Optimization of Intrusion Detection Systems in Wireless Sensor Networks

DTEDI

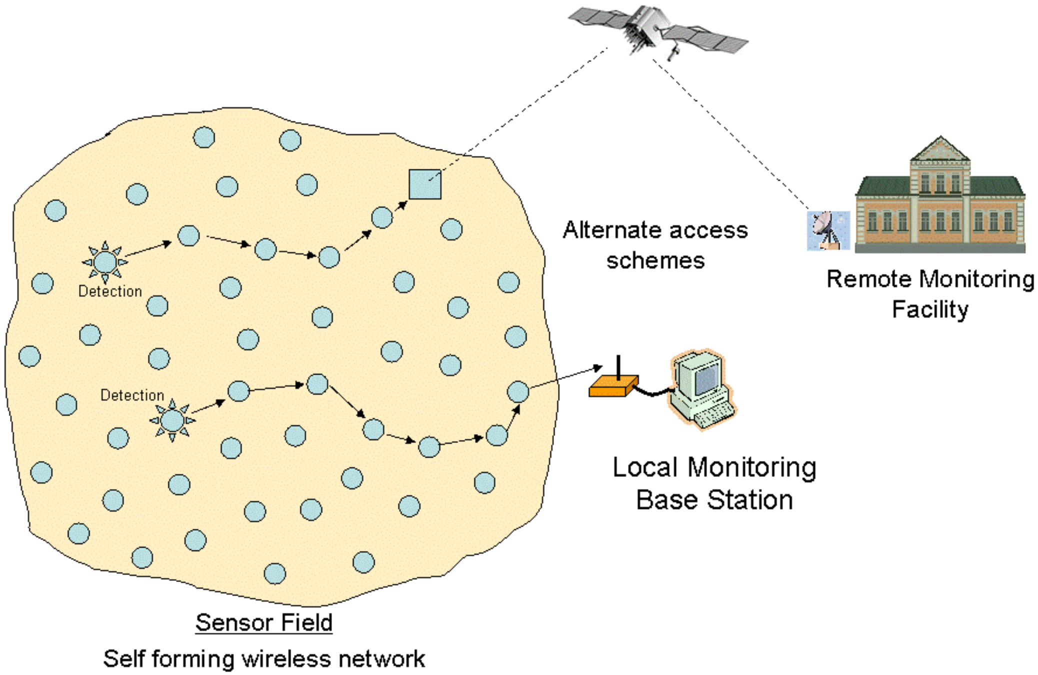
Martin Stehlík

DTEDI – Introduction to my research
Faculty of Informatics, Masaryk University, Brno

November 2012

Wireless Sensor Network (WSN)

- Highly distributed network which consists of many low-cost sensor nodes and a base station (or sink) which gathers the observed data for processing.



Source: http://embedsoftdev.com/embedded/wireless-sensor-network-wsn/

Sensor node (TelosB)

- Microcontroller
  - 8 MHz, 10 kB RAM
- External memory
  - 1 MB
- Radio
  - 2.4 GHz, 250 kbps
- Battery
  - 2 x AA (3 V)
- Sensors
  - Temperature, light, humidity, …

Security

- WSNs are more vulnerable than conventional networks by their nature.
- Sensor nodes:
  - Have lower computational capabilities.
  - Have limited energy supply.
  - Can be easily captured.
  - Are not tamper-resistant.
- WSNs are deployed in hostile environment.

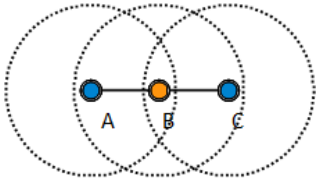
Attacker model

- Passive attacker
  - Eavesdrops transmissions.
- Active attacker
  - Alters data.
  - Drops or selectively forwards packets.
  - Replays packets.
  - Injects packets.
  - Jams the network.

=> can be detected by *Intrusion Detection System* ☺

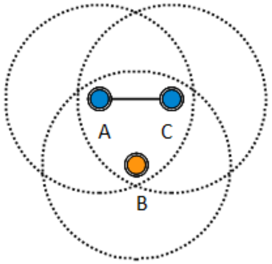
Intrusion detection system (IDS)

- IDS can monitor packets addressed to itself.



Local monitoring

- IDS can overhear and monitor communication of its neighbors.



Watchdog monitoring

Figures: Andriy Stetsko

IDS techniques

- Many techniques have been proposed to detect different attacks.
- We can measure:
  - Packet send & delivery ratio.
  - Packet sending & receiving rate.
  - Carrier sensing time.
  - Sending power.
- And monitor:
  - Packet alteration.

IDS optimization

- Sensor nodes are limited in its energy and memory.
- Better IDS accuracy usually requires:
  - *Energy* (network lifetime).
  - *Memory* (restriction to other applications).

⇒ **Trade-off** between *IDS accuracy* and *WSN performance*.

- Parameters of IDS can be optimized!

IDS optimization framework

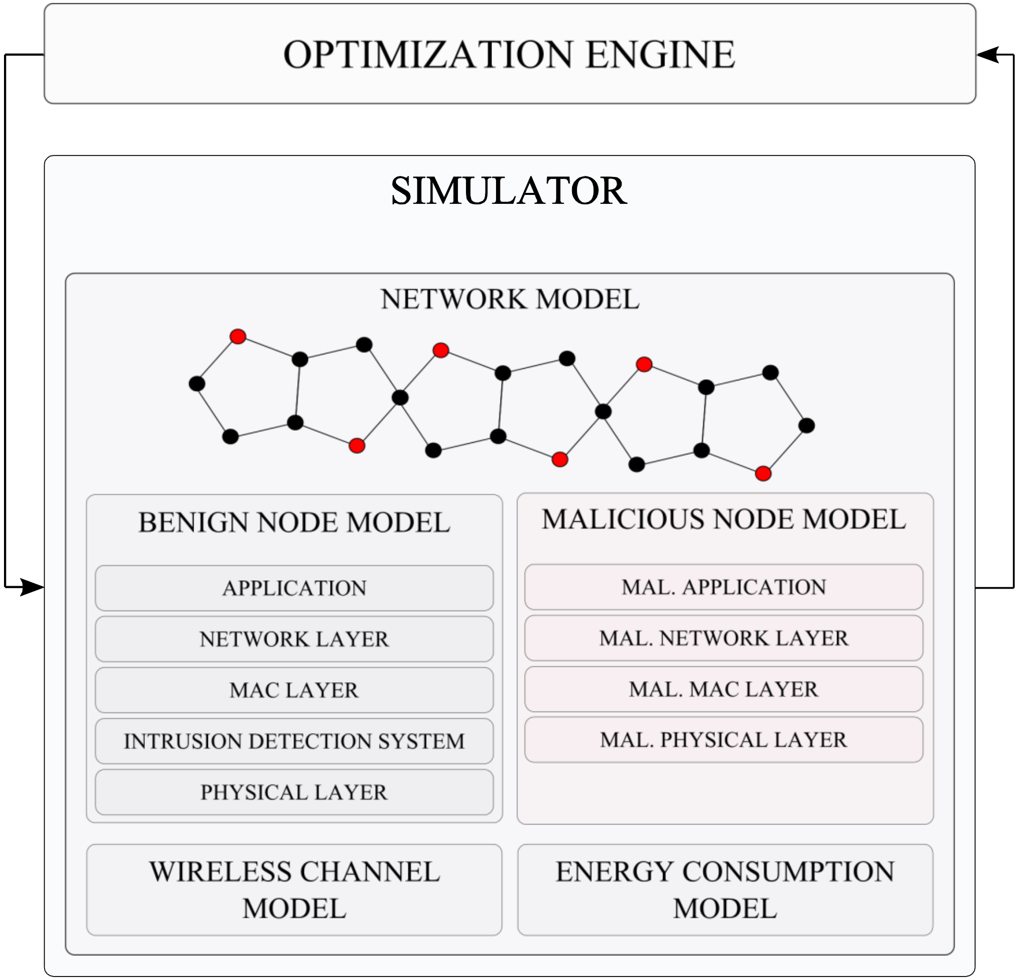


Figure: Andriy Stetsko

Why do we simulate WSN?

- New protocols and security approaches are being developed rapidly => need to investigate and explore their functionality.
- Time of implementation and runtime (e.g. battery depletion).
- Simulation of hundreds or thousands of sensor nodes during development of new WSN solutions.
- Verifiability of results.
- Repeatability of tests.

Simulation of WSN

- Accurate simulation of wireless channel and energy consumption is important to verify our proposals.
- Protocols which work during simulations may fail in real environment because of simplicity of the model.
- Many simulators of different quality are available.
- Some of them are developed specifically for wireless networks or even for WSN, others are generic or generic with specific extension/framework.

Simulation of WSN

- Model should represent:
  - Environment.
  - Radio signal propagation.
  - Topology.
  - Physical properties of sensor nodes (radio chips and batteries).
  - Protocols (PHY and MAC).
- We performed comprehensive comparison in the past.
- Currently we use:
  - MiXiM.
  - TOSSIM.

IDS optimization framework

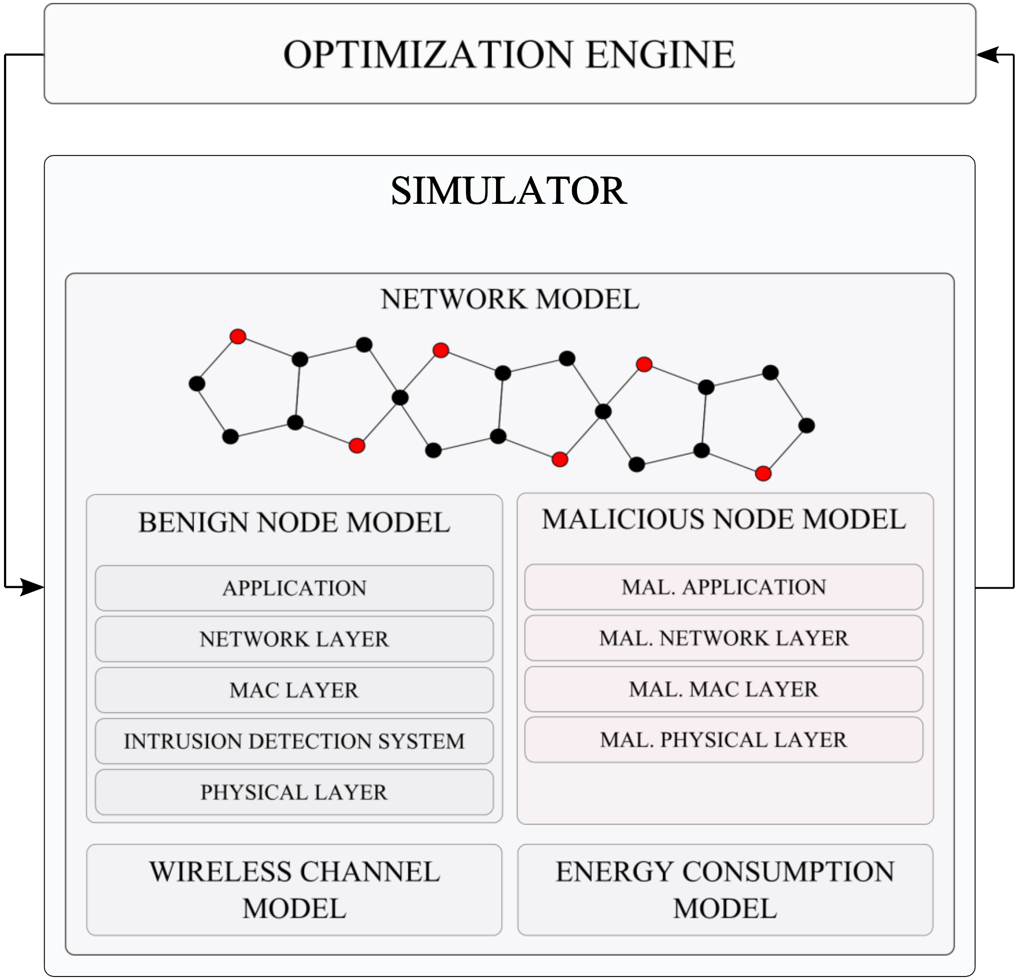


Figure: Andriy Stetsko

Simulator

- *Input:* candidate solution represented as a simulation configuration.
  - Number of monitored neighbors.
  - Max. number of buffered packets.
- *Output:* statistics of a simulation.
  - Detection accuracy.
  - Memory and energy consumption.
- *Simulation:* specific WSN running predefined time configured according to the candidate solution.

IDS optimization framework

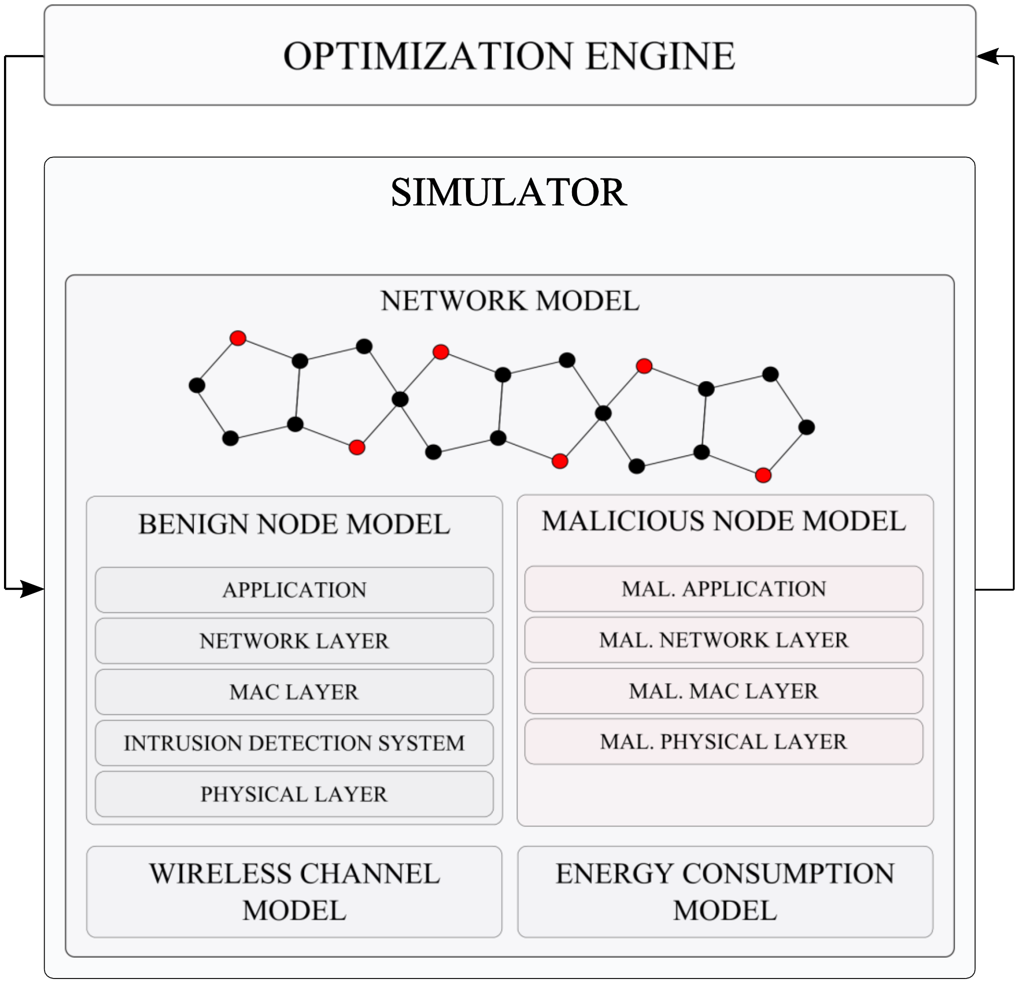


Figure: Andriy Stetsko

Optimization engine

- *Input:* statistics from the simulator.
    - Detection accuracy.
    - Memory and energy consumption.
- *Output:* new candidate solution(s) in form of simulation configurations.
    - Number of monitored neighbors.
    - Max. number of buffered packets.
- *Algorithms:* evolutionary algorithms, particle swarm optimization, ant colony optimization, ...

Multi-objective evolutionary algorithms

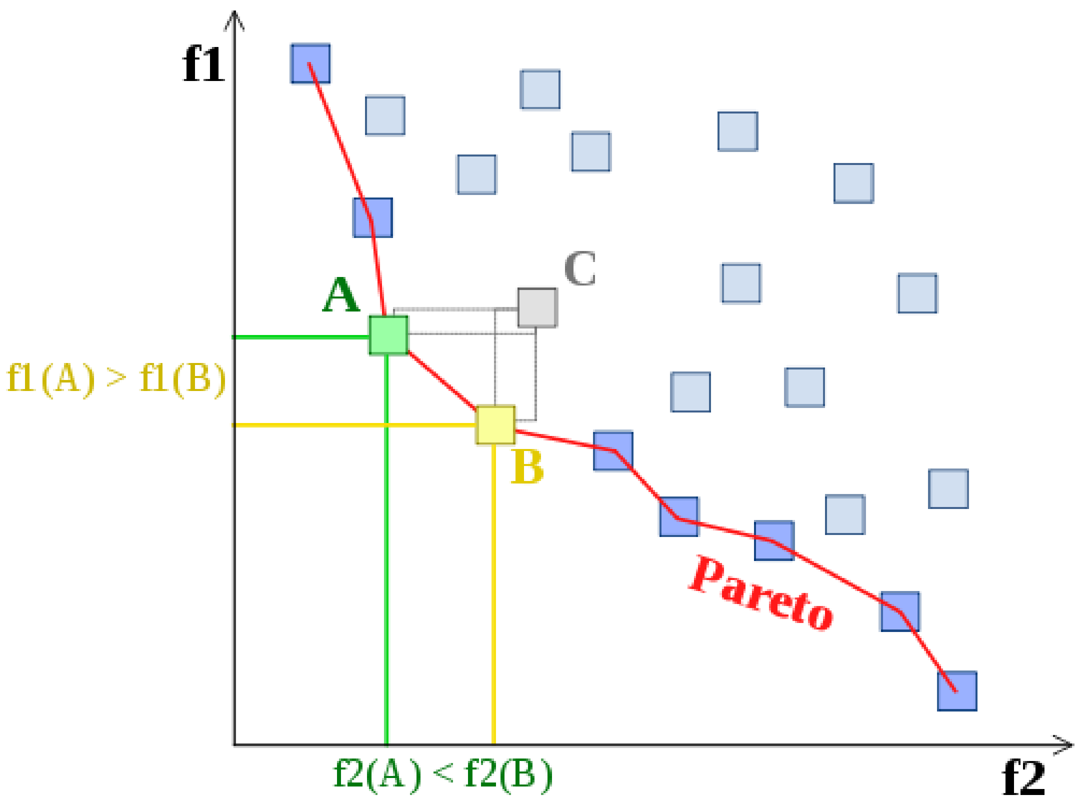
- Single aggregate objective function

$$\frac{1}{2|C|} * \sum_{c_i \in C} \frac{x(c_i)}{n(c_i)} + \frac{1}{2|A|} * \sum_{a_i \in A} \frac{y(a_i)}{n(a_i)} + 0.1 \frac{1}{|C|} * \sum_{c_i \in C} \frac{1}{1+m(c_i)}$$

- Pareto-based ranking schemes.
  - Set of non-dominated solutions.

Pareto front

- Set of non-dominated solutions.



Source: http://en.wikipedia.org/wiki/Pareto_efficiency

Comparison of MOEA

- Quality of Pareto front approximation.
- Diversity of found solutions.
- Speed of convergence.

=> All based on:

- Algorithms (NSGA-II, SPEA2).
- Mutation and crossover probabilities.
- Population size.
- Number of generations.

Thesis proposal

- Examination of the optimization techniques.
  - Evolutionary algorithms, multi-objective evolutionary algorithms, coevolutionary algorithms.
- Optimization of IDS for specific attacks.
  - Selective forwarding attack, delay attack, data modification attack, jamming attack and Sybil attack.
- Impact of topology, wireless channel model, traffic and environment on the optimization.
  - Robust solutions for complex changing environments.

Thesis proposal

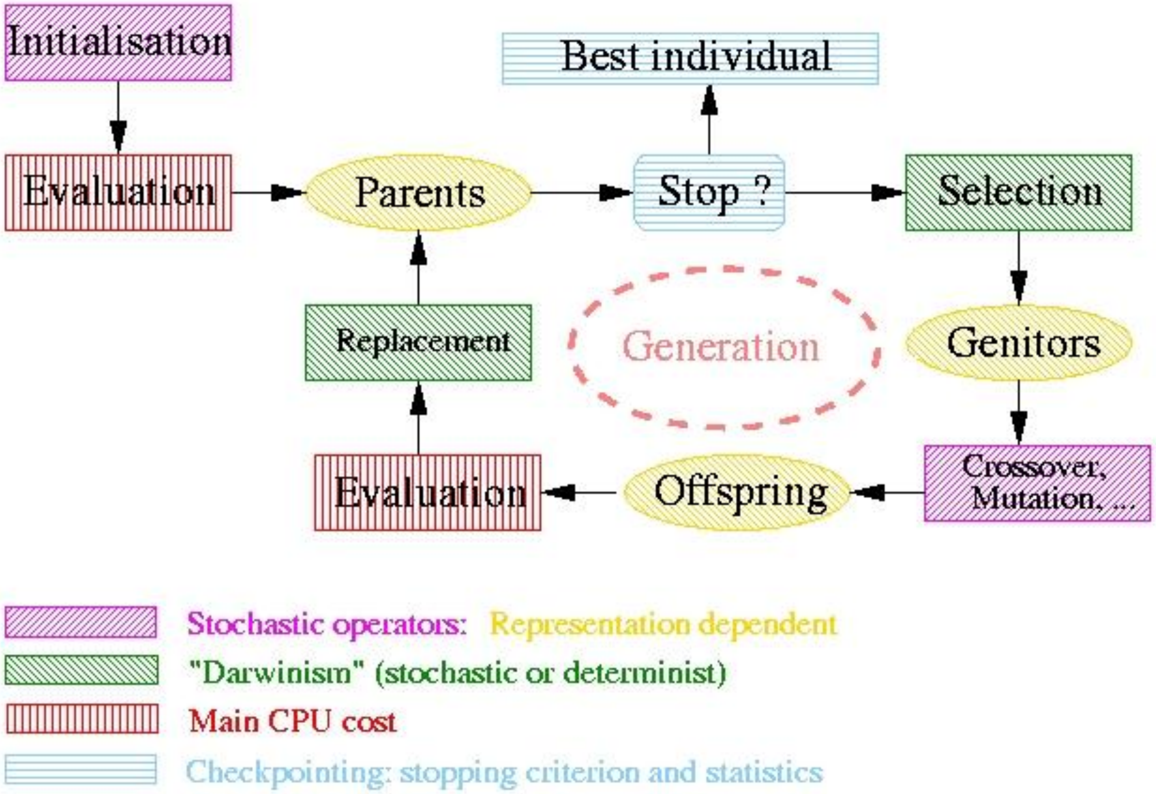
- Investigation of options for configurations of the whole network stack using optimization techniques or semi-automatically.
  - Application, network, MAC and PHY layer. Intrusion detection system.
- Integration of the found solutions into a working IDS design framework for wireless sensor networks.
  - Framework will be tested in our laboratory testbed and released under a suitable open access license.
- Interdisciplinary research.
  - WSN, security, optimizations.

Thank you for your attention.

Questions?

Evolutionary algorithms

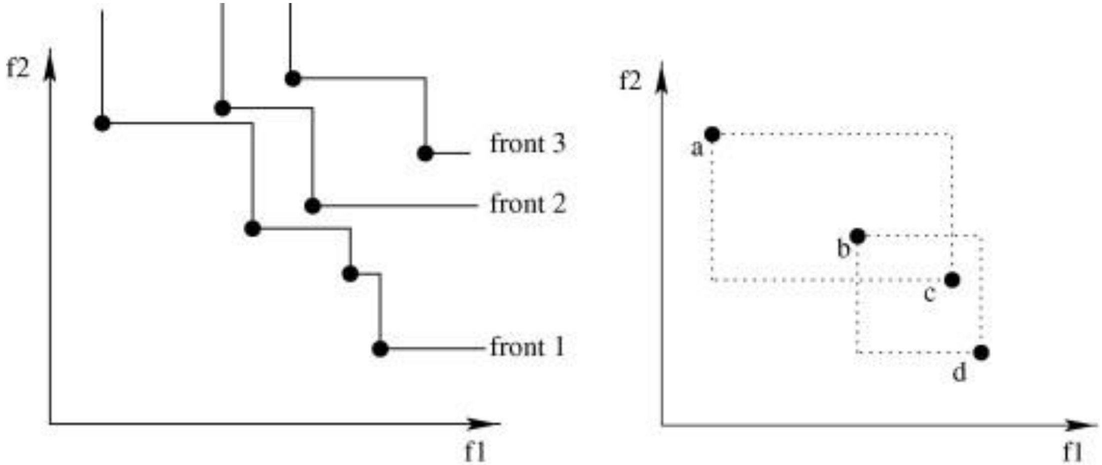
- Inspired in nature.



Source: http://eodev.sourceforge.net/eo/tutorial/html/EA_tutorial.jpg

NSGA-II

- Nondominated Sorting Genetic Algorithm II.
- Two criteria:
  - Ranking using nondominance concept *(convergence)*.
  - Crowding distance *(diversification)*.



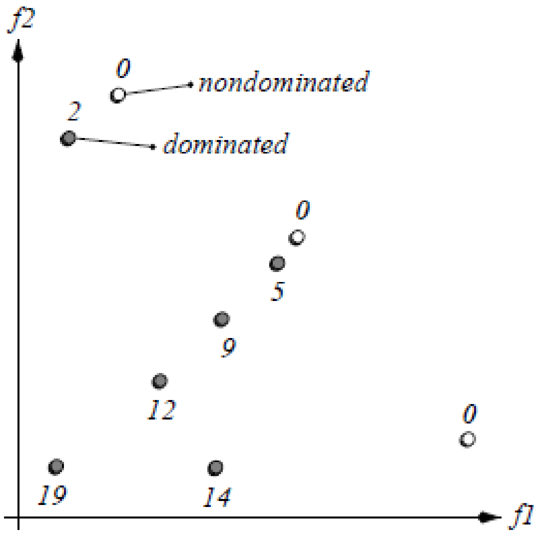
(a) Non-dominance sorting to determine fronts

(b) Crowding distance calculation within a front

Source: J. Branke, B. Scheckenbach, M. Stein, K. Deb, H. Schmeck,
Portfolio optimization with an envelope-based multi-objective evolutionary algorithm, 2009.

SPEA2

- Strength Pareto Evolutionary Algorithm 2.
- Fitness value based on:
  - Number of dominating solutions and their strength of dominance.
  - Density estimation.



Source: E, Zitzler, M. Laumanns, L. Thiele,
SPEA2: Improving the Strength Pareto Evolutionary Algorithm, 2001.

Our test case

- Tools:
  - Simulator MiXiM, ParadisEO, BOINC.
- Wireless channel model:
  - Based on own results for outdoor environment.
- MAC layer.
  - CSMA.
- Topology:
  - 100, 250 and 500 uniformly distributed sensor nodes.
  - Topology corresponding to the lab testbed.

Our test case

- IDS:
  - Detection of selective forwarding and dropping based on watchdog monitoring.
- Optimized parameters:
  - p1 – number of nodes to be monitored. Influences accuracy and memory usage.
  - p2 – number of packets stored in a buffer. Influences accuracy and memory usage.
  - p3 – number of packets received. Influences accuracy.
  - p4 – detection threshold. Influences accuracy.

Coevolutionary algorithms

- Competitive Coevolutionary Algorithms.
  - Individuals are rewarded at the expense of those with which they interact.
- Cooperative Coevolutionary Algorithms.
  - Individuals are rewarded when they work well with other individuals.
- Would it be possible to use coevolutionary algorithms to optimize the IDS?
- The first population would aim to produce the best IDS while the second population would produce more and more sophisticated attacks.