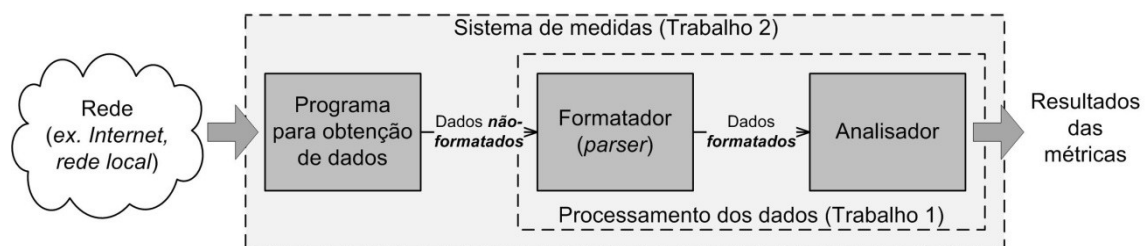


=== Trabalho ===

Este trabalho é dividido em duas partes, cada uma com data de entrega e avaliação separadas. O objetivo é criar um sistema de medidas capaz de obter dados não necessariamente formatados e extrair desses dados as métricas de interesse. Para isso, é preciso criar um sistema como o ilustrado abaixo:



O primeiro passo para a realização deste trabalho é definir qual problema investigar. Uma vez definido o problema, pode-se determinar que tipo de informações serão obtidas da rede e que tipo de métricas serão calculadas. Nesse sentido, recomenda-se ler artigos técnicos de conferências e revistas de medidas em rede (p.ex., a conferência ACM Internet Measurement Conference - IMC). Um artigo interessante da conferência IMC que pode ser usado de inspiração é o seguinte:

Liang Wang, Antonio Nappa, Juan Caballero, Thomas Ristenpart, and Aditya Akella, "WhoWas: A Platform for Measuring Web Deployments on IaaS Clouds", ACM Internet Measurement Conference (IMC'14), pages 101-114, 2014.

O passo preliminar de escolha do problema a investigar deve ser executado rapidamente e a escolha deve ser comunicada ao Professor em aula. **O prazo é até o dia 10/04/15.**

Após a definição do problema, segue a etapa de desenvolvimento do sistema de medidas. **No Trabalho 1, deve-se criar o bloco de processamento de dados.** Esse bloco assume como entrada um arquivo de dados não-formatado, ou seja, um arquivo que não possui estrutura de organização de dados bem definida, e oferece como saída as métricas de interesse. Portanto, o primeiro passo é escrever um programa que

formate um arquivo de entrada não-formatado e o transforme em um arquivo formatado. Note que esse procedimento consiste em definir quais informações são importantes que estejam presentes no arquivo formatado. Como exemplo, podemos ter um arquivo contendo informações diversas de todos os sites visitados em um dia. Já o arquivo formatado conterá apenas as informações sobre o endereço IP dos servidores web e as datas das últimas atualizações das páginas. O segundo passo é implementar o programa analisador para calcular as métricas de interesse usando como entrada o arquivo formatado. Novamente, considerando o exemplo do arquivo contendo endereços IP de servidores web e data de modificação das páginas, uma métrica de interesse poderia ser a relação das faixas de endereço IP que mais realizam atualizações em suas páginas.

Note que tanto o programa formatador quanto o analisador devem ser implementados em C++, podendo incluir bibliotecas ou APIs de outras linguagens de programação. No caso do programa formatador, o programa pode realizar análise de expressão regular ou usar APIs de outras linguagens de programação que façam esse tipo de análise nativamente (p.ex., Perl). O arquivo de entrada pode ser um pronto, sem que haja ainda um programa para obtê-lo. **O prazo de entrega do programa funcionando e de um relatório sucinto explicando o problema abordado e um guia de uso do programa é até o dia 06/05/2014.**

No Trabalho 2, o programa para obtenção dos dados deve ser desenvolvido. Este programa deve ser capaz de gerar um arquivo de dados não-formatado através de requisições a algum tipo de servidor em rede. O programa deve ser escrito também em C++ e pode usar bibliotecas ou APIs de outras linguagens. **O prazo de entrega do sistema completo (programas do Trabalho 1 e 2) mais um relatório sucinto explicando o problema abordado, um guia de uso completo do sistema e os resultados das métricas investigadas é até o dia 03/06/2014. O relatório desenvolvido para a entrega do Trabalho 1 pode ser reaproveitado.**

Obs.: Os programas deverão ser entregues em formato digital contendo todos os arquivos necessários para a sua correta compilação e execução. Um modelo para elaboração de relatórios pode ser encontrado em <http://www.gta.ufrj.br/~miguel/docs/lingprog/modeloRelatorio.pdf>. Os alunos poderão ser escolhidos para explicar oralmente o trabalho realizado. **Os trabalhos copiados receberão nota ZERO. O trabalho é individual.**
