Self-organizing Sensor Networks: State of the art, new solutions and applications

Gabriele Monti
DEIS – Università di Bologna
Area Progetti Integrati Cesena (APICe)
Via Venezia 52, I-47023 Cesena (FC)
gabriele.monti4@unibo.it
www-db.deis.unibo.it/~gmonti

Research activities carried out during the Ph. D. Course (2004-2007)
Scientific tutor: prof. Gianluca Moro (gianluca.moro@unibo.it)
Main international collaboration: University of Illinois at Chicago

Introduction

- Small and powerful devices equipped with CPU, RAM and communication interfaces (i.e. Bluetooth and Wi-Fi)
- Realization of Ad-hoc networks has become possible:
  - Deployment on the fly
  - No fixed and pre-configured infrastructure
  - Connectivity achieved by multi-hop transmissions
  - Devices act as **peers**, consumers and signal repeaters at the same time

Introduction to ad-hoc and sensor networks (i)

- Composed of small and powerful devices equipped with CPU, RAM and communication interfaces (i.e. Bluetooth and Wi-Fi)
- Deployed without needing any fixed and pre-configured infrastructure
  - Emergency communications in case of a natural disaster
  - On the fly network of portable devices (i.e. University campus, Industrial logistic and manufacturing process)
  - Environmental monitoring
- Communication relies on multi-hop transmissions
- Require cooperation, devices act both as users and as signal repeaters at the same time (peers)

Introduction to ad-hoc and sensor networks (ii)

- New generation of sensornets forming distributed systems and distributed databases
- Devices have scarce resources
  - Battery power, low range connectivity
  - Memory and computational capability
- Protocols must be designed with those constraints in mind
  - Simple, lightweight, robust to frequent failures
- In sensor networks data management should be integrated in the protocols

Example: Enviromental monitoring

- La mancanza di stazioni di rilevamento meteorologiche su grandi aree del pianeta causa elevate imprecisioni nelle previsioni (es: deserti, oceani, vaste catene montuose)
- Reti di sensori distribuiti nelle zone remote del pianeta che assolvono a questo compito con comunicazioni via satellite, due tipi di soluzioni:
  - rete di sensori tradizionale (senza intelligenza, trasmette tutti i dati elementari
  - rete di sensori avanzata che pre-elabora localmente i dati e trasmette solo informazioni di alto livello

Related works

- Main tasks in ad-hoc and sensor networks:
  - Routing protocol
  - Data management
- Classification of existing routing protocols according to:
  - Routing strategy
  - Structure of the network
  - Geographic routing protocols
- Data management protocols for sensor networks
  - DHT on top of routing protocols
  - Space partitioning protocols

Routing approach

- Table-driven (a.k.a. proactive) routing
  - Constantly updated routing-information to all nodes
  - Routes are calculated before they are needed, and even if not needed
  - Scalability problems, need lots of storage
- On-Demand (a.k.a. reactive) routing
  - Does not aim to keep global and up-to-date view at nodes
  - Requires message flooding causing latency and network overhead

Network structure

- Hierarchical networks
  - Nodes/sensors are clustered and some of them act as super-nodes
  - Intra-cluster communication are managed by normal nodes
  - Super nodes manage inter-cluster communication (gateway function)
- Flat networks
  - All nodes/sensors perform the same tasks
  - No clustering
  - No gateways

Geographic routing protocols

- Nodes/sensors must be aware of their physical location [2]
- Requires GPS or position estimation techniques
- Packets are forwarded trying to reduce as much as possible the physical distance to the destination
  - Good scalability (no flooding)
- Dead-ends may occur under low density environment or in case of obstacles
- Positioning can be expensive, inaccurate or unavailable

Data management protocols

- DHT on top of routing protocols [5,6]
  - Build a distributed index on top of a structured network (routing protocol)
  - Scarce integration between the protocols causes overhead (time and resources)
- Space partitioning protocols [3,4]
  - Require sensors localization
  - Data load unbalanced in case of low density zones

Most popular algorithms

<table>
<tr><th colspan="2">Routing</th><th colspan="2">Network Structure</th><th rowspan="2">Geographical</th></tr>
<tr><th>Reactive</th><th>Proactive</th><th>Hierarchical</th><th>Flat</th></tr>
<tr><td>DSDV
STAR
WRP</td><td>AODV
DSR
TORA</td><td rowspan="2">FSR
GSR
HARP</td><td rowspan="2">DSDV STAR
DSR WRP
TORA, etc.</td><td rowspan="2">GPSR
DREAM
LAR
GSL</td></tr>
<tr><td colspan="2">HSLS
ZRP</td></tr>
</table>

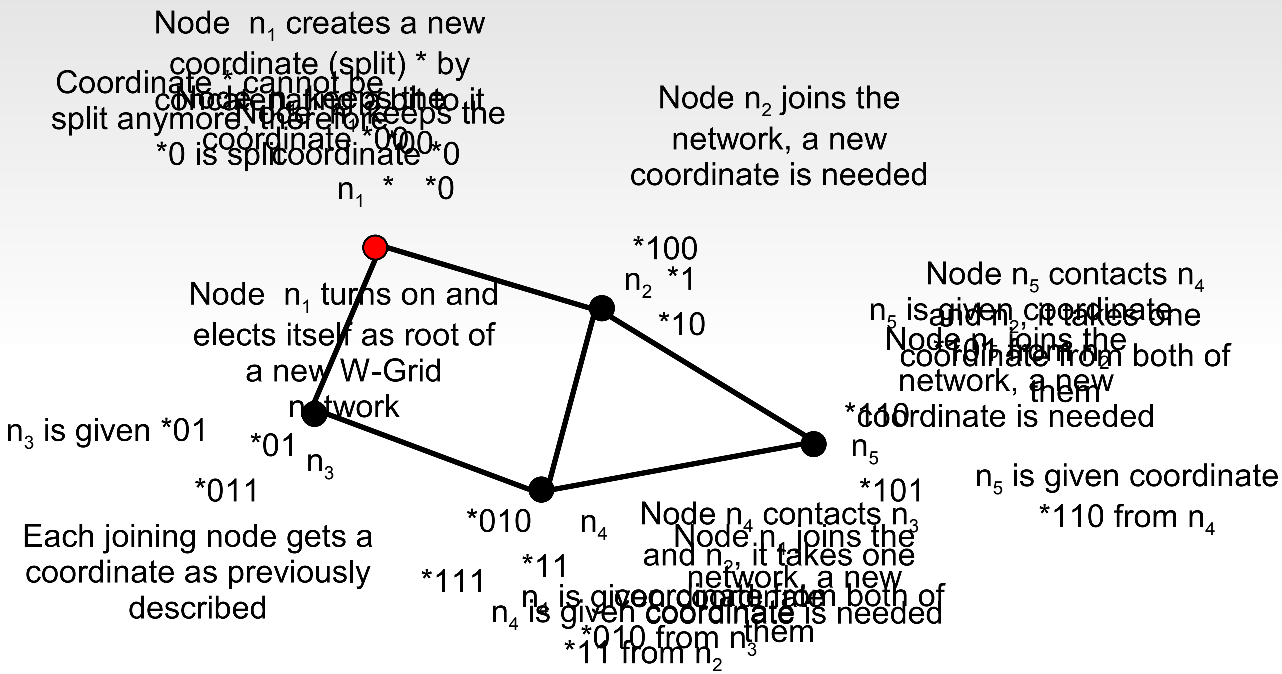
W-Grid: a new solution

- A cross-layer infrastructure which self-organizes routing and data management in wireless Ad-Hoc networks
  - Mobile devices (laptops, PDAs, etc.), communication and data sharing
  - Last generation Sensornets
- Cross Layer: The network structure is used for addressing both routing and data management
  - Different from usual underlying-overlay approach in Ad-Hoc networks
- Self-organization: no sensor has global information
- Currently working mainly on static networks

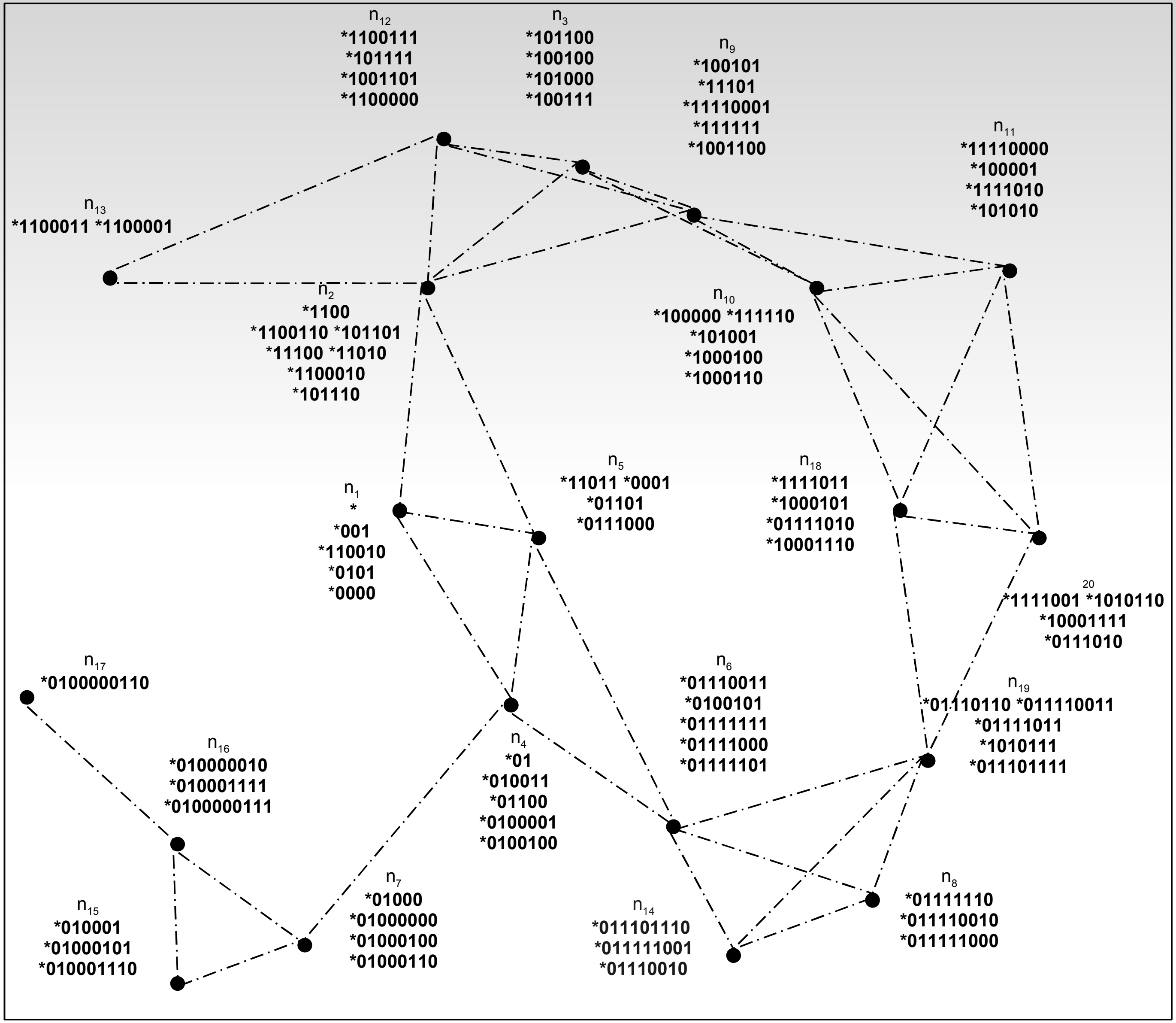
W-Grid Main Characteristics (I)

- Network organization
- Does not rely on nodes localization (expensive)
- Uses virtual coordinates
  - Does not try to approximate nodes physical position
  - VCs are binary string
- Devices are logically mapped on a binary tree structure but there are no super-nodes
  - No hierarchy, all the nodes perform the same tasks
  - Nodes are assigned several coordinates

W-Grid: Coordinates generation



Example of W-Grid Network



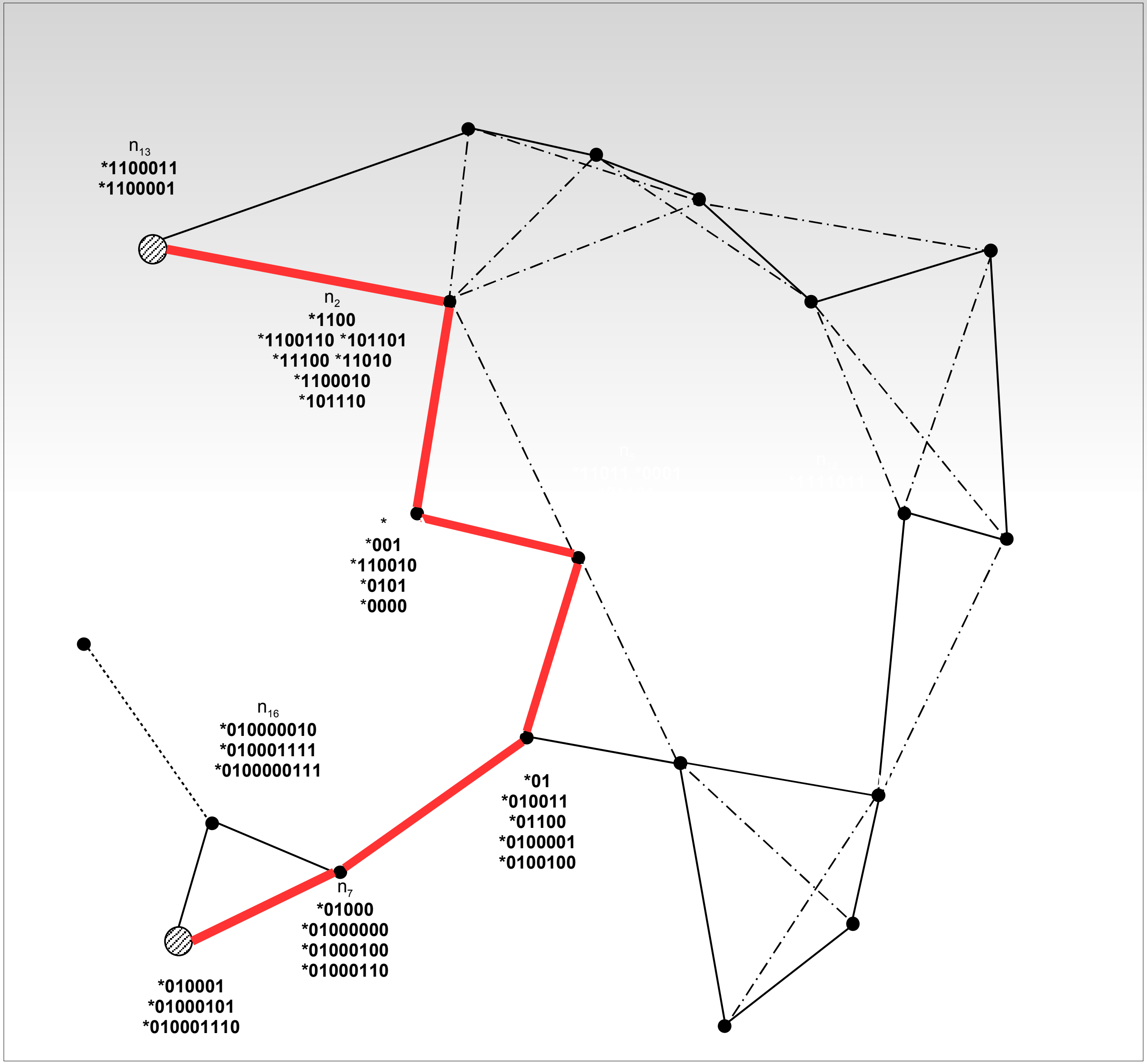
W-Grid Main Characteristics (II)

- Routing
  - Nodes routing table contains information about one-hop neighbours (no global knowledge)
  - No broadcasting is required
- Based on the concept of distance between coordinates, example:

$$d(*\mathbf{0}011,*\mathbf{0}11) = 5$$

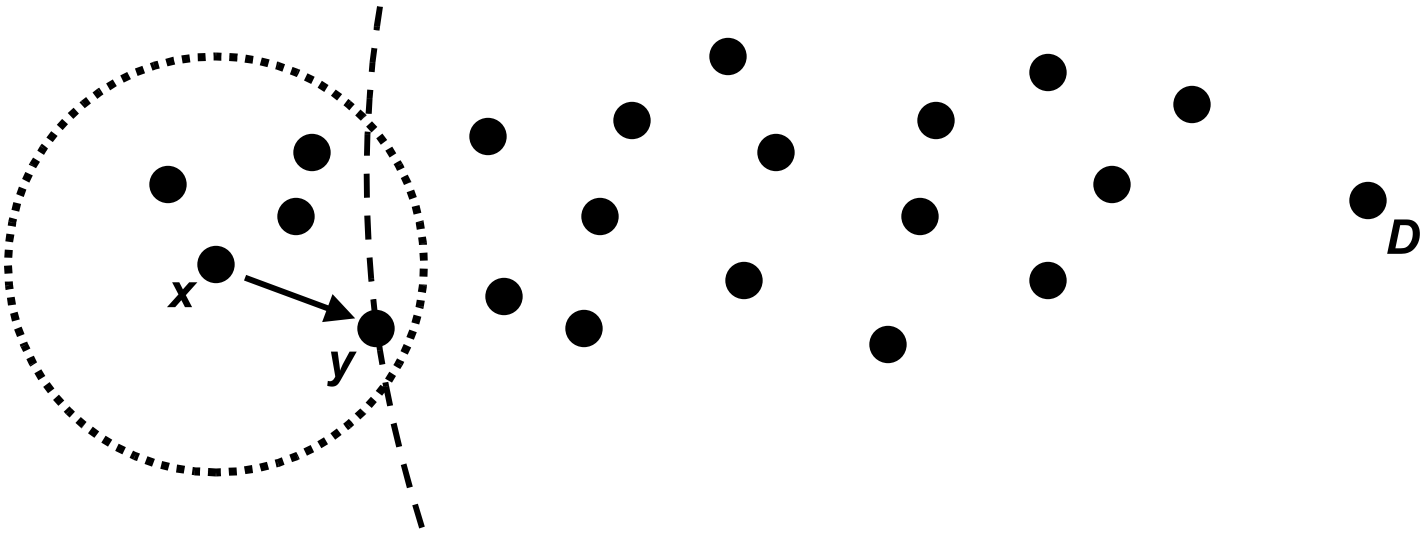
- Given a destination coordinate, messages travel the network hop-to-hop by choosing the neighbour which mostly reduces the distance

W-Grid: Routing example



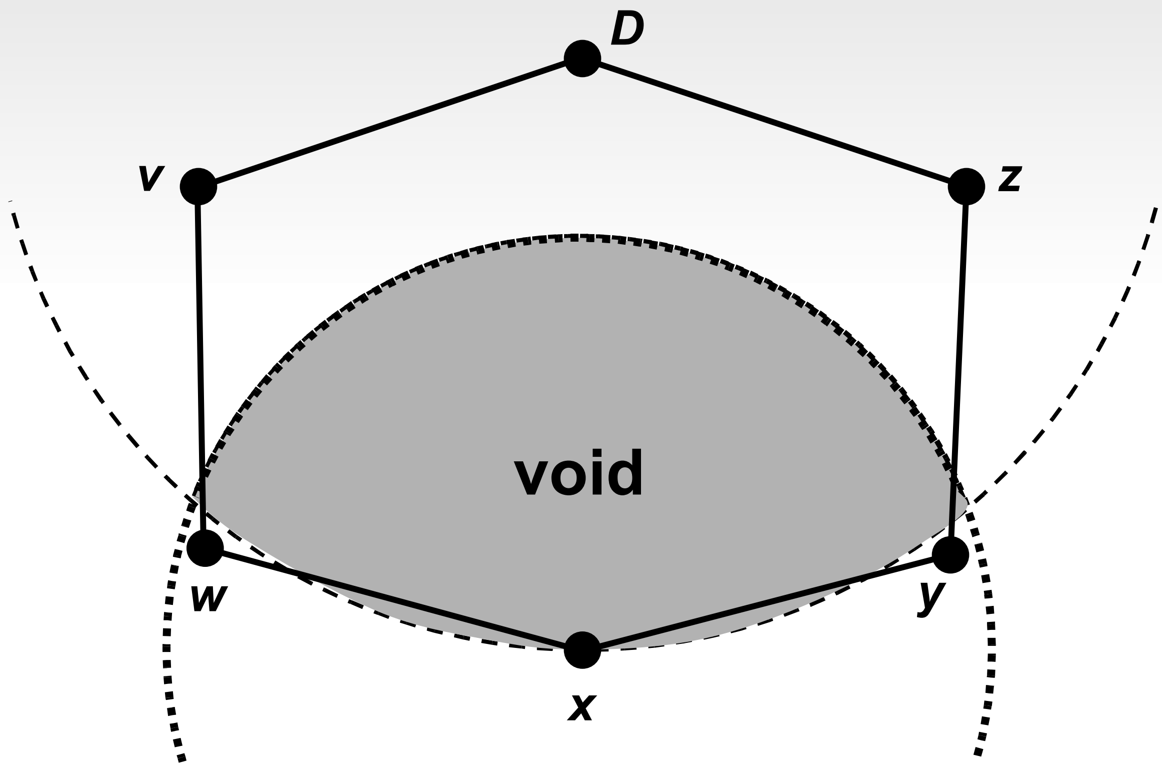
GPSR - Greedy Forwarding

- Nodes learn immediate neighbors' positions from beaconing/piggybacking on data packets
- Locally optimal, greedy next hop choice:
  - Neighbor geographically nearest destination



Greedy Forwarding Failure

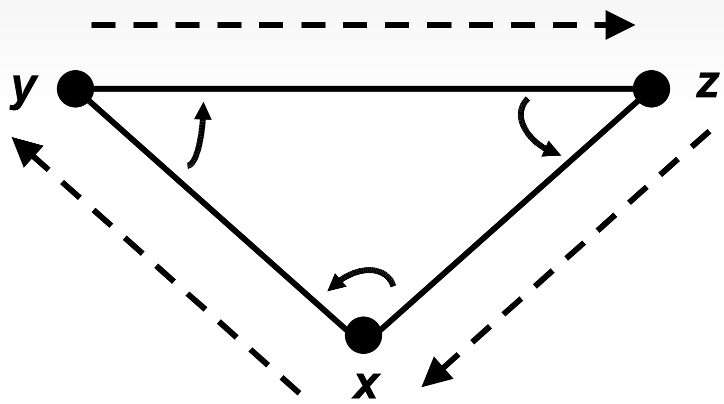
Greedy forwarding not always possible! Consider:



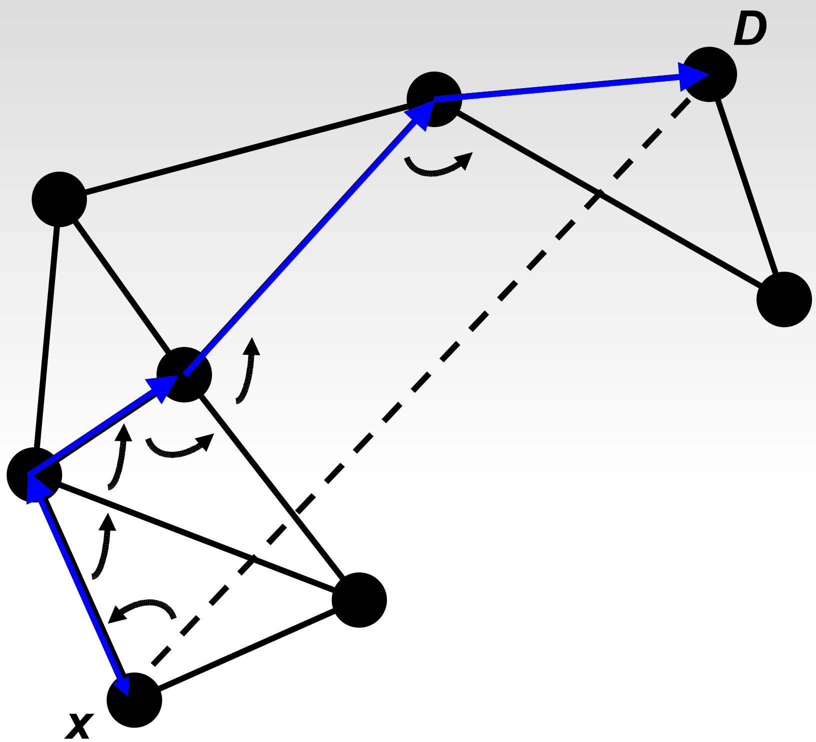
Void Traversal: The Right-hand Rule

Well-known graph traversal: right-hand rule

Requires only neighbors' positions



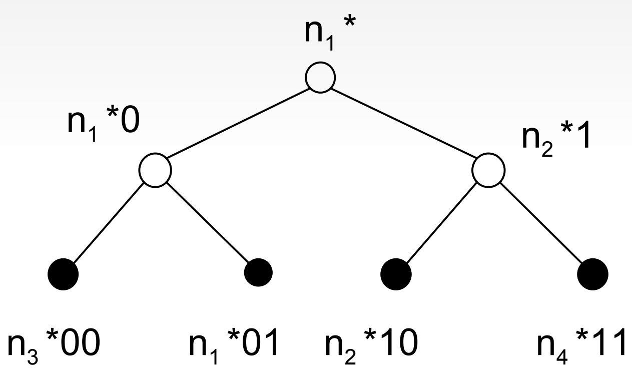
Perimeter Mode Forwarding

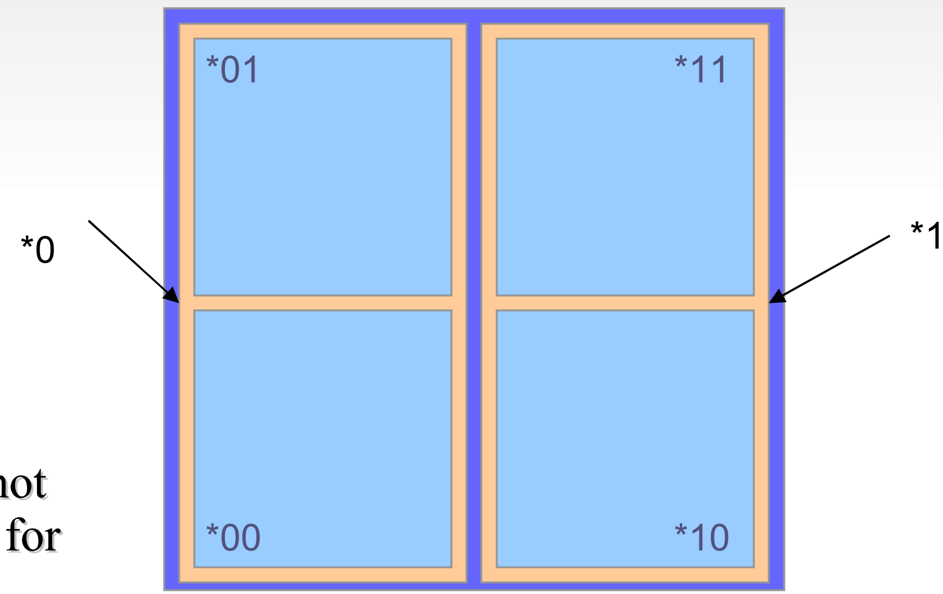


- Traverse face closer to *D* along $\overline{xD}$ by right-hand rule, until crossing *xD*
- Repeat with next-closer face, *&c.*

W-Grid Data management

- Virtual coordinates implicitly generate a distributed database





○ Regions that have been split. They do not manage any data but they may be used for routing

● Regions that can manage data and also used for routing

Application Scenario Exact match query

Example of an environment monitoring application in which sensors survey temperature (T) and pressure (P), we will refer to T and P as d_1 and d_2

- We define a domain for d_1 and d_2

$$Dom(d_1) = [-40, 60];\ Dom(d_2) = [700, 1100]$$

- Return the times at which sensors surveyed a temperature ranging of 26 Celsius degrees and pressure of 1013 mbar
- We must calculate the binary string corresponding to (26,1013)

$$c = {*}11011000$$

- Now all we have to do is querying the sensor whose coordinate is *11011000

Application Scenario
Range query

Considering the same sensor network, return the times at which sensors surveyed a temperature ranging from 26 to 30 Celsius degrees and pressure ranging from 1013 to 1025mbar.

- We must calculate the correspondent binary string for the four corner of the range query, namely (26,1013) (26,1025) (30,1013) (30,1025):

$$c_1 = \mathbf{*110110}00 \quad c_2 = \mathbf{*110110}01$$

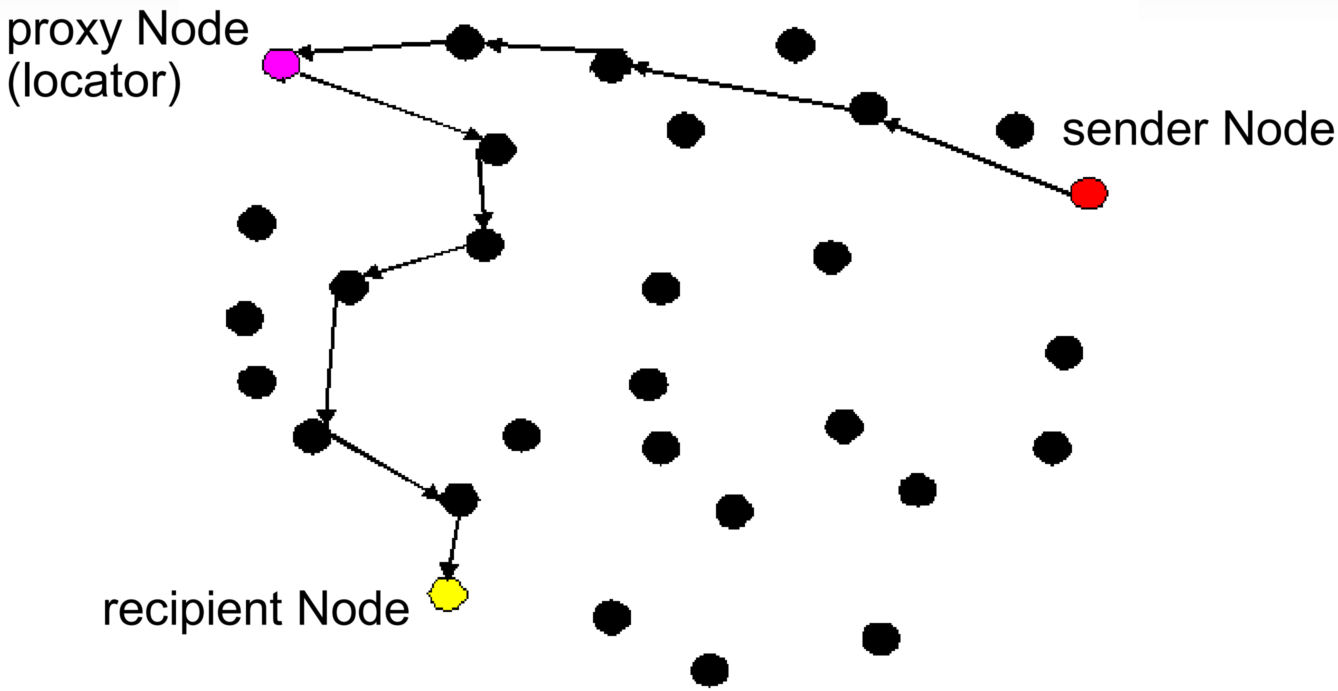
$$c_3 = \mathbf{*110110}10 \quad c_4 = \mathbf{*110110}11$$

- Now all we have to do is querying the sensors whose coordinate have 110110 as prefix

Data management: example of location service

In an Ad-Hoc Network we can build a location service by indexing information about nodes (i.e. nickname, Peer ID) and storing their virtual coordinates

Nodes act as proxies for devices whose IDs fall in their managing region



W-Grid Experimental results

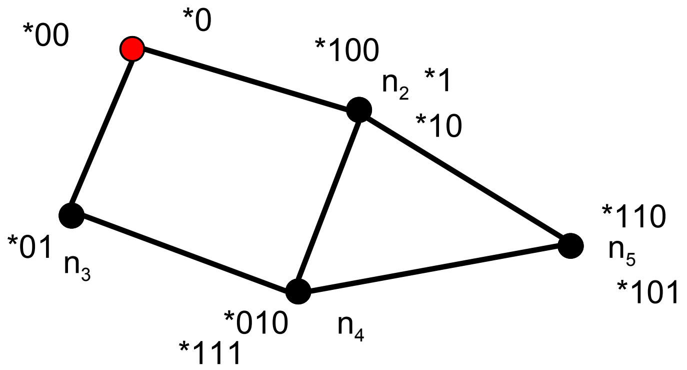
Routing performances of W-Grid have been compared to GPSR routing algorithm, obtaining absolutely good results

	APL(in hops)		RMSE	
Area(nodes number)	W-Grid	GPSR	W-Grid	GPSR
800×800(120)	6,13	7,49	3,11	8,44
1000×1000(200)	8,05	9,02	4,45	13,00
1200×1200(290)	9,75	9,64	4,47	12,74
1400×1400(400)	11,54	10,87	4,99	14,52
1600×1600(520)	13,96	13,71	5,86	14,99
1800×1800(660)	14,81	14,14	6,41	12,15
2000×2000(820)	17,43	16,57	8,44	13,20

W*-Grid enhancing robustness

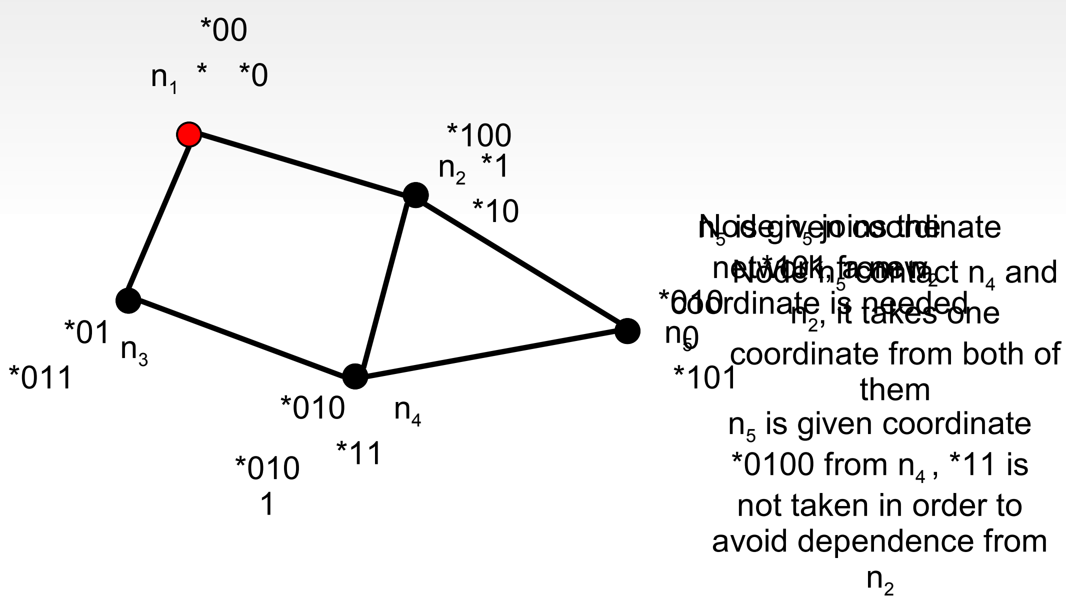
- The infrastructure may deal with nodes disconnections and failures
- Robustness to failures represent a step toward nodes mobility
  - Nodes movements can be seen as couple of disconnection/reconnection
  - We assume single disconnections

Both n_5 coordinates have prefix *1, we say that n_5 depends on n_2. This could be a problem in case of n_2 disconnection (failure)



W*-Grid Coordinates generation

- The goal is avoid dependencies

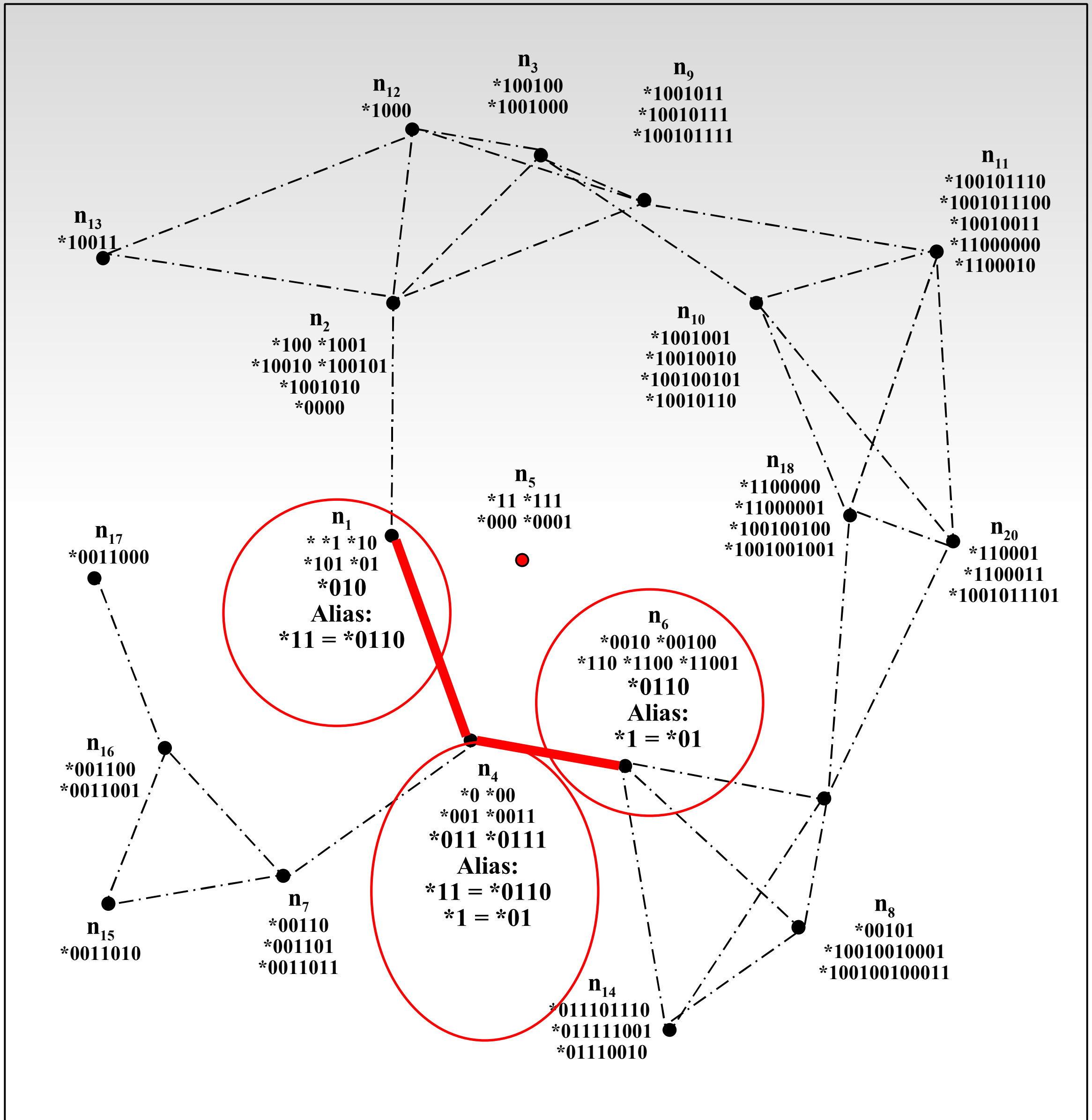


- Independent coordinates introduce more heterogeneity (Multiple paths)

W*-Grid Failure recovery (I)

- Performed whenever a node n_l loses connection with the father n_f of one of its coordinates c_l, obtained by c_f
- The goal is contact the closest relative of the failed node (usually the father of the father of n_l)
  - If n_l is not fully dependent from n_f it starts recovery sending a message and searching for the father of n_f
    - The destination coordinate is c_l
    - The message must avoid to use every coordinate having prefix c_f
  - The found node n_g and the nodes $n_1 .. n_i$ crossed by the message generate new coordinates until n_l is reached
  - n_l, n_g, *and* $n_1 .. n_i$, set aliases of the form $c_{old} = c_{new}$ in order to adjust routing

Failure Recovery: An example



- Node n_6 lost its father n_5
- It is not fully dependent from n_5, recovery starts
- n_6 uses its other coordinates to find its closest relative (* in this case)
- n_1 and the other nodes crossed by the message generates new coordinates to reach n_6
- Both n_6 and its relative n_1 set aliases to adjust routing

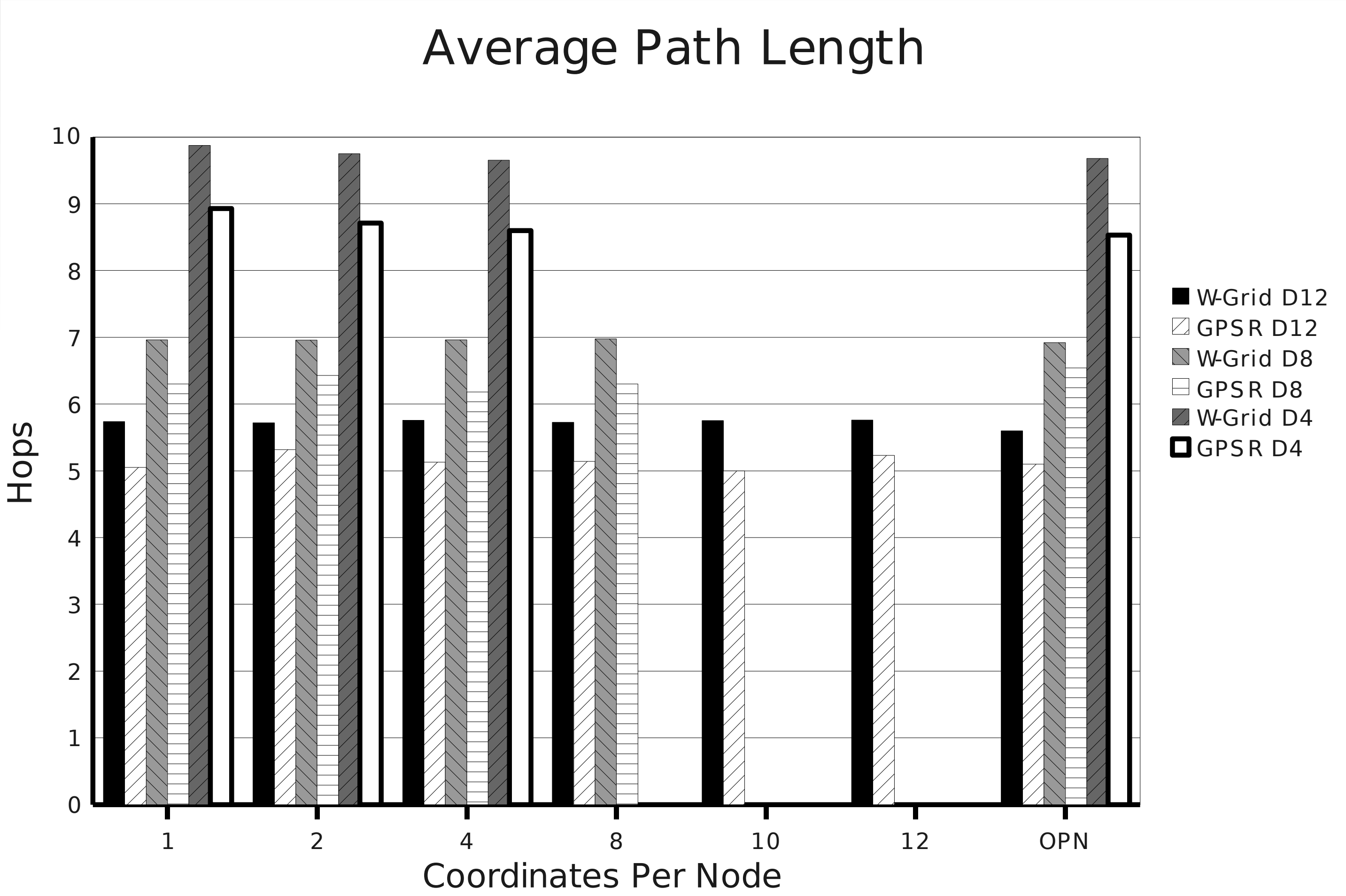
W*-Grid Failure recovery (II)

- Nodes independence allows to recovery from nodes failure without using broadcast
- Recovery is performed by the node(s) that received coordinates from the failed node (descendant)
- Recovery is guaranteed (100% success in simulations) whenever:
  - The failure do not partition the network
  - The node which performs the recovery or at least one of its neighbors is not fully dependent from the failed node

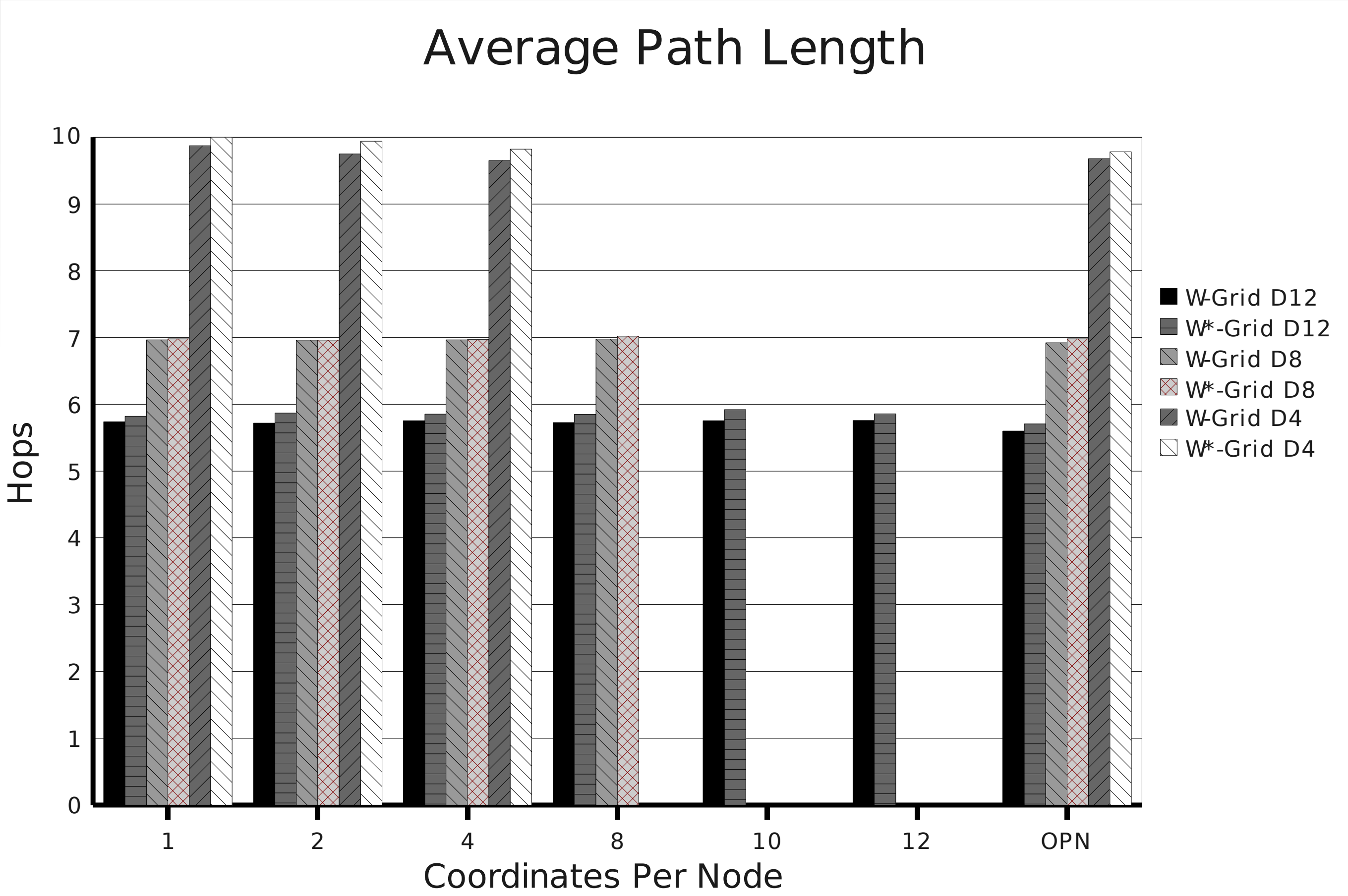
W*-Grid Simulation results

- The variation in queries APL (Average Path Length), the number of hops necessary to resolve a query, between W*-Grid, W-Grid and GPSR
- The number of coordinates in the system in W*-Grid and W-Grid scenarios
- The ratio of succeeded failure recoveries in W*-Grid and in a scenario in which broadcast is used to discover the closest relative

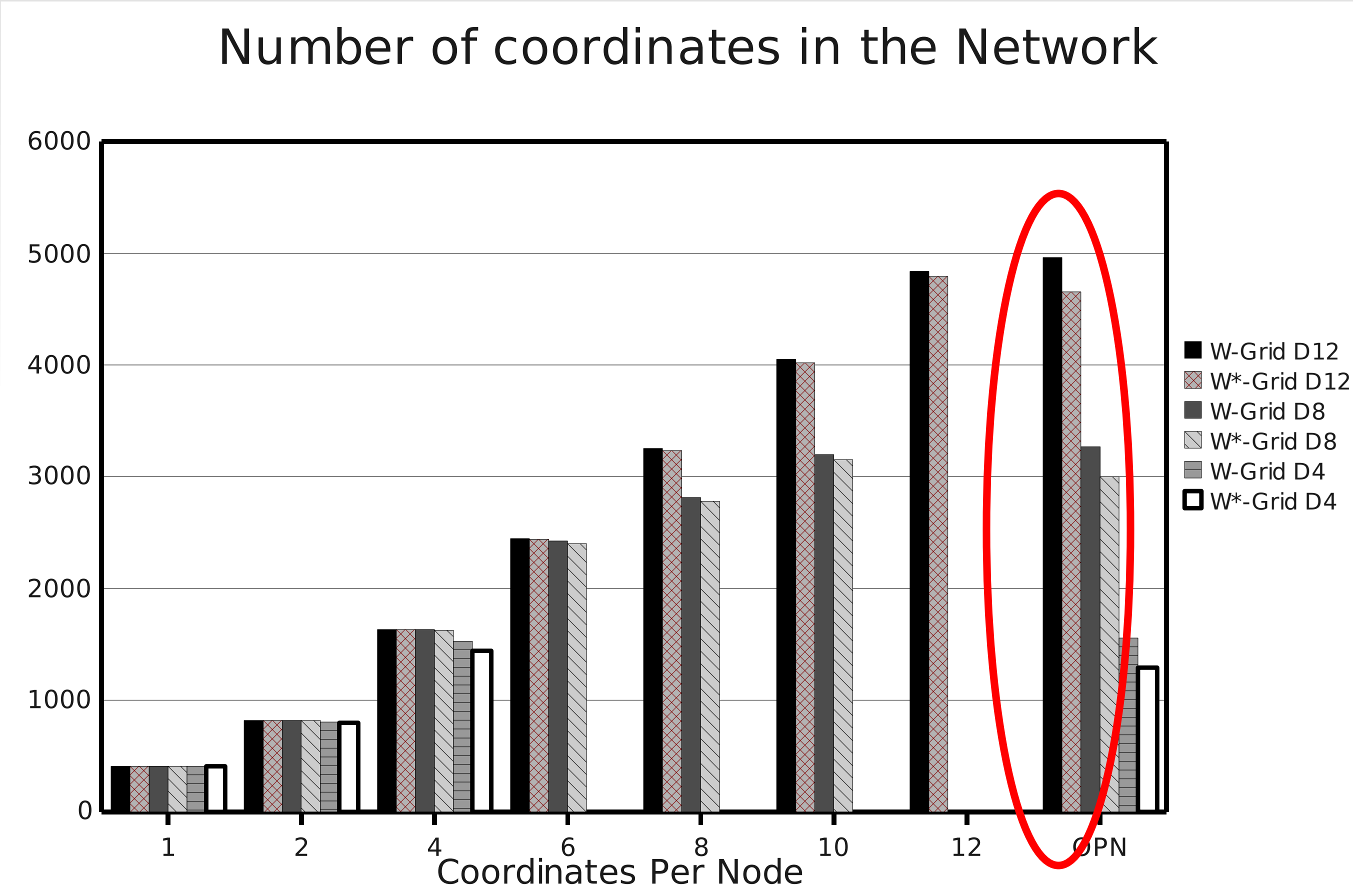
Average Path Length (APL) W-Grid vs. GPSR



Average Path Length (APL) W*-Grid vs. W-Grid



Number of coordinates in the network



Collaborative Signal Processing and Data Fusion

Collaborative signal processing

- Gathering useful information about the physical world
- Sensors' sensing modality
  - seismic, acoustic, thermal, IR, etc.
- Current detection and classification algorithms are based on single-node processing
- Single processing is not always enough
  - Localization and tracking algorithms require collaboration between nodes

Here comes Data fusion

- Infer global information over a certain space-time region from local observations
- Data fusion can be used to allow each sensor to send quantized data (decision) to a fusion center
  - reduce data size
  - prevent overloading the wireless network
  - conserve energy
- Data Fusion it is not just aggregation
  - detection (boolean decision)
  - classification (multiple values)

Data Fusion

- Fusion application is challenging in general
  - requires time-correlation and synchronization of data streams coming from different sources.
  - usually requires parallelism
- Sensornets add another level of complexity
  - power-constrains
- Fusion is sometimes confused with aggregation
  - Aggregation is typically performed on data of same type in order to minimize communication
  - Fusion is performed on data of possibly different types in order to derive a decision

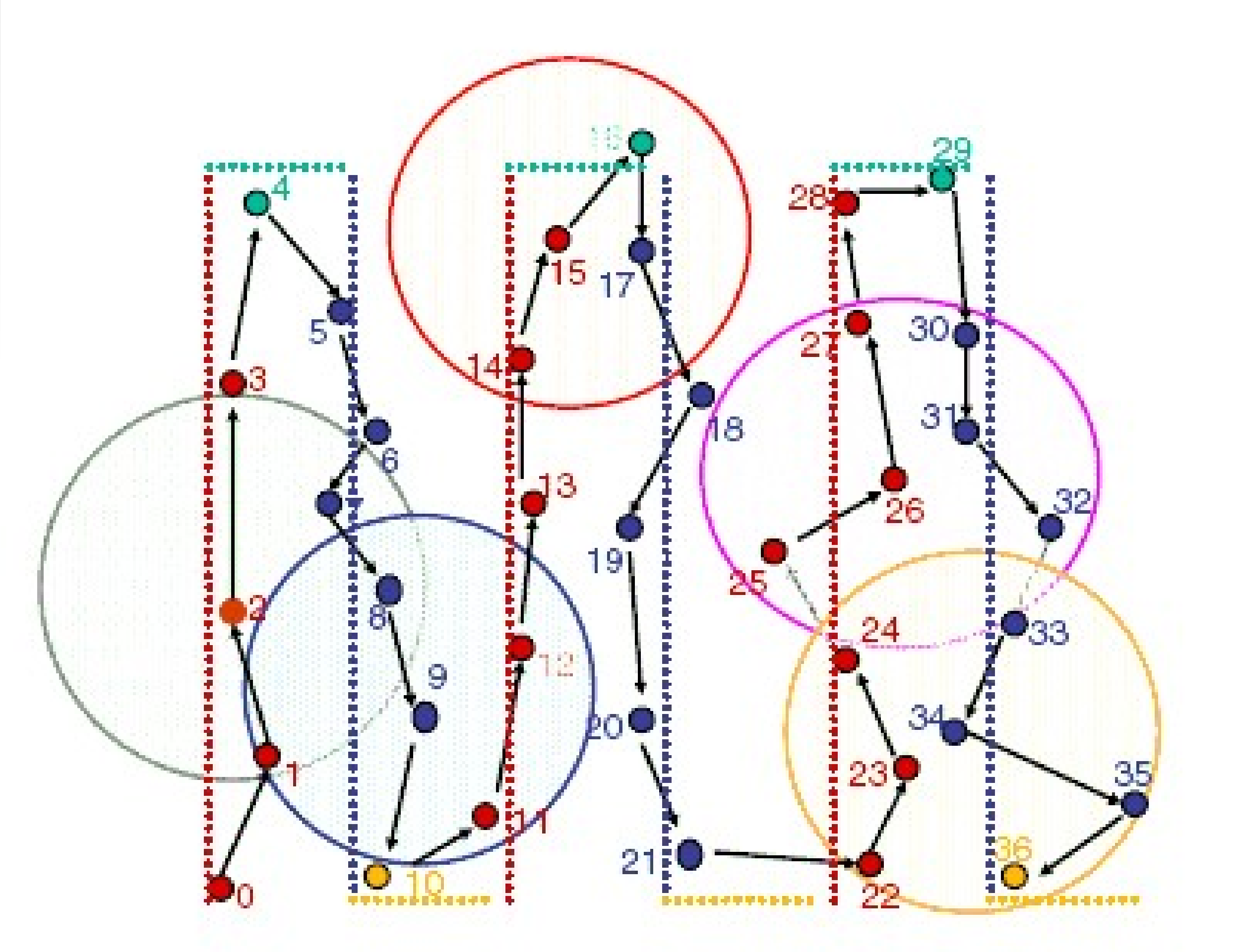
Serial Fusion

- Requires computing a routing path in the network that travels through each sensor
    - network graph traversal problem
    - need to find the minimum number of hops
- Must develop practical traversal techniques in sensor networks
    - use of a geographically-based technique based on the general notion of space-filling curve
- Depth-first traversal
    - geographically based intelligence as to which unvisited neighbor should be traversed next, where more than one is available

Nodes Traversal(I)

- Space-filling curves
  - mapping of any point in the space to a corresponding point on the curve
- The traversal starts from a specific node
  - use of a message or agent
  - each node knows its neighbors(identifications and coordinates)
- At each node, let say P
  - the agent orders all unvisited neighbors of P by their curve indices
  - the unvisited neighbor which is next in this order is visited
- If the agent (or message) finds itself at a node P with no unvisited neighbor it backtracks

Nodes Traversal(II)



Self-organizing Sensor Networks: State of the art, new solutions and applications

Papers

- Moro, G., Monti, G. *W-Grid: a Cross-Layer Infrastructure for Multi-Dimensional Indexing, Querying and Routing in Ad-Hoc and Sensor Networks*. In: P2P 2006: Sixth IEEE International Conference on Peer-To-Peer Computing, Cambridge, UK, IEEE Computer Society (2006) 210–220
- Moro, G., Monti, G., Ouksel, A. *Routing and localization services in self-organizing wireless ad-hoc and sensor networks using virtual coordinates*. In: ICPS'06: International Conference on Pervasive Services. IEEE Computer Society (2006) 243–246
- Monti, G., Moro, G., *W-grid: A P2P Self-organizing Infrastructure For Routing In Ad-hoc Wireless Networks.* In AICA 2006, Congresso Annuale Associazione Italiana per l'Informatica ed il Calcolo Automatico.
- Monti, G., Moro, G., Sartori C. *W^R-Grid: A Scalable Cross-Layer Infrastructure for Routing, Multi-Dimensional Data Management and Replication in Wireless Sensor Networks*. In: ISPA'06, International Symposium on parallel and distributed processes and applications

Papers

- Monti G., Moro G. and Lodi S., *W*-Grid: A Robust Decentralized Cross-layer Infrastructure for Routing and Multi-Dimensional Data Management in Wireless Ad-Hoc Sensor Networks*. In P2P 2007: Seventh IEEE international conference on Peer-to-Peer computing, Galway, Ireland, 2007
- Monti G., Moro G, *W*-Grid: A Robust Decentralized Cross-layer Infrastructure for Routing and Multi-Dimensional Data Management in Wireless Ad-Hoc Sensor Networks.* Submitted to IEEE Transactions on parallel and distributed systems

Bibliography

[1] Aris M. Ouksel, The Interpolation-Based Grid File. PODS1985, 90-105, 1985.

[2] Karp, B., Kung, H.: GPRS: greedy perimeter stateless routing for wireless networks. In: MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking, ACM Press (2000) 243–254

[3] Li, X., Kim, Y., Govindan, R., Hong, W.: Multi-dimensional range queries in sensor networks. In: SenSys '03: Proceedings of the 1st international conference on Embedded networked sensor systems, New York, NY, USA, ACM Press (2003) 63–75

[4] Xiao, L., Ouksel, A.: Tolerance of localization imprecision in efficiently managing mobile sensor databases. In: MobiDE '05: Proceedings of the 4th ACM international workshop on Data engineering for wireless and mobile access, New York, NY, USA, ACM Press (2005) 25–32

[5] Ratnasamy, S., Karp, B., Shenker, S., Estrin, D., Govindan, R., Yin, L., Yu, F.: Data-centric storage in sensornets with ght, a geographic hash table. Mob. Netw. Appl. 8(4) (2003) 427–442

[6] Greenstein, B., Estrin, D., Govindan, R., Ratnasamy, S., Shenker, S.: Difs: A distributed index for features in sensor networks. In: Proceedings of first IEEE WSNA, IEEE Computer Society (2003) 163–173

Alcune possibili applicazioni

- Outdoor
  - Ambientali Estreme
    - riconoscere/prevedere tsunami dal moto ondoso e segnalazione via satellite solo se la probabilità supera una soglia fissata
    - idem per eruzioni vulcaniche, tornado etc.
  - Agricoltura di precisione
    - monitoraggio di piantagioni per riconoscere/prevedere insorgenza malattie, deterioramento/maturazione prodotti, precipitazioni etc.
- Indoor
  - Controllo dei processi produttivi
  - Domotica
    - risparmio energetico tramite anche apprendimento delle abitudini delle persone, sicurezza, intrusioni, wellness

Sviluppo della ricerca

- Studio e implementazione di tecniche per l'ottimizzazione del sistema:
  - Analisi degli effetti della presenza di alias all'aumentare del tempo di vita della rete
  - Consentire il riutilizzo di coordinate appartenute a nodi usciti dalla rete
- Introdurre nel simulatore mobilità dei sensori e/o di attori che attraversano la rete statica

Sviluppo della ricerca

- Simulare reti P2P mobili costituite da sensori, palmari, computer che formano reti metropolitane
  - informazioni immesse in rete alla p2p per scopi sociali, commerciali, di emergenza e pubblica utilità
  - Implementare nel simulatore diversi scenari
    - Numerosità dei partecipanti
    - Densità e distribuzione
- Aggiungere al simulatore nuovi algoritmi con cui confrontarsi
  - Es. AODV, DSR

Grazie!