

MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY



Evolutionary optimization of intrusion detection system in wireless sensor networks

DIPLOMA THESIS

Adam Saleh

Brno, spring 2008

Declaration

Hereby I declare, that this paper is my original authorial work, which I have worked out by my own. All sources, references and literature used or excerpted during elaboration of this work are properly cited and listed in complete reference to the due source.

Adam Saleh

Advisor: Andryi Stetsko, Ph.D.

Acknowledgement

I would like to thank my supervisor ...

Abstract

Keywords

Wsn,ids,spea2,nsgaii

Contents

1	Introduction	1
2	Evolutionary Optimization of Wireles Sensor Networks	3
2.1	<i>Basic principles of metaheuristics</i>	3
2.1.1	Diversification	4
2.1.2	Intensification	4
2.2	<i>Single objective evolutionary algorithms</i>	5
2.2.1	Template of an evolutionary algorithm. (b Talbi)	5
2.3	<i>Multiobjective evolution algorithms</i>	6
2.4	<i>NsgaII</i>	8
2.4.1	Non-dominated sorting	8
2.4.2	Crowding distance sorting	9
2.5	<i>Spea2</i>	9
2.5.1	Fitness assigment	10
2.5.2	Enviromental selection	11
2.5.3	Spea2+	11
2.6	<i>Ibea</i>	12
2.6.1	ϵ -indicator	12
2.6.2	Hypervolume based indicator	12
2.7	<i>Choosing the algorithm</i>	14
3	Our Framework	15
3.1	<i>Paradiseo MOEO</i>	15
3.1.1	Object-oriented design of MOEA	15
3.1.2	Population object	16
3.1.3	Fitness and diversity assigment	17
3.1.4	Creation of new generation	17
3.1.5	Stopping criteria	17
3.2	<i>Entry points</i>	18
3.2.1	Input parameters	18
3.2.2	Objectives	19
3.2.3	Evaluation function	19
3.2.4	Boinc Assisted Evolution	19
3.2.5	Mutation and Crossover	20
3.2.6	Reporting	21
3.3	<i>Example problem</i>	21
4	Calibration	23

4.1	<i>Exhaustive search</i>	23
4.2	<i>Iteratively</i>	24
4.3	<i>Measuring convergence</i>	24
4.3.1	Hypervolume indicator	24
4.3.2	Distance indicator	25
4.3.3	Average distance indicator	25
4.4	<i>Measuring diversity</i>	26
4.5	<i>Population size and Number of generations</i>	27
4.6	<i>Mutation and Crossover</i>	27
5	Experiments	28
5.1	<i>Wireless sensor network</i>	28
5.2	<i>Intrusion Detetection System</i>	28
5.3	<i>Setup</i>	30
5.4	<i>Calibration</i>	30
5.4.1	Mutation function	30
5.4.2	Crossover function	30
5.5	<i>Evolution with aid of BOINC</i>	32
6	Results	34

1 Introduction

Wireless sensor networks are relatively new concept describing a wirelessly communicating network of battery powered nodes, that contain some computational capacity and are able to gather information with attached sensors. Current research hints at applications in monitoring wild-life and gathering information in otherwise unattainable places. They are an interesting topic for security research because of their possible military and intelligence gathering applications, their distributed nature and the constraints on their hardware.

For example, setting up intrusion detection system on network of nodes so that it satisfies constraints on accuracy, battery life and memory can be a timeconsuming process. With the usage of simulation frameworks, such as MiXiM, effort spent on optimization is directly proportional to invested computational resources, with realistic simulations taking up to several hours of processor time, based on its parameters.

Achieving high accuracy and low memory footprint are in many cases inversely proportional. This means that the network operator has to choose a tradeoff between desired objectives. We believe that this process of optimization and evaluation of different tradeoffs could be greatly improved by using multi-criteria algorithms. Multi-criteria algorithms instead of a single solution return a set of candidates covering different tradeoffs between the objectives. This allows for better evaluation of said objectives.

Evolutionary algorithms help immensely with reducing demands for computational resources while optimizing problems. Because evolutionary optimization works by iterating evaluations and modification of a set of candidates, algorithms have been successfully modified to solve multi-optimization problems.

We explain the basic ideas behind evolutionary heuristics in second chapter, as well as their applicability on concepts of multi-criteria optimization in terms of pareto optimality, as well as the main principles of three state of the art multicriteria evolutionary algorithms.

In third chapter we introduce our black-box micro-framework that is based on paradiseo evolutionary framework. Because our framework is partially written in paradiseo, we cover some of its design decisions in

grater detail. Then we have provided a walkthrough of all the configuration entry points of our microframework, complete with examples in Python.

We discuss calibration of the framework in fourth chapter. Because evolutionary algorithms are heuristic in nature, we provide several metrics that can be used to guide the settings.

In the last chapter we discuss intrusion detection systems for wireless sensor networks and provide a model example of optimizing such wireless sensor network application, its setup, analysis and calibration of the framework and parallel evaluation on Boinc cluster. We conclude with short comparison of used evolutionary algorithms.

2 Evolutionary Optimization of Wireless Sensor Networks

As the title states, we can afford to be quite specific in the description of our optimization problem. Usually more formal definitions of optimization problems deal with inputs and outputs in terms of words over alphabet and appropriate cost function. We think that defining our optimization problem in terms of n -tuples of real numbers will be better at representing evaluation of a WSN configuration.

Definition 1 (Optimization problem). *We define optimization problem as a triplet (X, Z, f) , where $X \subseteq \mathbb{R}^n$ is the decision space, $Z \subseteq \mathbb{R}^m$ is the objective space and $f : X \rightarrow Z$ is the evaluation function. Without loss of generality it is assumed, that we want to minimize objective.*

This definition allows the evaluation function f to output incomparable objective vectors, which is a key concept for multi-criteria optimization. Optimization problem can be easily made compatible with single-criteria algorithms by defining cost-function $c : Z \rightarrow \mathbb{R}$, that creates an aggregated metric from the objective vector.

Because of the nature of wireless communication, where interference in shared medium has to be taken into account, it is hard to predict how will a change in configuration impact the result. Therefore thorough and expensive simulation is required for every configuration that we would want to evaluate. This renders the usage of simple exhaustive search for optimization purposes impractical. We could say, that our optimization problem has problem space that is hard to characterize and expensive to search, which are good fit for heuristic approaches[5].

[copy figure 1.9 from talbi]

2.1 Basic principles of metaheuristics

Most of the heuristics used are variants on search through problem space, with different approaches to avoid undesirable local optima. There are two basic principles that are used when designing a meta-heuristic:

2. EVOLUTIONARY OPTIMIZATION OF WIRELES SENSOR NETWORKS

2.1.1 Diversification

An algorithm should explore as much of the search space as possible. This principle is called diversification.

Random search can be considered an extreme application of diversification, where in every round we generate next solution randomly without using any information from previous (good) solutions. Less extreme example is the family of population based heuristics, where in each iteration we are trying to improve a set of candidates, called a population.

Problem spaces with many local optima are less likely to cause problems for algorithms that have strong diversification component.

2.1.2 Intensification

An algorithm should exploit information from the solutions it previously found. This principle is called intensification, because it covers the notion of trying to intensify a good solution to find a better one.

Local search can be considered an extreme application of intensification, where next instance to evaluate is always selected deterministically, based on the previous solution. Local search and its variants belong to a family of single-solution based heuristics, where in each iteration we are basing our search of the problem-space of single solution.

Algorithms with strong intensification component are usually more efficient with regards to number of unique evaluations.

Because of their nature, heuristic approaches are hard to evaluate and reason about, therefore they are often considered to be a method of last resort. (b Talbi) warns against using heuristics lazily on problems where more deterministic approaches (such as using a SAT-solver or linear programming) would be more efficient. Heuristic approaches are usually considered when we need to solve our problem in order of magnitude faster than possible using conventional approaches, while sacrificing guarantees on optimality.

In our case, we are constrained by number of configuration evaluations we are able to perform in a given timeframe. We are focusing on evolutionary heuristics, because they seem to have reasonable performance with regard of optimizing WSN. These are modeled natural process of evolution of species. It is a family of stochastic, population-

2. EVOLUTIONARY OPTIMIZATION OF WIRELES SENSOR NETWORKS

based heuristics (as opposed to deterministic, single-solution based).

2.2 Single objective evolutionary algorithms

Evolutionary algorithms are based on the fact, that natural process of evolution can be considered an algorithm solving an optimization problem of adapting a species to its environment.

2.2.1 Template of an evolutionary algorithm. (b Talbi)

Basic principles behind evolutionary algorithms can be shown on this template.

Step 0: Initialization: Generate an initial population of individuals

Step 1: Fitness assignment: calculate the fitness values of individuals in population

Step 2: Environmental selection: select the best individuals of a population to "survive" to the next generation

Step 3: Termination: check if termination criterion holds (usually based on number of generations), if it does, return the result and stop.

Step 4: Variation: apply mutation and crossover parameters on the surviving individuals, increment the generation counter and proceed with step 1.

If we continue with this metaphor, we can find these parallels:

- Population of individuals is a set of configurations. Their environment is the optimization problem.
- We decide surviving individuals based on their evaluation function score.
- From surviving individuals we generate next generation, either by mutation or crossover.

2. EVOLUTIONARY OPTIMIZATION OF WIRELES SENSOR NETWORKS

- Mutation of individual usually consists of randomly selecting some configuration from its neighbourhood as its offspring. This step provides intensification for evolution heuristic.
- Crossover usually consists of creating an offspring by permutation of two individuals. This step provides diversification for evolution heuristic.

In algorithms, there are variations on this template, for example, instead of first selecting the best and then generating the rest of population, *variation* can precede *enviromental selection*, and enviromental selection then trims the population to the confines of preset population limit.

In our framework selection and replacement steps are part of the evolutionary algorithm, while crossover and mutation functions are to be calibrated on case by case basis.

2.3 Multiobjective evolution algorithms

Problem that we are trying to optimize is unfortunately illsuited for single objective evolution. Objectives such as memory consumption and IDS accuracy are ortoghonal, therefore it is hard to determine single criteria that would satisfactory cover both of them. A usual solution is to provide some sort of weighted average. In this case, because output of our algorithm would be only a single solution, we need to know emphasis on different criteria before we run our optimization.

Better approach is to recognize multi-objective nature of our problem.

Multi objective algorithms are based on idea of Pareto optimality and domination. We say that solution A dominates solution B, if A has no objective "worse" than B and at least one objective "better". We say that A is pareto-optimal, if there is no other solution that would dominate A. This means that no objective of A can be improved without deterioration of another objective.

Definition 2. *Pareto dominance* An objective vector $u = (u_1, \dots, u_n)$ is said to dominate $v = (v_1, \dots, v_n)$ (denoted by $u \prec v$) if and only if

2. EVOLUTIONARY OPTIMIZATION OF WIRELES SENSOR NETWORKS

no component of v is smaller¹ than the corresponding component of u and at least one component of u is strictly smaller:

$$\forall i \in \langle 1, n \rangle : u_i \leq v_i \wedge \exists i \in \langle 1, n \rangle : u_i < v_i$$

Definition 3. *Pareto optimality* A solution $x \in S$ is Pareto optimal if for every $x' \in S$, $F(x')$ does not dominate $F(x)$, that is

$$\forall x' \in S, F(x) \not\prec F(x')$$

Subset of solutions where no solution from the superset dominate the one in the set is called Pareto optimal set.

Definition 4. *Pareto optimal set* For given $MOP(F, S)$, the pareto optimal set is defined as $\mathcal{P}^* = \{x \in S \mid \nexists x' \in S, \mathcal{F}(x') \prec \mathcal{F}(x)\}$

Pareto front is the largest Pareto optimal set in the solution space. It is the set of all non-dominated solutions the solution space.

Definition 5. *Pareto front* For given $MOP(F, S)$, the pareto optimal set is defined as $\mathcal{PF} = \{x \in S \mid \nexists x' \in S, \mathcal{F}(x') \prec \mathcal{F}(x)\}$

We will call a subset that contains only solutions that don't dominate each other a Pareto approximation.

Definition 6. *Pareto approximation* For given $MOP(F, S)$, we call set $A \subset S$ a pareto approximation if $\forall x \in A \nexists x' \in A, F(x') \prec F(x)$

Multiobjective heuristics then try to obtain best approximation of Pareto front. To gauge how good an approximation of Pareto front is, we usually use two criteria, first to measure convergence to Pareto optimal front and second to measure diversity in found pareto set.

Calibration of multi objective evolution algorithms is harder than their single criteria counter-parts, mainly because it is hard to reason about the convergence of output without having the real Pareto optimal front in the first place. Pareto optimal sets usually aren't directly comparable, but there are several metrics that can be used to compare them, some of them used in following algorithms. One of the more popular metric is the hypervolume indicator. We will discuss using these metrics for calibration in greater detail in later chapters.

1. we assume minimization

2.4 NsgaII

Second generation of the nondominated sorting genetic algorithm[2] is currently the most popular multiobjective heuristic. It improved on its previous iteration in several ways, most notable is the exclusion of sharing parameter δ_{share} , that previously needed to be specified to maintain good diversity in final population.

Step 0: Initialization: Generate an initial population P_0

Step 1: Fitness assigment: calculate the fitness values of individuals P_0

Step 2: Variation: Generate temporary population P'_t by applying crossover and mutation operators on P_t

Step 3: Fitness assigment: calculate the fitness of P'_t

Step 4: Nondominated sort: sort the $P_t \cup P'_t$, first by domination rank, second by crowding metric

Step 5: Enviromental selection: create P_{t+1} as the first N individuals of the sorted $P_t \cup P'_t$

Step 6: Termination: If $t \geq T$, or other stopping criterion is satisfied, return nondominated individuals from P_{t+1} , else increment t and continue with step 2.

We can separate step 4, the nondominated sort into two phases:

2.4.1 Non-dominated sorting

Every individual is assigned a rank based on order of nondominated front it belongs to. The first front is the Pareto front of the population:

$$\mathcal{F}_\infty = \{ \mathbf{x} \in \mathcal{P} \mid \nexists \mathbf{x}' \in \mathcal{P}, \mathbf{x}' \prec \mathbf{x} \}$$

Other can be defined recursively as the paretofront of the remaining population not containing previous fronts $R_i = P - \bigcup_{1 \leq j < i} \mathcal{F}_j$:

$$\mathcal{F}_\infty = \{ \mathbf{x} \in \mathcal{R}_i \mid \nexists \mathbf{x}' \in \mathcal{R}_i, \mathbf{x}' \prec \mathbf{x} \}$$

2. EVOLUTIONARY OPTIMIZATION OF WIRELES SENSOR NETWORKS

2.4.2 Crowding distance sorting

Individuals in the same rank are then sorted by their crowding distance. First crowding distance per objective of each individual is calculated. Aggregated crowding distance of the individual is then the sum through all the objectives.

Let $\mathbb{P}_{>}$ be a nondominated set ordered by objective m , and $\mathbb{P}_{>}[i]$ the value of objective m of individual i , then we can recursively define individuals crowding distance C_i :

$$C_i = \sum_m C_{i_m}$$

, where

$$C_{i_m} = \frac{\mathbb{P}_{>}[i+1] - \mathbb{P}_{>}[i-1]}{f_m^{max} - f_m^{min}}$$

It has to be noted, that for aggregation to work, each C_{i_m} is normalized to be a fraction between 0 and 1.

We can see that first sorting helps with general convergence to real Pareto front, while second improves diversity in the nondominated part of the population. To get the result we just need to extract nondominated results from the final population.

2.5 Spea2

In Strength Pareto Evolutionary Algorithm[6], unlike previous NSGAii, stores non-dominated individuals in archive separate from the rest of the population. Therefore after new population is generated, new contents of archive are determined in two passes, first using procedure to ensure convergence and then procedure to ensure diversity.

Let us denote:

P_t to be the population in generation t , \overline{P}_t the corresponding archive of nondominated individuals, N the population size, \overline{N} the archive size and T the maximum number of generations.

The overall algorithm then is as follows:

Step 1: Initialization: Generate an initial population P_0 and create the empty archive \overline{P}_0 , set $t = 0$

2. EVOLUTIONARY OPTIMIZATION OF WIRELES SENSOR NETWORKS

Step 2: Fitness assigment: calculate the fitness values of individuals in $P_t \cup \overline{P}_t$

Step 3: Enviromental selection: Copy all nondominated individuals in $P_t \cup \overline{P}_t$ to \overline{P}_{t+1} . If size of \overline{P}_{t+1} exceeds \overline{N} , then remove exceeding individuals with worst fitness, otherwise, if size of \overline{P}_{t+1} is less than \overline{N} , fill \overline{P}_{t+1} with dominated individuals in with the best fitness $P_t \cup \overline{P}_t$

Step 3: Termination: if $t \geq T$ or another stopping criterion is satisfied, then stop and return the result as the nondominated individuals from \overline{P}_{t+1}

Step 5: Mating selection: generate the mating pool M

Step 6: Variation: apply crossover and mutation operators to the mating pool and set P_{t+1} to the resulting population. Increment generation counter and go to Step 1.

The final result then contains just the nondominated individuals of the archive.

2.5.1 Fitness assigment

Fitness of an individual is an agregated metric composed of strength of individual and density estimation. Strength of an individual i is the number of individuals from $P_t \cup \overline{P}_t$ it dominates:

$$S(i) = |\{j | j \in P_t \cup \overline{P}_t \wedge i \prec j\}|$$

Raw fitness of i is then agregated from all strengths of individuals that dominate i :

$$R(i) = \sum_{j \in P_t \cup \overline{P}_t, j \prec i} S(j)$$

To distinguish between individuals that do not dominate each other a density metric based on the distance of k -th nearest neighbour of i , denoted δ_i^k . As a common setting, k equal to the square root of the sample size is used. Thus density is defined by:

2. EVOLUTIONARY OPTIMIZATION OF WIRELES SENSOR NETWORKS

$$D(i) = \frac{1}{\delta_i^k + 2}$$

Because of the two added in the denominator $0 < D(i) < 1$. Aggregate fitness of an individual is then the sum of its raw fitness and density:

$$F(i) = R(i) + D(i)$$

Because the raw fitness has integer values, sorting individuals by the fitness metric yields similar results to the two pass sorting of NSGAii.

2.5.2 Enviromental selection

Enviromental selecton step depends on the number of nondominated individuals in the union. This is the set of individuals with fitness lower than one:

$$P'_{t+1} = \{i | i \in P_t \cup \overline{P}_t \wedge F(i) < 1\}$$

If $|P'_{t+1}| < \overline{N}$ it is sufficient to set $\overline{P}_t + 1$ to contain the best \overline{N} individuals from $P_t \cup \overline{P}_t$ based on the fitness.

If $|P'_{t+1}| > \overline{N}$ a trimming procedure is used that is based on the density in the P'_{t+1} set, as opposed to $P_t \cup \overline{P}_t$ used for fitness metric.

Iteratively, individual that has minimum distance to another individual is chosen to be removed. If there are several individuals with minimum distance, tie is broken by considering their second smallest distances and so forth.

2.5.3 Spea2+

Spea2+ a modification of Spea2, that lays additional heuristic on top of it. The main changes are:

1. there are two archives, one maintaining diversity in solution space, the other one maintaining diversity in decision space
2. mating selection for cross-over is done based on individuals neighbouring each other in solution space

2. EVOLUTIONARY OPTIMIZATION OF WIRELES SENSOR NETWORKS

3. instead of binary tournament to select the mating population, it uses the whole archive of nondominated solutions

It seems, that in some problem domains Spea2+ outperforms both NS-GAII and Spea2, unfortunately, paradiseo doesn't provide direct implementation in its current version.

2.6 Ibea

Both NsgaII and Spea2 differ mostly in their approach to characterize the convergence to Pareto front and diversity. Indicator-Based evolutionary algorithm takes more abstract approach, based on a concept of binary quality indicators. Indicator is a function, that takes two pareto sets of the same domain and outputs some quantification of a difference between the two.

Zitzler and Kunzli proposed two indicators:

2.6.1 ϵ -indicator

Additive ϵ -indicator that quantifies the minimal distance first pareto set has to be moved in each dimension in objective space such that the second pareto set is weakly dominated. Formal definition:

$$I_{\epsilon+}(A, B) = \min_{\epsilon} \{ \epsilon | \forall x_1 \in A, \forall x_2 \in B, \forall m \in M : P_m[x_1] + \epsilon < P_m[x_2] \}$$

If A dominates B , resulting indicator will be negative.

2.6.2 Hypervolume based indicator

Hypervolume-distance indicator, that quantifies how much volume is dominated by the first pareto set but not dominated by the second, with respect to predefined reference point Z . Hypervolume $H(A)$ of a pareto approximation A is the total volume of n -dimensional space that is "contained" between the individual results and the reference point Z . That is, hypervolume of a set is the total volume of space dominated by the sets individuals.

Hypervolume indicator then compares two pareto approximations:

2. EVOLUTIONARY OPTIMIZATION OF WIRELES SENSOR NETWORKS

$$I_{HD}(A, B) = \begin{cases} H(B) - H(A) & \text{if } \forall x_1 \in A, \forall x_2 \in B, x_2 \prec x_1 \\ H(B \cup A) - H(A) & \text{else} \end{cases}$$

Hypervolume is considered to be the best single value indicator of how good a pareto aproximation is. (c FUNKY HYPERVOLUME THESIIS)

Both of these indicators are dominance preserving.

Definition 7. A binary quality indicator is denoted as dominance preserving if $\forall x_1, x_2, x_3 \in Z$:

$$(i) \ x_1 \prec x_2 \Rightarrow I(x_1, x_2) < I(x_2, x_1), \text{ and}$$

$$(ii) \ x_1 \prec x_2 \Rightarrow I(x_3, x_1) > I(x_3, x_2)$$

Because any single individual can be considered a pareto set of size one, IBEA uses given indicator to quantify fitness of an individual:

$$F(x_1) = \sum_{x_2 \in P - x_1} a(I(x_1, x_2), k)$$

, where $a(i, k) = -e^{-\frac{i}{k}}$ is function that amplifies fitness of dominating individuals.

Algorithm then goes as follows:

Step 1: Initialization: Generate an initial population P_0 , set $t = 0$

Step 2: Fitness assigment: calculate the fitness values of individuals in P_t based on the chosen indicator

Step 3: Enviromental selection: iterate the folowing steps untill $|P_t| < N$:

1. find $x \in P_t$ that has minimal fitness and remove it from P_t
2. recalculate the fitness of the remaining population,

Step 3: Termination: if $t \geq T$ or another stopping criterion is satisfied, then stop and return the result as the nondominated individuals from P_{t+1}

2. EVOLUTIONARY OPTIMIZATION OF WIRELES SENSOR NETWORKS

Step 5: Mating selection: generate the temporary mating pool Ma with binary tournament selection on P

Step 6: Variation: apply crossover and mutation operators to the mating pool and set P_{t+1} to the resulting population. Increment generation counter and go to Step 2.

Paradiseo implementation of IBEA uses adaptive scaling of the amplification function parameter, therefore calibration of the k is not as important. On the other hand, chosing appropriate indicator is crucial.

2.7 Choosing the algorithm

Design of both of these algorithms give some guarantees on convergence to Pareto front. Both of them are incremental improvements on their predecessor.

Right now, NSGAII is considered to be the standard benchmark for MOEA algorithms, both Spea2, Spea2+ and IBEA are comparing . It is the oldest of the three and therefore the most widely used one. Because there are no algorithm specific variables, it has the easiest calibration.

While Spea2 is newer, it doesn't mean it is better in every instance. It has to be noted, that even comparison between these two algorithms with regards to optimizing IDS for WSN came out inconclusive.[4] On the other hand, several papers claim they achieved better result with spea2 than NSGAii. Similiary to NSGAii it doesn't have any algorithm specific configuration.

Of the more interesting concepts that IBEA uses is the hyper-volume distance indicator, which might provide an interesting alternative to heuristics used in more popular NSGA and SPEA. Unfortunately experimental results in (citacion) show, that I_{HD} based IBEA is very sensitive to miscalibration of reference point Z . If provided with good reference point, it consistently outperformed the other tho algorithms, but optimal value of Z varied greatly based on the problem set. We will talk more about hyper-volume as a metric of pareto set convergence in chapter on calibration.

3 Our Framework

Using multi-criteria evolutionary algorithms is a viable and efficient way of optimizing Wireless Sensor networks[4]. Therefore we have decided to create a micro-framework, that would simplify the setup of evolutionary optimization for particular problem. Our combines the paradiseo framework and system for distributed computation boinc with python as a glue language. Emphasis is on ease of configuration, ability to evaluate population in parallel and having good facilities for result analysis.

3.1 Paradiseo MOEO

ParadisEO is a highly configurable framework for metaheuristics written in C++.[3]

Out of the seven algorithms provided in Paradiseo, we include three of them: NSGAII, SPEA2 and IBEA. We do not include the predecessors, NSGA and SPEA, because in general, they have worse performance. SEEA, or simple elitist evolutionary algorithm is best suited for problems with inexpensive evaluation functions, where the most time consuming process is the selection, but evaluating WSN configuration is seldom cheap. Should for any reason be an algorithm not included in our micro-framework be better suited than the ones that are, it is easy to add one. Main modifications required are to enable parallel evaluation for the algorithm.

Even if the algorithm we'd wanted to use wasn't present in Paradiseo framework, due to its whitebox nature and modular design and large library of integrated metrics it could be easily added.

3.1.1 Object-oriented design of MOEA

Every multiobjective evolutionary algorithm in Paradiseo extends the `moeoEA`. The implementation of such algorithm is best explained on a `moeoNSGAII` class, probably the most popular multi-objective algorithm. Because `moeoEA` implements the unary functor interface, to define new algorithm it is sufficient to define the `operator()`.

Listing 3.1: NSGAII algorithm in Paradiseo

```

virtual void operator () (eoPop <MOEOT> &_pop){
    eoPop<MOEOT> offspring, empty_pop;
    popEval(empty_pop, _pop); // a first eval of _pop
    // evaluate fitness and diversity
    fitnessAssignment(_pop);
    diversityAssignment(_pop);
    do{
        //generate offspring, worths are recalculated if
        necessary
        breed (_pop, offspring);
        //eval of offspring
        popEval (_pop, offspring);
        //after replace, the new pop is in _pop. Worths
        are recalculated if necessary
        replace (_pop, offspring);
    }while (continuator (_pop));
}

```

As you can see, the algorithm itself is a template defined by implementation of the fitnessAssignment,diversityAssignment,breed,popEval,replace and continuator functors and eoPop population object. Therefore most of the implementation itself is done in defining various classes that implement. We can showcase how this composing works on an example of a continuator:

Listing 3.2: Object composition in Paradiseo

```

/** a continuator based on the number of generations*/
eoGenContinue < MOEOT > defaultGenContinuator;
/** stopping criteria */
eoContinue < MOEOT > & continuator;

moeoNSGAI1 (unsigned int _maxGen, ...):
    defaultGenContinuator(_maxGen), continuator(
        defaultGenContinuator)

```

If we use constructor that ends evaluation after certain number of generation, it first initializes defaultGenContinuator with number of generations, and then passes it as a continuator to use.

3.1.2 Population object

Population in Paradiseo is in its core an ordered vector of individuals. Because our individuals are represented by vectors of real numbers, we

will describe the `moeoRealVector` implementation. Each individual has its value, and contains variables for fitness and diversity metrics. It should be noted, that `moeoRealVector` doesn't contain any bounds on specific parameters of individuals configuration. In `paradiseo`, we define boundaries externally, with specification in initialization function and crossover and mutation functions.

3.1.3 Fitness and diversity assignment

Class `moeoDominanceDepthFitnessAssignment` is used to sort the population into nondominated ranks, by updating the fitness variable of each individual. Class `moeoFrontByFrontCrowdingDiversityAssignment` is similarly used to set the diversity variable. The replace operator uses these to evaluate the children of the old population and creating the new generation.

3.1.4 Creation of new generation

Breed operator is a standard `eoBreed` class that gets supplied with transform-object that aggregates mutation and cross-over functions. Similarly, the `popEval` function just evaluates all the new individuals in populations and assigns them objective values. The specifics of the algorithm are hidden in the replace operator of `moeoElitistReplacement` class:

1. recalculate fitness and diversity
2. sort the population with standard `moeoFitnessThenDiversityComparator`
3. remove exceeding population

As you can see, this implementation slightly differs from the sketch of algorithm provided in previous chapter, namely that the NSGAI in paper never exceeded the population size.

3.1.5 Stopping criteria

Continuator operator provides the basic stopping criteria. The default implementation is based on limiting maximum number of generations¹,

1. As hinted by the class-name *defaultGenContinuator*

but in theory, continuator can use population fitness or another metric to decide whether to proceed with another step. This feature is often utilized in evolutionary implementation of decision problem algorithms, because it can help evaluate whether population contains enough information to carry out the decision. In multicriteria optimization problems we usually don't have a definition of "good enough" solution, but if we had, we could use it to further limit the number of expensive evaluations.

We chose python for these reasons:

1. it is an efficient glue language. Most often you will want to optimize already written application with minimal changes. Python is well suited for running external applications, specifying their inputs and parsing their outputs.
2. it has good facilities for statistical analysis. That allows us to include analysis of every optimization run into the executable itself. Having automatically generated experimental log has proven invaluable especially in calibration of algorithms for our specific problem.
3. it is easily integrated with C++. We strive to make the integration with ParadisEO simple and extensible.

Our microframework consists of a single executable and a configuration file in python. Configuration itself consists of defining selected entry points, in particular the number and bounds of input parameters, the number of objectives to optimize, definitions of cross-over and mutation functions and either the definition of evaluation function, or functions for scheduling evaluations in parallel on boinc. We provide reporting function entry-point as well, to automatically collect and analyze results of the algorithm run.

3.2 Entry points

3.2.1 Input parameters

For inputs you specify their bounds, which are then used to initialize the population and to supply constraints for mutation and crossover parameters

3.2.2 Objectives

Because we focus on multiobjective evolution, you need to specify number of objectives. To simplify configuration, minimization of all the objectives is hardcoded.

It is advisable to have three objectives at maximum, because number of solutions on the Pareto optimal dramatically increase with new objectives. At minimum, $(n + 1)$ objective problem will contain all the solutions of an n -objective problem. [5]

Then there is the problem of visualizing and evaluating more than 3-dimensional data.

3.2.3 Evaluation function

To simplify configuration, inputs and outputs of evaluation function are allways an n -tuple of floating point numbers. We believe this is general enough, and that most WSN optimisation problems can be fitted to this constraint. Only problem might pose translation of combinatorial problems to continuous ones, fortunately so far we have been quite successfull with using just a simple rounding techniques. We wanted to avoid usage of binary strings as input vectors.

Many evolutionary algorithms support them and if we were using them we wouldn't need to concern ourselves with implementation details of mutation and crossover functions. On the other hand, with generalized mutation functions on binary strings, there are no guarantees that mutated individual will even be valid. CITATION In our opinion, vectors of reals are esier to reason about, and their convergence may provide valuable insight to structure of the problem.

3.2.4 Boinc Assisted Evolution

Because evolution is population-based heuristic, it is well suited for parallel evaluation of its individuals. ParadisEO already has two methods included, either by use of MPI protocol, or by using SMP on multi-core machine[3]. Unfortunately, SMP module of ParadisEO is not yet stable on all platforms, and MPI has specific demands on infrastructure and is usually hard to retro-fit on already written program.

Boinc on the other hand can simply utilize machines already preasent to form a simple computational grid. With its simple archhitecture con-

sisting of Management server and several worker nodes it can be deployed in most settings. Its simple architecture prohibits cooperation between worker nodes, but that doesn't concern us while evaluating individuals.

We have decided to give the evolution algorithms ability to use a simple scheduler for creating Boinc work units, pluggable from python.

In our micro-framework it consists of three functions, `schedule_work_unit`, `wait_for_completion` and `gather_result`.

Even though the evaluations themselves will be done in parallel, we wanted to avoid any threaded code in our framework itself. Because we have written it with research in mind, we understand that accessing auxiliary data is important to evaluate experiments. In threaded environment, this could lead to deadlocks.

3.2.5 Mutation and Crossover

This directly influences the intensification of heuristic. There are several popular mutation and crossover functions for individuals based on bounded vector of real numbers. These are based on their implementation in Paradiseo. To provide convenience, if mutation function is not specified, `eoUniformMutation` will be used and similarly, if no crossover function is provided, `eoRealUXover` will be used.

Mutation is in essence some function $m : X \rightarrow X$ that on input of individual outputs another individual.

Listing 3.3: Mutation example

```
d = 0.1
def mutate(x):
    return x + random.uniform(-d,d) * (max_bounds-min_bounds)
```

Crossover is some function $c : X \times X \rightarrow X \times X$ that on input of tuple of parent individuals outputs a tuple of children individuals.

Listing 3.4: Crossover example

```
def mutate(x_1,y_1,x_2,y_2):
    return ((x_1,y_2),(x_2,y_1))
```

3.2.6 Reporting

Reporting function takes two arguments, both strings. First string is the char * args string passed to the executable. Second string is the console printout or the Paradiseo result-archive.

This way, we could integrate the creation of experimental log directly into our executable, computing different . This helps especially with calibration of evolution algorithm parameters. We provide examples of different metrics and indicators that help with evaluation of result in next section.

Integrating python with C++ by passing strings is not particularly elegant. We have chosen this particular implementation, because it is easiest to extend and debug.

3.3 Example problem

We have chosen the first Schaffer’s bi-objective problem SCH1, to provide an example of minimal python file.

The goal of the SCH1 problem is to minimize the following two objectives objectives of a function with single parameter x :

$$f(x) = [x^2, (x - 2)^2]$$

Listing 3.5: Minimal working example

```
# -*- coding: utf-8 -*-
import logging
logging.basicConfig(filename=
)

#minimizing everything
def nObjectives():
    return 2

def nParams():
    return 1
def minimumBounds(i):
    return 0.0
def maximumBounds(i):
    return 2.0

def evaluate(x):
```

3. OUR FRAMEWORK

```
    return (x**2, (x-2)**2)
def report(inp, rep):
    logging.info(inp+    +rep)
```

4 Calibration

We can think of calibration on an evolutionary algorithm as a meta-optimization problem with two objectives,

1. maximizing the fitness of result
2. minimizing the number of evaluations

With single-objective algorithms, optimizing for these two objectives is easy, because both of them are usually represented by numbers. With multi-objective problems, results are Pareto sets, which are often not directly comparable.

Quality of a multiobjective result can be judged based on several criteria:

1. number of results in Pareto set.
2. convergence to true Pareto front
3. regularity in distance between results

4.1 Exhaustive search

Because the main reason we use multi-objective algorithms is to avoid the need to do the expensive computation of true Pareto front, we could calibrate it on a easier variant of the problem. Underlying assumption is, that the particular variant of the problem will have similar solution space. We will show such an example in following chapter.

Computing an exhaustive search for a subset of our problem can help in several ways:

- we can check our assumptions about the shape and homogeneity of solution space
- we can use the Pareto front as a reference point
- for any approximation, we can easily compute how many individuals it dominates and how many individuals dominate it

This is probably the most expensive, but on the other hand the most precise way to estimate, how will an evolutionary algorithm behave in a particular problem domain.

4.2 Iteratively

For iterative improvements we have used different convergence criteria. To measure convergence of Pareto approximations to the Pareto optimal set, we can either specify a global reference point, with respect to which all approximations are measured, or we specify a comparison function that can provide ordering. Indicator functions from IBEA algorithm are particularly suited for this purpose.

4.3 Measuring convergence

4.3.1 Hypervolume indicator

One of the more popular is maximizing the hypervolume of the objective space dominated by the resulting approximation. Hypervolume indicator is based on idea of measuring the volume that a given individual dominates based on some reference point in the solution-space. This works with assumption that the solution-space is homogenous, compact, and that we can specify a reasonable reference point. A good reference point is a solution that is dominated by all the other individuals. [1]. We usually don't know how will this individual look like, by guessing this reference point we are expressing certain biases (or expectations) on the shape of the solution space as well.

Computation of a hypervolume difference between two individuals as used in IBEA algorithm can be implemented as follows:

Listing 4.1: Hypervolume implementation

```
def hypervolume(o1, o2, R, bounds):
    v1 = o1.pop()
    v2 = o2.pop()
    max = R.pop()
    r = range.pop()
    // computation of the volume
    if o1 == []:
        if v1 < v2
```

```

        return (v2 - v1) / r;
    else:
        return = 0;
if v1 < v2:
    max_o2 = list(o2).pop().put(R.peak())
    return ( hypervolume(o1, o2,R,range) * (v2 - v1
        ) / r )
        + ( hypervolume(o1, max_o2, R,range) * (
            max - v2) / r );
else:
    result = hypervolume(_o1, _o2,R,range) * (max -
        v2) / r;

```

Here, the *o1* and *o2* are two lists of reals, representing the compared individuals, while *R* provides the reference point and *bounds* keeps the list of ranges for each objective, to help normalize the resulting value.

4.3.2 Distance indicator

Distance indicator $I_{\epsilon^+} : Z \times Z \rightarrow \mathbb{R}$ is one of possible choices for indicator that IBEA algorithm uses for comparing two pareto-approximations. I_{ϵ^+} gives the minimum distance, by which Pareto set approximation needs to or can be translated in each dimension in objective space, such that the other approximation is weakly dominated. If we know, that all objectives will have minimal value of 0, implementation of this metric may look like this:

Listing 4.2: Distance indicator implementation

```

def distance(o1, o2,maximum):
    return max([(v2-v1)/max for (v1,v2,max)
        in zip(o1,o2,maximum)])

```

4.3.3 Average distance indicator

Given pareto approximation *A* and *B* it calculates the average of distances from all $a \in A$ to the nearest solution $b \in B$. This indicator is not dominance preserving, because euclidean distances will always be positive. There have been achieved good results in measuring convergence when using this indicator to compare pareto approximation to pareto front[4].

4.4 Measuring diversity

Diversity metrics help us avoid pareto approximations, that have too similar solutions. We used metric based on the the spacing metric taken from Spea2, which we have discussed in greater detail in previous chapter. For the sake of simplicity, we simply average the distance agregate betwee the closest neighbours. We use the term distance-agregate, because Spea2 calculates its metric as a sum of normalized differences between every objective.

Listing 4.3: Spacing indicator implementation

```
def distance_agregate(indi, nb, range):
    sum([(i-n)/r for (i,n,r) in zip(indi, nb, range)])

def spacing(approximation, range):
    dist_indi = []
    for indi in approximation:
        neighbour_dist = [distance_agregate(indi, nb, range)
                           for nb in approximation]
        dist_indi.append(min(neighbour_dist))
    return sum(dist_indi)/len(dist_indi)
```

Similarly we can utilize the individuals crowding distance from NSGAI. We calculate the value of each individual with regards to the rest of pareto approximation and then create the compound metric by averaging the distances.

Listing 4.4: Spacing indicator implementation

```
def (approximation, range):
    split = zip(*approximation)
    agregate = 0
    for (objective, r) in zip(split, range):
        agregate += sum([(prev-next)/r
                          for (prev, next) in zip(
                              [prev for prev in objective[:-2]],
                              [nxt for nxt in objective[2:]]))])
    return agregate/len(split)
```

4.5 Population size and Number of generations

These two parameters set the boundaries on minimal and maximal number of evaluations. Lets mark the size of population N and number of generations G Because the whole initial population needs to be evaluated, number of evaluations will at least the size of populations. Because at each new generation, at most N new individuals are generated, therefore in one optimization run, at most $N \times G$ evaluations.

4.6 Mutation and Crossover

Probability of applying mutation, or crossover directly influence how many new individual evaluations are there to be computed. By calibrating this parameter we aim to find the optimal balace between diversification and computational cost of evaluating one generation. More importantly, the implementation of mutation and crossover specify how are these new individuals generated.

With standard mutation operator, the value of δ parameter can have significant impact on the speed of convergence, because it specifies the bounds of neighbourhood where the individual can mutate. Experimenting with mutation operator implementation can be useful if we hve some aditional domain knowledge that we can use to choose the mutated individual beyond simple uniform randomization on its parameters.

5 Experiments

To test our framework we have decided to optimize a simple intrusion detection system on a model wireless sensor network based on our laboratory WSN testing setup.

5.1 Wireless sensor network

Implementation of WSN was reused from [4], with just a slight modifications to decouple it from previous optimisation framework. It uses MiXiM simulator that provides complex communication models, including capabilities for precise simulation of interference on physical layer and various energy consumption models. Simulation this precise is often costly enough to varant the usage of evolutionary heuristics.

Test application running on each of the nodes periodically sends a packet containing arbitrary information through the network to a base station. We have marked some of the nodes as droppers, so that instead of forwarding all of the packets through the network, they drop certain percentage. In our case, dropping ratio was set to 0.5. Network has a static routing tree, as described in Fig.5.1. The routing was set up so that individual nodes can eavesdrop their neighbours and that there will be some benign nodes that drop packets unintentionally.

5.2 Intrusion Detetection System

Our nodes implement a simple IDS capable of detecting nodes that seem to be intentionally dropping packets. If we look at the wireless network stack of our node, IDS is part of the medium access controll sublayer. There it can eavesdrop on neighbouring nodes and check whether they behave accordingly. For each neighbour it updates the number of packets recieved and the number of packets forwarded.

We wanted to use IDS that is simple, but highly configurable. We have these four parameters to optimize:

1. number of monitored nodes p_n
2. size of buffer for eavesdropped packets p_b

3. threshold for minimal number of recieved packets for a node to be considered malicious p_m
4. threshold for ratio between forwarded and recieved packets for a node to be considered to be malicious p_t

At the end of simulation, based on a preset treshhold and ratio between forwarded and reieved packed it decides whether node is malicious or benign. More specificaly, the set of malicious neighbours of a node i is:

$$M_i = \{n \in N_i | r_{in} \geq p_m \wedge \frac{f_{in}}{r_{in}} \geq p_t\}$$

And the set of benign nodes is:

$$B_i = \{n \in N | r_{in} < p_m \vee \frac{f_{in}}{r_{in}} < p_t\}$$

You can see, that both sets are disjoint, and their union is the set of neighbours N_i .

We want to minimize these three objective functions:

False negative ratio that calculates for each malicous node the percentage of its neighbours that claimed it is benign and returns the average.

$$fn(x) = \frac{1}{|M|} * \sum_{m_k \in M} \frac{|\{i | m_k \in B_i\}|}{|\{i | m_k \in N_i\}|}$$

False positive ratio that calculates for each benign node the percentage of its neighbours that claimed it is malicious and returns the average.

$$fn(x) = \frac{1}{|B|} * \sum_{b_k \in B} \frac{|\{i | b_k \in M_i\}|}{|\{i | m_i \in N_i\}|}$$

Memory footprint that agregates the impact of number of monitored nodes and the size of buffer on the memory consumption.

$$m(x) = 8 * p_n + 16 * p_b$$

5. EXPERIMENTS

	Name	Description	Range	Step
Parameters	p_n	Max monitored nodes	$\langle 1, 28 \rangle$	1
	p_m	Min received packets	$\langle 1, 100 \rangle$	1
	p_t	Detection threshold	$\langle 0, 1 \rangle$	0.01
	p_b	Buffer size	$\langle 0, 50 \rangle$	1
Scenario	Δ_t	Period for sending packets ¹	$\{0.1, 0.5, 1, 5, 10\}$	

Table 5.1: List of parameters to be optimized for each IDS scenario

5.3 Setup

We want to use the framework to optimize the four parameters of our IDS as described in 5.1. Because this experiment is aimed primarily on showcasing the functionality of our framework, we have created several optimization scenarios based on the rate of network traffic.

First we have calculated two sample exhaustive searches, with period Δ_t in which each of the nodes sends a packet to be rooted to base station set to 10 seconds and 5 seconds. Because simulation of a network with low network traffic is not as source intensive, it allowed us to gauge the problem space and run calibration on it.

5.4 Calibration

We want to compare the performance of all four algorithms, NSGAI, Spea2, MOGA and IBEA. All the algorithms will share mutation and cross-over functions:

5.4.1 Mutation function

We have used the default mutation function. It has two parameters, probability that individual will be mutated $pMut$ and relative δ neighbourhood around the old value, that the new value will come from.

5.4.2 Crossover function

We have used the default cross-over function as well, with parameters $pCross$ (probability that two individuals will be crossed over)

and *crossProb* (probability that params of the two individuals will be swapped)

For algorithms NSGAii, Spea2 and MOGA, calibration was done on fixed number of generations and population size, with optimizing for values *popSize*, *pMut*, δ , and *pCross*:

Population size to be set to following values: *popSize* $\in \{50, 75, 100\}$

Mutation will be done with probability of *pCross* $\in \{0.01, 0.1, 0.25, 0.5\}$, with its mutation parameter $\delta \in \{0.05, 0.1, 0.2\}$

Crossover will have probability set *pCross* $\in \{0.01, 0.1, 0.25, 0.5\}$, with *crossProb* = 0.5

Number of generations is be set to *NGen* = 200

For IBEA we have chosen the I_{HD} indicator, that we use as a comparison metric for other algorithms as well. Because IBEA with I_{HD} requires calibration of reference point Z , we haven't used crossover operator with IBEA at all (*pCross* = 0). We know that a good reference point is dominated by all solutions, therefore we will use $Z = (f_p, f_n, m)$, $f_p, f_n \in \{0.8, 1.1, 1.5\}$, $m \in \{1500, 2000, 2200\}$, to see the different behavior, when the reference can dominate some of the individuals, when it is tightly dominated, and where it is worse by a large margin.

Because we have computed the exhaustive search for both of the scenarios, we know the pareto front \mathcal{PF} of each their solutions. We have used these metrics, that we have already discussed in previous chapter:

ϵ_+ **indicator:** computes the smallest distance pareto-approximation has to be translated to dominate the pareto front.

HD indicator computes the difference in hypervolumes between pareto approximation and pareto front.

Average distance from PF computes the average euclidean distances from between every individual of pareto approximation and its closest neighbour from the pareto front

Diversification computes the diversification metric of NSGAI algorithm

Spacing computes the spacing metric of Spea2 for the closest neighbour

Number of evaluations will be used to evaluate cost of running the optimization

This gives us enough data not only to decide the best algorithm for the task at hand, but to further discuss relevance of different convergence metrics.

5.5 Evolution with aid of BOINC

After comparison of problem spaces, we have chosen the best settings and algorithm. We then apply it to the three of the more resource intensive scenarios, with $\Delta_t \in \{0.1, 0.5, 1\}$ seconds. With the help of Boinc-assisted evolution we could assess whether each MOEA evolution algorithm can provide reasonable and consistent optimisation of IDS on WSN.

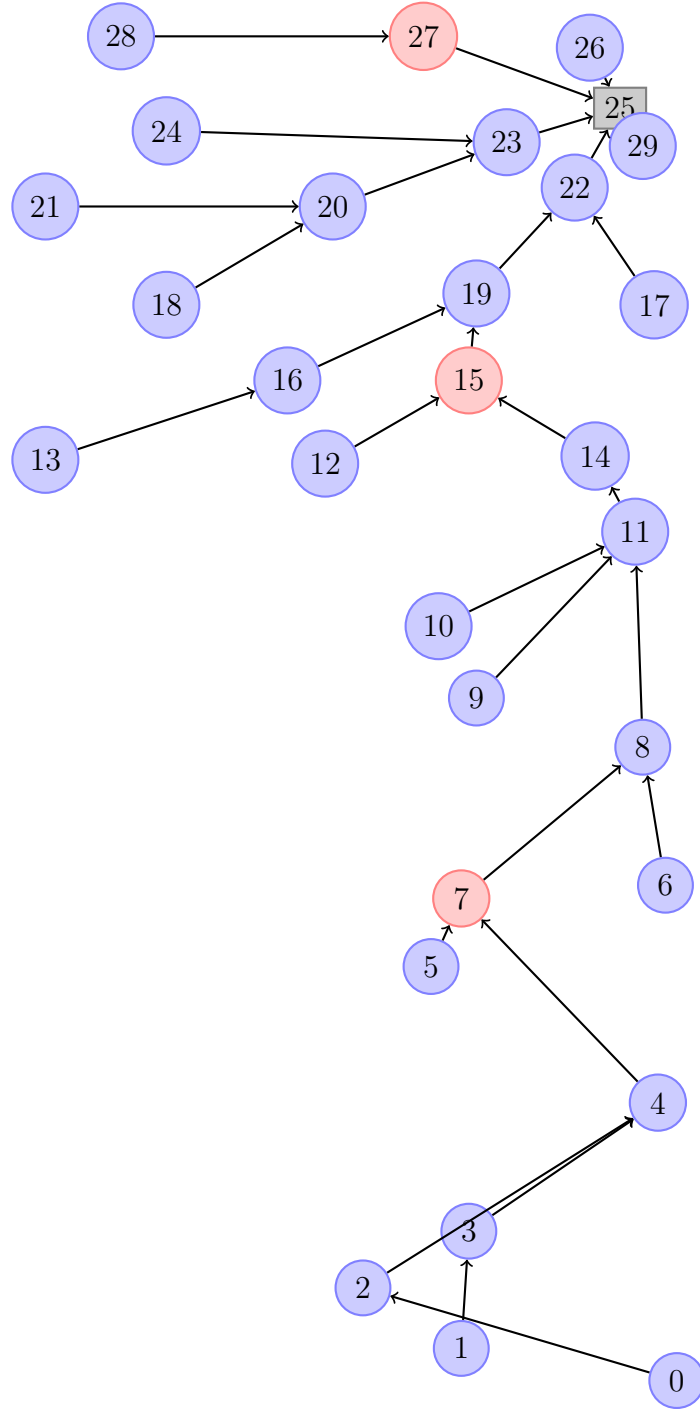


Figure 5.1: Topology of the network used in experiments, with grey base station 25 and red droppers 7, 15 and 27

6 Results

Bibliography

- [1] Anne Auger, Johannes Bader, Dimo Brockhoff, and Eckart Zitzler. Theory of the hypervolume indicator: optimal μ -distributions and the choice of the reference point. In *Proceedings of the tenth ACM SIGEVO workshop on Foundations of genetic algorithms*, pages 87–102. ACM, 2009.
- [2] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *Evolutionary Computation, IEEE Transactions on*, 6(2):182–197, 2002.
- [3] Arnaud Liefooghe, Matthieu Basseur, Laetitia Jourdan, and El-Ghazali Talbi. Paradiseo-moeo: A framework for evolutionary multi-objective optimization. In *Evolutionary multi-criterion optimization*, pages 386–400. Springer, 2007.
- [4] Martin Stehlik, Adam Saleh, Andriy Stetsko, and Vashek Matyash. Multi-objective optimization of intrusion detection systems for wireless sensor networks, 2013.
- [5] El-Ghazali Talbi. *Metaheuristics: from design to implementation*, volume 74. Wiley, 2009.
- [6] Eckart Zitzler, Marco Laumanns, Lothar Thiele, Eckart Zitzler, Eckart Zitzler, Lothar Thiele, and Lothar Thiele. Spea2: Improving the strength pareto evolutionary algorithm, 2001.