Scaling Forensics to the Jungle

Pieter Hijma\textsuperscript{1} \hspace{1cm} Ben van Werkhoven\textsuperscript{2} \hspace{1cm} Jason Maassen\textsuperscript{2} 
Zeno Geradts\textsuperscript{3} \hspace{1cm} Ceriel Jacobs\textsuperscript{4}

1University of Amsterdam \hspace{1cm} p.hijma@uva.nl 
2Netherlands eScience Center \hspace{1cm} \{b.vanwerkhoven,j.maassen\}@esciencecenter.nl 
3Netherlands Forensic Institute \hspace{1cm} z.geradts@nfi.minvenj.nl 
4Vrije Universiteit Amsterdam \hspace{1cm} ceriel@cs.vu.nl

**Problem**

Hansken, the main forensics tool for the NFI, is very successful. However, the expected growth in data volumes and the need for deep analysis of multimedia traces put increased demands on the Hansken platform, especially when considering deadlines to keep a suspect in custody.

**Goal**

In this project we aim to realize a Jungle Computing (computing on many different types of hardware) enabled version of Hansken to increase scalability for future demands in the forensics domain.

**Architecture**

- PRNU Compare
- MCL
- Cashmere
- Constellation

PRNU Compare identifies camera's based on the noise patterns of images. Photo-response non-uniformity (PRNU) noise is caused by the variation in sensitivity of pixels to light. Since this application is compute-intensive and potentially a large dataset of images needs to be investigated, it is a good target to accelerate with many-cores and distribute on a cluster.

The computational kernels of PRNU are written in the Many-Core Levels (MCL) language. MCL allows kernels to be written on multiple levels of abstraction. The lower the abstraction level, the more control the programmer has over performance. The challenge is to reason about optimizations of a pipeline of kernels.

Cashmere is a library that efficiently schedules MCL kernels on heterogeneous many-core clusters: MCL generates kernels, optimized for different many-core devices with different performance capabilities. Cashmere ensures that computation is overlapped with data-transfers.

Constellation is a run-time system that allows programmers to define in a precise way how an application is mapped to the hardware. This allows programmers to map certain tasks to certain nodes, keep data in certain nodes and move the computation to those nodes.

**eScience Projects**

- A Jungle Computing Approach to Large-Scale Online Forensics Analysis - Henri Bal
- Generic eScience Technologies - Cees de Laat