

# Escalonamento Dinâmico Não-Preemptivo para Sistemas Distribuídos de Impressão

Victoria Ramos Pires, Carolina Marques Fonseca,  
Rafael Nemetz, Mateus Raeder,  
Luiz Gustavo Fernandes

Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)  
Av. Ipiranga, 6681 - Prédio 32 - PPGCC  
{victoria.pires, carolina.fonseca, rafael.nemetz, mateus.raeder}@acad.pucrs.br,  
luiz.fernandes@pucrs.br

## Resumo

*Com o crescimento da personalização de documentos, formatos foram criados para descrevê-los (como o formato PDF, por exemplo). Todavia, tais documentos não são diretamente reconhecidos pelas impressoras, e uma fase de pré-processamento deve ser realizada. Esta fase, entretanto, necessita de muitos recursos computacionais, prejudicando a produção das PSPs (Print Shop Providers). Neste contexto, o presente trabalho objetiva propor melhorias no processo de pré-impressão, escalonando dinamicamente a carga de trabalho nos clusters das PSPs.*

## Introdução

O avanço tecnológico das impressoras industriais (capazes de imprimir grande quantidade de páginas com alta resolução em curtos espaços de tempo) tem por consequência um aumento da demanda pela publicação impressa. Aproveitando-se disto, formatos digitais foram criados com o intuito de facilitar a descrição do conteúdo de documentos a serem publicados, tais como PS (*Post Script*) e PDF (*Portable Document Format*), que apresentam um alto nível de abstração, permitindo a definição de áreas estáticas e dinâmicas de um documento [GFT 06].

Atualmente, a maioria das impressoras digitais não é capaz de interpretar este tipo de documento. Assim, tornam-se necessários procedimentos adicionais que possibilitem sua impressão de forma correta. Tais procedimentos têm como objetivo fornecer ao dispositivo de impressão o documento no formato necessário para que a impressão seja realizada corretamente (formato *bitmap*).

A Figura 1 ilustra de maneira simplificada o fluxo de uma casa de impressão, do cliente ao distribuidor. Desde a especificação do produto até o produto impresso, o *job* passa pelas fases de pré-impressão (preparação do documento para impressão), impressão e pós-impressão (geração do produto final a partir das folhas impressas). Os MIS (*Management Information System*) são máquinas que controlam estes processos.

Neste trabalho será utilizado o formato PDF, devido a sua ampla utilização no contexto da impressão digital e por apresentar diversas características vantajosas. O PDF é obtido através do processo de renderização, processo este altamente custoso, que foi paralelizado em trabalhos anteriores [NRK 09]. A conversão de documentos PDF para arquivos de imagem *bitmap* recebe o nome de rasterização ou *RIPping*

(*Raster Image Processing*) e este processo também demanda alto poder computacional [NGK 09]. Tal fator acaba tornando-se um redutor na produção das casas de impressão, por ser um processo mais lento que os demais.

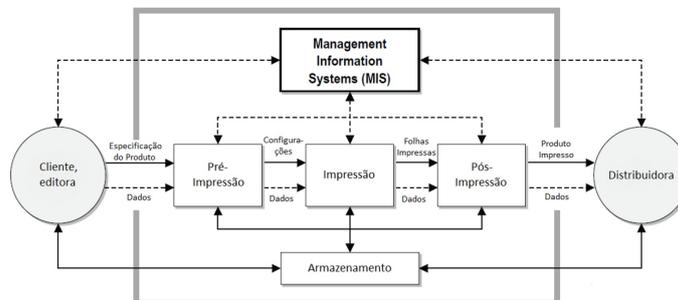


Figura 1: Fluxo de produção em uma casa de impressão [BUC 06].

Para estas etapas, então, casas de impressão possuem diversos *clusters* [FNR 07]. Geralmente, existe mais de um *cluster* responsável pelo processamento (podendo ser comparado a uma pequena grade computacional [BFH 03]) e o escalonamento dos *jobs* ainda é feito de forma manual. Através da análise das características dos *jobs*, a automatização e a otimização do escalonamento e da gerência das filas de documentos que esperam pela rasterização podem trazer um ganho considerável de desempenho. Desta forma, um aumento de velocidade de todo o sistema pode ser alcançado, trazendo benefícios para todo o fluxo de *jobs* em um sistema de impressão.

Neste contexto, existem duas etapas principais para o escalonamento dos *jobs*. A primeira delas é o escalonamento global, que tem por intuito distribuir os *jobs* entre os diferentes *clusters* das casas de impressão. A segunda é o escalonamento local, que tem como escopo um *cluster* específico, distribuindo os *jobs* internamente entre os diferentes núcleos de processamento.

## Escalonamento Global

O escalonamento global é a distribuição de *jobs* (documentos PDF) entre os *clusters* presentes em uma casa de impressão. Todavia, o escalonamento de *jobs* para os *clusters* ainda é uma das tarefas realizadas manualmente. Dependendo das necessidades e urgências no processamento dos *jobs*, estas são repassadas (por usuários) aos diferentes *clusters*. No contexto global, fazer divisões do mesmo *job* para diferentes aglomerados de computadores não é a melhor alternativa, pois suas localizações físicas podem ser distantes e as folhas impressas após o processamento ficariam espalhadas em locais distintos. O escalonador global distribui *jobs* inteiros e é capaz de lidar com as prioridades, restrições de tempo e outras necessidades que possam ser impostas pelos diferentes clientes aos seus *jobs*. A fim de prover informações úteis ao escalonador, um mecanismo de previsão estima o tempo que determinado *job* leva para ser rasterizado em determinado *cluster*. Este mecanismo utiliza a complexidade do *job* e informações sobre recursos disponíveis nos *clusters* como entrada.

A Figura 2 ilustra o funcionamento do sistema de escalonamento, o qual possui três módulos: o próprio escalonador, o mecanismo de previsão (estima o tempo que o *job* levará para ser rasterizado) e o sistema de busca de recursos (aplicação mestre-escravo que busca os recursos disponíveis nos *clusters*). As necessidades dos clientes

a serem consideradas pelo escalonador englobam fatores que podem ser de interesse quando especificadas características do *job*, tais como restrições de tempo e tipo de saída.

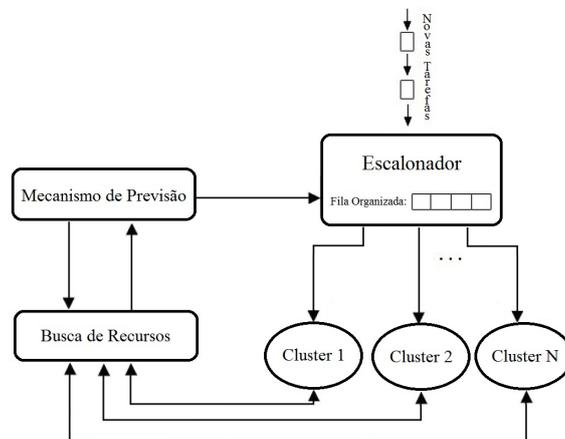


Figura 2: Ilustração simplificada do sistema.

Podem existir, também, máquinas otimizadas para características mais custosas (transparência, por exemplo), podendo ser vantajoso aumentar a carga de *jobs* como estes. Assim, mantendo a fila de *jobs* organizada de forma inteligente, o escalonador distribui os trabalhos de uma forma mais adequada.

## Escalonamento Local

Ao receber *jobs* do escalonador global, primeiramente, realiza-se a análise sobre cada um deles, buscando as características presentes através da ferramenta *Profiler* [NGK 09] [FNK 10]. Esta ferramenta gera um arquivo XML contendo todas as informações do *job*, que será utilizado pela fase seguinte. Na segunda etapa, realiza-se a divisão do *job* de forma balanceada conforme o número de RIPs (máquinas responsáveis pelo processamento). Esta divisão é feita através do Roteador Adaptativo de *Jobs*, o qual tem a capacidade de particionar os *jobs* de forma inteligente, baseando-se em diferentes estratégias.

As estratégias levam em consideração as características existentes, como transparência e reusabilidade. Entende-se por transparência a propriedade que permite que se visualize de forma sobreposta duas ou mais imagens que ocupam regiões iguais ou o fundo do documento através da imagem. Já a reusabilidade ocorre quando instâncias da mesma imagem são reutilizadas mais de uma vez na mesma página ou em páginas diferentes. Tais características são mais custosas do que outras (como quantidade de texto, por exemplo) [NGK 09], e algumas estratégias foram definidas para aproveitar ao máximo os recursos computacionais de forma balanceada. As estratégias, então, agrupam os documentos PDF de acordo com os aspectos que apresentam. São elas: transparência, reusabilidade, mais transparência e menos reusabilidade, menos transparência e mais reusabilidade, sem transparência e sem reusabilidade.

A Figura 3 apresenta o funcionamento do escalonamento local, que possui dois módulos: *Profiler* (analisa a tarefa buscando todas as suas informações) e o Roteador Adaptativo de *Jobs* (particiona os *jobs* de forma inteligente com a melhor estratégia).

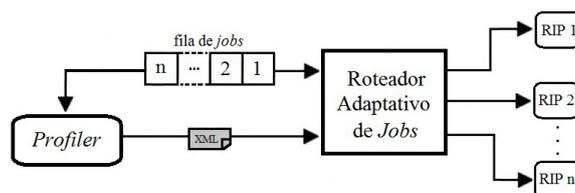


Figura 3: Roteador Adaptativo de jobs.

## Considerações Finais

O processo de impressão de documentos personalizados realizado pelas PSPs trata-se de um *job* computacionalmente custoso. Assim, este trabalho propõe algumas práticas para acelerar este processo, utilizando os *clusters* presentes nas casas de impressão. Primeiramente, um mecanismo de previsão provê informações sobre a estimativa de tempo necessário para conclusão dos *jobs* pelo escalonador de maneira global. Além disto, serve como uma ferramenta de planejamento de capacidade, auxiliando no trabalho de dimensionamento dos *clusters*. Em seguida, dentro de cada *cluster*, os *jobs* recebidos pelo processo de escalonamento global são analisadas localmente dentro de cada *cluster*, utilizando algumas estratégias desenvolvidas, de acordo com as características presentes nos documentos.

Ambas as etapas estão sendo desenvolvidas e testadas. No nível global, alguns testes preliminares mostram um comportamento interessante sobre o mecanismo de previsão, derivando uma equação que aproxima ainda mais o resultado da previsão com a realidade. Já no nível local, as estratégias desenvolvidas foram comparadas com implementações anteriores, e apresentaram um ganho de até 41%. Como próximos passos, as implementações serão finalizadas, e os dois tipos de escalonamento serão integrados, desenvolvendo um sistema completo desde a entrada do documento até sua impressão final.

## Referências

- [GFT 06] Giannetti, F.; Fernandes, L. G.; Timmers, R.; Nunes, T.; Raeder, M.; Castro, M.. High performance XSL-FO rendering for variable data printing. In: ACM-SAC'06: Proceedings of the 2006 ACM Symposium on Applied Computing. Nova Iorque, NY, EUA: ACM Press, 2006. p. 811-817.
- [NRK 09] Nunes, T.; Raeder, M.; Kolberg, M.; Fernandes, L. G.; Cabeda, A.; Giannetti, F. High Performance Printing: Increasing Personalized Documents Rendering through PPML Jobs Profiling and Scheduling. Computational Science and Engineering, IEEE International Conference on, vol. 1, pp. 285-291, 2009 International Conference on Computational Science and Engineering, 2009.
- [NGK 09] Nunes, T., Giannetti, F., Kolberg, M., Nemetz, R., Cabeda, A., Fernandes, L. G. Job Profiling in High Performance Printing. In: 9th ACM Symposium on Document Engineering, 2009, Munique. Proceedings of the 9th ACM Symposium on Document Engineering (ACM DOCENG). Nova Iorque, NY (EUA): ACM Press, 2009. p. 109-118.
- [BUC 06] Buckwalter, C. Integrating Systems in the Print Production Workflow: Aspects of Implementing JDF. Digital Media Division, 2006.
- [FNR 07] Fernandes, L. G.; Nunes, T.; Raeder, M.; Giannetti, F.; Cabeda, A.; Bedin, G. An Improved Parallel XSL-FO Rendering for Personalized Documents. In: Euro PVM/MPI'07. Recent Advances in Parallel Virtual Machine and Message Passing Interface (LNCS). Heidelberg: Springer Berlin, 2007. v.4757. p.56-63.
- [BFH 03] Berman, F.; Fox, G.; Hey, A. J.. Grid Computing: Making the Global Infrastructure a Reality. John Wiley and Sons, Inc. 2003.
- [FNK 10] Fernandes, L. G.; Nunes, T.; Kolberg, M.; Giannetti, F.; Nemetz, R.; Cabeda, A.. Job Profiling and Queue Management in High Performance Printing. Computer Science - Research and Development, v. 26, p. 1-20, 2010.