Combining Heuristics for Optimizing and Scaling the Placement of IoT applications in the Fog

Ye XIA[1,2]

supervisors: Thierry COUPAYE[1] Frédéric DESPREZ[2]
Xavier ETCHEVERS[1] Loïc LETONDEUR[1] Adrien LEBRE[2]

[1]Orange Labs
first.last@orange.com

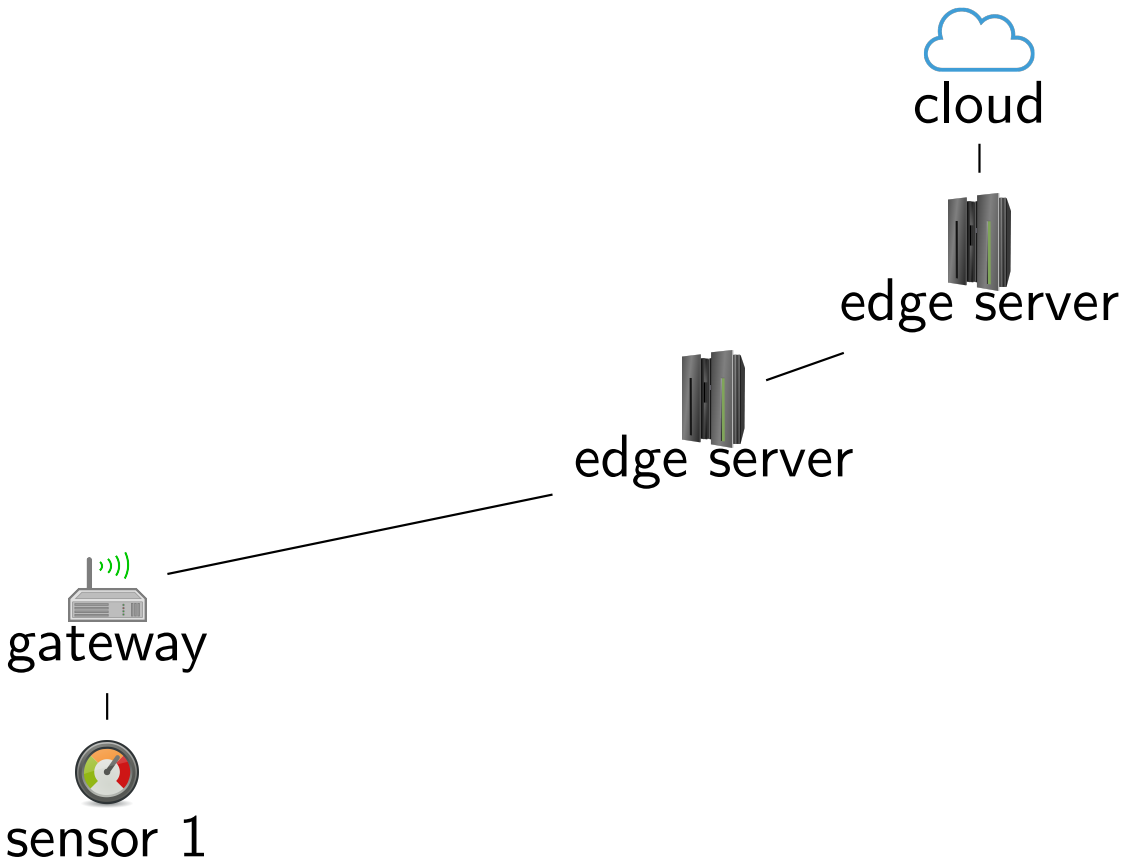
[2]INRIA
first.last@inria.fr

19 / 06 / 2018

Outline

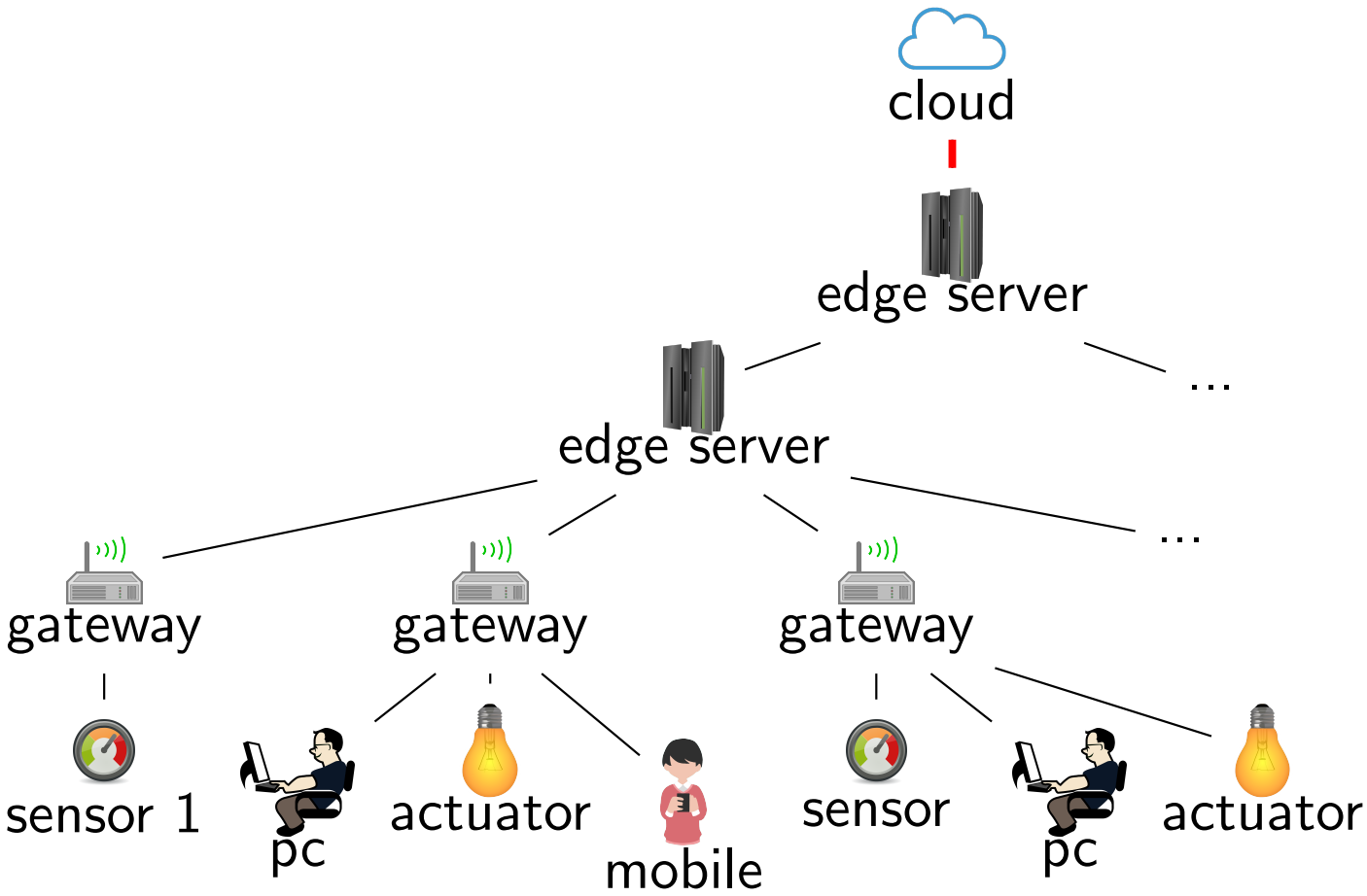
from Cloud to Fog

How to satisfy a time-sensitive IoT application ?

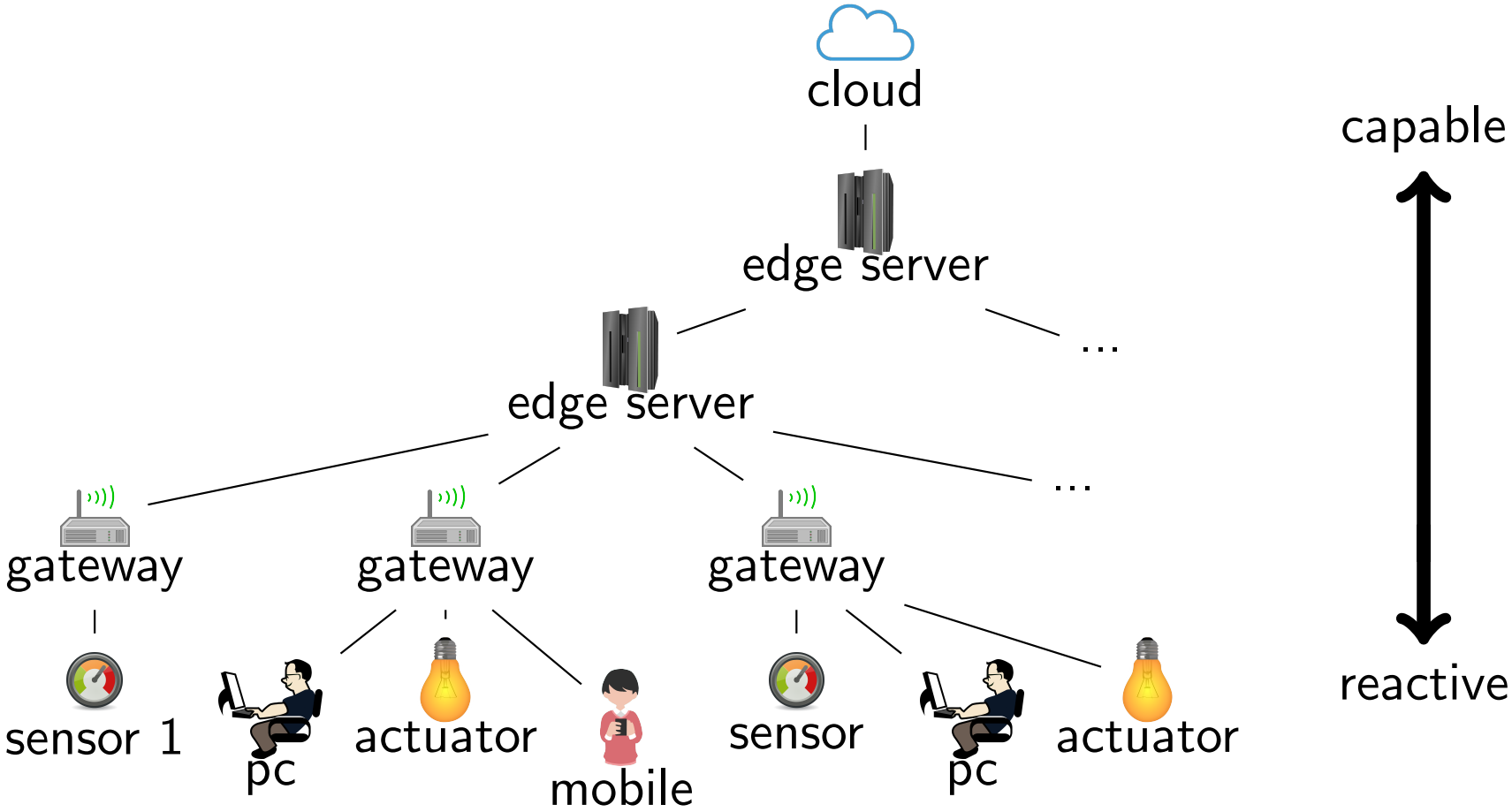


from Cloud to Fog

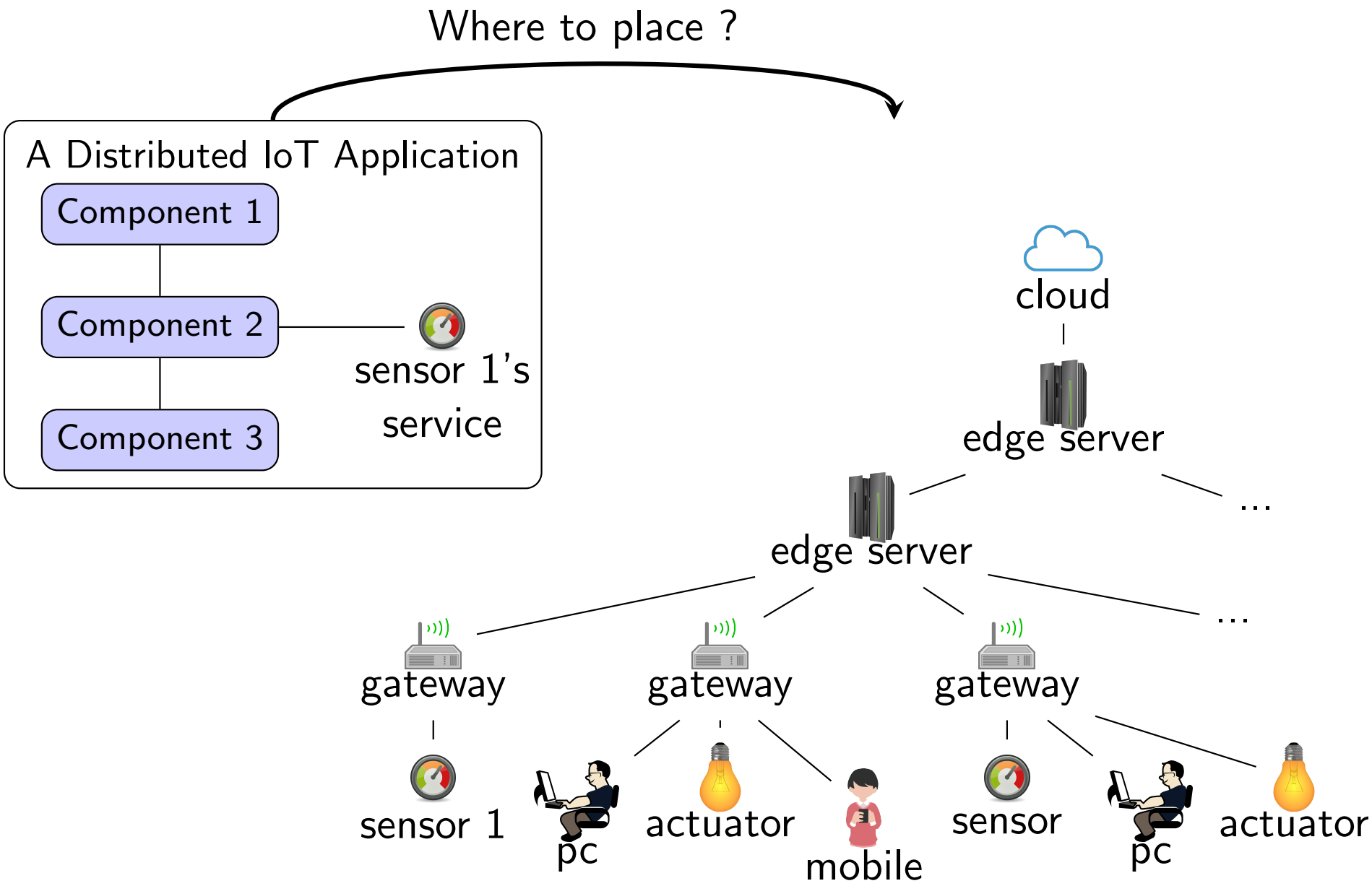
How to avoid congestions in the core network ?



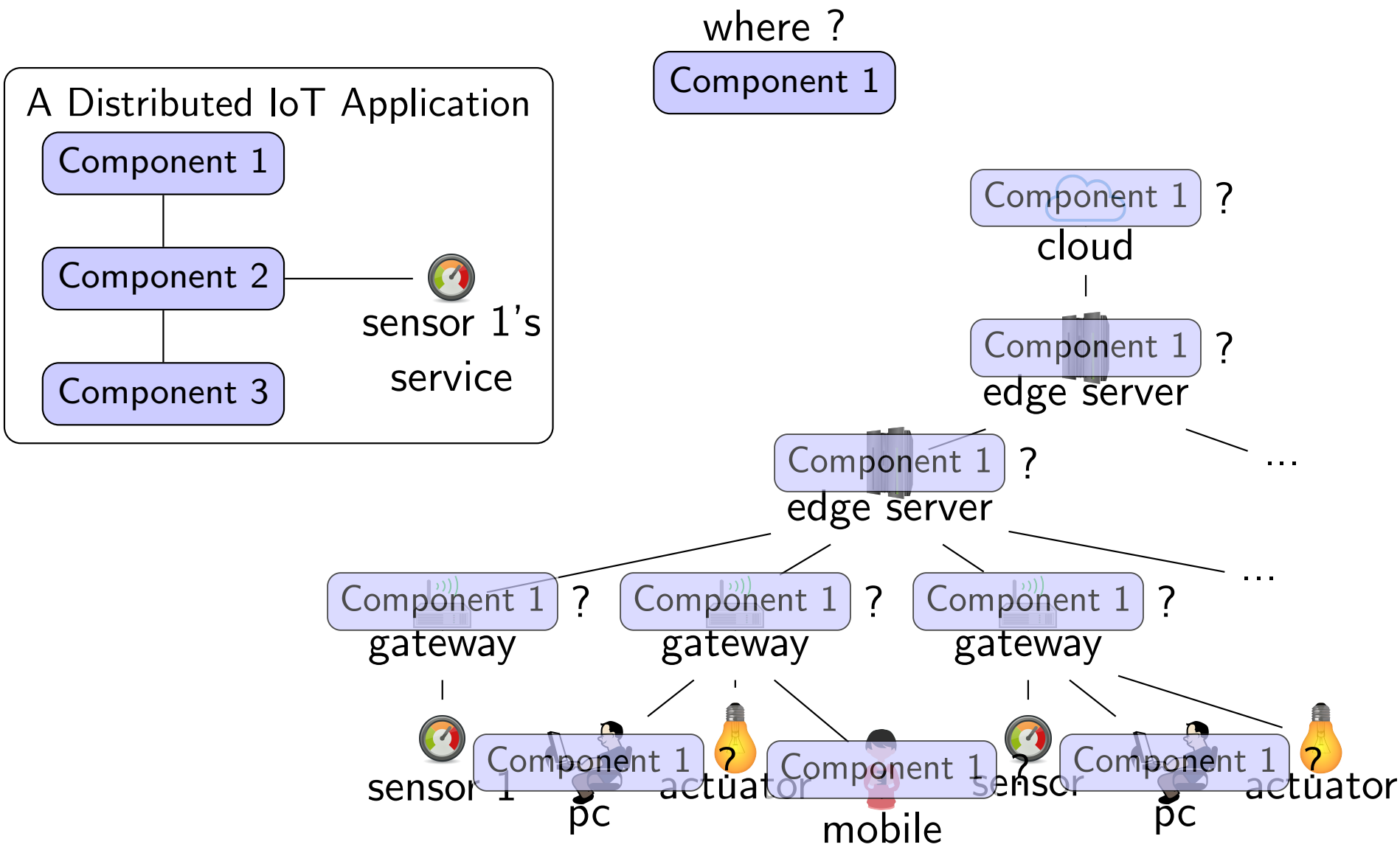
from Cloud to Fog



Placement Problem Description



Placement Problem Description



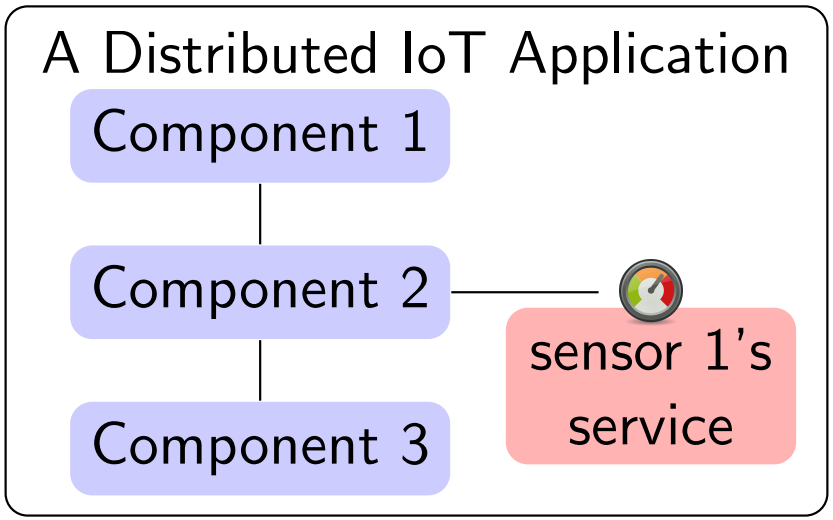
Outline

Model — Applications

Apps : the set of applications to place

Application

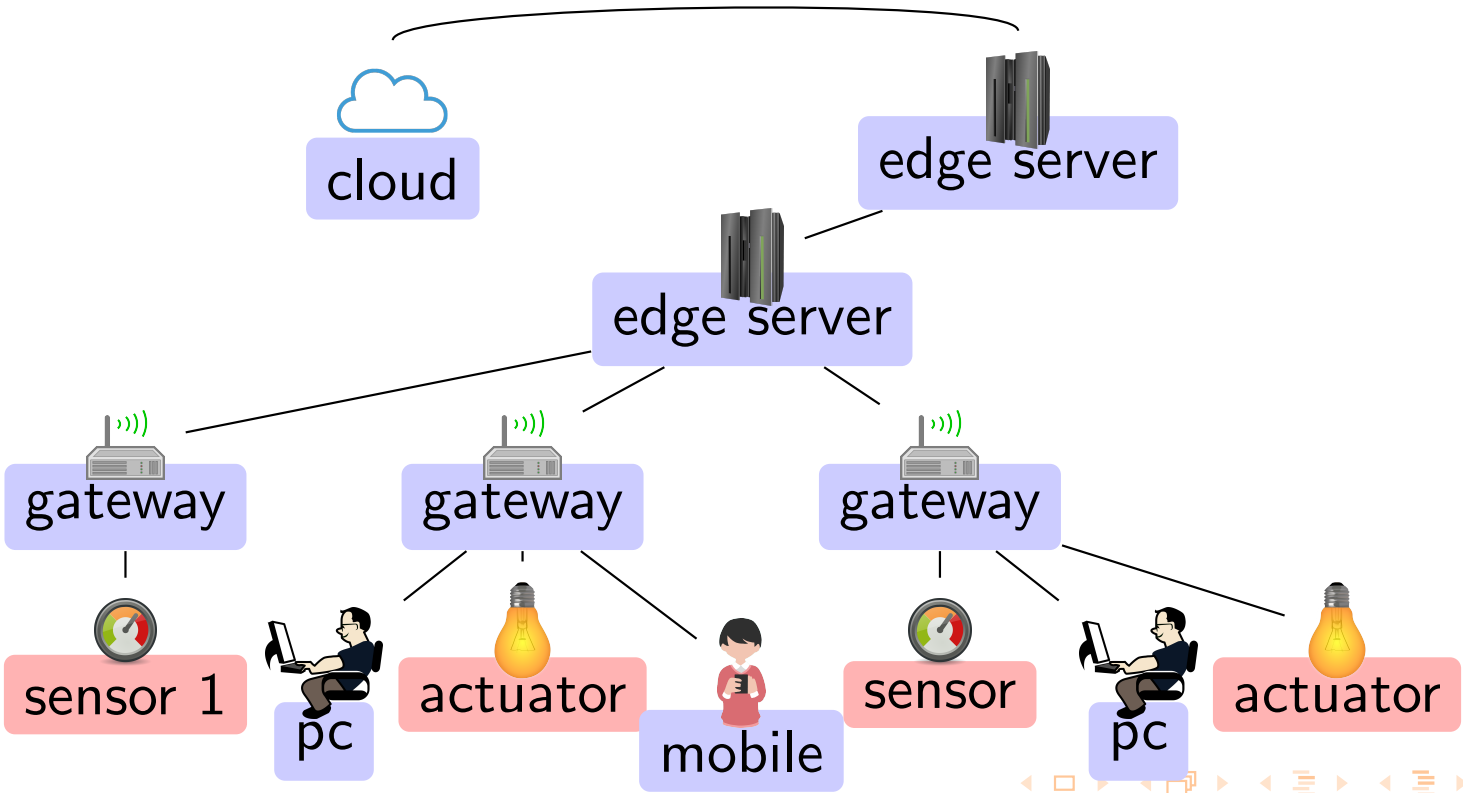
- component : reqCPU, reqRAM, reqDISK
- pinned service (i.e., services provided by sensors / actuators)
- binding (i.e., communication channel) : reqBW, maximal latency



Model — Infrastructure

Infrastructure

- fog node : CPU, RAM, DISK
- appliance (i.e., sensors / actuators without hosting features)
- link : network latency (Lat), bandwidth (BW)



Constraints

Placement

- each component -≻ one fog node

$$a\ placement\ = \begin{pmatrix} comp_1, node_i \\ comp_2, node_j \\ \dots \\ comp_n, node_k \end{pmatrix}$$

A Solution must conform to :

- each Fog Node's CPU / RAM / DISK consumption
- each Link's Bandwidth consumption
- each Binding's Maximal Latency

Solution Selection

How to select one solution as the final placement decision?
Optimization Objective : minimizing applications' response times

Objective Function

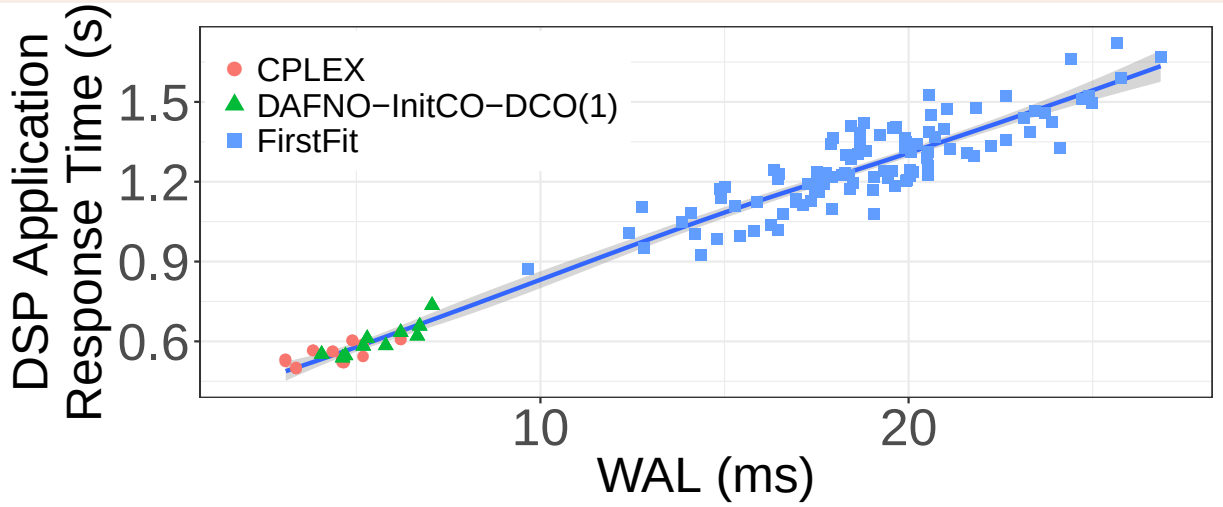
min : Weighted Average Latency (WAL)

$$WAL = \sum_{bind} \frac{bind.reqBW}{total\_BW} \times bind.Lat$$

Objective Function Evaluation

SIMGRID [3]

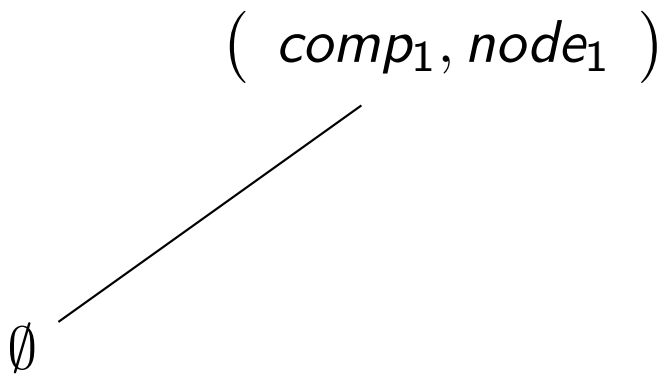
correlation : 0.97



Outline

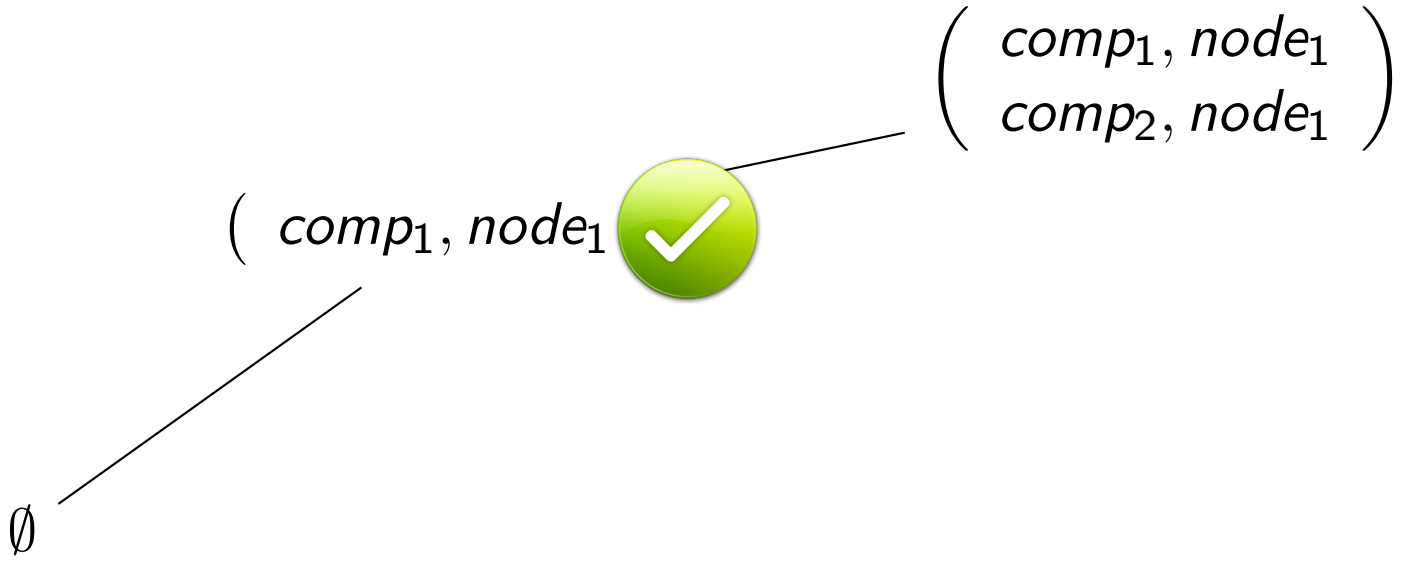
Example – Backtrack Search-Based FirstFit Algorithm

$\{comp_1, comp_2\}$ -$\succ$ $\{node_1, node_2\}$



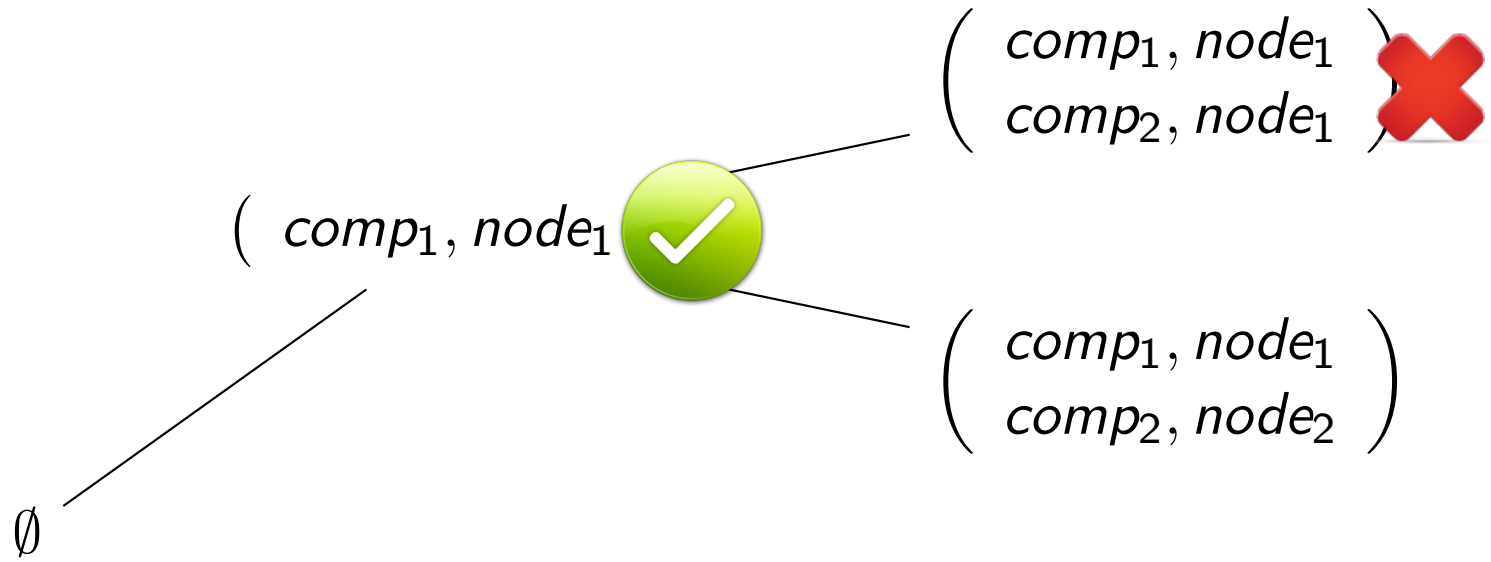
Example – Backtrack Search-Based FirstFit Algorithm

$\{comp_1, comp_2\}$ -$\succ$ $\{node_1, node_2\}$



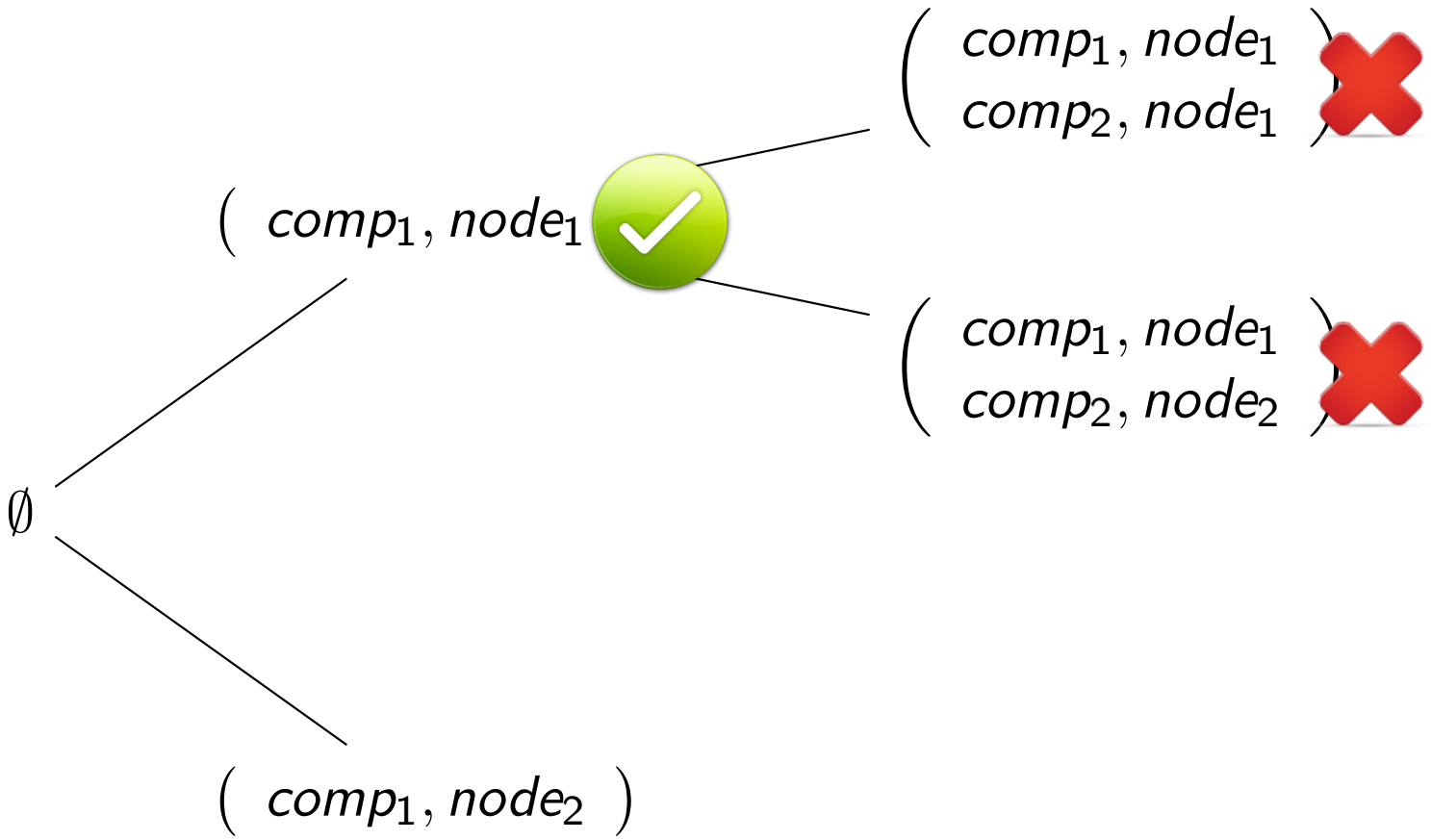
Example – Backtrack Search-Based FirstFit Algorithm

$\{comp_1, comp_2\}$ -$\succ$ $\{node_1, node_2\}$



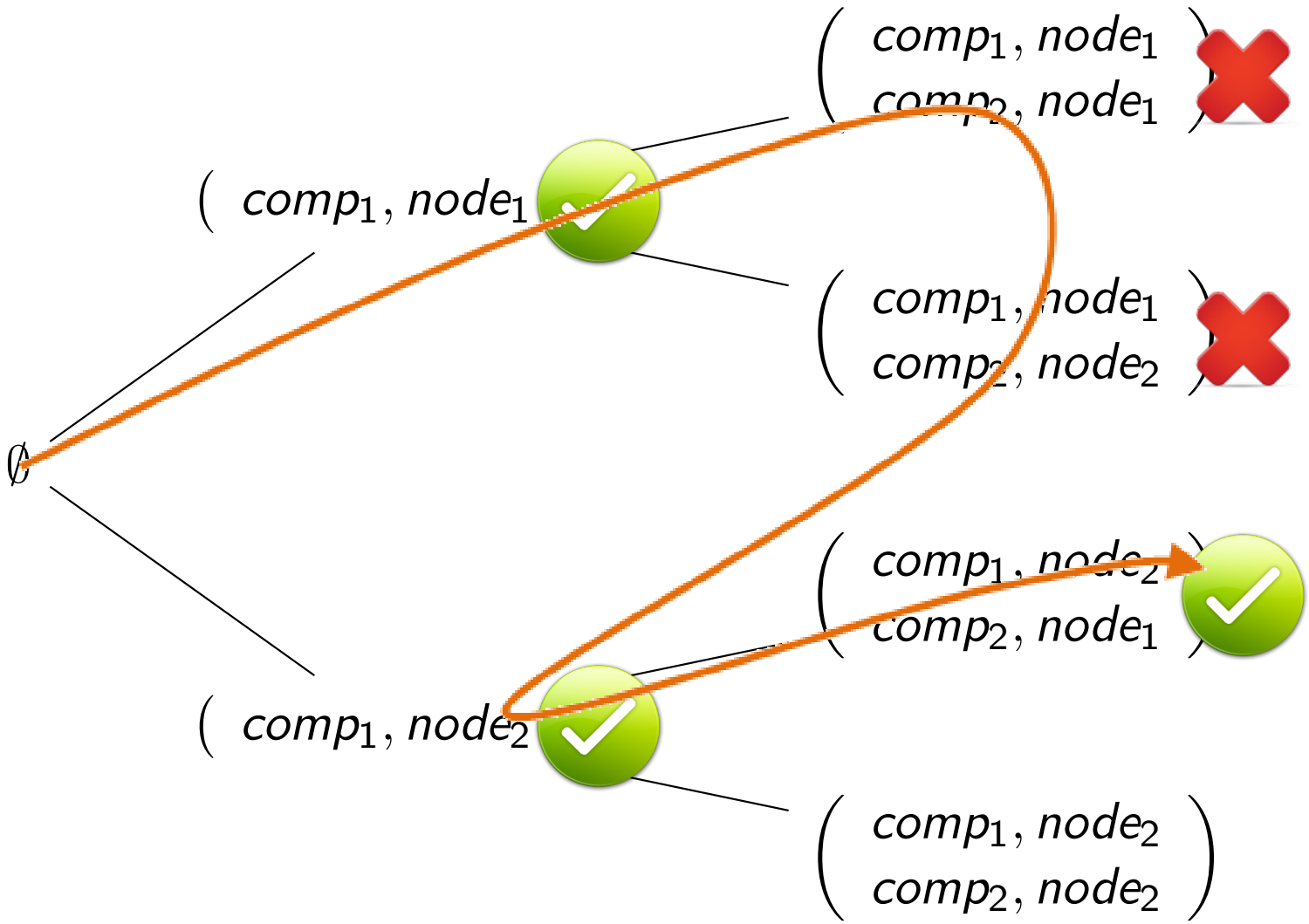
Example – Backtrack Search-Based FirstFit Algorithm

$\{comp_1, comp_2\}$ -$\succ$ $\{node_1, node_2\}$



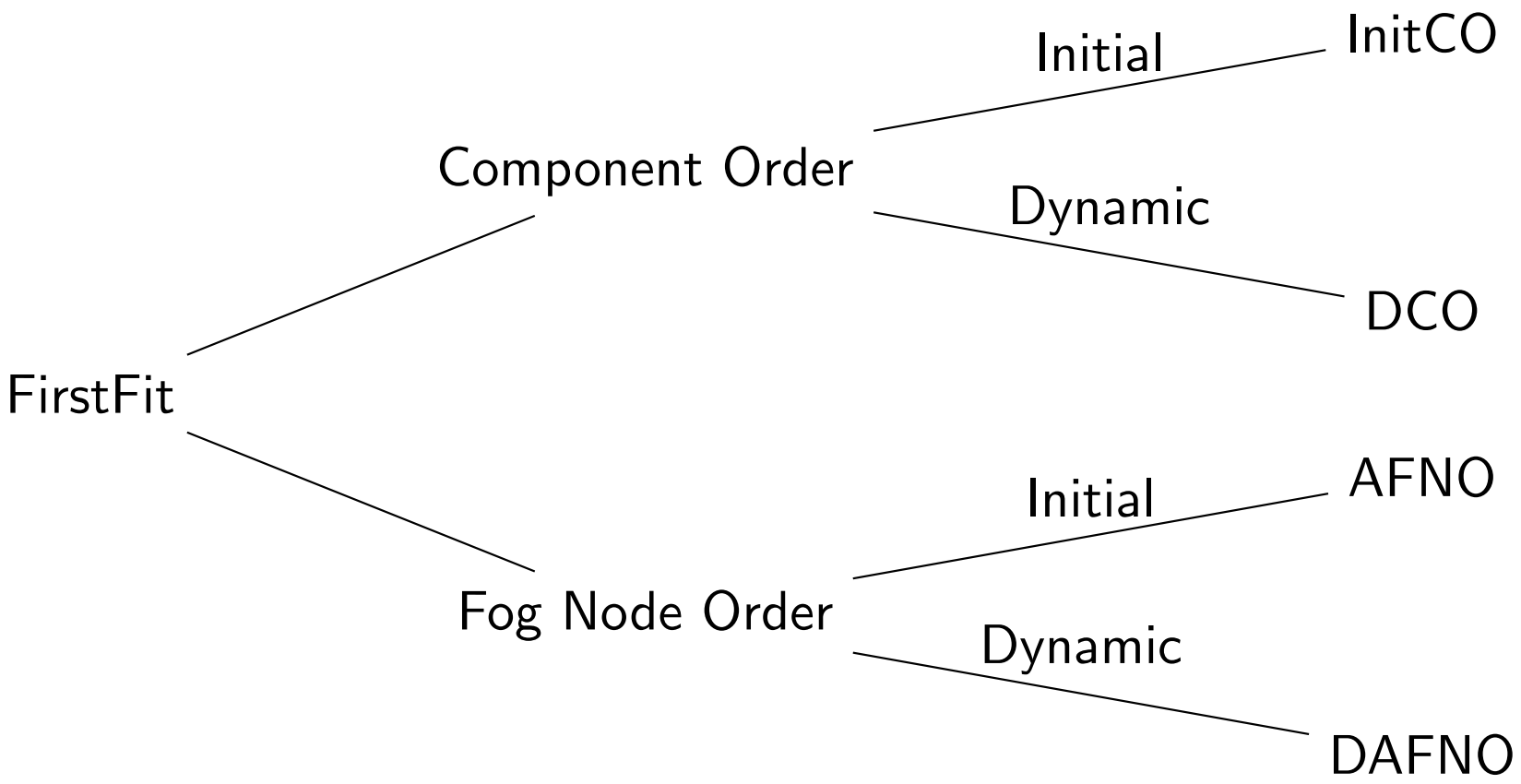
Example – Backtrack Search-Based FirstFit Algorithm

$\{comp_1, comp_2\}$ -$\succ$ $\{node_1, node_2\}$



Outline

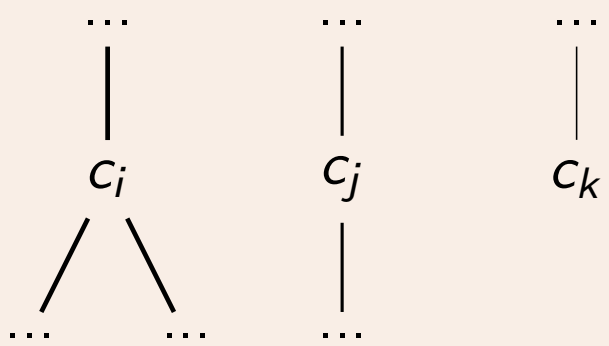
Heuristics Positioning



Components Ordering-Based Heuristics

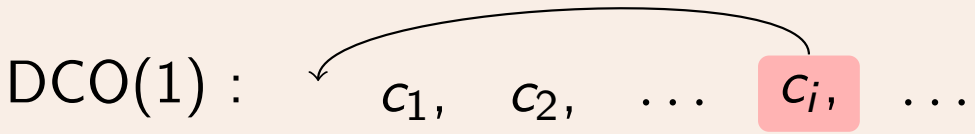
Initial Components Ordering (InitCO)

- sort components in descending order of their bandwidth requirements



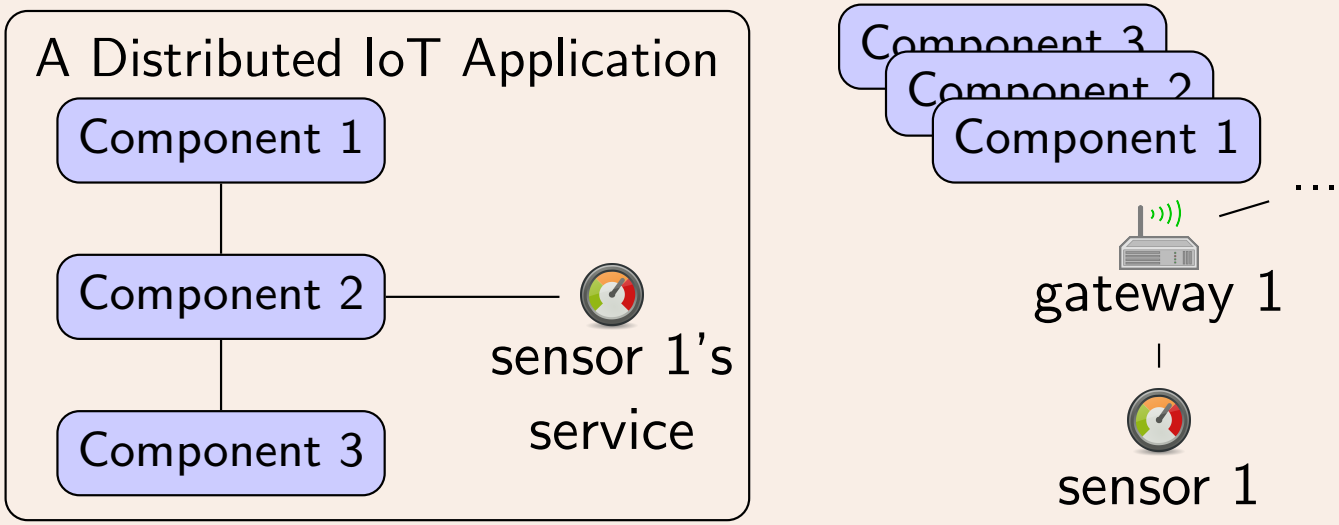
Dynamic Components Ordering (DCO)

- when failing to place a component c, set c as the first to place
- adjust components' order for deriving to a direct searching trajectory



Fog Nodes Ordering-Based Heuristics

Anchor-based Fog Nodes Ordering (AFNO)



- a component's anchor : the fog node minimizing WAL without considering constraints
- for placing a component, test priorly fog nodes close to its anchor

Dynamic Anchor-based Fog Nodes Ordering (DAFNO)

- dynamically update anchors during the search

Outline

Small-Scale Problems

To compare with CPLEX, which guarantees to return the optimal solution.

Evaluation Setup

- infra : randomly generated infrastructures with 27 fog nodes
- app : a Data Stream Processing application with 16 components

Evaluation Result

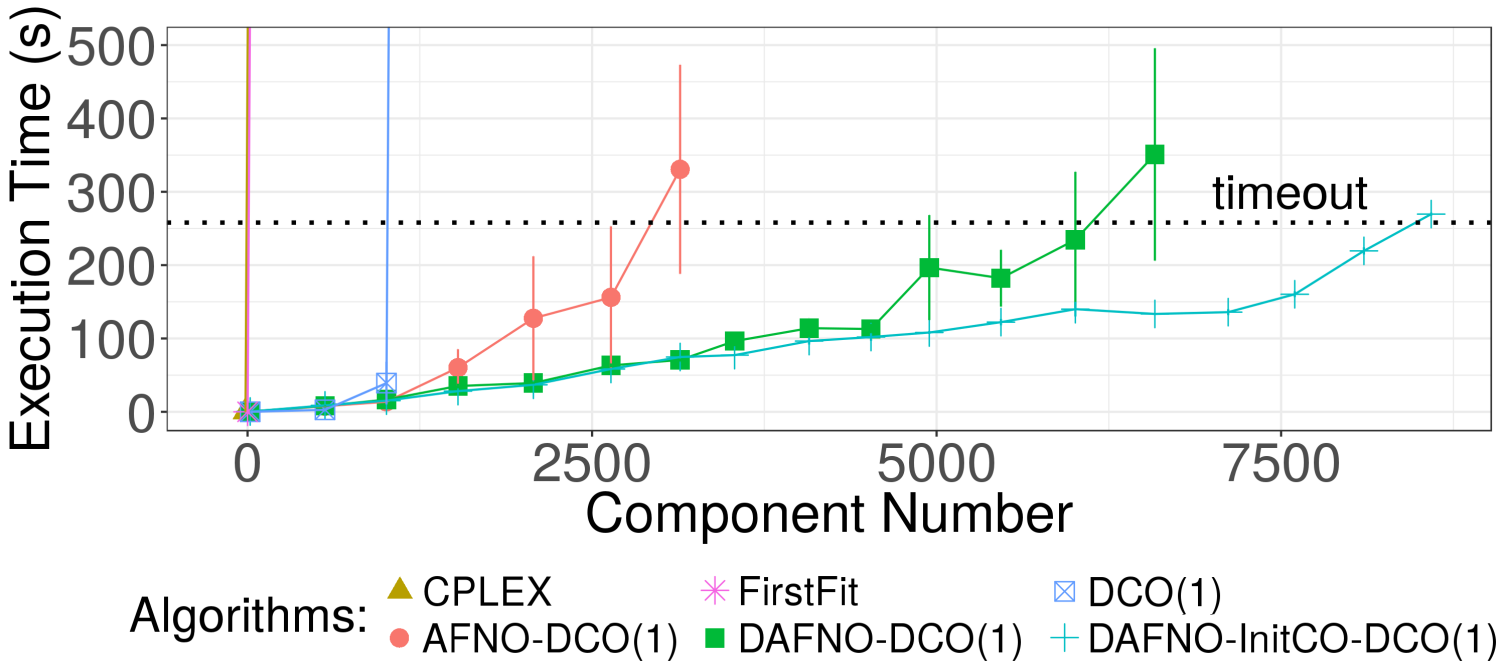
Algorithm	**Execution Time** (s)	**Simulated App Response Time**
CPLEX	258	100%
DAFNO-InitCO-DCO(1)	0.022	110%
FirstFit	756	230%

Large-Scale Problems

Given a timeout, how many applications / components each algorithm can deal with in a large-scale infrastructure ?

Evaluation Setup

- infra : 10561 fog nodes (i.e., cloud DCs, edge servers, end devices)
- app : continuously add randomly generated DSP applications
- timeout : 258s (execution time of CPLEX in the previous evaluation)

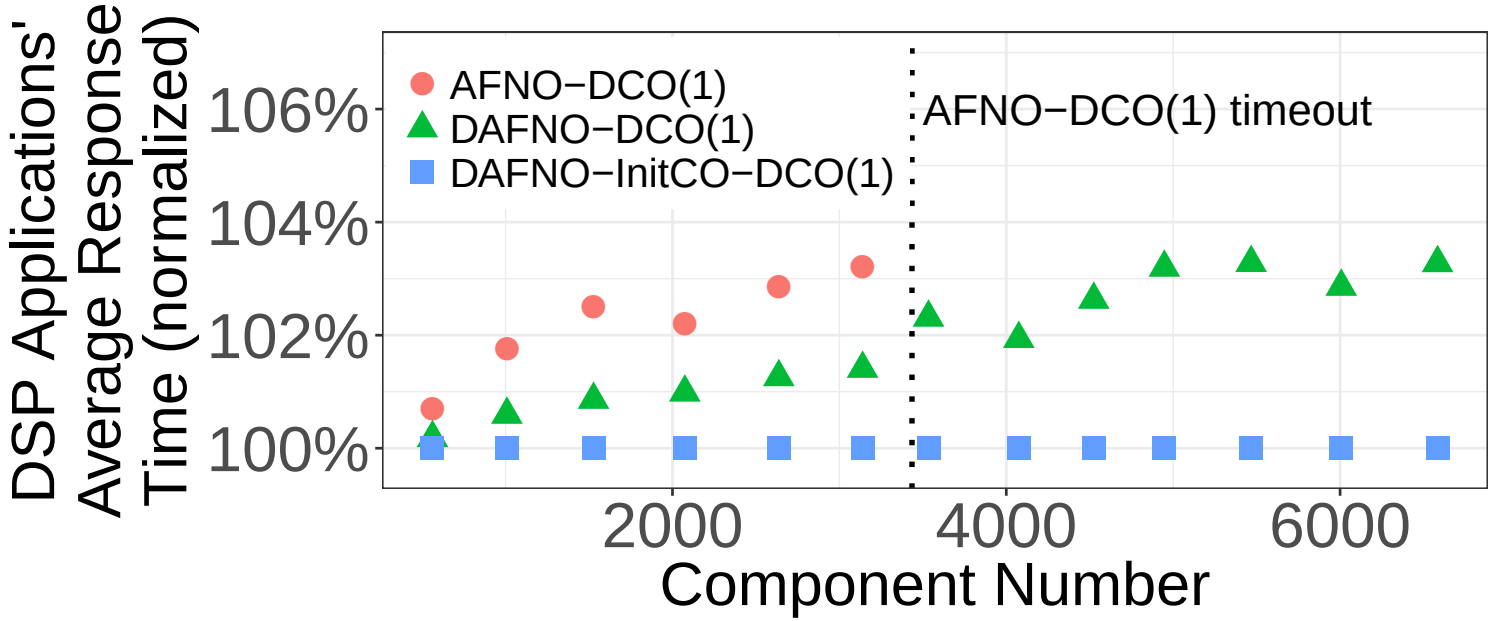


Large-Scale Problems

Given a timeout, how many applications / components each algorithm can deal with in a large-scale infrastructure ?

Evaluation Setup

- infra : 10561 fog nodes (i.e., cloud DCs, edge servers, end devices)
- app : continuously add randomly generated DSP applications
- timeout : 258s (execution time of CPLEX in the previous evaluation)



Outline

Conclusion

Contribution

- Model, Objective Function
- Four heuristics (AFNO, DAFNO, InitCO, DCO) based on FirstFit
- Combined heuristics : highly scalable & near-optimal result

Future Work

- Infrastructure Dynamicity (e.g., churn & mobility)

Conclusion

Contribution

- Model, Objective Function
- Four heuristics (AFNO, DAFNO, InitCO, DCO) based on FirstFit
- Combined heuristics : highly scalable & near-optimal result

Future Work

- Infrastructure Dynamicity (e.g., churn & mobility)

Thanks for your attention!

References

[1] Ye Xia, Xavier Etchevers, Loïc Letondeur, Thierry Coupaye and Frédéric Desprez
Combining Hardware Nodes and Software Components Ordering-based Heuristics for Optimizing the Placement of Distributed IoT Applications in the Fog.
In: The 33rd ACM/SIGAPP Symposium On Applied Computing. ACM. 2018.

[2] Ye Xia, Xavier Etchevers, Loïc Letondeur, Adrien Lebre, Thierry Coupaye and Frédéric Desprez
Combining Heuristics to Optimize and Scale the Placement of IoT Applications in the Fog.
In: 11th IEEE/ACM International Conference on Utility and Cloud Computing. IEEE/ACM. 2018.

[3] Casanova Henri, Giersch Arnaud, Legrand Arnaud, Quinson Martin and Suter Frédéric
Versatile, scalable, and accurate simulation of distributed applications and platforms.
In: Journal of Parallel and Distributed Computing 74.10 (2014), pp. 2899–2917.