

Análise de algoritmos distribuídos no cenário do projeto de Mineração de dados em tempo real para tomada de decisão no trânsito de uma cidade (Olhos da Cidade).

Lucas Schmidt Corrêa Franco, Ricardo Augusto Rabelo Oliveira
PPGCC - Programa de Pós-Graduação em Ciência da Computação
UFOP - Universidade Federal de Ouro Preto
Ouro Preto, Minas Gerais, Brasil
email: lucasscf@gmail.com, rrabelo@gmail.com

Resumo—O projeto Olhos da Cidade em execução no DECOM / UFOP visa a mineração de dados em tempo real para tomada de decisão no trânsito de uma cidade. Neste contexto, há a necessidade do desenvolvimento de sistemas que consigam processar os dados colhidos por câmeras distribuídas em campo, em tempo real, recuperando informações importantes para tomada de decisão, com o menor atraso possível. Para recuperação deste tipo de informação, em tempo hábil, é necessário a utilização e abordagem de algoritmos distribuídos como base para a recuperação de informação distribuída.

Neste trabalho, serão utilizados algoritmos distribuídos que serão apresentados no contexto do projeto Olhos da Cidade. Estes algoritmos serão executados através de uma ferramenta de simulação de algoritmos distribuídos, chamada SimJava, e posteriormente terão as suas devidas análises de complexidade apresentadas.

Keywords—recuperação de informação, algoritmos distribuídos, análise de complexidade.

I. INTRODUÇÃO

Com o crescimento da frota de veículos e consequentemente o aumento do congestionamento no trânsito das principais capitais do país, houve-se a necessidade e a preocupação no investimento de estudos e pesquisas que pudessem contribuir para a melhoria ou solução de tais problemas de trânsito. Este aumento na frota de veículos deve-se a: facilidade de crédito na venda de automóveis, conforto, suposta sensação de segurança e a facilidade de locomoção.

Os problemas de trânsito, resultantes do aumento da frota de veículos nas principais capitais, só podem ser melhorados ou solucionados com informações importantes capazes de indicar pontos críticos no trânsito da cidade estudada.

Para a realização de pesquisas e estudos, que envolvam recuperação deste tipo de informação, várias áreas são envolvidas, principalmente a área de Ciência da Computação. Esta, permite fornecer sistemas e técnicas que consigam processar os dados colhidos, usando mineração de dados de imagens obtidas de câmeras em campo, em tempo real, descobrindo informações importantes com tempo hábil na apresentação dos resultados. Esta é a proposta do projeto Olhos da Cidade em execução no DECOM / UFOP.

No contexto deste projeto, pode-se destacar como principais recursos e técnicas da computação as abordagens de recuperação de informação, mineração de dados, processamento e recuperação de imagens e algoritmos distribuídos, este, tema principal deste trabalho, a fim de apresentar resultados para o problema que permitam a tomada de decisão em tempo factível.

Basicamente, os algoritmos distribuídos permitem que vários processos executem um mesmo algoritmo usando por meio da comunicação entre si, realizando trocas de mensagens (informações), entre os processos [1]. Os objetivos de um algoritmo distribuído é possibilitar um alto processamento em baixo tempo de execução, usando por meio do compartilhamento de recursos [1]. Esta abordagem é semelhante ao contexto de sistemas distribuídos, porém menos abrangente. Um sistema distribuído pode ser definido como um conjunto de unidades de processamento independentes que podem processar uma aplicação em diferentes localidades, além de assumir elementos de rede diferenciados, como cliente e servidor [2].

Em uma analogia, um algoritmo distribuído pode ser efetivamente representado através de um grafo. Seja um grafo $G(V, E)$, onde V são os nós (vértices), representados por cada tarefa e E os meios de comunicação (arestas), representados pelas trocas de mensagens [1].

De acordo com estas informações, as pesquisas relacionadas, encontradas até o momento, são:

d-Agent: An Approach to Mobile Agent Planning for Distributed Information Retrieval [3], que propõe um novo algoritmo de planejamento de novos agentes chamado *d-Agent* capaz de realizar recuperação distribuída de informações de forma prática e realista. Em analogia aos grafos, para este algoritmo, existem n nós onde a informação distribuída reside. Em cada nó, um tempo de computação (recuperação da informação) é necessário para um agente móvel realizar a tarefa de recuperação naquele nó. Neste caso, o problema de planejamento é encontrar o número de agentes móveis e o itinerário de cada agente de modo que o tempo de rotação não seja maior do que o tempo determinado com o menor gasto de recurso de sistema. Pelo fato deste problema de planejamento ser NP-difícil, a solução é baseada em

heurística.

A outra proposta conhecida como *Distributed Information Retrieval Service for Ubiquitous Services* [4] propõe um serviço de recuperação distribuída de informação utilizando tecnologia *Peer-to-Peer (P2P)*, que permite uma flexível forma de recuperação de informação para uma rede de pequena escala.

Com isso, este trabalho propõe um estudo dos dois algoritmos citados acima e suas devidas análises de complexidade, tendo como cenário de teste o projeto Olhos da Cidade, a fim de propor uma solução a nível de algoritmos distribuídos que poderão ser utilizados. Além disso, os algoritmos serão implementados através de uma ferramenta chamada SimJava, que representará o cenário proposto neste trabalho. O SimJava é um pacote para Java que permite a simulação de processos e eventos capaz de representar animações de estruturas de algoritmos distribuídos.

Sendo assim, esta proposta está organizada da seguinte forma. A Seção II apresenta as justificativas e relevâncias que consistem na apresentação das razões que justificam a realização deste projeto, além da relevância de contribuição deste trabalho. Os objetivos concretos do trabalho são apresentados na Seção III. A Seção IV apresenta a seção metodologia do trabalho que detalha como será desenvolvido. A última seção, V, descreve quais os resultados que são esperados com a realização desse projeto.

II. JUSTIFICATIVAS

Devido à grande complexidade e necessidade de grandes análises do projeto Olhos da Cidade, este estudo visa contribuir com o andamento do projeto, apresentando novas propostas de algoritmos distribuídos e eficientes capazes de ajudar na solução do problema, assim como suas devidas análises de complexidade e conclusões. Este trabalho também visa o estudo de algoritmos distribuídos no contexto da disciplina Projeto e Análise de Algoritmos a fim de apresentar, para os membros da turma, conceitos práticos e teóricos sobre o tema abordado.

III. OBJETIVOS

Este trabalho visa como objetivo geral o estudo dos algoritmos distribuídos em um cenário de teste real aplicado para recuperação da informação distribuída e como objetivo específico, visa analisar as características dos algoritmos apresentados propondo melhorias nas complexidades de troca de mensagens.

IV. METODOLOGIA

O trabalho será realizado da seguinte forma:

1 - Comparação dos algoritmos distribuídos básicos encontrados na literatura com os algoritmos propostos nas pesquisas relacionadas, encontradas até o momento.

Estes algoritmos distribuídos básicos encontrados na literatura são: *PI* (Propagação de Informação), *Test Connectivity* (Conectividade) e *Compute Distances* (Menor Distância) [1].

Podemos dizer que o algoritmo *PI* representa um grafo G não dirigido e dissemina a informação *inf*. n é o número de nós em G e m é o número de vértices em G [1]. O problema do algoritmo *PI* é difundir em G informações presentes num subconjunto dos nós de G . Como solução para o problema, supondo que n_i seja um nó que tenha *inf*, é necessário calcular uma árvore geradora em G e utilizar esta árvore para fazer a difusão de *inf*. Posteriormente n_i envia *inf* em todas as arestas da árvore geradora que são incidentes a n_i . Este processo é repetido por todos os outros nós, exceto na aresta onde *inf* foi recebida. Com isso, se a árvore geradora foi computada anteriormente, um algoritmo assíncrono baseado nessa estratégia tem complexidade de tempo e mensagem de $O(n)$ [1].

O algoritmo *Test Connectivity* representa um grafo G e seu problema consiste em cada nó (vértice) em N que deve descobrir as identificações de todos os outros nós aos quais está conectado por um caminho em G [1]. No caso de sistemas sujeitos a falhas, é importante saber os identificadores dos nós de um componente conectado. No contexto de solução, este algoritmo não é adequado para casos onde G altera-se dinamicamente. No caso, basicamente, ele mostra o uso da técnica do algoritmo *PI* [1].

Já o algoritmo *Compute Distances* possui como problema determinar a menor distância em G entre todos os pares de nós (vértices), onde esta distância é medida em número de arestas através de pulsos síncronos [1]. Neste caso cada nó (vértice) possui uma identificação onde se $s = 0$, cada nó envia sua identificação para os seus vizinhos, se $s = i$, sendo $i > 0$, cada nó envia para os vizinhos as identificações recebidas durante $s - 1$ e $N_0 = N$ [1].

As outras etapas de metodologia estão caracterizadas por:

- 2 - Modelo analítico, através da análise de complexidade dos algoritmos;
- 3 - Implementação dos algoritmos no simulador SimJava, para análise de desempenho.

V. RESULTADOS ESPERADOS

Comparações entre os algoritmos apresentados, assim como suas devidas análises, testes e devidas mudanças ou novas propostas.

REFERÊNCIAS

- [1] V. C. Barbosa, *An Introduction to Distributed Algorithms*. The MIT Press, 1996, ISBN-13: 978-0-262-51442-2.
- [2] A. S. Tanenbaum, *Distributed operating systems*. Prentice Hall, 1995, ISBN-13: 978-0-132-19908-7.
- [3] J.-W. Baek and H.-Y. Yeom, *d-Agent: An Approach to Mobile Agent Planning for Distributed Information Retrieval*. Consumer Electronics, IEEE Transactions on, pp. 115-122, 2003.
- [4] Y. H. K. K. Tsuchiya Takeshi, Mark Lihan, *Distributed Information Retrieval Service for Ubiquitous Services*. Proceedings of IEEE 3rd International Conference on Availability, Reliability and Security, pp. 842-850, 2008.