International Workshop on *Intelligent Multi-core Systems*

Amit Kumar Singh

School of Computer Science and Electronic Engineering

University of Essex

United Kingdom

W: http://aksingh.co.uk/ E: a.k.singh@essex.ac.uk

Intelligence to Address Computation and Communication Issues

Already Discussed - Intelligence

- For computation
  - Homogeneously distribute loads of tasks on the cores
  - Bring heterogeneous cores or ASIC accelerators
  - Apply dynamic voltage and frequency scaling (DVFS)
  - Offloading to cloud server

- For communication
  - Homogeneously distribute communication loads on the links
  - Bring heterogeneous links
  - Apply dynamic voltage and frequency scaling (DVFS)

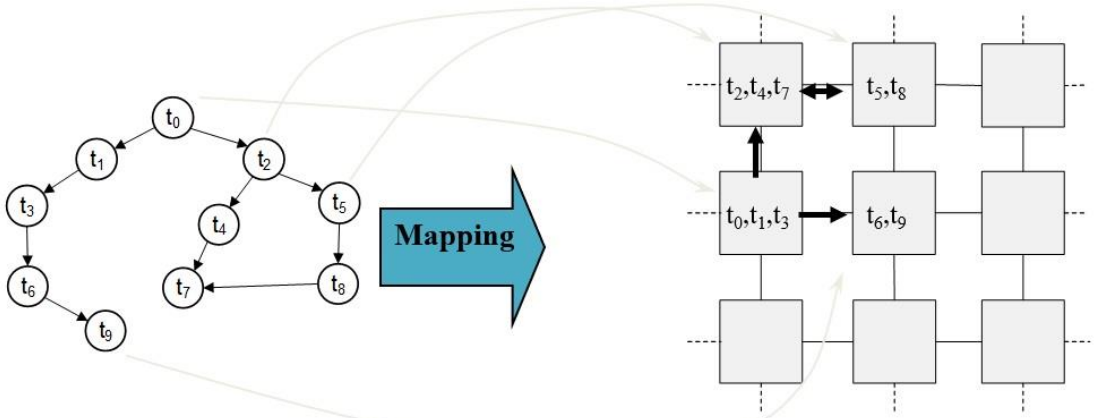
Computation and Communication Consideration

Cost function – to homogeneously distribute loads of tasks and communications on the cores and links, respectively

$$cost(c,t)=c_1*P(c,t) + c_2*L(c,t)$$

P(c,t): the normalized processor load when assigning task t to core c;

L(c,t): the ratio of allocated links on core c when t is assigned to c;



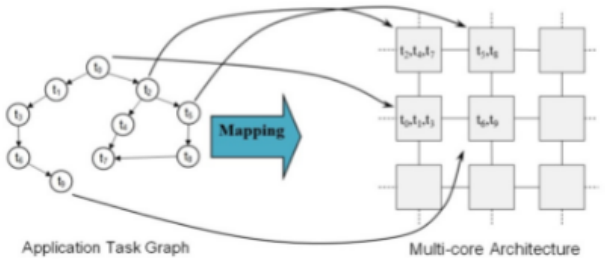
What will be the change in the cost function if independent tasks are considered?

Cost function – to homogeneously distribute loads of tasks and communications on the cores and links, respectively

$$cost(c,t)=c_1*P(c,t) + c_2*L(c,t)$$

$P(c,t)$: the normalized processor load when assigning task t to core c;

$L(c,t)$: the ratio of allocated links on core c when t is assigned to c;



4

No P(c,t) term

No L(c,t) term

Same as earlier, i.e. keep both terms

None of the above

Heuristic-based Intelligence

```
Sort all tasks on criticality, descending;
for all sorted tasks
        Sort all cores on cost(c, t), ascending
        for all sorted cores
                if assigning task t to core c is feasible
                        Assign task t to core c;
                        Assign connections to/from t; break;
                end
        end
        if task is not assigned
                return "Unable to find feasible assignment";
        end
end
```

Which line(s) to remove if tasks are independent?

Heuristic-based Intelligence

```
Sort all tasks on criticality, descending;
for all sorted tasks
        Sort all cores on cost(c, t), ascending
        for all sorted cores
                if assigning task t to core c is feasible
                        Assign task t to core c;
                        Assign connections to/from t; break;
                end
        end
        if task is not assigned
                return "Unable to find feasible assignment";
        end
end
```

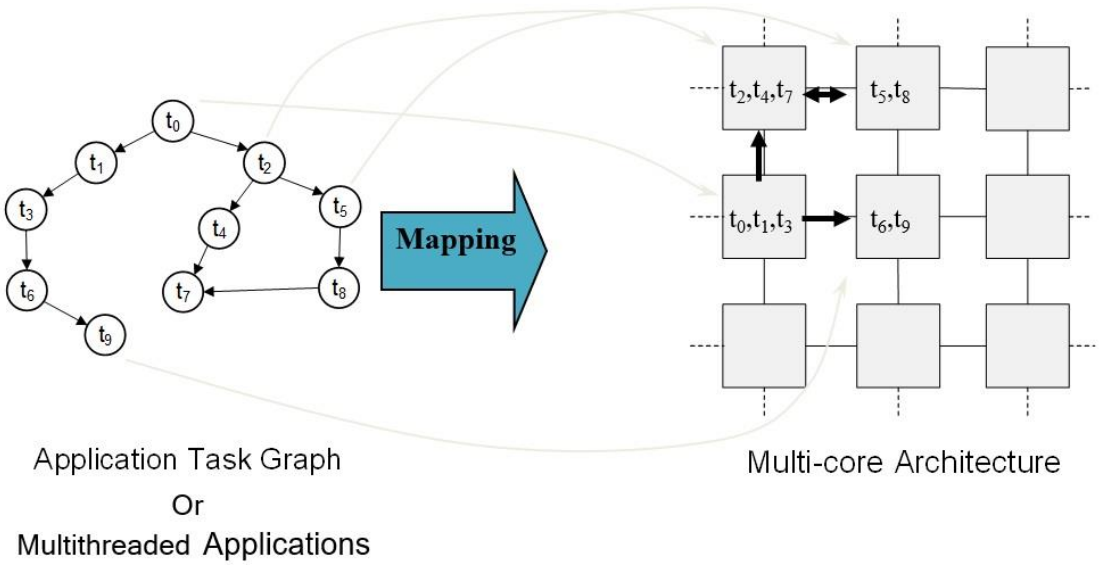
Heterogeneity Consideration

Cost function – to homogeneously distribute loads of tasks and communications on the **heterogeneous** cores and **heterogeneous** links, respectively

$$cost(c,t)=c_1*P(c,t) + c_2*L(c,t)$$

P(c,t): the normalized processor load when assigning task t to core c;

L(c,t): **the normalised average latency of all the links to/from t** when t is assigned to c;



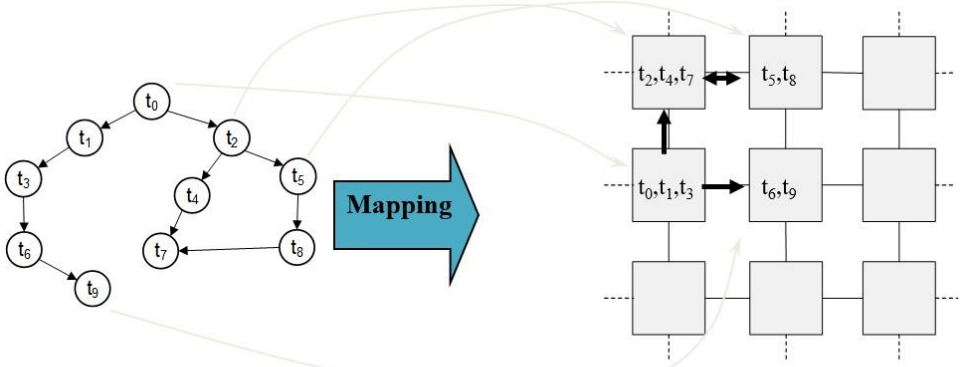
Heterogeneity+DVFS Consideration

Cost function – to homogeneously distribute loads of tasks and communications on **DVFS supported heterogeneous** cores and **heterogeneous** links, respectively

$$cost(c,t)=c_1*P(c,t) + c_2*L(c,t)$$

P(c,t): the normalized processor load when assigning task t to core c **that operates at certain V/f level**;

L(c,t): the normalised average latency of all the links to/from t when t is assigned to c, **and links operate at certain V/f level**;



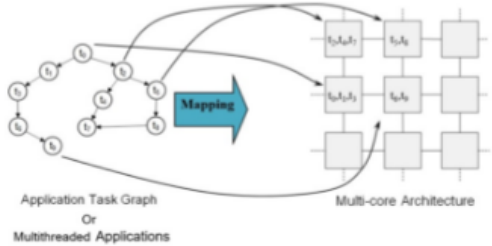
How to consider reliability and security?

Cost function – to homogeneously distribute loads of tasks and communications on **DVFS supported heterogeneous** cores and **heterogeneous** links, respectively

$$cost(c,t)=c_1*P(c,t) + c_2*L(c,t)$$

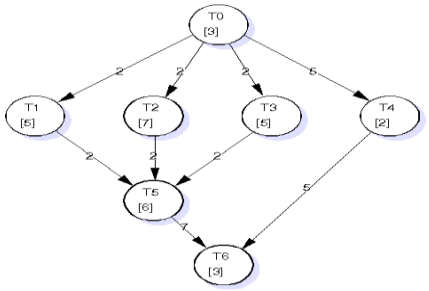
$P(c,t)$: the normalized processor load when assigning task t to core c **that operates at certain V/f level**;

$L(c,t)$: **the normalised average latency of all the links to/from t** when t is assigned to c, **and links operate at certain V/f level**;

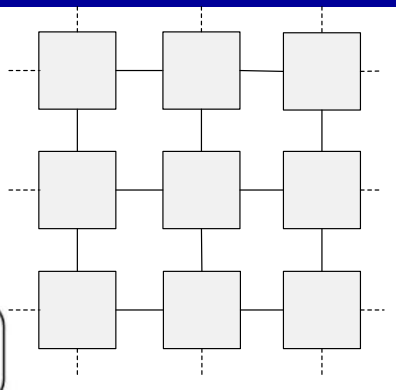


9

Genetic Algorithm (GA) based Intelligence

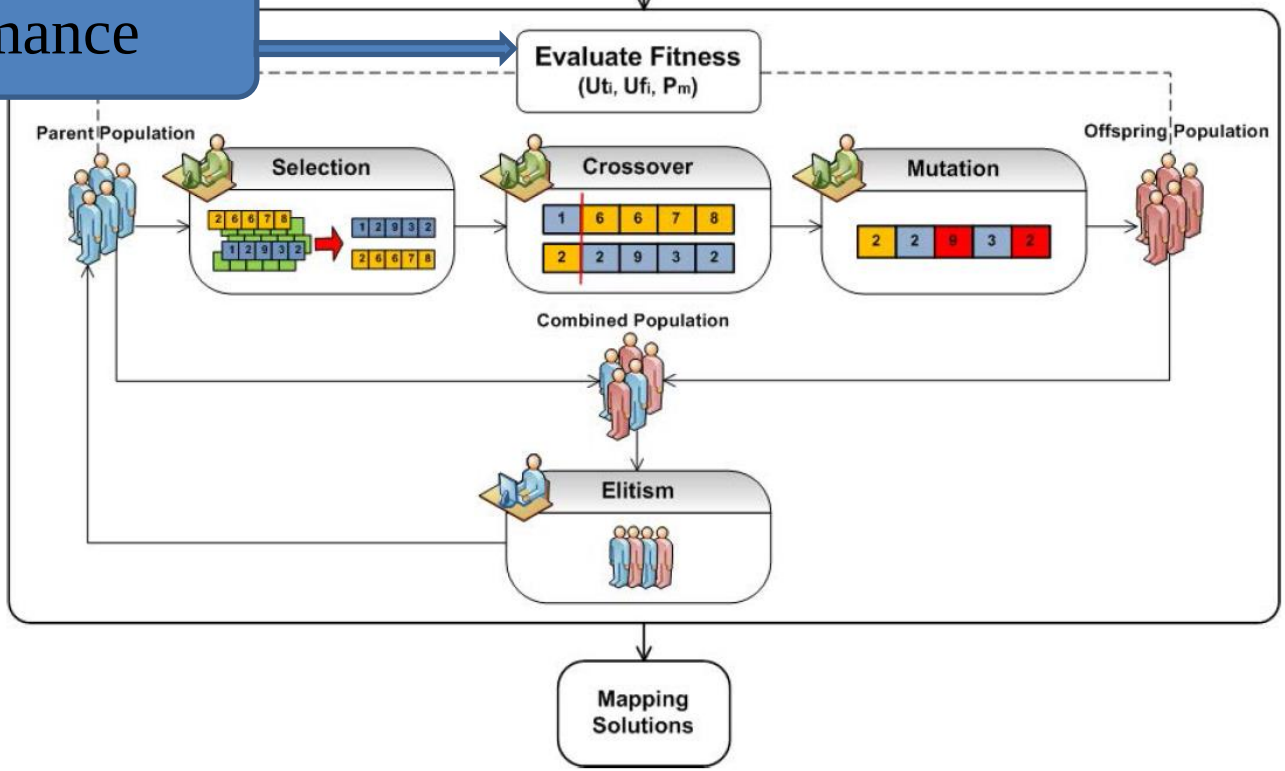


Typical GA



Multi-core Architecture

Performance



What do you think is the problem with covered heuristic and GA based Intelligence?

Problems-Summary

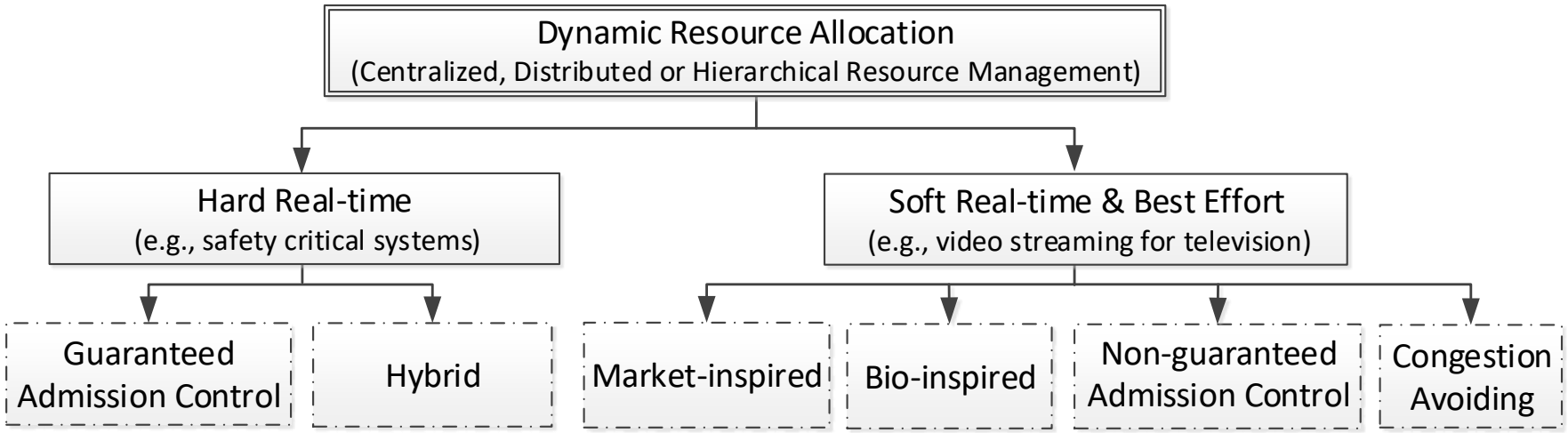
- Cannot handle time varying workload efficiently
  - It could be by a single application or varying number of applications over the time.
- Cannot find efficient solution quickly for large problem sizes.
- Cannot handle environmental changes.
- …

Possible Solution(s)

- Online heuristic-based intelligence
- Online (machine) learning-based intelligence
  - Reinforcement Learning

- Online Adaptations needed in all types of solutions

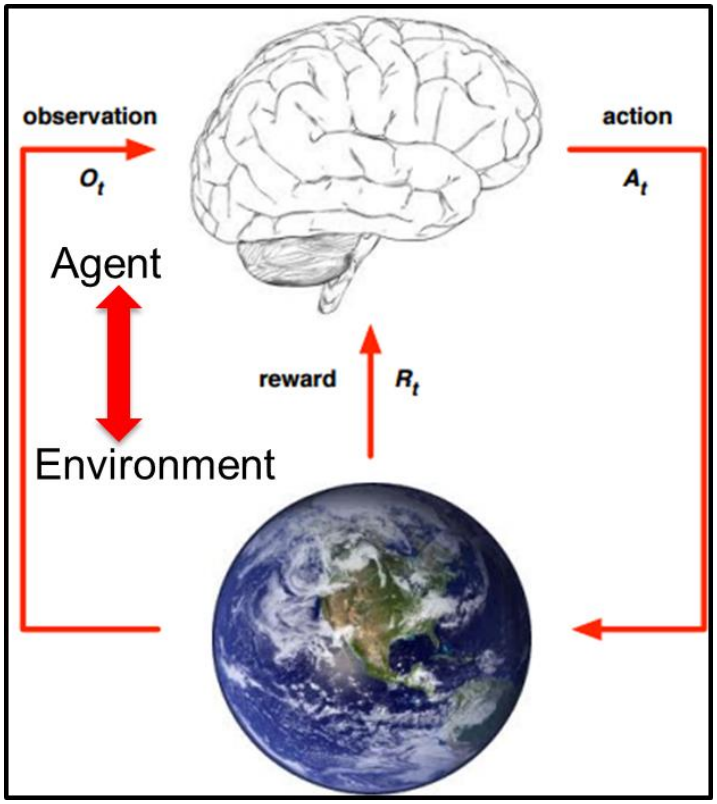
Online heuristic-based intelligence



A Survey and Comparative Study of Hard and Soft Real-time Dynamic Resource Allocation Strategies for Multi/Many-core Systems
Amit Kumar Singh, Piotr Dziurzanski, Hashan Roshantha Mendis, Leandro Soares Indrusiak
ACM Computing Surveys (CSUR), 2017

Spectrum of Run-time Management for Modern and Next Generation Multi/Many-core Systems
Amit Kumar Singh, Geoff Merrett, Amir Rahmani, Akash Kumar
IEEE/ACM Embedded Systems Week (ESWeek) – Tutorial, 2018

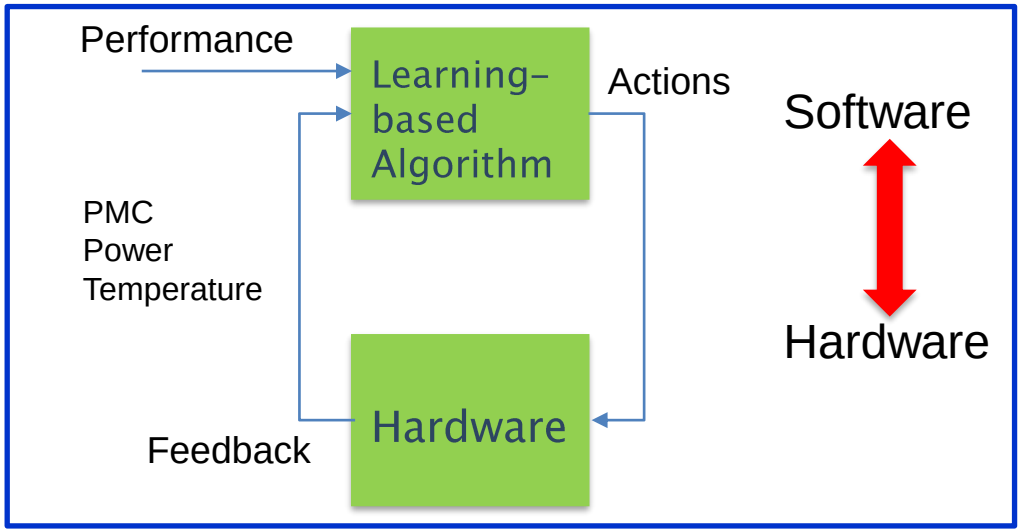
Online (machine) learning-based intelligence



- States (workloads),
- Actions (V-F pairs),
- Payoffs (Rewards/Penalty)

- Observes the current system state
- Selects an action
- Changes the state
- Leads to a payoff

Reinforcement learning (RL)



Challenges for online ML-based Intelligence

- Need to take accurate system status into account
- May take long time to find a good solution when number of states and actions are quite high
- It may not find an efficient solution during the application execution
- Doesn’t scale well with heterogeneity and number of cores

Possible Solutions

- Support with offline learning or training
  - Approach that ingests all the data at one time to build a model, e.g. CNN.
- Carefully consider (limited) number of states in online learning process

Other possible solution(s) or consideration(s)?

Summarising

- Intelligence to address computation and communication issues in multi-core systems
  - Homogeneous workload distribution
  - Heterogeneity consideration
  - DVFS consideration

- Compute intensive heuristic and GA based Intelligence
  - Not suitable for dynamic workloads

- Online light-weight heuristic
- Online ML based Intelligence
  - Next-> Possible solutions

Further Questions?