

Proyecto con OpenMP / C++

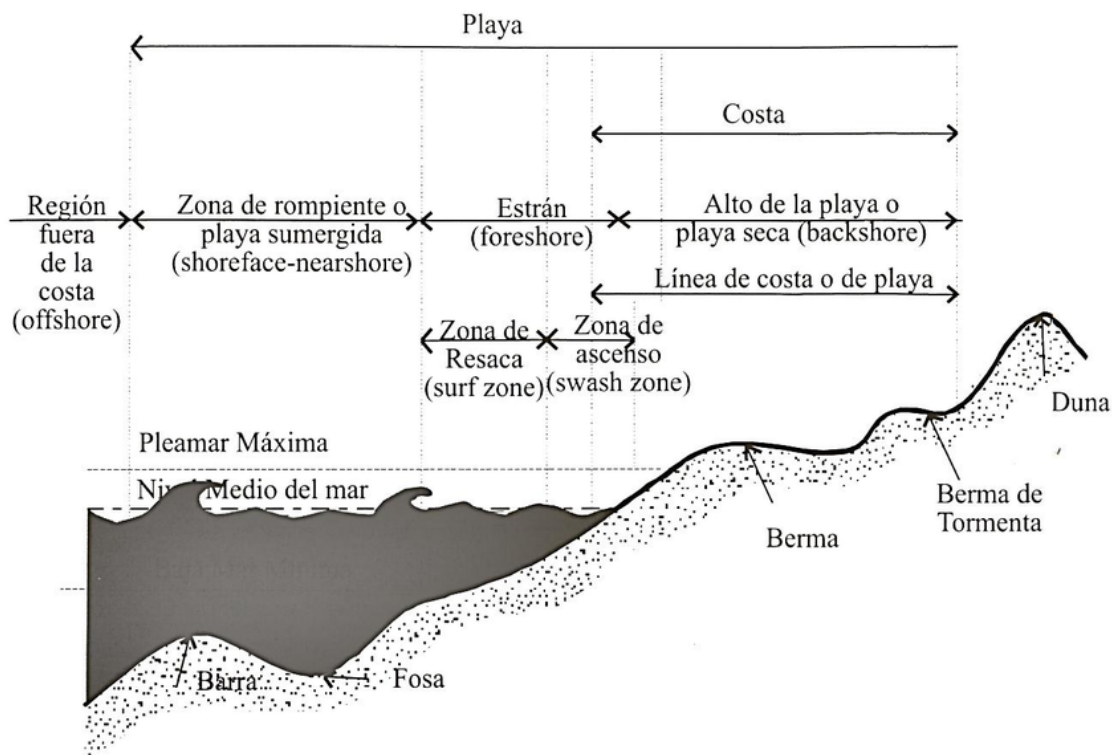
Objetivos

Que el estudiante profundice en:

1. El diseño de programas “escalables” basados en concurrencia por hilos y memoria compartida.
2. La programación, depuración y comprobación con OpenMP en C++.
3. La evaluación del desempeño de un programa con OpenMP en C++ utilizando conceptos simples como “tiempo pared” y “aceleración”.

Descripción del problema

El conteo de tortugas en una arribada no es fácil por las condiciones en que usualmente debe realizarse (de noche) y por lo masivo de las mismas. Por esta razón se utilizan tres métodos de muestreo conocidos en adelante como: método del transecto paralelo a la berma (TPB), método de los cuadrantes (C) y el método de los transectos verticales sobre la berma (TVB). Para aclarar el término transecto y berma, considérese la siguiente figura:



Obsérvese que los niveles de las tres fases de la marea se ubican por debajo de la “berma”: este montículo es por tanto producto de la acción de la marea. Obsérvese que en una tormenta el nivel de la marea puede subir y generar una “berma de tormenta” más alto. Finalmente, no se debe confundir estas bermas con la duna eólica que se forman por la acción del viento y no de la marea.

Los transectos son franjas sobre la playa, ya sea el horizontal o paralelo a la berma, o los verticales ubicados perpendicularmente a la berma, entre ésta y las dunas eólicas. De acuerdo con esto, los métodos de muestreo se describen a continuación:

1. El TPB consiste en contar las tortugas en el transecto que inicia en la línea de costa (ver la figura, inicia aproximadamente en el punto donde llega la marea media) y se extiende en un ancho típico de 15 metros por una longitud de hasta un kilómetro cada cierta cantidad de minutos.

2. El TVB consiste en contar las tortugas que están anidando en franjas verticales o perpendiculares a la berma y sobre ésta. Estos transectos típicamente miden dos metros (aproximadamente lo que mide una persona de mano a mano con los brazos extendidos al máximo hacia los lados) de ancho y su longitud va desde la berma hasta las dunas eólicas. El muestreo se hace cada cierta cantidad de minutos, se pueden incluir desde dos hasta cinco transectos cada cien metros de playa y no deben escrutarse más de 20 en un mismo muestreo.

3. El C consiste en contar las tortugas que están anidando (por ende se ubican sobre la berma, entre ésta y las dunas) o que han terminado de anidar y están todavía sobre el nido en cuadrantes previamente identificados y de cierta dimensión DxD en metros. El muestreo se aplica cada cierta cantidad de minutos.

El problema de simulación consiste en modelar una arribada de N tortugas y los tres métodos de muestreo con el propósito de determinar cuál puede ser el mejor método y bajo qué condiciones es el más adecuado. Se trata entonces de realizar K experimentos de simulación cada uno con N tortugas arribando, y comparar los promedios de cada una de las estimaciones correspondientes a cada tipo de muestreo con el valor N, para intentar encontrar el mejor método de muestreo.

El problema de programación surge porque esta simulación duró hasta seis horas reales con $N = 7,000$, simulando 12 horas de arribada, lo que es una arribada pequeña considerando que hay arribadas hasta de cientos de miles de tortugas y que se prolongan por días. El objetivo fundamental en nuestro proyecto (además de la implementación correcta de la simulación) es reducir al máximo el “tiempo pared” (tp), así como experimentar con 8, 16, 32 y 64 hilos en las estaciones del laboratorio #101 para determinar si la aceleración lograda es significativa. Para tal efecto se deberán realizar los siguientes experimentos (en “experimentos.csv”):

Experimento de desempeño #1:

Cantidad de simulaciones: 10.

Tiempo simulado: 6 horas.

Cantidad de tortugas: 7 mil.

Cantidad de hilos: 8, 16, 32 y 64.

Experimento de desempeño #2:

Cantidad de simulaciones: 10.

Tiempo simulado: 6 horas.

Cantidad de tortugas: 70 mil.

Cantidad de hilos: 8, 16, 32 y 64.

Experimento de desempeño #3:

Cantidad de simulaciones: 10.

Tiempo simulado: 6 horas.

Cantidad de tortugas: 700 mil.

Cantidad de hilos: 8, 16, 32 y 64.

Cuadro comparativo de desempeño que se deberá generar:

# hilos	Experimento #1		Experimento #2		Experimento #3	
8	tp_{11}	$ac_{11} = tp_{11}/tp_{11}$	tp_{12}	$ac_{12} = tp_{12}/tp_{12}$	tp_{13}	$ac_{13} = tp_{13}/tp_{13}$
16	tp_{21}	$ac_{21} = tp_{11}/tp_{21}$	tp_{22}	$ac_{22} = tp_{12}/tp_{22}$	tp_{23}	$ac_{23} = tp_{13}/tp_{23}$
32	tp_{31}	$ac_{31} = tp_{21}/tp_{31}$	tp_{32}	$ac_{32} = tp_{22}/tp_{32}$	tp_{33}	$ac_{33} = tp_{23}/tp_{33}$
64	tp_{41}	$ac_{41} = tp_{31}/tp_{41}$	tp_{42}	$ac_{42} = tp_{32}/tp_{42}$	tp_{43}	$ac_{43} = tp_{33}/tp_{43}$

Debe interpretarse los tp_{ij} como los promedios de los tp de las 10 diez simulaciones en cada experimento. El tiempo pared sólo debe contabilizar el tiempo neto de la simulación, omitiendo el tiempo para cargar datos y generar la salida.

Datos de entrada de la simulación

1. El archivo “terreno.csv” con información sobre los sectores en que se divide la playa sobre la cual se realizará la simulación. Por cada sector este archivo contiene una fila. En cada fila aparecen 4 números separados por comas:

- 1.1 longitud en metros del sector,
- 1.2 distancia del nivel de marea media a la berma en metros,
- 1.3 la altura de la berma respecto del nivel de marea media,
- 1.4 distancia de la berma a las dunas en metros.

2. El archivo “marea.csv” con información tres ciclos de aproximadamente 6 horas cada uno. Contiene tres filas cada una con 3 números separados por comas:

- 2.1 altura de la marea baja o alta en metros,
- 2.2 altura de la marea alta o baja en metros,
- 2.3 cantidad de minutos del ciclo a simular (nunca es exactamente 6 horas).

3. El archivo “comportamiento_tortugas.csv” con información sobre el comportamiento estocástico de las tortugas. Contiene una sola fila de 9 números todos entre 0 y 1. Cada número representa la probabilidad de que la tortuga se desactive después de completar alguna fase o estado de su comportamiento:

- 3.1 probabilidad de desactivarse después de “vagar”,
- 3.2 probabilidad de desactivarse después de “camar”,
- 3.3 probabilidad de desactivarse después de “excavar”,
- 3.4 probabilidad de desactivarse después de “poner”,
- 3.5 probabilidad de desactivarse después de “tapar”,
- 3.6 probabilidad de desactivarse después de “camuflar”,
- 3.7 velocidad promedio estimada,
- 3.8 desviación estándar de la velocidad,
- 3.9 parámetro s (escala) para la distribución logística de la arribada.

NOTA: el parámetro μ debe ser igual a cero.

4. El archivo “cuadrantes.csv” con información sobre la aplicación del método C y los cuadrantes a usar. Contiene una primera fila con dos datos: la cantidad de contadores asignados y cada cuántos minutos se repetirá el conteo. Luego contiene una fila por cada cuadrante. En cada fila aparecen 4 números separados por comas:

- 4.1 coordenada X de la esquina inferior izquierda,
- 4.2 coordenada Y de la esquina inferior izquierda,
- 4.3 coordenada X de la esquina superior derecha,
- 4.4 coordenada Y de la esquina superior derecha.

5. El archivo “transectos_verticales.csv” con información sobre la aplicación del método TVB y los transectos verticales ubicados entre la berma y las dunas. Contiene una primera fila con dos datos: la cantidad de contadores asignados y cada cuántos minutos se repetirá el conteo. Luego contiene una fila por cada transecto vertical. En cada fila aparecen tres números:

- 5.1 coordenada X de la esquina inferior izquierda (X_{inf}),
- 5.2 coordenada Y de la esquina inferior izquierda,
- 5.3 coordenada Y de la esquina superior derecha.
- 5.4 La coordenada X de la esquina superior derecha NO se incluye pues es $X_{inf} + 2$.

6. El archivo “transecto_paralelo_berma.csv” con información sobre la aplicación del método TPB y el transecto horizontal paralelo a la berma. Contiene una primera fila con dos datos: la cantidad de contadores asignados y cada cuántos minutos se repetirá el conteo. Luego contiene una fila con dos números:

- 6.1 el ancho del transecto en metros,
- 6.2 la longitud del transecto en metros.

Resultados o salida de la simulación

- 1. Cantidad total de tortugas N.
- 2. Cantidad de tortugas que efectivamente anidaron y por ende ovopositaron A.
- 3. Estimación por TPB de N.
- 4. Estimación por TVB de A.
- 5. Estimación por C de A.

Tipos de agentes de la simulación

Esta es una simulación basada en dos tipos de agentes: “tortuga” y “contador”. Toda simulación basada en agentes requiere además un objeto “Simulador” que ejecute la simulación. De acuerdo con lo anterior usted deberá programar las siguientes clases en C++:

Nombre de la clase	Función que cumple
Tortuga	Representar los atributos y el comportamiento de las tortugas.
Contador	Representar los atributos y el comportamiento de los contadores.
Simulador	Ejecutar la simulación

Se le provee un código base para cada clase que usted podrá modificar agregando atributos privados, métodos públicos y privados. Si es necesario modificar o eliminar algún método provisto se hará en acuerdo con el docente.

Las funciones del programa main() son:

1. Cargar y validar los datos del archivo “experimentos.csv”.
2. Cargar y validar los archivos de datos de entrada.
3. Ejecutar cada experimento indicado en el archivo “experimentos.csv”:
 - 3.1 Asignar la instancia de Simulador con los datos de entrada.
 - 3.2 Ejecutar la simulación invocando Simulador::simular(...).
 - 3.3 Generar el archivo con los resultados de la simulación.
 - 3.4 Generar la tabla de desempeño del experimento (dos columnas por cuatro filas).

Escalas del modelo de simulación

1. Cada tic de la simulación representará un minuto.
2. Cada posición (X,Y) representará un metro cuadrado.
3. Cada tortuga ocupa una posición de un metro cuadrado.

Reglas que rigen el comportamiento de los agentes y la marea

El comportamiento de las tortugas pasa por la secuencia de estados indicada en “EstadoTortuga”: {vagar, camar, huequear, poner, tapar, camuflar, inactiva} en ese orden empezando por “vagar”. La tortuga puede desactivarse sin completar toda la secuencia, para lo cual se usan las probabilidades correspondientes en “comportamiento_tortugas.csv”.

La velocidad de cada tortuga se determinará al azar usando la distribución normal (ver en cplusplus “normal_distribution”) con los parámetros de promedio y desviación estándar dados en el archivo correspondiente.

La posición inicial de cada tortuga en la playa se determinará al azar usando una distribución uniforme (ver en cplusplus “uniform_int_distribution”) para ambas coordenadas X,Y, tomando en cuenta que la coordenada Y puede variar entre la línea de la marea baja y la línea de la marea en el tic en que está entrando la tortuga a la playa.

El comportamiento de los contadores pasa por el ciclo indicado en “EstadoContador”: {contar, esperar}, lo que implica que estos dos estados se repiten varias veces durante una simulación para cada contador. Si los contadores se desplazan lo hacen todos a velocidad constante de 6 kilómetros por hora.

El comportamiento de la arribada se modelará en un periodo de aproximadamente 360 minutos contados a partir de la marea baja hasta la alta (ver archivo “marea.csv”) y mediante una distribución logística (ver https://en.wikipedia.org/wiki/Logistic_distribution) que alcanza su pico en la marea media. Los parámetros para la distribución vienen en el archivo “comportamiento_tortugas.csv”

El comportamiento de la altura de la marea se modelará por medio de una función lineal:

$$a = p \cdot t + c$$

donde “a” es la altura en metros y “t” el tiempo en minutos. Para calcular “p” y “c”, deben usarse los valores del archivo “marea.csv”, bajo el supuesto de que por ejemplo, según la primera línea de dicho archivo, cuando $t = 0$, a debe ser 0.6 y cuando $t = 372$ a debe ser 2.2. Cada línea por tanto tendrá sus propios valores p y c.

Fórmula de cálculo para cada tipo de muestreo

Cada tipo de muestreo implica una fórmula de cálculo del total estimado de tortugas.

Mecanismo de conteo y fórmula de estimación para el método TPB

$T_e = N_c \cdot i / (4.2 \cdot m)$, en la que:

1. T_{ec} es el total estimado de tortugas caminando hacia la zona de anidación,
2. N_c es la sumatoria de la cantidad contada de tortugas en cada muestreo TPB,
3. i es la cantidad de minutos entre muestreos,
4. m es la cantidad de muestreos que se aplicaron durante el periodo simulado.

Para el TPB es necesario simular el desplazamiento de los contadores, cuya cantidad aparece en el archivo correspondiente, a una velocidad de 6 kms por hora o 100 mts por minuto. La longitud total del transecto paralelo es de 1.5 kms.

Mecanismo de conteo y fórmula de estimación para el método TVB

$$T_p = (A \cdot d / (2 \cdot w \cdot m \cdot \sum l_j)) \cdot (N_o / pr - t_o)$$

1. T_p es el total estimado de tortugas que anidaron,
2. A es el área de anidación total en metros cuadrados (entra la berma y las dunas),
2. d es la duración en minutos del fenómeno simulado,
3. w es el ancho en metros de cada transecto,
4. m es la cantidad de muestreos que se aplicaron durante el periodo simulado,
5. $\sum l_j$ es la sumatoria de las longitudes de todos los transectos,
6. N_o es la sumatoria de las cantidades de tortugas que se detectaron ovopositando,
7. $pr - t_o$ representa el promedio en minutos de cuánto duran las tortugas ovopositando.

En el archivo correspondiente aparecen 15 transectos verticales y un contador por cada transecto. Esto se debe repetir dos veces, cada una de 750 mts, para completar toda la playa a simular que, según el archivo “terreno.csv”, es de 1.5 kms.

Mecanismo de conteo y fórmula de estimación para el método C

$$T_p = ((N_{oc} + 0,94*N_e + 0,47*N_v)*1,25*(A_c/A_{ci})*((d/60)/64,8*m))$$

1. T_p es el total estimado de tortugas que muy probablemente anidaron,
2. N_{oc} conteo de tortugas vistas anidando o con el nido completo en un cuadrante,
3. N_e conteo de tortugas vistas excavando en un cuadrante,
4. N_v conteo de tortugas vistas vagando en un cuadrante,
5. A_c es el área por cuadrante en metros cuadrados,
6. A_{ci} es ??????????????????????
7. d es la duración en minutos del fenómeno simulado,
8. m es la cantidad de muestreos que se aplicaron durante el periodo simulado.

Sobre las constantes:

0,94 es la probabilidad de que una tortuga vista excavando haya anidado finalmente,
0,47 es la probabilidad de que una tortuga vista vagando haya anidado finalmente,
1,25 porque aproximadamente un 25% de T_p se estima que anidan debajo de la berma,
64,8 es el tiempo promedio estimado en que una tortuga construye el nido y pone.

En el archivo correspondiente ("cuadrantes.csv") aparecen los datos de tres cuadrantes para una playa de 750 metros. Para una playa de 1.5 kms se pueden repetir esos datos dos veces, además de incluir un contador más, de tal manera que haya un contador por cada tres cuadrantes.

Criterios de evaluación o rúbrica de evaluación

La calidad de su programa se valorará con base en los siguientes criterios:

1. **Escalabilidad demostrada:** su trabajo deberá funcionar correctamente y distribuir adecuadamente la carga con 8, 16, 32 y 64 hilos. Sin embargo, si la arquitectura en que se ejecuta tiene menos de 16 núcleos sugerirá al usuario una cantidad adecuada de hilos partiendo de la regla bien aceptada de 2 hasta 5 hilos por núcleo.
2. **Desempeño demostrado:** su trabajo será comparado con aquél que muestre el **mejor desempeño por tic** tomando en cuenta los experimentos descritos.
3. **Eficacia demostrada:** su trabajo deberá generar todas las salidas indicadas.
4. Eficiencia en el uso de memoria, basado en las estructuras de datos utilizadas.
5. Simplicidad del código.
6. Forma y estilo del código: sangrado de bloques de código, nombres de objetos empiezan en minúsculas, nombres de clases empiezan en mayúsculas, nombres de métodos también empiezan en minúscula pero se usan mayúsculas para concatenar palabras, comentarios para los atributos y variables de métodos.
7. División de responsabilidades entre main-modelo: el main() sólo se ocupa de la entrada de datos, generar mensajes de error, el despliegue de ciertos resultados por la consola, invocación a los demás objetos para que realicen todos los procesamientos necesarios.

Fecha de entrega: domingo 12 de mayo a las 23:55 por medio del enlace en el sitio del curso. SÓLO DEBERÁ SUBIR LOS ARCHIVOS DE CÓDIGO FUENTE (*.h y *.cpp) los archivos de salida de los tres experimentos y la tabla comparativa de desempeño.

Criterio	%Prc
Escalabilidad demostrada	20
Desempeño demostrado	30
Eficacia demostrada	50
Hasta 10/100 puntos extra por un buen reporte de errores cuando NO funcione	10

Notas importantes:

1. Si el programa NO funciona, tendrá cero puntos en todos los rubros.
2. Este proyecto deberá realizarse idealmente y a lo más en parejas. NO SE ACEPTARÁ NINGÚN TRABAJO ELABORADO POR MÁS DE DOS PERSONAS.
3. Cada hora de atraso en la entrega se penalizará con -1/100, lo que se aplicará a la nota obtenida.
4. A TODOS LOS ESTUDIANTES INVOLUCRADOS EN UN FRAUDE SE LES APLICARÁ EL ARTÍCULO #5 INCISO C DEL "Reglamento de Orden y Disciplina de los Estudiantes de la Universidad de Costa Rica".
5. NO SUBA ningún otro archivo que no sea de código fuente (*.h y *.cpp) o de datos para evitar la transmisión de virus.