, es posible construir contenedores de *Singularity* a partir de contenedores de *Docker*.

Adicionalmente, este *software* soporta la generación de imágenes *custom* a través de sus archivos de definición, permitiendo envolver todas las imágenes necesarias en un sólo archivo que provee facilidad de manejo, además de simplificar el despliegue de los conte-

nedores [15]. La interoperabilidad que ofrece *Singularity* permite también su portabilidad a través de distintas arquitecturas HPC.

4.2.2. *OpenMP*

OpenMP es una API de modelo escalable que ofrece una interfaz flexible y sencilla de manejar para programación paralela [16]. Provee soporte para los lenguajes C, C++ y Fortran, y permite el desarrollo de aplicaciones portables, encontrándose orientado principalmente a la programación para multiprocesadores de memoria compartida.

Esta API consiste en una especificación para un set de directivas de compilación, rutinas de librería y variables de ambiente que pueden ser usadas para especificar paralelismo de alto nivel. Actualmente, se encuentra en la versión 5.2.

4.2.3. SUMO y *libtraci*

La librería *libtraci* de SUMO es una herramienta que compatibiliza el uso de TraCI con códigos de C++ [17], de manera que no sólo se pueda ejecutar una simulación, sino que también pueda controlarse y monitorearse sin necesidad de una GUI. *Libtraci* provee, además de las funcionalidades de TraCI, soporte para el uso de múltiples clientes, lo cual resulta conveniente al momento de pensar en la paralelización de las simulaciones.

4.3. Partición de mapas de Barcelona y Viladecans

Para realizar la división de los mapas originales de Barcelona y Viladecans se utilizó una implementación en Java ya existente del algoritmo de *SPartSim* [13], el cual ejecuta una división por zonas geográficas eligiendo puntos distantes en el mapa dada una heurística inicial y va expandiendo las particiones con los nodos cercanos en cada iteración, para luego balancear la carga de las particiones moviendo aquellos nodos con más peso a particiones vecinas que tengan un menor peso. El objetivo de esto es entregar un mapa particionado de manera relativamente equitativa, de forma que los hilos de ejecución de cada sub-simulación no se vean sobrecargados.

En orden de mantener la compatibilidad con los archivos de entrada de las simulaciones con SUMO, a dicha implementación se le agregó un módulo de compatibilidad con archivos XML, ya que el código original sólo trabajaba con archivos GeoJSON, los cuales además perdían información importante al momento de convertir desde archivos XML. A raíz de esto, se debió implementar una clase que conservase dicha información, para después restaurarla al momento de hacer la conversión desde los grafos generados por el *software* hacia la clase de grafos compatibles con la estructura XML requerida por las simulaciones. La siguiente figura ilustra el proceso de conversión implementado para la compatibilidad del *software* de particionamiento de rutas con archivos XML.

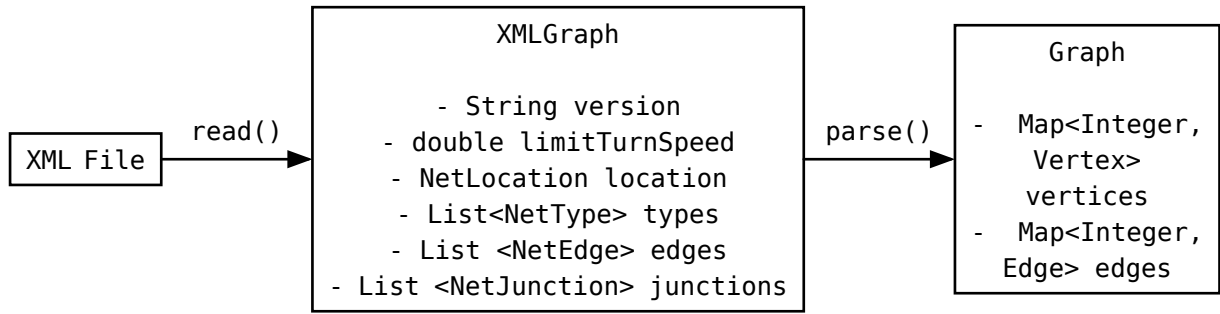


Figura 7: Diagrama de conversión de archivos XML a estructuras de tipo Graph.

Como es posible observar, la estructura `XMLGraph` conserva del mapa original atributos tales como la localización, los límites de velocidad y la versión del archivo, los cuales se mantienen al momento de realizar la conversión de las particiones desde la estructura `Graph` a `XMLGraph` para su posterior escritura en los distintos archivos XML correspondientes a cada partición.

4.4. Generación y partición de rutas vehiculares

Para la generación de las rutas vehiculares se consideraron, en primer lugar, dos enfoques distintos: por un lado, el generar dichas rutas a partir de datos reales de movilidad urbana, y por otro lado, generarlas de manera aleatoria o pseudo-aleatoria por medio de la herramienta *randomTrips.py*, provista por SUMO y utilizada para realizar las pruebas de escalabilidad de las simulaciones secuenciales.

Respecto al primer enfoque, se obtuvieron los datos de movilidad de la población de Barcelona en formato de matrices de Origen-Destino obtenidas a partir de los datos recolectados de las redes móviles por parte del ayuntamiento de Barcelona. Sin embargo, dada la complejidad implicada por la conversión de dichas matrices a los archivos XML que describen las rutas correspondientes, en contraste con la simplicidad de generar las rutas de manera pseudo-aleatoria con una herramienta ya integrada en el *software*, y considerando los objetivos planteados, se decidió tomar el segundo enfoque.

Para la partición de las rutas generadas se hizo uso de la herramienta *cutRoutes.py* de SUMO, la cual corta las rutas acorde a la partición del mapa que se busca simular. No obstante, dicha herramienta presenta problemas en tanto termina descartando ciertas rutas particionadas y recalculando los tiempos de partida para cada una; si bien existe una extensión que soluciona este problema para la futura paralelización de los procesos de simulación [18], no fue posible acceder a ella, por lo que sólo se consideraron aquellas rutas resultantes de *cutRoutes.py* para la posterior reconstrucción.

4.5. Contenerización de simulaciones secuenciales

En una primera instancia de exploración del uso de *Singularity Containers*, se procedió a implementar la contenerización de simulaciones sin paralelizar. El resultado de esto fue la creación de un archivo de definición de *containers* con extensión `.def`, el cual construye el entorno necesario para la ejecución de las simulaciones de manera aislada. A continuación, se muestra un ejemplo de estos archivos de definición para el levantamiento de los contenedores.

```
Bootstrap: debootstrap
OSVersion: jammy
From: ubuntu:22.04
MirrorURL: http://us.archive.ubuntu.com/ubuntu/

%files
    partition.net.xml /home/
    viladecans.poly.xml /home/
    partition.rou.xml /home/
    sim.sumocfg /home/
    traci_simulation /home/

%post
    chmod 755 /root
    apt update
    apt-get install -y software-properties-common build-essential python3
    add-apt-repository universe
    add-apt-repository multiverse
    apt-get update
    apt-get install -y sumo sumo-tools sumo-doc

%environment
    export SUMO_HOME=/usr/share/sumo

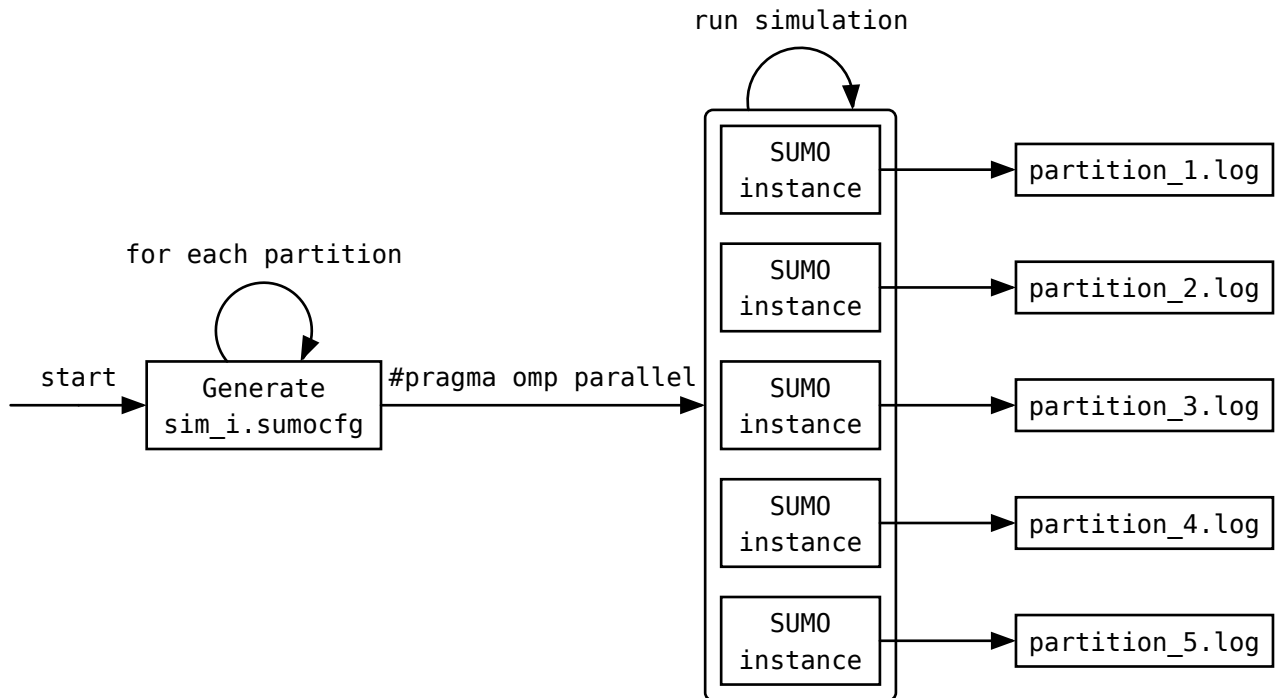
%runscript
    cd /home
    ./traci_simulation
```

El *script* de C++ *traci_simulation* tiene por objetivo ejecutar simulaciones aisladas de forma secuencial. Para esto, luego de configurar el entorno de la simulación, el código llama a una instancia de SUMO con el archivo de configuración previamente definido a partir de un *script* en Python, el cual genera un archivo XML especificando, entre otras cosas, cuáles son los archivos a utilizar para las redes de caminos y las rutas de los vehículos.

4.6. Paralelización de las simulaciones particionadas

La paralelización de las simulaciones particionadas consistió en la implementación de un código en C++ con la ayuda de *OpenMP* para las directivas de compilación. Este código implementa una arquitectura de *multi-threading*, en la cual cada hilo de ejecución posee una instancia de SUMO contenerizada que simula una partición dada del mapa original, con las rutas propiamente particionadas de acuerdo con el algoritmo de *cutRoutes.py*.

En la siguiente figura, se muestra el funcionamiento general de la paralelización realizada con 5 *threads*. Primero se realiza un proceso de configuración de forma secuencial, en el que se generan los archivos de configuración *sim.sumocfg* a partir de los cuales se ejecutan las simulaciones en forma paralela mediante una instancia de SUMO contenerizada con *Singularity*. De esta manera, una vez que se crean los archivos de configuración, se realiza un llamado con el comando `#pragma omp parallel`, configurando el número de threads a ejecutarse como el número de particiones hechas previamente con *SPartSim*.



4.7. Sincronización de las simulaciones particionadas

4.7.1. Reensamblaje de rutas

En cuanto a la reconstrucción de las rutas, se implementó un código en `Python` que toma tanto las rutas originales como las particionadas y, por cada ruta particionada, busca la ruta original y las agrupa en un diccionario indicando sus aristas, la partición a la que pertenece la ruta, y su índice de ocurrencia en la ruta original. De esta manera, fue posible realizar una reconstrucción parcial de los viajes generados por *randomTrips.py*.

El resultado de esta reconstrucción se guarda finalmente en un archivo `.json`, con la estructura que se muestra a continuación:

```
{
  "routes": [
    {
      "original_route": str,
      "cut_routes": [
        {
          "partition": int,
          "id": str,
          "cut_route": str,
          "next_partition": int,
          "next_route": str
        },
        ...
      ]
    },
    ...
  ]
}
```

4.7.2. Comunicación entre nodos

Dado que la implementación sigue un enfoque para *shared-memory devices*, la comunicación entre los nodos de ejecución se realizó mediante la implementación de una cola global en la cual, a cada paso de la simulación, cada hilo escribe los vehículos salientes de su partición con la partición a la que corresponde insertarse después, mientras que al paso siguiente lee e inserta en la simulación aquellos vehículos cuya entrada corresponde a su partición. Para esto, se hace uso de exclusión mutua por medio de la implementación de secciones críticas para la escritura en la cola de tráfico, mientras que para la adición o remoción de vehículos en cada partición se hace uso de las herramientas provistas por *libtraci*.

4.7.3. Impedimentos para la sincronización

Como se menciona en la Sección 4.4, la herramienta de particionamiento de rutas *cutRoutes.py* presenta limitaciones al momento de considerarla para la sincronización de simulaciones de tráfico urbano mediante SUMO, principalmente al momento de definir los tiempos de partida de los vehículos (*departure times*), redefiniendo aquellas rutas que cruzan más de una partición como rutas independientes con el mismo tiempo de partida; esto radica en problemas tales como que al ingreso de nuevas rutas, no se reconozca el vehículo a insertar dado que el mismo ya ha salido de la simulación.

4.8. Diseño e implementación de test de carga

4.8.1. Test versión secuencial

Para poder realizar la comparación en cuanto al crecimiento de los tiempos de ejecución para las simulaciones secuenciales, se implementó un módulo de test en el lenguaje C++, el que contempla una primera fase de *setup*, donde se definen los períodos de generación de vehículos a partir de los cuales se generan los archivos de rutas necesarios para el mapa definido. En total, se definen veinte períodos para los cuales se definen también los archivos de configuración a utilizar para cada simulación.

La segunda etapa contempla la ejecución de cada simulación de forma secuencial, midiendo y registrando el tiempo que toma cada una en finalizar. Para esto, cada simulación se ejecutó un total de cincuenta veces, para luego obtener el tiempo total como el promedio de los tiempos de cada iteración.

4.8.2. Test versión paralelizada

Para probar el escalamiento de la solución implementada, se siguió un procedimiento similar a los tests secuenciales, resultando en la implementación de los siguientes módulos:

- ***setup***: este módulo genera los períodos, rutas, particiones y archivos de configuración necesarios para cada simulación, y organiza los archivos en sus directorios correspondientes.
- ***start_test***: por cada set de particiones realizadas, este módulo se encarga de entrar a cada directorio e iniciar el test de carga para cada período dado.
- ***load_test***: este módulo implementa el test de carga, el cual, a partir de los archivos de configuración y la imagen de *Singularity* construida para las instancias de SUMO, ejecuta las simulaciones llamando al módulo encargado de paralelizar los procesos de simulación según el número de particiones realizadas. Tal como en el test para la versión secuencial, cada simulación se ejecuta un total de cincuenta veces, para luego promediar los tiempos de cada iteración.

Capítulo 5

Marco teórico

5.1. Simuladores de tráfico urbano

Dentro del campo de las simulaciones computacionales, se destacan las simulaciones de tráfico urbano en tanto son capaces de proveer información útil para la movilidad en distintos entornos y bajo diversas condiciones. Pablo López et. al. realizan una clasificación de los tipos de simuladores existentes según el nivel de granularidad que poseen [7]:

1. Macroscópicas, donde se modelan dinámicas a gran escala tales como la densidad total de tráfico en una ciudad.
2. Microscópicas, en las cuales se modela cada vehículo o agente como una entidad individual, con sus correspondientes interacciones.
3. Mesoscópicas, que contemplan características tanto de las simulaciones macroscópicas como microscópicas, y
4. Submicroscópicas, en las cuales no sólo se simula cada vehículo como un agente individual, sino que es posible simular dinámicas o funciones dentro de ellos.

Respecto a la variedad de simuladores de tráfico microscópicos existentes, [1] realiza una comparación entre los simuladores SUMO, Quadstone Paramic Modeler, Treiber's Microsimulation of Road Traffic, Aimsun, Trafficware SimTraffic y CORSIM TRAFVU, considerando entre sus criterios de comparación:

- Apertura de código (*open source*)
- Portabilidad de sistema operativo
- Documentación
- Escalabilidad
- *Performance*

En cuanto al criterio de apertura de código, solamente SUMO y Treiber's Microsimulation of Road Traffic resultaron ser *open source*. Esto resulta un criterio importante a considerar, dado que al poder ser modificado por otros programadores, el *software* tiene el potencial de poder paralelizar modelos de simulación y desarrollar paquetes orientados a la alta disponibilidad. Ambos simuladores también coinciden en la portabilidad que ofrecen para

sistemas operativos Linux, mientras que el resto de *software* ofrece únicamente funcionamiento para sistemas Windows.

Respecto a la disponibilidad de documentación, la mayoría de los simuladores resultaron poseer documentación disponible al usuario, mientras que en términos de escalabilidad y performance, SUMO resulta ser el simulador con mayor capacidad de soportar grandes redes de caminos

Adicionalmente, Ejército et. al. [2] realiza una comparación entre los simuladores Matsim, SUMO, Aimsun y PTV Vissim utilizando los mismos criterios ya descritos. De estos trabajos es posible extraer la siguiente tabla comparativa entre todos los *software* presentados:

Simulador	Apertura de código	Portabilidad de SO	Documentación	Escalabilidad
SUMO	✓	Linux, Windows, MacOS	✓	Alta
Treiber's Micro-simulation	✓	Linux, Windows, MacOS	✓	Baja
Trafficware Sim-Traffic		Windows	✓	Media
CORSIM TRAF-VU		Windows	✓	Baja
Matsim	✓	Linux, Windows	✓	Media
Aimsun		Linux, Windows, MacOS		Alta
PVT Vissim		Windows	✓	Alta

Tabla 1: Comparación entre simuladores de tráfico vehicular (adaptada de [1] y [2])

En cuanto a los aspectos de *performance*, se presenta la siguiente tabla que reúne la información obtenida de [1] y [2]:

Simulador	Uso de CPU	Uso de memoria
SUMO	30-40%	12-16 MB para redes promedio
Paramics Modeler	50%	40-140 MB
Aimsun	30-40% dependiendo del número de vehículos	300-400 MB dependiendo de la red de tráfico
SimTraffic	50%	35 MB
CORSIM TRAFVU	50%	28-31 MB dependiendo de la red de tráfico
Vissim	50-60%	720 MB

Tabla 2: Comparación entre *performance* de simuladores de tráfico vehicular (adpatada de [1] y [2])

De ambas tablas presentadas, es posible concluir que SUMO parece ser una opción eficiente en cuanto a escalabilidad y performance, sumando a esto la capacidad de paralelización que implica su apertura de código y la amplia documentación que posee para el usuario.

Por otro lado, y en una comparación más detallada, Diallo et. al. [3] realiza un *benchmarking* entre los *software* Matsim, SUMO, Aimsun Next, PTV Vissim y GAMA según los criterios de naturaleza del *software*, creación de redes de caminos y demanda de transporte, realismo en las simulaciones, documentación e interfaz de usuario (GUI), y especificaciones del modelador. Respecto a este último aspecto, se desprende la siguiente tabla comparativa:

Simulador	Modelo micro/meso	Escala-miento	Salida de estadísticas	Intermodalidad	Calibración	API	Acceso a código fuente
SUMO	✓	✓	✓	✓	✓	✓	✓
Matsim	✓	✓	✓	✓	✓		✓
Aimsun Next	✓	✓	✓	✓	✓		
PTV Vissim	✓	✓	✓	✓	✓		
GAMA	✓	✓	✓				✓

Tabla 3: Comparación entre características de simuladores de tráfico vehicular (Adaptada de [3])

Dos aspectos que además destacan la diferencia entre los simuladores Matsim y SUMO radican en el modelamiento de redes multimodales y la simulación de interacciones microscópicas tales como los cruces de peatones, los cuales son atributos importantes a considerar al momento de querer recrear entornos fieles a la realidad. En este sentido, SUMO presenta ventaja sobre Matsim dado que las redes multimodales se encuentran mejor modeladas

con este *software*, además de que SUMO sí presenta la capacidad de simular interacciones microscópicas en sus simulaciones.

5.2. Simulation of Urban MObility (SUMO)

Simulation of Urban MObility (SUMO) es un *software* consistente en una herramienta de simulación de tráfico continuo y microscópico diseñado para manejar grandes redes de caminos [19] y simulaciones multiagente, modelando cada vehículo de manera explícita, con rutas individuales a través de dicha red. Produce simulaciones deterministas por defecto, aunque posee opciones para introducir aleatoriedad en éstas. Este *software* provee una plataforma para simular y visualizar variados escenarios de transporte, incluyendo transporte público y privado, además de peatones y ciclistas. SUMO se encuentra diseñado para simular las redes de caminos bajo diferentes condiciones de tráfico, tales como congestión, accidentes y cambios en la infraestructura de la red. También permite el modelamiento de diferentes estrategias de gestión de tráfico e impactos de nuevas tecnologías de transporte.

Adicionalmente, este *software* se caracteriza por ser altamente portable [4], además de poseer la capacidad de importar redes de caminos de diversas fuentes, tales como OpenStreetMap (OSM), OpenDRIVE, Matsim, Vissim y otros simuladores [7]. En cuanto a la generación de rutas, SUMO permite la obtención de éstas a través de datos reales (matrices de Origen-Destino), así como la generación aleatoria de rutas a través de su herramienta *randomTrips.py*. La siguiente figura muestra el funcionamiento básico del simulador, desde los parámetros de entrada que debe obtener hasta los posibles *outputs* que dispone.

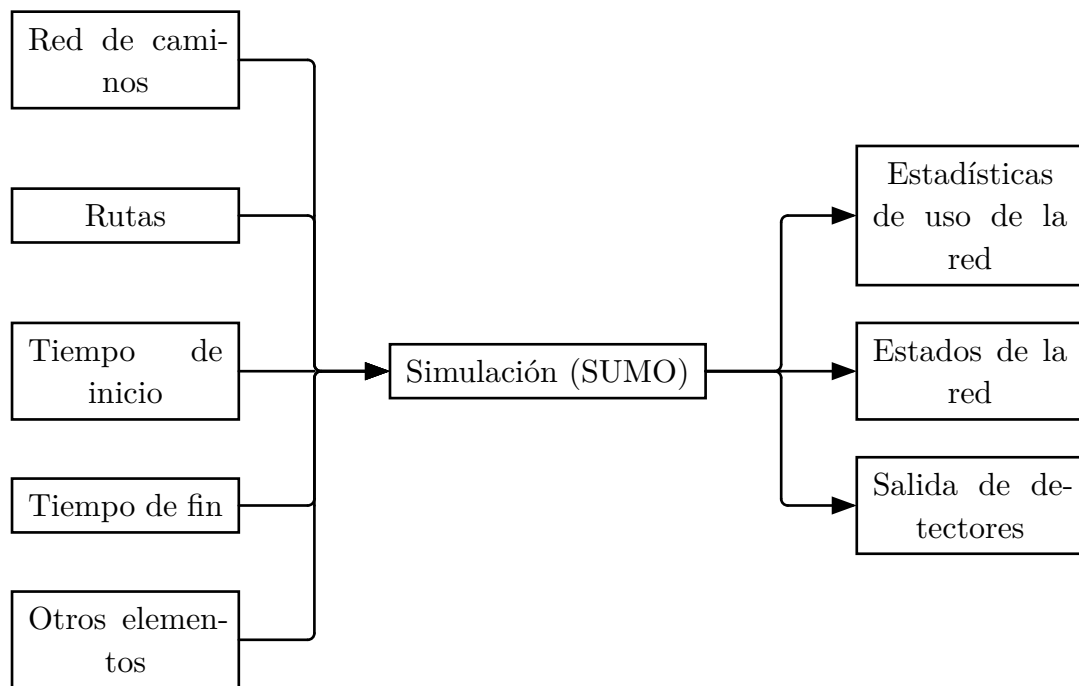


Figura 9: Entradas y salidas del simulador SUMO (adaptado de [4])

Las redes de caminos de SUMO consisten en grafos de nodos y aristas unidireccionales que representan calles y veredas. Cada arista posee una geometría descrita por un arreglo de segmentos y consisten en una o más vías corriendo en paralelo [7]. Además de estas redes, SUMO puede modelar otros elementos, tales como detectores en la red, elementos avanzados como luces de tráfico, infraestructura adicional como paradas de buses, y polígonos y puntos de interés (POI).

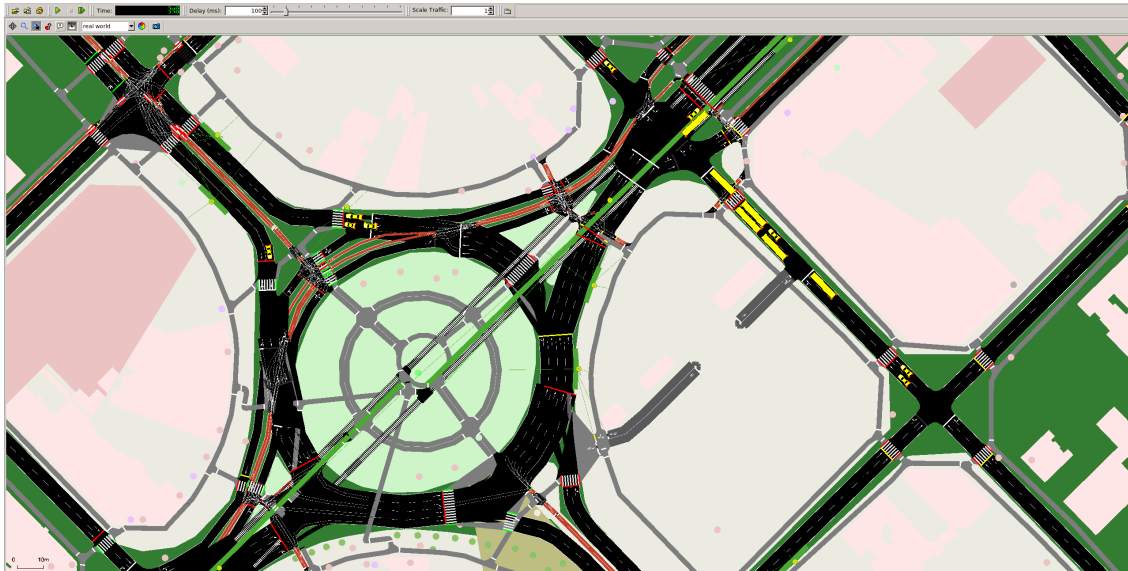


Figura 10: Ejemplo de simulación con SUMO de la ciudad de Barcelona.

Para la gestión y ejecución de las simulaciones, SUMO posee una interfaz de control de tráfico (TraCI), la cual permite obtener valores de los objetos simulados y manipular su comportamiento durante la ejecución de las simulaciones [20].

En la actualidad, SUMO también permite manejar simulaciones microscópicas de áreas extensas; diversos estudios han logrado simular ciudades tales como Luxemburgo [21] y Bologna [22]. A pesar de que estos avances en las simulaciones densas y de alta granularidad presentan sus limitaciones en cuanto a la escala de lo que se puede simular, dichos trabajos muestran que el *software* SUMO posee la capacidad de modelar entornos realistas con alta precisión.

5.3. Gemelos Digitales (*Digital Twins*) para ciudades

La revolución digital y el crecimiento global de la Internet han levantado el fenómeno de las ciudades manejadas por datos (DDC o *Data-Driven City*) y ciudades inteligentes (*Smart Cities*) [23], las cuales integran datos y tecnologías digitales para asegurar la sustentabilidad, el bienestar ciudadano y el desarrollo económico en el sector urbano.

Por otro lado, el concepto de Gemelo Digital (*Digital Twin*) conlleva el desarrollo y soporte de modelos de objetos y procesos del mundo real. En [24] se describe como «una simulación probabilística integrada multiescala de un objeto complejo que utiliza modelos

físicos, matemáticos y simulativos para obtener la representación más precisa del objeto real correspondiente, basado en el análisis de datos desde diversas fuentes».

En este contexto, un Gemelo Digital de una ciudad se vale de datos de ésta, recolectados desde distintas fuentes para modelar su estructura, dinámicas y procesos. Adicionalmente, este modelo requiere la capacidad de simular escenarios hipotéticos sobre el modelo físico, ofreciendo funcionalidades que soporten cambios de estado en la infraestructura urbana y siendo capaz de proveer soluciones mediante el análisis de los datos obtenidos de los procesos simulados.

Como se define en [23], un Gemelo Digital de una ciudad es un sistema interconectado de Gemelos Digitales que representan ciertos aspectos del funcionamiento y desarrollo del entorno urbano. Por otro lado, [25] menciona que el desarrollo de los Gemelos Digitales para ciudades ha sido relativamente rápido, y se espera que provea potencial para el mejoramiento de la gestión de las ciudades y la calidad de vida de sus ciudadanos.

Parte importante de lo anterior es el poder comprender los patrones de movilidad de la gente, lo cual puede aportar a la planificación urbana y operaciones de transporte, tales como los sistemas de control de tráfico. Para esto, se hace necesario el correcto modelamiento del tráfico urbano para el desarrollo de una réplica precisa a nivel digital.

5.4. Simulaciones de tráfico con SUMO para el desarrollo de gemelos digitales

Dentro del uso de SUMO para el desarrollo de gemelos digitales, destaca el trabajo titulado *«Getting Real: the Challenge of Building and Validating a Large-Scale Digital Twin of Barcelona's Traffic with Empirical Data»* de Javier Argota Sánchez-Vaquerizo [26], el cual busca implementar una microsimulación de tráfico basada en agentes a gran escala de la ciudad de Barcelona utilizando SUMO, para mostrar las posibilidades y desafíos de construir estos escenarios basados en grandes cantidades de datos de alta granularidad, concluyendo que si bien es posible obtener simulaciones de estos escenarios, se hace necesario obtener más y mejores métodos de recolección y análisis de datos, mientras que el desarrollo de modelos más precisos y realistas sigue siendo un desafío.

Por otro lado, [27] se enfoca en la aplicación de simulaciones microscópicas en SUMO con datos reales de alta granularidad para el Gemelo Digital de la autopista de Geneva. En este trabajo, los autores destacan el mecanismo de calibración dinámica del flujo vehicular y el re-enrutamiento dinámico por medio de TraCI, características necesarias para el modelamiento de un Gemelo Digital.

Otro resultado relevante es el que presenta [28], trabajo que utiliza Gemelos Digitales para enfrentar los desafíos de performance que presenta la integración de diversas tecnologías y plataformas de datos en torno a la movilidad vehicular. Azevedo et. al. recurren

al *software* SUMO para evaluar la eficiencia y confiabilidad de la arquitectura de datos vehiculares, destacando como resultado relevante la observación de que hubo un aumento significativo del uso de CPU durante la simulación, alcanzando valores máximos, como se puede observar en la siguiente tabla:

Tiempo de simulación	Uso promedio de CPU	Uso máximo de CPU	Uso promedio de memoria
Antes	2.3%	2.6%	1.7%
Durante	63.43%	99.9%	1.86%
Después	2.3%	2.6%	1.7%

A partir de esto, surge la necesidad de explorar vías de mejoramiento en la *performance* de las simulaciones, de manera que se optimice el uso de recursos computacionales para el funcionamiento esperado del Gemelo Digital.

5.5. Simulaciones de tráfico vehicular urbano en High-Performance Computing

Respecto al uso de simuladores de tráfico vehicular urbano en ambientes de High-Performance Computing, existe un precedente puesto por Klefstad et. al. [29] presentando ParamGrid, un framework sincronizado y escalable que distribuye una simulación a través de un *cluster* de *performance* ordinaria, compuesto por computadores personales de bajo costo. Este trabajo utiliza el simulador PARAMICS y se muestra que el *speedup* de las simulaciones aumenta de forma lineal de acuerdo al número de computadores presentes en el *cluster*. Sin embargo, y dadas las limitaciones que presenta, se señala como trabajo futuro el mejoramiento del algoritmo de partición de las redes de caminos basado en la carga computacional anticipada, además de establecer como necesario el estudio sobre la distribución dinámica de carga vehicular y la implementación de enrutamientos globales más eficientes.

Más adelante, Suzumura et. al. presentan una plataforma de simulaciones de tráfico multi-modal y altamente escalable [30]. Si bien el estudio de *performance* correspondiente demuestra que el simulador es 15.5 veces más rápido que el tiempo real con 12 *threads* en paralelo, se propone como trabajo futuro una mejora dentro de la precisión de las simulaciones por medio de la calibración de los agentes y la introducción de la búsqueda del camino más corto en relación al tiempo de viaje para los algoritmos de enrutamiento. Fuera de ello, no se hace mención a la portabilidad del simulador para otros entornos de High-Performance Computing.

Por otro lado, Umemoto et. al. realizan una serie de simulaciones implementadas en SUMO sobre la ciudad de Kobe para su ejecución en el supercomputador K [31]. Los resultados arrojan que una demanda alta de simulaciones requiere un mayor tiempo para

empezar a ejecutarse, mientras que los tiempos de simulación llegaron a ser incluso el doble de lo que toma la misma simulación en una CPU convencional.

Finalmente, durante el año 2021 se presenta el trabajo de Franchi et. al., el cual implementa un pipeline para simulación de vehículos autónomos utilizando Webots y SUMO en paralelo en un *cluster* HPC (Palmetto Cluster) [32]. Si bien los resultados muestran una mejora en la performance de las simulaciones multimodales, se destaca que es necesario tener en cuenta el *overhead* que implica el levantamiento de las simulaciones y la escritura de resultados por cada simulación ejecutada en paralelo.

Capítulo 6

Resultados

A continuación, se muestran los resultados obtenidos para el test de carga de la versión paralelizada de las simulaciones implementadas en SUMO, considerando para ello las distintas frecuencias de inserción de vehículos definidas para cada *set* de simulaciones y los tiempos de ejecución promedio para cada una. Para estas mediciones, se consideraron cinco conjuntos de particiones de diferentes tamaños: 4, 8, 16, 32 y 64 particiones. Para cada uno de estos conjuntos, se ejecutaron 1000 simulaciones, a saber, 50 simulaciones por cada frecuencia de inserción de vehículos definida.

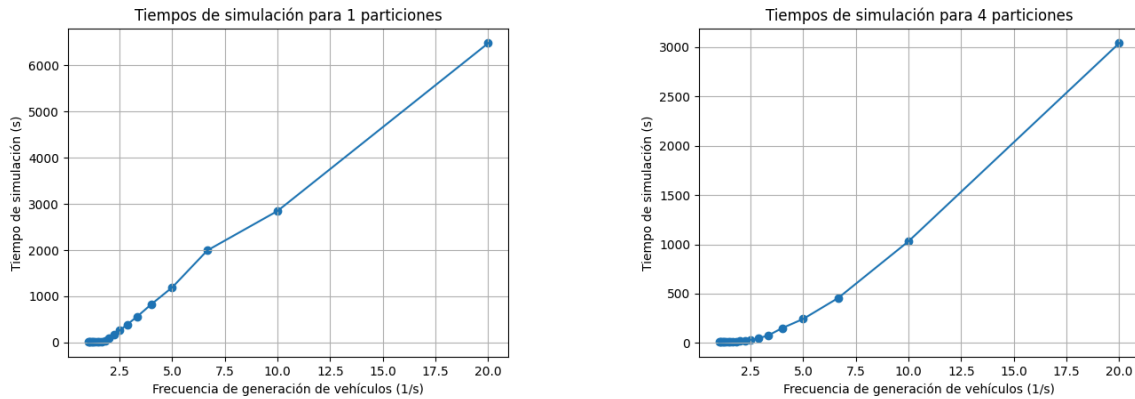


Figura 11: Resultados para la prueba de escalabilidad basada en la carga de tráfico vehicular para la versión paralelizada de SUMO para 1 y 4 particiones.

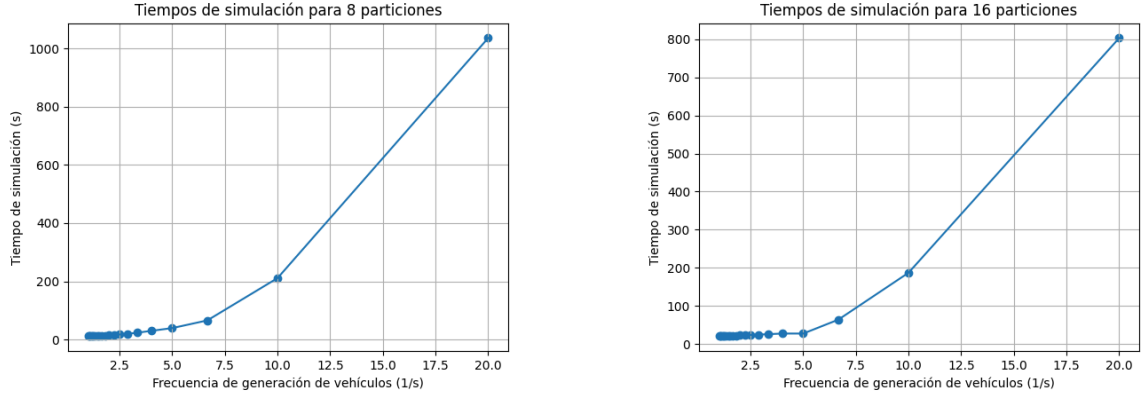


Figura 12: Resultados para la prueba de escalabilidad basada en la carga de tráfico vehicular para la versión paralelizada de SUMO para 8 y 16 particiones.

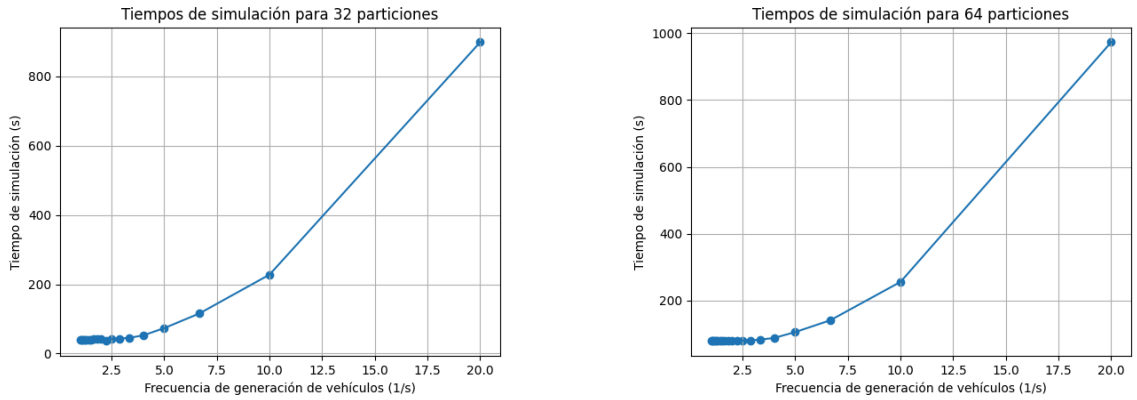


Figura 13: Resultados para la prueba de escalabilidad basada en la carga de tráfico vehicular para la versión paralelizada de SUMO para 32 y 64 particiones.

Como es posible observar, la cota superior para el crecimiento de los tiempos de simulación para cada conjunto de particiones disminuye significativamente en comparación a una versión secuencial de éstas (es decir, una simulación que considera una única partición). Un detalle importante a observar dentro de estos resultados radica en la cota superior para las ejecuciones de los *sets* de 64 particiones, ya que se muestra ligeramente mayor respecto a la cota para los *sets* de 32 particiones.

Para estudiar más a fondo la escalabilidad de la solución propuesta, resulta necesario realizar una comparación entre los crecimientos de los tiempos de simulación para cada conjunto de particiones, además de contrastarlos con la versión secuencial. El siguiente gráfico muestra dichas diferencias, mostrando el comportamiento de cada conjunto de particiones al momento de ejecutar cada simulación para cada frecuencia de inserción definida:

Comparación del crecimiento de los tiempos de simulación para distintas cantidades de particiones

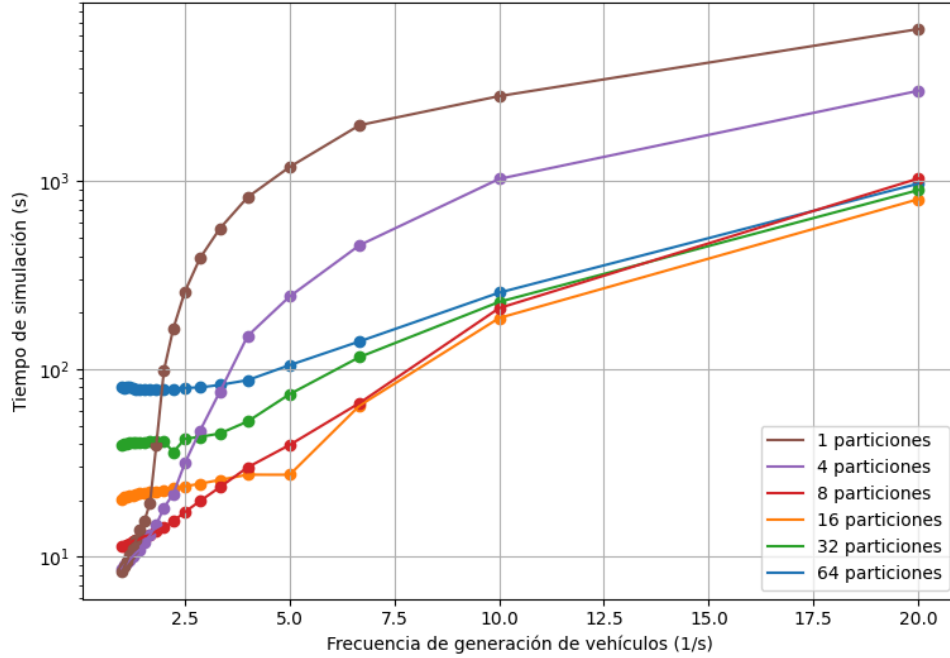


Figura 14: Comparación de escalabilidad de simulaciones con su versión secuencial.

De aquí es posible observar que, si bien para frecuencias pequeñas no resulta eficiente una paralelización con 32 y 64 particiones, a medida que se aumenta el número de particiones los tiempos de simulación exhiben un comportamiento mucho más lineal, mientras que para altas tasas de inserción de vehículos, en la mayoría de los casos los tiempos de simulación tienden a converger, lo cual deja interpretar que, en casos de simulaciones con alta congestión de tráfico vehicular, no resulta más eficiente la solución que posea mayor número de particiones.

Por otro lado, el *speedup* para cada *set* de simulaciones, es decir, la medida de aceleración de la solución, se calcula mediante la razón entre los tiempos originales de ejecución y los tiempos de ejecución de la solución propuesta. Mediante dicho cálculo, se obtuvieron los resultados que se muestran en las siguientes tablas, los cuales muestran que en gran parte de los casos, logró obtenerse un *speedup* superior al 5% esperado:

period	time	speedup
0.05	3039.8	2.1324166063556813
0.1	1031.12	2.76264644270308
0.15	456.98	4.363254409383343
0.2	243.66	4.899778379709431
0.25	150.36	5.487629688747007
0.3	75.66	7.411842453079567
0.35	46.72	8.321061643835616
0.4	31.54	8.206721623335447
0.45	21.6	7.589814814814814
0.5	18.24	5.400219298245615
0.55	14.82	2.6653171390013495
0.6	13.0	1.4830769230769232
0.65	11.98	1.3038397328881468
0.7	10.92	1.2673992673992673
0.75	10.48	1.166030534351145
0.8	10.0	1.1019999999999999
0.85	9.6	1.0625
0.9	9.22	1.021691973969631
0.95	8.96	0.9866071428571428
1.0	8.62	0.9651972157772623

Tabla 5: Speedup para 4 particiones

period	time	speedup
0.05	1035.76	6.258322391287557
0.1	211.2	13.487784090909091
0.15	65.86	30.275129061645917
0.2	39.4	30.3015228426396
0.25	29.94	27.559118236472944
0.3	23.44	23.924061433447097
0.35	19.76	19.67408906882591
0.4	17.3	14.961849710982657
0.45	15.44	10.617875647668393
0.5	14.42	6.830790568654646
0.55	13.68	2.8874269005847952
0.6	13.06	1.4762633996937213
0.65	12.84	1.2165109034267911
0.7	12.42	1.1143317230273753
0.75	12.24	0.9983660130718954
0.8	12.12	0.9092409240924093
0.85	11.76	0.8673469387755102
0.9	11.62	0.810671256454389
0.95	11.44	0.7727272727272727
1.0	11.44	0.7272727272727273

Tabla 6: Speedup para 8 particiones

period	time	speedup
0.05	803.4	8.068359472242967
0.1	187.06	15.22837592216401
0.15	63.82	31.242870573487934
0.2	27.32	43.69985358711567
0.25	27.34	30.17995610826628
0.3	25.64	21.87129485179407
0.35	24.46	15.893704006541292
0.4	23.7	10.921518987341772
0.45	23.08	7.1031195840554595
0.5	22.64	4.3507067137809186
0.55	22.26	1.7744833782569631
0.6	22.08	0.8731884057971016
0.65	21.9	0.7132420091324201
0.7	21.66	0.6389658356417359
0.75	21.34	0.5726335520149953
0.8	21.16	0.5207939508506616
0.85	21.14	0.48249763481551555
0.9	20.92	0.45028680688336514
0.95	20.9	0.4229665071770335
1.0	20.28	0.41025641025641024

Tabla 7: Speedup para 16 particiones

period	time	speedup
0.05	898.52	7.214218937808841
0.1	227.92	12.498332748332748
0.15	115.78	17.221627224045605
0.2	73.46	16.25210999183229
0.25	52.58	15.692658805629517
0.3	45.3	12.37924944812362
0.35	43.5	8.937011494252873
0.4	42.48	6.093220338983051
0.45	36.06	4.546311702717692
0.5	41.22	2.3896166909267347
0.55	41.18	0.9592034968431278
0.6	40.98	0.4704734016593461
0.65	40.72	0.3835952848722986
0.7	40.46	0.34206623826000987
0.75	40.4	0.3024752475247525
0.8	40.42	0.272637308263236
0.85	40.34	0.2528507684680218
0.9	40.26	0.23397913561847988
0.95	40.16	0.22011952191235062
1.0	39.2	0.21224489795918366

Tabla 8: Speedup para 32 particiones

period	time	speedup
0.05	973.4	6.6592562153277175
0.1	255.34	11.156183911647215
0.15	140.34	14.207781103035485
0.2	104.92	11.378955394586352
0.25	87.3	9.451546391752577
0.3	82.24	6.818822957198444
0.35	79.78	4.872900476309852
0.4	79.32	3.263237518910741
0.45	78.32	2.09320735444331
0.5	78.22	1.259268729225262
0.55	78.4	0.5038265306122449
0.6	77.9	0.24749679075738126
0.65	77.96	0.2003591585428425
0.7	78.04	0.17734495130702202
0.75	78.14	0.15638597389301254
0.8	79.14	0.13924690422036895
0.85	79.96	0.12756378189094547
0.9	79.92	0.11786786786786786
0.95	79.58	0.11108318673033425
1.0	79.82	0.10423452768729644

Tabla 9: Speedup para 64 particiones

El siguiente gráfico muestra con mayor claridad la variación del *speedup* para cada set de simulaciones. Aquí es posible observar que, a medida que aumentamos el número de particiones, la tendencia general del *speedup* de las simulaciones para cada punto mejorará; sin embargo, para simulaciones con alta carga de tráfico vehicular, la *performance* de la solución propuesta disminuirá, llegando a ser similar a la solución secuencial:

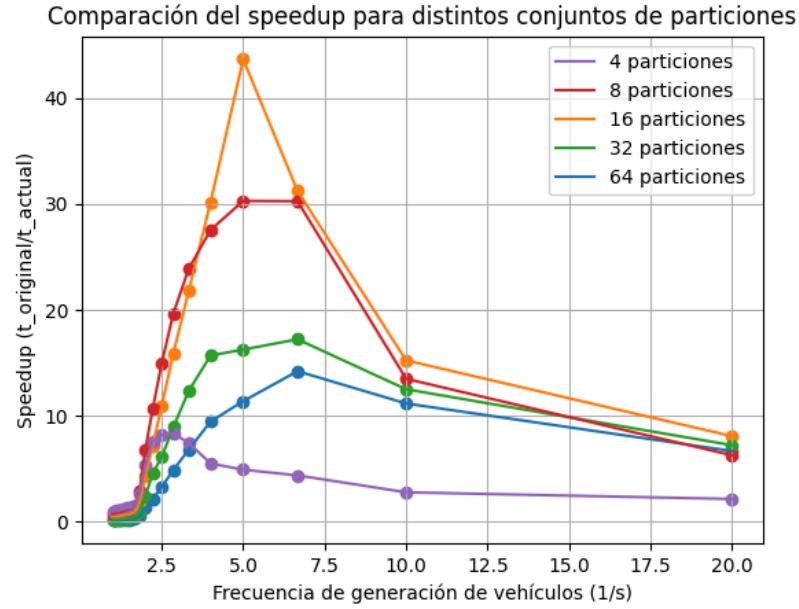


Figura 15: Comparación del *speedup* para distintos conjuntos de particiones.

Por otro lado, es posible calcular la eficiencia de la paralelización mediante la razón entre el *speedup* y el número de particiones realizadas. Esto nos da una idea de qué tan eficiente es la solución en relación a la versión secuencial (ejecutada mediante un solo nodo). Los resultados obtenidos de este cálculo se resumen en el siguiente gráfico, el cual muestra que la mayor eficiencia se obtiene con números bajos de particiones (menor a 32):

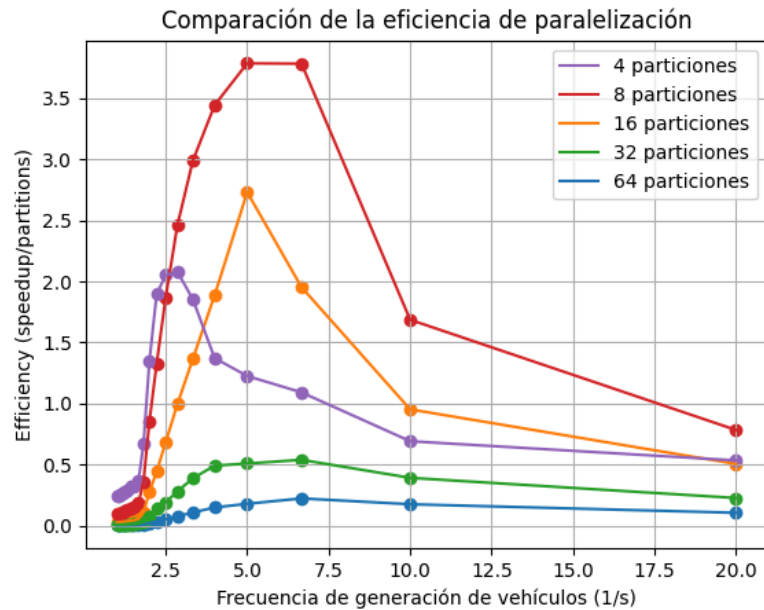


Figura 16: Comparación de la eficiencia de paralelización para distintos conjuntos de particiones.

En cuanto al uso de CPU para cada set de simulaciones, se obtiene el siguiente gráfico comparativo entre la escalabilidad de cada número de particiones realizadas:

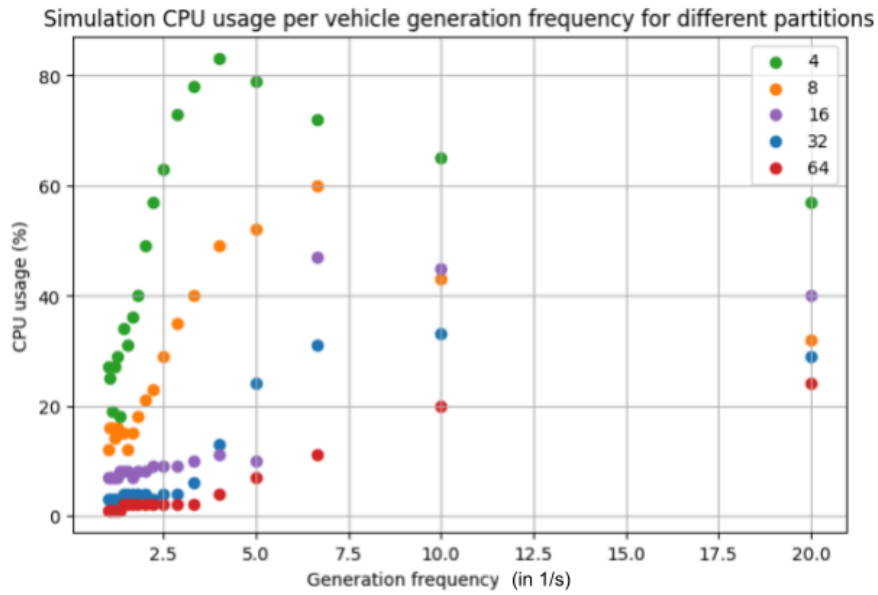


Figura 17: Comparación de escalabilidad de uso de CPU por simulaciones ejecutadas con diferentes números de particiones.

Capítulo 7

Discusión

Como es posible observar de la Figura 11, la Figura 12 y la Figura 13, existe una gran diferencia entre la escalabilidad de la versión secuencial de una simulación y las versiones paralelizadas, teniendo estas últimas una cota superior considerablemente inferior a la cota superior de tiempo de las simulaciones no paralelizadas, lo cual muestra que las versiones paralelas de las simulaciones son más eficientes en cuanto al tiempo de ejecución de éstas.

Adicionalmente, es posible extraer de la Figura 14 que, a medida que se aumenta la cantidad de particiones, el escalamiento de las simulaciones aparenta ser cada vez más lineal, implicando también un mayor costo para las simulaciones con menor tasa de inserción de vehículos para simulaciones de grandes cantidades de particiones, pero un menor crecimiento en cuanto a los tiempos de ejecución de las simulaciones.

Asimismo, el aumento en el *speedup* mostrado en la Figura 15 nos muestra que, si bien es posible llegar a tener tiempos hasta 40 veces mejores que en una versión secuencial para escenarios con una inserción de autos moderada, para escenarios con altas tasas de tráfico la *performance* tiende a descender. En cuanto a la eficiencia de la solución planteada, la Figura 16 nos muestra que para un número de particiones también moderada (es decir, menor a 32 particiones), se tiene una alta eficiencia sobre los tiempos de simulación para frecuencias de generación de vehículos que se encuentran alrededor y bajo la mediana, mientras que para escenarios con mayor congestión de tráfico la eficiencia de la solución propuesta tiende a ser más baja.

Cabe considerar en estos resultados dos aspectos importantes: el balance de las cargas en las particiones y la pérdida de información al realizar el corte de rutas.

Por un lado, el algoritmo aplicado para el particionamiento de los grafos correspondientes a la representación de los mapas conlleva un desbalance entre la cantidad de nodos presentes en las particiones que es importante considerar, ya que existen particiones que superan considerablemente a otras en este aspecto. Esto puede provocar que la escalabilidad

de la solución se vea sujeta principalmente a la escalabilidad que ofrezcan los nodos con mayor carga, en vez de cada una de las particiones realizadas.

Por otro lado, el proceso de recorte de rutas según las particiones realizadas a los mapas implica una pérdida de información necesaria a considerar, dado que implica una menor carga para los nodos de ejecución, al eliminar vehículos cuyas rutas no cumplen con los criterios definidos por el algoritmo de particionamiento de rutas provisto por *cutRoutes.py*. Un posible trabajo a futuro consiste en visitar este algoritmo para poder manejar de otra manera el corte de las rutas con los tiempos de partida de los vehículos y, de esta manera, evitar dicha pérdida de información.

En cuanto a los procesos de sincronización, se hace necesario implementar una manera de gestionar los tiempos de partida de los vehículos, de forma que no se produzca el fenómeno descrito en la Sección 4.7.3 y las rutas particionadas no queden como rutas independientes, sino que posean dependencia de unas con otras entre los nodos de ejecución. Dentro de la implementación también resulta necesario considerar el *overhead* que implican los procesos de comunicación entre los nodos, lo cual podría afectar de manera considerable a la escalabilidad de la solución ya implementada.

En cuanto al uso de CPU, es posible observar que hasta en las condiciones menos óptimas se llega a un uso de alrededor del 80%, lo cual sugiere una mejora en cuanto al uso de recursos computacionales respecto a la solución secuencial, que ocupa, en el general de los casos, el 100% de la CPU que se le asigna para el proceso de simulación.

Cabe considerar, además, que la ejecución de estos experimentos fue realizada en un computador Intel® Core™ i7-106G7 CPU @2,30GHz, dado que por problemas que escapan al alcance de esta tesis, no fue posible llevarla a cabo en un ambiente de supercomputación. Sin embargo, es posible replicar estos experimentos en supercomputadores dada la contenerización de los procesos de SUMO en *Singularity*, lo cual mantiene al *software* portable para estos entornos.

Capítulo 8

Conclusiones

Durante este trabajo, se desarrolló un sistema de paralelización de simulaciones de tráfico vehicular urbano a partir del uso del *software* SUMO, orientado a su uso en supercomputadores por medio de la herramienta de paralelización *OpenMP*. Esto, con el objetivo de estudiar la escalabilidad de una solución al problema de los costos en CPU y tiempos de ejecución para estas simulaciones, de manera que puedan insertarse de manera eficiente a proyectos tales como el desarrollo de *Gemelos Digitales* para ciudades.

Dentro de la investigación, se plantearon preguntas en torno al mejoramiento en la *performance* de las simulaciones; en particular, si es que es posible optimizarla y cómo influyen el área a simular y la carga vehicular en el rendimiento de las ejecuciones. A partir de los resultados obtenidos y lo discutido en el Capítulo 7, es posible concluir que si bien es posible mejorar la *performance* de las simulaciones de tráfico vehicular urbano mediante la paralelización de sus procesos, es necesario tener en cuenta el *overhead* que implica la necesidad de procesos de sincronización entre los nodos de ejecución, dado que en el *pipeline* para llegar a la paralelización ocurre pérdida de información que se debe manejar si se quiere mantener la precisión y la alta granularidad de las simulaciones.

Adicionalmente, se concluye que únicamente con paralelización, el escalamiento de las simulaciones en cuanto a tiempos de ejecución se *lineariza* a medida que se aumenta el número de particiones; es decir, que pasan de mostrar un comportamiento exponencial a uno más lineal. Mientras tanto, el uso de CPU pasa de ser constante en un 100% de la CPU asignada a las simulaciones a un máximo del 80% de ésta, manteniéndose variable en función del número de particiones realizadas y la frecuencia de inserción de vehículos en la simulación. Respecto al comportamiento de los tiempos de ejecución en función del tamaño del mapa, se tiene que estos crecen de forma lineal a medida que la red de caminos lo hace; por lo tanto, se concluye que este no resulta ser un factor determinante al momento de considerar las variables que afectan de manera significativa la escalabilidad tanto de las simulaciones secuenciales como las paralelizadas.

Finalmente, se consideran diferentes vías de trabajo futuro para el mejoramiento de las simulaciones en cuanto a su precisión y la pérdida de información en el proceso de corte de rutas para su paralelización.

En primer lugar, se plantea la necesidad del desarrollo de un mecanismo que maneje de manera apropiada los tiempos de partida de los vehículos que se insertan o salen de cada nodo de ejecución, además de determinar el destino de aquellas rutas que por defecto descarta la herramienta *cutRoutes.py*, con el objetivo de obtener una mayor precisión y una menor pérdida de información al momento de particionar las simulaciones.

Por otro lado, se presenta como un posible trabajo a futuro el estudio de la escalabilidad de simulaciones sincronizadas considerando el *overhead* que implica la comunicación entre los nodos, además del estudio y comparación de estas soluciones en un ambiente de supercomputación.

Otro aspecto importante a considerar es la elección del algoritmo de particionamiento de grafos a utilizar para preparar la paralelización de las simulaciones; si bien en el presente trabajo se hizo uso únicamente del algoritmo de *SPartSim*, es posible realizar comparaciones sobre la *performance* de las simulaciones en función del algoritmo de particionamiento elegido.

El trabajo aquí desarrollado deja la posibilidad de continuar expandiendo la línea del conocimiento sobre la paralelización de simulaciones que, en un futuro próximo, podrá aportar una gran utilidad para el desarrollo de modelos tales como los *Gemelos Digitales* para ciudades, los cuales pretenden ser una herramienta de mejoramiento en la calidad de vida de millones de personas alrededor del mundo por medio de la implementación de dinámicas y políticas públicas de manera más informada, eficiente y segura, por medio del modelamiento computacional de los entornos y sus dinámicas.

Bibliografía

- [1] Gligor Kotusevski y Ken A. Hawick, «A review of traffic simulation software», *Research Letters in the Information and Mathematical Sciences*, vol. 13, pp. 35-54.
- [2] Paolo M. Ejercito, Kristine Gayle E. Nebrija, Rommel P. Feria, y Ligaya Leah Lara-Figueroa, «Traffic simulation software review», *IEEE*.
- [3] Azise Oumar Diallo, Guillaume Lozenguez, Arnaud Doniec, y René Mandiau, «Comparative evaluation of road traffic simulators based on modeler's specifications An application to intermodal mobility behaviors», *SCITEPRESS - Science and Technology Publications*.
- [4] Daniel Krajzewicz *et al.*, «The 'Simulation of Urban MObility' package: An open source traffic simulation».
- [5] Rainer Stark y Thomas Damerau, «Digital Twin», *Springer Berlin Heidelberg*. [En línea]. Disponible en: https://doi.org/10.1007/978-3-642-35950-7_16870-1
- [6] Cheng Zhou *et al.*, «Digital Twin Network, Concepts and Reference Architecture». [En línea]. Disponible en: <https://www.ietf.org/archive/id/draft-irtf-nmrg-network-digital-twin-arch-01.html>
- [7] Pablo López Álvarez *et al.*, «Microscopic traffic simulation using SUMO», *IEEE*.
- [8] German Aerospace Center, «Duarouter». [En línea]. Disponible en: <https://sumo.dlr.de/docs/duarouter.html>
- [9] German Aerospace Center, «Randomtrips». [En línea]. Disponible en: <https://sumo.dlr.de/docs/Tools/Trip.html>
- [10] Nicolas Arroyave, Andrés Acosta, Jorge Espinosa Oviedo, y Jairo Espinosa Oviedo, «A new strategy for synchronizing traffic flow on a distributed simulation using SUMO».
- [11] Hao Chen *et al.*, «QarSUMO: A Parallel, Congestion-optimized Traffic Simulator», *ACM*.
- [12] George Karypis y Vipin Kumar, «A Fast and High-Quality Multilevel Scheme for Partitioning Irregular Graphs», *SIAM Journal on Scientific Computing*, vol. 20, pp. 359-392.
- [13] Anthony Ventresque *et al.*, «SPartSim : A Space Partitioning Guided by Road Network for Distributed Traffic Simulations».

- [14] Sylabs Inc., «Introduction to Singularity». [En línea]. Disponible en: <https://docs.sylabs.io/guides/3.5/user-guide/introduction.html>
- [15] Andrew J. Younge, Kevin Pedretti, Ryan E. Grant, y Ron Brightwell, «A Tale of Two Systems: Using Containers to Deploy HPC Applications on Supercomputers and Clouds», *IEEE*.
- [16] OpenMP Architecture Review Board, *OpenMP 5.2 Reference Guide*.
- [17] German Aerospace Center, «Libtraci». [En línea]. Disponible en: <https://sumo.dlr.de/docs/Libtraci.html>
- [18] Andrés Acosta, Jairo Espinosa, y Jorge Espinosa, «Distributed Simulation in SUMO Revisited: Strategies for Network Partitioning and Border Edges Management».
- [19] German Aerospace Center, «Introduction». [En línea]. Disponible en: <https://sumo.dlr.de/docs/index.html>.
- [20] German Aerospace Center, «TraCI». [En línea]. Disponible en: <https://sumo.dlr.de/docs/TraCI.html>
- [21] Lara Codeca, Raphaël Frank, y Thomas Engel, «Luxembourg SUMO Traffic (LUST) scenario: 24 hours of mobility for vehicular networking research», *IEEE*.
- [22] Laura Bieker, Daniel Krajzewicz, Antonio Pio Morra, Carlo Michelacci, y Fabio Cartolano, «Traffic Simulation for all: a real world traffic scenario from the city of Bologna», *Springer*.
- [23] Sergey Ivanov, Ksenia Nikolskaya, Gleb Radchenko, Leonid Sokolinsky, y Mikhail Zymbler, «Digital twin of city: concept overview», *IEEE*.
- [24] Edward Glaessgen y David Stargel, «The digital twin paradigm for future NASA and US air force vehicles».
- [25] Ehab Shahat, Chang T Hyun, y Chunho Yeom, «City digital twin potentials: A review and research agenda», *Sustainability*, vol. 13, p. 3386.
- [26] Javier Argota Sánchez-Vaquero, «Getting Real: the Challenge of Building and Validating a Large-Scale Digital Twin of Barcelona's Traffic with Empirical Data», *ISPRS International Journal of Geo-Information*, vol. 11, p. 24.

- [27] Krešimir Kušić, Rene Schumann, y Edouard Ivanjko, «Building a motorway digital twin in SUMO: Real-time simulation of continuous data stream from traffic counters», *IEEE*.
- [28] Mariana Azevedo *et al.*, *Optimizing Vehicle IoT Systems: SUMO-Digital Twin Performance Analysis*. en 2024 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0 & IoT). IEEE, pp. 204-209.
- [29] Raymond Klefstad, Yue Zhang, Mingjie Lai, R. Jayakrishnan, y Riju Lavanya, «A distributed, scalable, and synchronized framework for large-scale microscopic traffic simulation», *IEEE*.
- [30] Toyotaro Suzumura, Gavin McArdle, y Hiroki Kanezashi, «A high performance multi-modal traffic simulation platform and its case study with the Dublin city», *IEEE*.
- [31] Daigo Umemoto y Nobuyasu Ito, «Large-scale parallel execution of urban-scale traffic simulation and its performance on K computer», *Journal of Computational Social Science*, vol. 2, pp. 97-101.
- [32] Matthew Franchi *et al.*, «PAVSP: a parallel autonomous vehicle simulation pipeline on High-Performance Computing».

Anexo A

Ejemplos de archivos

1.1. Archivo XML para la descripción de redes de caminos en SUMO

```
<?xml version="1.0" encoding="UTF-8"?>

<!-- generated on 2025-01-11 20:15:28 by Eclipse SUMO netgenerate Version
v1_17_0+0507-8eb8f3bd1d0
<configuration                                xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:noNamespaceSchemaLocation="http://sumo.dlr.de/xsd/netgenerateConfig
uration.xsd">

    <grid_network>
        <grid value="true"/>
        <grid.number value="2"/>
    </grid_network>

</configuration>
-->

<net
version="1.16" junctionCornerDetail="5" limitTurnSpeed="5.50" xmlns:xsi="http://
www.w3.org/2001/XMLSchema-instance"    xsi:noNamespaceSchemaLocation="http://sumo.
dlr.de/xsd/net_file.xsd">

<location        netOffset="0.00,0.00"        convBoundary="0.00,0.00,100.00,100.00"
origBoundary="0.00,0.00,100.00,100.00" projParameter="!"/>

    <edge id=":A0_0" function="internal">
        <lane id=":A0_0_0" index="0" speed="6.08" length="7.74" shape="-1.60,3.20
-1.30,1.10 -0.40,-0.40 1.10,-1.30 3.20,-1.60"/>
    </edge>
    <edge id=":A0_1" function="internal">
        <lane id=":A0_1_0" index="0" speed="3.90" length="2.58" shape="3.20,1.60
2.50,1.70 2.00,2.00 1.70,2.50 1.60,3.20"/>
    </edge>
    <edge id=":A1_0" function="internal">
```

```

        <lane id=":A1_0_0" index="0" speed="6.08" length="7.74" shape="3.20,101.60
1.10,101.30 -0.40,100.40 -1.30,98.90 -1.60,96.80"/>
    </edge>
    <edge id=":A1_1" function="internal">
        <lane id=":A1_1_0" index="0" speed="3.90" length="2.58" shape="1.60,96.80
1.70,97.50 2.00,98.00 2.50,98.30 3.20,98.40"/>
    </edge>
    <edge id=":B0_0" function="internal">
        <lane id=":B0_0_0" index="0" speed="3.90" length="2.58" shape="98.40,3.20
98.30,2.50 98.00,2.00 97.50,1.70 96.80,1.60"/>
    </edge>
    <edge id=":B0_1" function="internal">
        <lane id=":B0_1_0" index="0" speed="6.08" length="7.74" shape="96.80,-1.60
98.90,-1.30 100.40,-0.40 101.30,1.10 101.60,3.20"/>
    </edge>
    <edge id=":B1_0" function="internal">
        <lane id=":B1_0_0" index="0" speed="6.08" length="7.74" shape="101.60,96.80
101.30,98.90 100.40,100.40 98.90,101.30 96.80,101.60"/>
    </edge>
    <edge id=":B1_1" function="internal">
        <lane id=":B1_1_0" index="0" speed="3.90" length="2.58" shape="96.80,98.40
97.50,98.30 98.00,98.00 98.30,97.50 98.40,96.80"/>
    </edge>

    <edge id="A0A1" from="A0" to="A1" priority="-1">
        <lane id="A0A1_0" index="0" speed="13.89" length="93.60" shape="1.60,3.20
1.60,96.80"/>
    </edge>
    <edge id="A0B0" from="A0" to="B0" priority="-1">
        <lane id="A0B0_0" index="0" speed="13.89" length="93.60" shape="3.20,-1.60
96.80,-1.60"/>
    </edge>
    <edge id="A1A0" from="A1" to="A0" priority="-1">
        <lane id="A1A0_0" index="0" speed="13.89" length="93.60" shape="-1.60,96.80
-1.60,3.20"/>
    </edge>
    <edge id="A1B1" from="A1" to="B1" priority="-1">
        <lane id="A1B1_0" index="0" speed="13.89" length="93.60" shape="3.20,98.40
96.80,98.40"/>
    </edge>
    <edge id="B0A0" from="B0" to="A0" priority="-1">
        <lane id="B0A0_0" index="0" speed="13.89" length="93.60" shape="96.80,1.60
3.20,1.60"/>
    </edge>
    <edge id="B0B1" from="B0" to="B1" priority="-1">
        <lane id="B0B1_0" index="0" speed="13.89" length="93.60" shape="101.60,3.20
101.60,96.80"/>
    </edge>
    <edge id="B1A1" from="B1" to="A1" priority="-1">
        <lane id="B1A1_0" index="0" speed="13.89" length="93.60" shape="96.80,101.60

```

```

3.20,101.60"/>
</edge>
<edge id="B1B0" from="B1" to="B0" priority="-1">
  <lane id="B1B0_0" index="0" speed="13.89" length="93.60" shape="98.40,96.80
98.40,3.20"/>
</edge>

  <junction id="A0" type="priority" x="0.00" y="0.00" incLanes="A1A0_0 B0A0_0"
intLanes=":A0_0_0 :A0_1_0" shape="-3.20,3.20 3.20,3.20 3.20,-3.20 -0.36,-2.49
-1.60,-1.60 -2.49,-0.36 -3.02,1.24">
  <request index="0" response="00" foes="00" cont="0"/>
  <request index="1" response="00" foes="00" cont="0"/>
</junction>
  <junction id="A1" type="priority" x="0.00" y="100.00" incLanes="B1A1_0 A0A1_0"
intLanes=":A1_0_0 :A1_1_0" shape="3.20,103.20 3.20,96.80 -3.20,96.80 -2.49,100.36
-1.60,101.60 -0.36,102.49 1.24,103.02">
  <request index="0" response="00" foes="00" cont="0"/>
  <request index="1" response="00" foes="00" cont="0"/>
</junction>
  <junction id="B0" type="priority" x="100.00" y="0.00" incLanes="B1B0_0 A0B0_0"
intLanes=":B0_0_0 :B0_1_0" shape="96.80,3.20 103.20,3.20 102.49,-0.36 101.60,-1.60
100.36,-2.49 98.76,-3.02 96.80,-3.20">
  <request index="0" response="00" foes="00" cont="0"/>
  <request index="1" response="00" foes="00" cont="0"/>
</junction>
  <junction id="B1" type="priority" x="100.00" y="100.00" incLanes="B0B1_0
A1B1_0" intLanes=":B1_0_0 :B1_1_0" shape="103.20,96.80 96.80,96.80 96.80,103.20
100.36,102.49 101.60,101.60 102.49,100.36 103.02,98.76">
  <request index="0" response="00" foes="00" cont="0"/>
  <request index="1" response="00" foes="00" cont="0"/>
</junction>

  <connection from="A0A1" to="A1B1" fromLane="0" toLane="0" via=":A1_1_0" dir="r"
state="M"/>
  <connection from="A0B0" to="B0B1" fromLane="0" toLane="0" via=":B0_1_0" dir="l"
state="M"/>
  <connection from="A1A0" to="A0B0" fromLane="0" toLane="0" via=":A0_0_0" dir="l"
state="M"/>
  <connection from="A1B1" to="B1B0" fromLane="0" toLane="0" via=":B1_1_0" dir="r"
state="M"/>
  <connection from="B0A0" to="A0A1" fromLane="0" toLane="0" via=":A0_1_0" dir="r"
state="M"/>
  <connection from="B0B1" to="B1A1" fromLane="0" toLane="0" via=":B1_0_0" dir="l"
state="M"/>
  <connection from="B1A1" to="A1A0" fromLane="0" toLane="0" via=":A1_0_0" dir="l"
state="M"/>
  <connection from="B1B0" to="B0A0" fromLane="0" toLane="0" via=":B0_0_0" dir="r"
state="M"/>

  <connection from=":A0_0" to="A0B0" fromLane="0" toLane="0" dir="l" state="M"/>

```

```
<connection from=":A0_1" to="A0A1" fromLane="0" toLane="0" dir="r" state="M"/>
<connection from=":A1_0" to="A1A0" fromLane="0" toLane="0" dir="l" state="M"/>
<connection from=":A1_1" to="A1B1" fromLane="0" toLane="0" dir="r" state="M"/>
<connection from=":B0_0" to="B0A0" fromLane="0" toLane="0" dir="r" state="M"/>
<connection from=":B0_1" to="B0B1" fromLane="0" toLane="0" dir="l" state="M"/>
<connection from=":B1_0" to="B1A1" fromLane="0" toLane="0" dir="l" state="M"/>
<connection from=":B1_1" to="B1B0" fromLane="0" toLane="0" dir="r" state="M"/>

</net>
```

1.2. Archivo XML para la descripción de rutas en SUMO

```
<?xml version="1.0" encoding="UTF-8"?>

<!-- generated on 2025-01-11 20:19:48 by Eclipse SUMO duarouter Version
v1_17_0+0507-8eb8f3bd1d0
-->
<configuration xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:noNamespaceSchemaLocation="http://sumo.dlr.de/xsd/duarouterConfigur
ation.xsd">

    <input>
        <net-file value="net.net.xml"/>
        <route-files value="trips.trips.xml"/>
    </input>

    <output>
        <output-file value="routes.rou.xml"/>
    </output>

    <report>
        <ignore-errors value="true"/>
    </report>

</configuration>
-->

<routes xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="http://sumo.dlr.de/xsd/routes_file.xsd">
    <vehicle id="1" depart="1.00">
        <route edges="A1A0 A0B0"/>
    </vehicle>
    <vehicle id="2" depart="2.00">
        <route edges="B0B1"/>
    </vehicle>
    <vehicle id="3" depart="3.00">
        <route edges="B1B0 B0A0 A0A1"/>
    </vehicle>
    <vehicle id="4" depart="4.00">
        <route edges="A1B1 B1B0 B0A0 A0A1"/>
    </vehicle>
    <vehicle id="10" depart="10.00">
        <route edges="B1A1 A1A0 A0B0 B0B1"/>
    </vehicle>
    <vehicle id="11" depart="11.00">
        <route edges="A1A0 A0B0"/>
    </vehicle>
    <vehicle id="13" depart="13.00">
        <route edges="A0A1"/>
    </vehicle>
    <vehicle id="15" depart="15.00">
        <route edges="B1A1 A1A0 A0B0 B0B1"/>
    </vehicle>
</routes>
```

```

</vehicle>
<vehicle id="16" depart="16.00">
  <route edges="B0A0 A0A1 A1B1 B1B0"/>
</vehicle>
<vehicle id="17" depart="17.00">
  <route edges="A1B1 B1B0 B0A0"/>
</vehicle>
<vehicle id="21" depart="21.00">
  <route edges="A0B0 B0B1 B1A1 A1A0"/>
</vehicle>
<vehicle id="24" depart="24.00">
  <route edges="B0B1 B1A1 A1A0"/>
</vehicle>
<vehicle id="25" depart="25.00">
  <route edges="A1A0 A0B0"/>
</vehicle>
<vehicle id="28" depart="28.00">
  <route edges="A0B0 B0B1"/>
</vehicle>
<vehicle id="32" depart="32.00">
  <route edges="B1A1"/>
</vehicle>
<vehicle id="34" depart="34.00">
  <route edges="A1A0"/>
</vehicle>
<vehicle id="38" depart="38.00">
  <route edges="B1B0 B0A0 A0A1 A1B1"/>
</vehicle>
<vehicle id="39" depart="39.00">
  <route edges="A1A0 A0B0"/>
</vehicle>
<vehicle id="41" depart="41.00">
  <route edges="B0A0 A0A1 A1B1 B1B0"/>
</vehicle>
<vehicle id="43" depart="43.00">
  <route edges="B1B0 B0A0"/>
</vehicle>
<vehicle id="44" depart="44.00">
  <route edges="A0A1"/>
</vehicle>
<vehicle id="47" depart="47.00">
  <route edges="A0A1 A1B1"/>
</vehicle>
<vehicle id="48" depart="48.00">
  <route edges="B1B0 B0A0"/>
</vehicle>
</routes>

```

1.3. Archivo XML para la configuración de una simulación en SUMO

```
<?xml version="1.0" ?>
<configuration xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="http://sumo.dlr.de/xsd/sumoConfiguration.xsd">
  <input>
    <net-file value="net.net.xml"/>
    <route-files value="routes.rou.xml"/>
  </input>
  <processing>
    <ignore-route-errors value="true"/>
  </processing>
  <routing>
    <device.rerouting.adaptation-steps value="18"/>
    <device.rerouting.adaptation-interval value="10"/>
  </routing>
  <report/>
  <gui_only/>
</configuration>
```


1.4. Resultados para tiempos de ejecución de simulaciones secuenciales

period	time
0.05	6482.12
0.1	2848.62
0.15	1993.92
0.2	1193.88
0.25	825.12
0.3	560.78
0.35	388.76
0.4	258.84
0.45	163.94
0.5	98.5
0.55	39.5
0.6	19.28
0.65	15.62
0.7	13.84
0.75	12.22
0.8	11.02
0.85	10.2
0.9	9.42
0.95	8.84
1.0	8.32

Tabla 10: Tiempos para simulación secuencial

1.5. Resultados para tiempos de ejecución de simulaciones paralelizadas

period	time
0.05	3039.8
0.1	1031.12
0.15	456.98
0.2	243.66
0.25	150.36
0.3	75.66
0.35	46.72
0.4	31.54
0.45	21.6
0.5	18.24
0.55	14.82
0.6	13.0
0.65	11.98
0.7	10.92
0.75	10.48
0.8	10.0
0.85	9.6
0.9	9.22
0.95	8.96
1.0	8.62

Tabla 11: 4 particiones

period	time
0.05	1035.76
0.1	211.2
0.15	65.86
0.2	39.4
0.25	29.94
0.3	23.44
0.35	19.76
0.4	17.3
0.45	15.44
0.5	14.42
0.55	13.68
0.6	13.06
0.65	12.84
0.7	12.42
0.75	12.24
0.8	12.12
0.85	11.76
0.9	11.62
0.95	11.44
1.0	11.44

Tabla 12: 8 particiones

period	time
0.05	803.4
0.1	187.06
0.15	63.82
0.2	27.32
0.25	27.34
0.3	25.64
0.35	24.46
0.4	23.7
0.45	23.08
0.5	22.64
0.55	22.26
0.6	22.08
0.65	21.9
0.7	21.66
0.75	21.34
0.8	21.16
0.85	21.14
0.9	20.92
0.95	20.9
1.0	20.28

Tabla 13: 16 particiones

period	time
0.05	898.52
0.1	227.92
0.15	115.78
0.2	73.46
0.25	52.58
0.3	45.3
0.35	43.5
0.4	42.48
0.45	36.06
0.5	41.22
0.55	41.18
0.6	40.98
0.65	40.72
0.7	40.46
0.75	40.4
0.8	40.42
0.85	40.34
0.9	40.26
0.95	40.16
1.0	39.2

Tabla 14: 32 particiones

period	time
0.05	973.4
0.1	255.34
0.15	140.34
0.2	104.92
0.25	87.3
0.3	82.24
0.35	79.78
0.4	79.32
0.45	78.32
0.5	78.22
0.55	78.4
0.6	77.9
0.65	77.96
0.7	78.04
0.75	78.14
0.8	79.14
0.85	79.96
0.9	79.92
0.95	79.58
1.0	79.82

Tabla 15: 64 particiones

1.6. Resultados para la eficiencia de paralelización de la solución propuesta

frequency	time	speedup	efficiency
20.0	3039.8	2.1324166063556813	0.5331041515889203
10.0	1031.12	2.76264644270308	0.69066161067577
6.666666666666667	456.98	4.363254409383343	1.0908136023458357
5.0	243.66	4.899778379709431	1.2249445949273579
4.0	150.36	5.487629688747007	1.3719074221867518
3.3333333333333335	75.66	7.411842453079567	1.8529606132698917
2.857142857142857	46.72	8.321061643835616	2.080265410958904
2.5	31.54	8.206721623335447	2.051680405833862
2.2222222222222223	21.6	7.589814814814814	1.8974537037037036
2.0	18.24	5.400219298245615	1.3500548245614037
1.8181818181818181	14.82	2.6653171390013495	0.6663292847503374
1.6666666666666667	13.0	1.4830769230769232	0.3707692307692308
1.5384615384615383	11.98	1.3038397328881468	0.3259599332220367
1.4285714285714286	10.92	1.2673992673992673	0.31684981684981683
1.3333333333333333	10.48	1.166030534351145	0.29150763358778625
1.25	10.0	1.1019999999999999	0.2754999999999997
1.1764705882352942	9.6	1.0625	0.265625
1.1111111111111112	9.22	1.021691973969631	0.25542299349240777
1.0526315789473684	8.96	0.9866071428571428	0.2466517857142857
1.0	8.62	0.9651972157772623	0.24129930394431556

Tabla 16: Eficiencia calculada para 4 particiones

frequency	time	speedup	efficiency
20.0	1035.76	6.258322391287557	0.7822902989109446
10.0	211.2	13.487784090909091	1.6859730113636364
6.666666666666667	65.86	30.275129061645917	3.7843911327057396
5.0	39.4	30.3015228426396	3.78769035532995
4.0	29.94	27.559118236472944	3.444889779559118
3.3333333333333335	23.44	23.924061433447097	2.990507679180887
2.857142857142857	19.76	19.67408906882591	2.459261133603239
2.5	17.3	14.961849710982657	1.870231213872832
2.2222222222222223	15.44	10.617875647668393	1.3272344559585492
2.0	14.42	6.830790568654646	0.8538488210818308
1.8181818181818181	13.68	2.8874269005847952	0.3609283625730994
1.6666666666666667	13.06	1.4762633996937213	0.18453292496171517
1.5384615384615383	12.84	1.2165109034267911	0.1520638629283489
1.4285714285714286	12.42	1.1143317230273753	0.1392914653784219
1.3333333333333333	12.24	0.9983660130718954	0.12479575163398693
1.25	12.12	0.9092409240924093	0.11365511551155116
1.1764705882352942	11.76	0.8673469387755102	0.10841836734693877
1.1111111111111112	11.62	0.810671256454389	0.10133390705679862
1.0526315789473684	11.44	0.7727272727272727	0.09659090909090909
1.0	11.44	0.7272727272727273	0.09090909090909091

Tabla 17: Eficiencia calculada para 8 particiones

frequency	time	speedup	efficiency
20.0	803.4	8.068359472242967	0.5042724670151855
10.0	187.06	15.22837592216401	0.9517734951352507
6.666666666666667	63.82	31.242870573487934	1.952679410842996
5.0	27.32	43.69985358711567	2.7312408491947293
4.0	27.34	30.17995610826628	1.8862472567666424
3.3333333333333335	25.64	21.87129485179407	1.3669559282371293
2.857142857142857	24.46	15.893704006541292	0.9933565004088307
2.5	23.7	10.921518987341772	0.6825949367088607
2.2222222222222223	23.08	7.1031195840554595	0.4439449740034662
2.0	22.64	4.3507067137809186	0.2719191696113074
1.8181818181818181	22.26	1.7744833782569631	0.1109052111410602
1.6666666666666667	22.08	0.8731884057971016	0.05457427536231885
1.5384615384615383	21.9	0.7132420091324201	0.044577625570776255
1.4285714285714286	21.66	0.6389658356417359	0.039935364727608495
1.3333333333333333	21.34	0.5726335520149953	0.03578959700093721
1.25	21.16	0.5207939508506616	0.03254962192816635
1.1764705882352942	21.14	0.48249763481551555	0.030156102175969722
1.1111111111111112	20.92	0.45028680688336514	0.02814292543021032
1.0526315789473684	20.9	0.4229665071770335	0.026435406698564593
1.0	20.28	0.41025641025641024	0.02564102564102564

Tabla 18: Eficiencia calculada para 16 particiones

frequency	time	speedup	efficiency
20.0	898.52	7.214218937808841	0.22544434180652628
10.0	227.92	12.498332748332748	0.3905728983853984
6.666666666666667	115.78	17.221627224045605	0.5381758507514252
5.0	73.46	16.25210999183229	0.5078784372447591
4.0	52.58	15.692658805629517	0.4903955876759224
3.3333333333333335	45.3	12.37924944812362	0.3868515452538631
2.857142857142857	43.5	8.937011494252873	0.2792816091954023
2.5	42.48	6.093220338983051	0.19041313559322035
2.2222222222222223	36.06	4.546311702717692	0.14207224070992788
2.0	41.22	2.3896166909267347	0.07467552159146046
1.8181818181818181	41.18	0.9592034968431278	0.029975109276347743
1.6666666666666667	40.98	0.4704734016593461	0.014702293801854565
1.5384615384615383	40.72	0.3835952848722986	0.011987352652259331
1.4285714285714286	40.46	0.34206623826000987	0.010689569945625308
1.3333333333333333	40.4	0.3024752475247525	0.009452351485148516
1.25	40.42	0.272637308263236	0.008519915883226125
1.1764705882352942	40.34	0.2528507684680218	0.007901586514625681
1.1111111111111112	40.26	0.23397913561847988	0.007311847988077496
1.0526315789473684	40.16	0.22011952191235062	0.006878735059760957
1.0	39.2	0.21224489795918366	0.0066326530612244895

Tabla 19: Eficiencia calculada para 32 particiones

frequency	time	speedup	efficiency
20.0	973.4	6.6592562153277175	0.10405087836449559
10.0	255.34	11.156183911647215	0.17431537361948773
6.666666666666667	140.34	14.207781103035485	0.22199657973492945
5.0	104.92	11.378955394586352	0.17779617804041176
4.0	87.3	9.451546391752577	0.14768041237113402
3.3333333333333335	82.24	6.818822957198444	0.10654410870622569
2.857142857142857	79.78	4.872900476309852	0.07613906994234143
2.5	79.32	3.263237518910741	0.05098808623298033
2.2222222222222223	78.32	2.09320735444331	0.032706364913176716
2.0	78.22	1.259268729225262	0.01967607389414472
1.8181818181818181	78.4	0.5038265306122449	0.007872289540816327
1.6666666666666667	77.9	0.24749679075738126	0.0038671373555840822
1.5384615384615383	77.96	0.2003591585428425	0.003130611852231914
1.4285714285714286	78.04	0.17734495130702202	0.002771014864172219
1.3333333333333333	78.14	0.15638597389301254	0.002443530842078321
1.25	79.14	0.13924690422036895	0.002175732878443265
1.1764705882352942	79.96	0.12756378189094547	0.001993184092046023
1.1111111111111112	79.92	0.11786786786786786	0.0018416854354354353
1.0526315789473684	79.58	0.11108318673033425	0.0017356747926614727
1.0	79.82	0.10423452768729644	0.0016286644951140068

Tabla 20: Eficiencia calculada para 64 particiones