6th International Symposium on Intelligent Distributed Computing - IDC 2012

September 24-26, 2012 - Calabria, Italy

A Distributed Allocation Strategy for Data Mining Tasks in Mobile Environments

Carmela Comito, **Deborah Falcone**, Domenico Talia, and Paolo Trunfio

University of Calabria & ICAR-CNR

Goals and motivations (1/2)

- **A variety of powerful mobile devices is available**
  - Smart phones, PDAs, laptops, netbooks, …


- **Wireless networks are always more end-user oriented**
- **Emergence of the *ubiquitous computing* paradigm**
  - Facilitates continuous access to data and information by mobile users with handheld devices
  - Mobile users perform intelligent analysis and monitoring of data
- **Applications**
  - Body-health monitoring
  - Vehicle control
  - Wireless security systems, …

Goals and motivations (2/2)

Data mining from such mobile/embedded devices faces various challenges:

- low-bandwidth networks;
- relatively small storage space;
- limited availability of battery power;
- slower processors;
- small displays to visualize the results.

➡ Optimally utilizing the limited resources

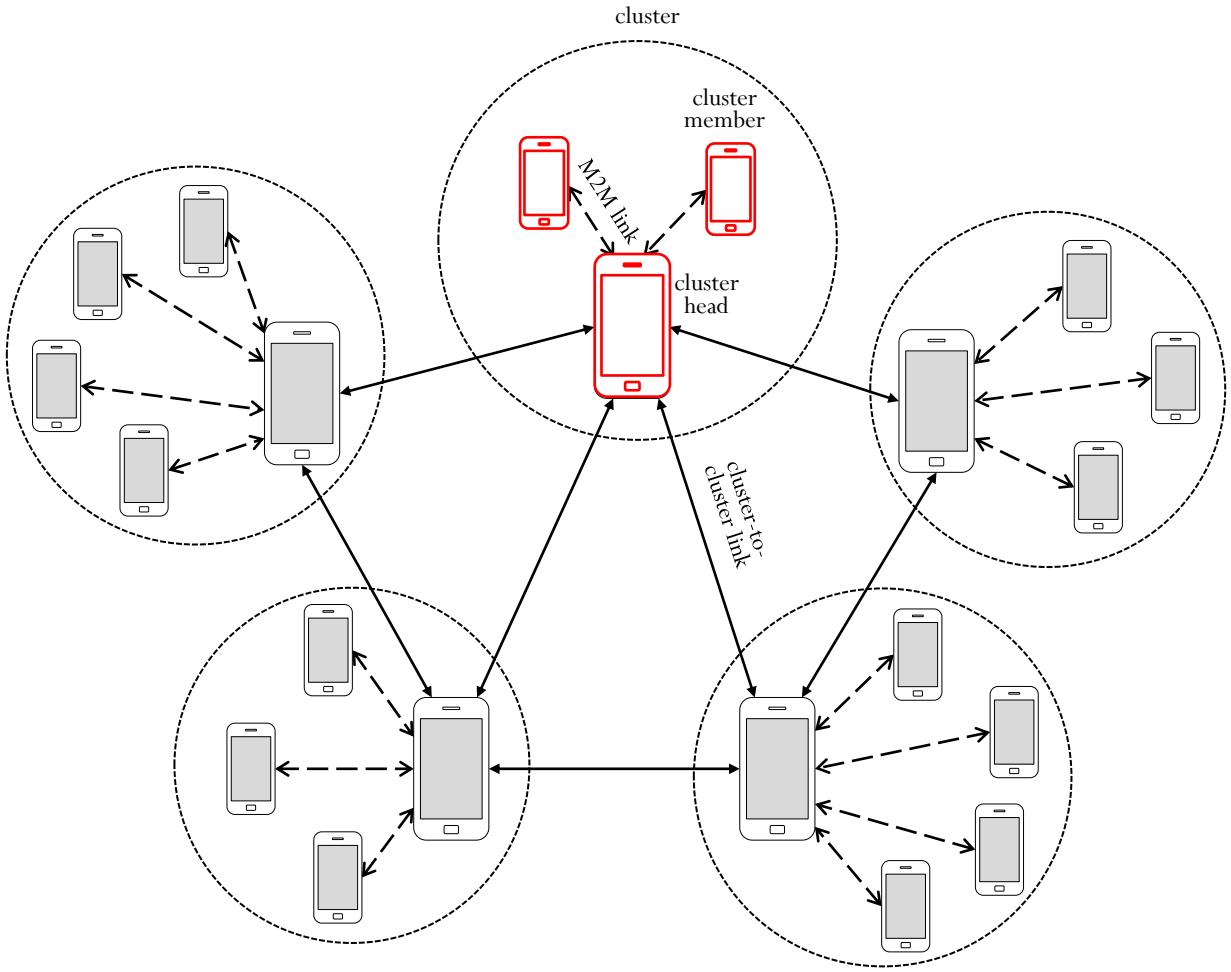
➡ E*nsuring energy efficiency*

Outline

- Mobile-to-Mobile Architecture
- Energy model
- Distributed task allocation strategy
  - Distribuited task allocation algorithm
- Performance Evaluation
  - Energy-Aware scheduler (EA) versus Round Robin (RR)
- Conclusion

Mobile-to-Mobile Architecture

The **M2M** architecture aims at exploiting Mobile-to-Mobile technologies to support pervasive and ubiquitous data analysis and mining through mobile devices



Energy Model (1/2)

Energy consumption of a mobile device depends on the *communication (ET)* and *computation (EC)* loads

Total energy consumption: $E = ET + EC$

- **COMMUNICATION ENERTGY (ET)**
  - In ad-hoc networks nodes are often in the idle mode
  - Nodes continuosly listen to the wireless channel
    - Nodes overhear every packet transmission within their range
    - Nodes consume energy uselessly referred to as ***overhearing***
  - Energy consumed for ***communication*** by a node **i**

$$ET_i = E_{send_i} + E_{receive_i} + E_{discard_i}$$

    - $E_{send_i} = |MSG|\ m_{send} + b_{send}$
    - $E_{receive_i} = E_{discard_i} = |MSG|\ m_{receive} + b_{receive}$

Energy Model (2/2)

- **COMPUTATION ENERTGY (EC)**
  - Identify the energy consumption characteristics of some commonly used statistical and data mining algorithms running on-board a mobile device
    - Experimentally quantified the performance of specific data mining algorithms
    - For low bandwidth lossy networks the high energy costs of communication often makes local on-board data mining a more energy efficient choice
  - Machine learning approach to predict energy consumption of mobile devices to perform data mining algorithms

EC

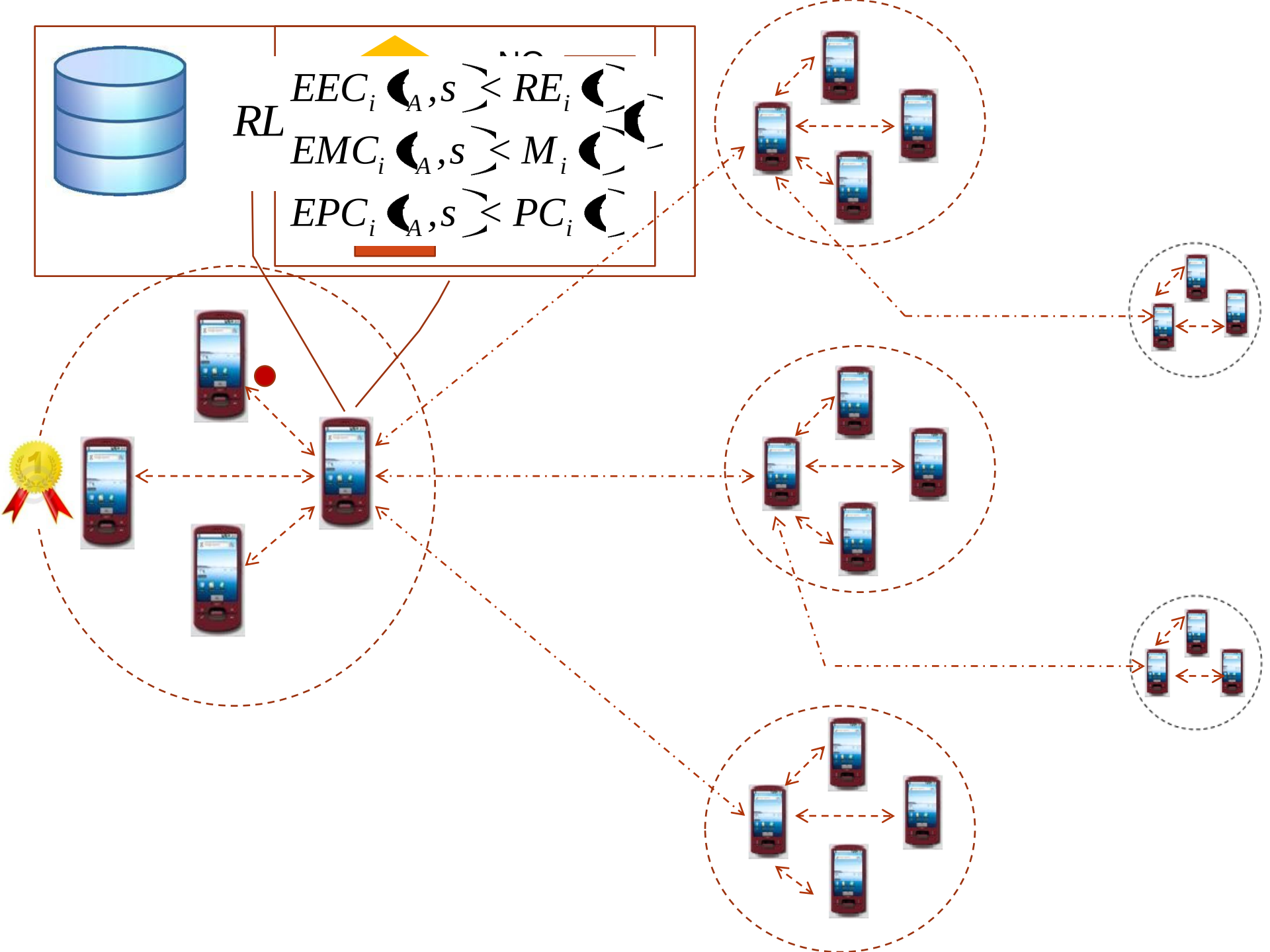
Method	*Algorithm*	*Data Set Size*	*RAM Memory (MByte)*	*Virtual Memory (MByte)*	*CPU (%)*	*Battery Charge Depletion (mAh)*	*Energy Consumption (J)*	*Time (sec)*
Association Rules								
Rule Induction	Apriori	**CENSUS_DISC.arff**						
		0,1 MB	15,86	95,19	96,92	0	0	6
		0,2 MB	16,97	105,36	98,03	0	0	12
		0,4 MB	18,06	104,95	98,24	0	0	26
		0,8 MB	19,87	102,75	98,13	2,7	35,964	73
		1,6 MB	23,32	103,99	96,87	13,5	179,82	300
		3,2 MB	26,92	100,01	95,44	23,3	310,356	3960
		6,4 MB	---	---	---	---	---	---
Classification								
Trees	J48	**COVERTYPE.arff**						
		0,1 MB	19,47	104,94	96,23	13,4	178,488	300
		0,2 MB	20,15	104,92	98,21	29,8	396,936	540
		0,4 MB	23,87	105,6	97,43	59,4	791,208	2040
		0,8 MB	27,68	103,87	97,36	194,64	2592,6048	8160
		1,6 MB	---	---	---	---	---	---
		3,2 MB	---	---	---	---	---	---
		6,4 MB	---	---	---	---	---	---
Clustering								
Instance-based/Lazy Learning	K-Means	**CENSUS.arff**						
		0,1 MB	16,73	96,56	98,03	6,75	89,91	55
		0,2 MB	17,95	102,05	97,65	8,1	107,892	150
		0,4 MB	19,72	102,16	97,02	18,9	251,748	300
		0,8 MB	23,08	101,86	97,97	18,9	251,748	600
		1,6 MB	26,4	95,96	97,82	43,2	575,424	1320
		3,2 MB	---	---	---	---	---	---
		6,4 MB	---	---	---	---	---	---

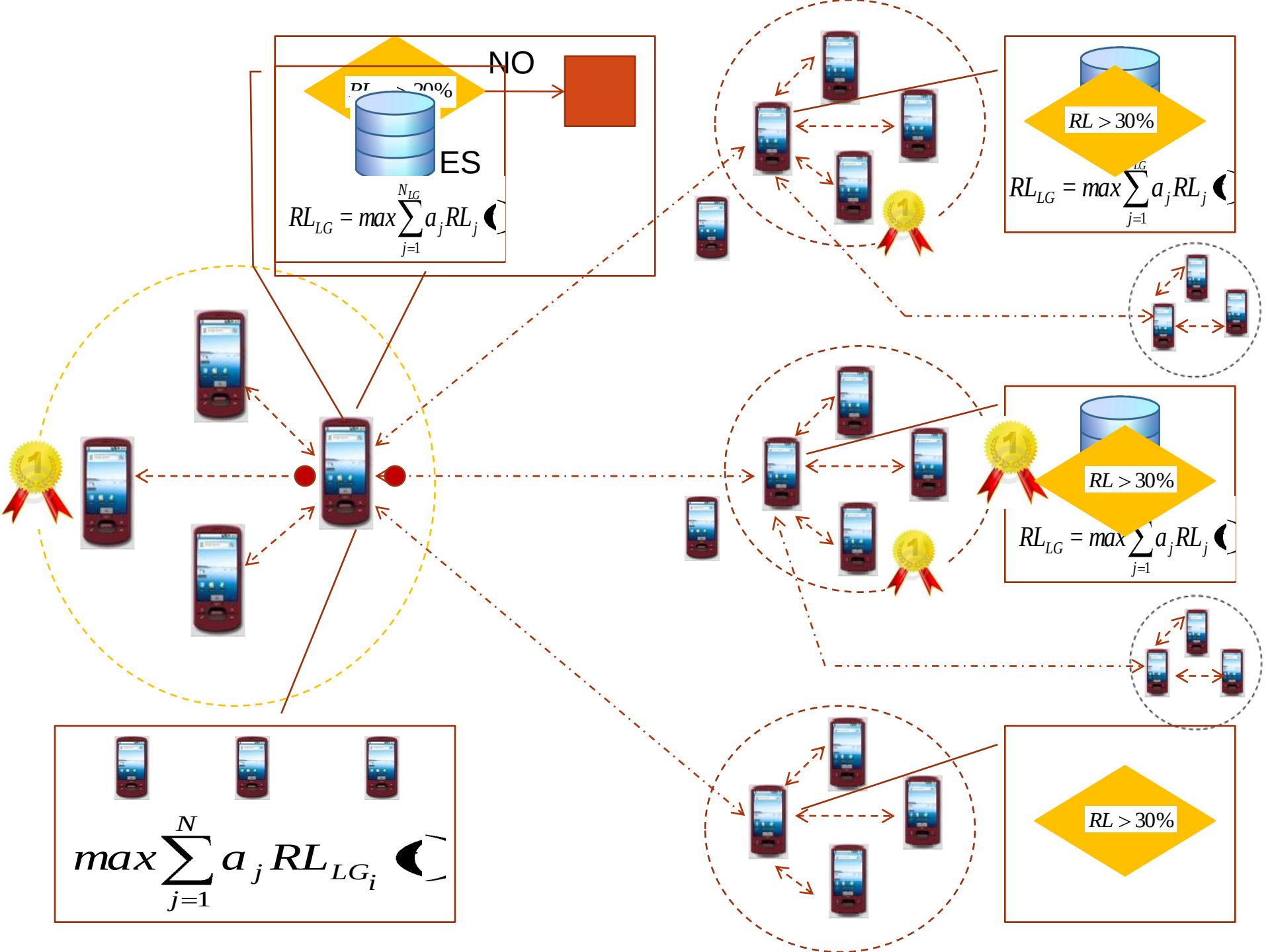
Distribuited Task Allocation Strategy

- Energy-efficient dynamic task allocation strategy over the cooperative M2M architecture
  - Whenever a task has to be executed, an efficient task assignment is found
    - the total consumed energy in the network is minimized
    - the network lifetime is prolonged by distributing energy consumption among mobile groups

- *Given a task model T and a device model D, determine a task allocation TA that maps each task to a device such as to maximize the network lifetime*
  - Independent tasks: atomic applications or tasks without dependencies
  - Residual life of a device:

$$RL(t)=RE(t)/P(t)$$


RL
$EEC_i(A,s) \le RE_i(\quad)$
$EMC_i(A,s) \le M_i(\quad)$
$EPC_i(A,s) \le PC_i(\quad)$

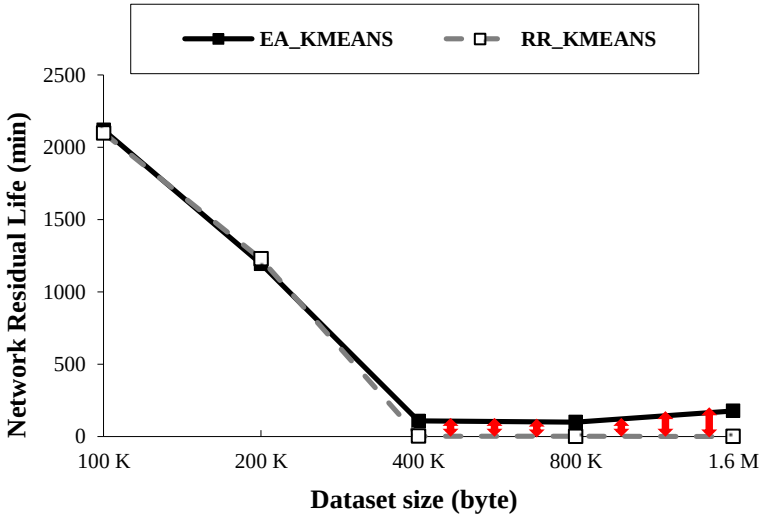

NO
ES
$RL_{LG} = max \sum_{j=1}^{N_{LG}} a_j RL_j$
$RL > 30\%$
$RL_{LG} = max \sum_{j=1}^{LG} a_j RL_j$
$RL > 30\%$
$RL_{LG} = max \sum_{j=1}^{} a_j RL_j$
$RL > 30\%$
$max \sum_{j=1}^{N} a_j RL_{LG_i}$

Performance Evaluation (1/5)

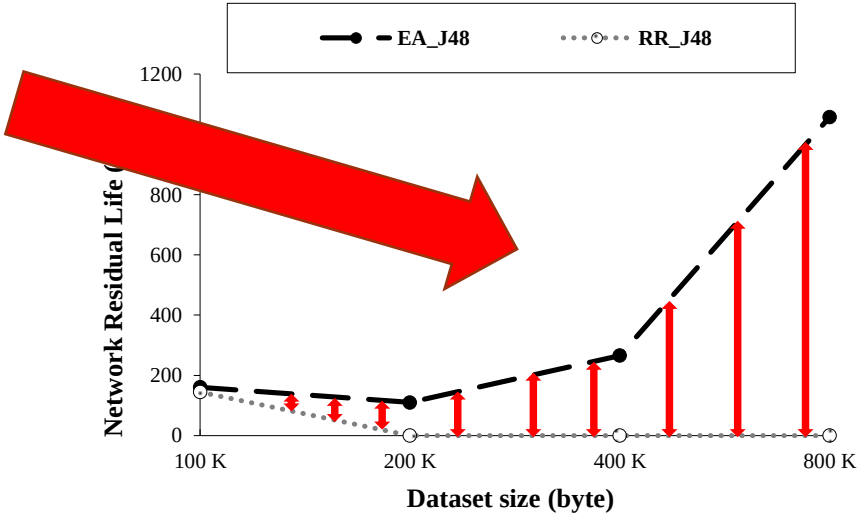
- **We developed a discrete-event simulator of the M2M architecture**
- **Implementation of the software components**
  - *Energy model*
  - *Energy-aware task allocation algorithm*
- **Simulation study of the energy-aware scheduler in terms of energy depletion and network lifetime**
  - *Performance metrics*
    - Residual life of the network
    - Number of alive devices
    - Number of completed tasks
  - *Energy-Aware scheduler (EA) versus Round Robin (RR)*
- **Parameters used in the simulation**
  - Network interface: 802.11 b/g with a bandwidth of 11 Mbps
  - Initial energy level on each device : ranging from 3,000 J to 11,000 J
  - Network size: 100 mobile nodes

Performance Evaluation (2/5)

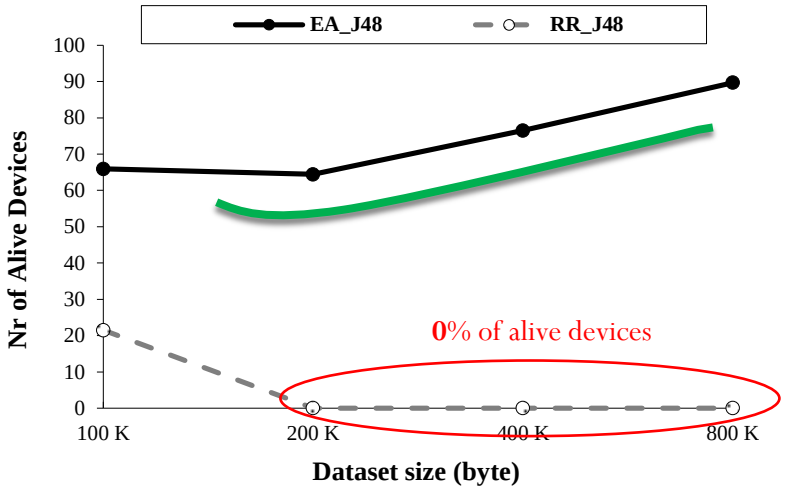
- *Data-mining task with different size of the dataset*
  - Simulation time: 30 hours
  - Variable dataset size: 100 kB - 3.2 MB
  - Task arrival rate: Poisson distribution with frequency of 160 tasks/hour

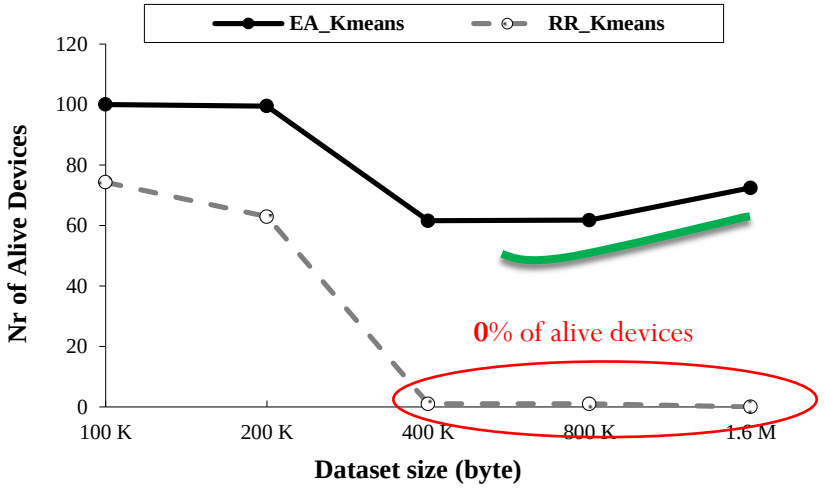


EA scheduler is effective in prolonging network lifetime compared to the RR algorithm.

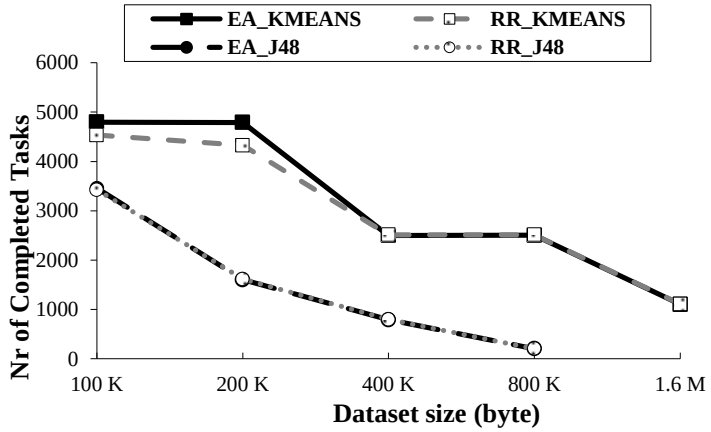


Performance Evaluation (3/5)





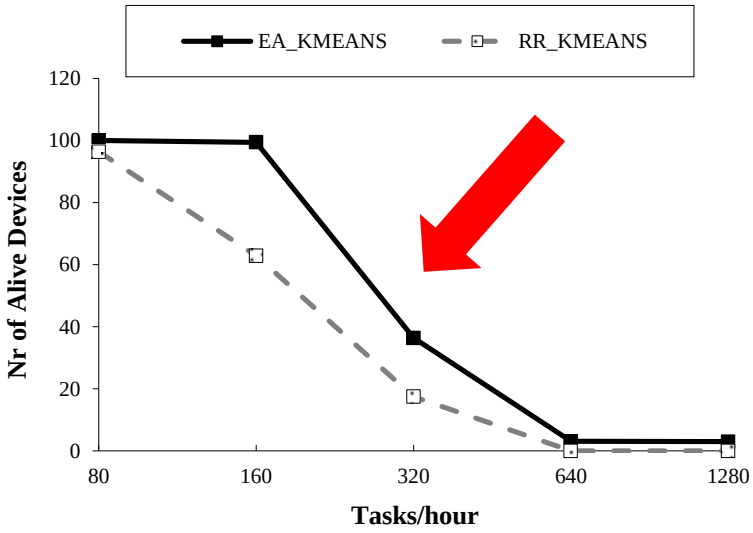
The number of alive devices with EA is greater than that achieved by RR

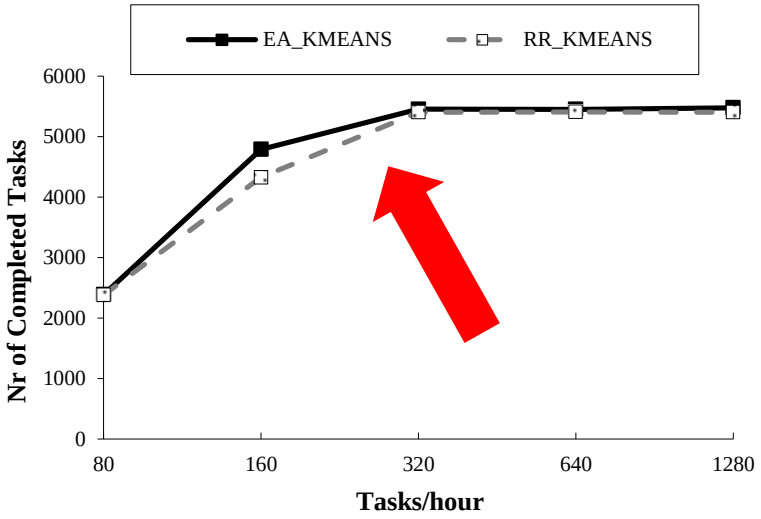


The higher number of alive devices ensured by EA compared to RR, is obtained without reducing the number of completed tasks

Performance Evaluation (4/5)

- *Data-mining task with different arrival rate*
  - Simulation time: 30 hours
  - Dataset size: 200 kB
  - Variable task arrival rate: Poisson distribution with frequency varying from 80 to 1280 tasks/hour





EA turn off a lower number of devices…

… without sacrificing the number of data mining tasks completed.

Conclusion

- **Energy-aware dynamic task allocation scheme over a cooperative architecture**
  - *Distributution of energy consumption among the available devices*
  - *A two-phase heuristic-based algorithm*
  - Results show that the proposed scheme enhances the energy efficiency of the system compared to RR scheduler
    - It is **effective in prolonging network lifetime,** without sacrificing the number of data mining tasks completed
    - It **keeps alive most of the mobile devices** thanks to its energy load balancing strategy

THANK YOU

FOR YOUR ATTENTION