An Uncertainty Quantification Algorithm for Performance Evaluation in Wireless Sensor Network Applications

Dong YU

University of Technology, Sydney, Australia
10743387@student.uts.edu.au

Abstract. The potential applications of Wireless Sensor Networks (WSNs) span a very wide range. One of the application domains is in healthcare industry. The diverse WSN application requirements signify the need of application specific methodology for system design and performance evaluation. Moreover the performance of typical wireless network is stochastic in nature, Probability is an essential instrument needed to assess the performance characteristics. Before WSNs widespread involvement in life or death critical applications, an urgent need is a generic systematic evaluation methodology for decision makers to evaluate performance among alternative solutions taking into account the cohesion characteristic. This paper offers a quantative decision making procedure to incorporate performance deviation as a target performance metric. Decision making is guided by goals and objectives for the particular application specified by application domain experts.

Keywords: Uncertainty Quantification, WSNs, Multi Criteria, Statistical Performance, Generic Methodology, Fair comparison

1 Introduction

We have witnessed in recent years the emergence of WSNs in healthcare. These applications aim to improve and expand the quality of care across wide variety settings and for different segments in the healthcare system. They range from real-time patient monitoring in hospitals, emergency care in large disasters through automatic electronic triage, improving the life quality of the elderly through smart environments, to large-scale field studies of human behavior and chronic diseases [1]. However, the barrier for wide spread adoption of the technology is still high. Fulfilling the potential of WSNs in healthcare requires addressing a multitude of technical challenges. These challenges reach beyond the resource limitations that all WSNs face in terms of limited network capacity and energy budget, processing and memory constraints. Particularly, healthcare applications impose stringent and diverse requirements on system response time, reliability, quality of service, and security.

The uniqueness of WSNs in its resources limitation, transient channel state and drastic different application requirements, brings in application specific system design methodology. From WSNs research kickoff at early stage, research efforts are mostly focused on the isolated programming issue of single layer protocols with little concern of other layer functionality. This leads to protocols that exist in a vacuum which perform well on theoretical basis, but have problems when deployed under real-life circumstances [2, 3]. Many of these protocols are further validated using ad-hoc experimental tests to the benefit of one specific solution, leaving little room for objective comparison with other protocols. Up to now, there exists no fixed set of accepted testing methods, scenarios, parameters or metrics to be applied to guarantee fair comparison between competing solutions. This lack of standardization significantly increases the difficulty for a developer to assess the relative performance of their protocols compared to the current state of the art**.**

Essentially, multiple-criteria evaluation is a well studied realm. There exist plenty of methodologies in multi criteria decision making domain [12]. But we can not apply these techniques directly to WSNs evaluation without considering the uniqueness of the target domain. In this paper, we try to apply analytic hierarchy process (AHP) to WSNs performance evaluation. It is a method to evaluate system performance according to application scenario and application expert preference, it is generic enough to provide a platform to fairly compare alternative solutions. Most importantly, we introduce statistical metrics to reflect uncertainty attribute impact on final performance.

The rest of the paper organized as such: Section 2 establishes a background understanding of performance uncertainty in WSNs that serves as the foundation for building our proposed solutions. The section presents uncertainty attribute of WSN performance, statistical concepts in performance evaluation. Section 3 introduces application specific evaluation based on AHP method. We emphasize that while application specific design is necessary to build efficient application based on limited energy budget, for a fair comparison of alternative design solutions, a generic evaluation algorithm is needed to deal with all of the components aforementioned. Section 4 provides the practical algorithm for uncertainty performance evaluation. Workflow and algorithm are given to get a single QoS performance index. We summarize our work in section 5.

2 Uncertainty Performance in WSNs

2.1 Source of Uncertainty of WSN performance

Several factors contribute to the fact that wireless sensor networks often do not work as expected when deployed in a real-world setting. Firstly, there is possibility of wrong expectation from system designer side: analytical model does not fit into the problem in hands. That is often a problem for inexperienced designers. For all simulation or other experiments methods, first step is to eliminate the possibility of this kind of profound design problem in preliminary stage. Secondly, there is possible wrong expectation from simulation results: Simulation modeling can not faithfully reflect the System Under Test (SUT).

Except the designer's preliminary problem of analytical model mismatching design target, we can further identify fault point of performance disagreement between expectation and real world implementation into components of WSNs hierarchy [11].

1. Environmental influences which may lead to non-deterministic behavior of radio transmission.
2. Node level problem: Malfunction of the sensors, or even the complete failure of a sensor node due to cheap manufacturing cost. Scarce resources and missing protection mechanisms on the sensor nodes may lead to program errors: operating system reliability and fault-tolerance.
3. Network level problem: Network protocols especially (MAC and Routing) are not robust to link failure, contention, topology dynamics.
4. Unforeseen traffic load pattern: A common cause for network problems is an unforeseen high traffic load. Such a traffic burst may occur for example, when a sensor network observes a physical phenomenon, and all nodes in the vicinity will try to report at once, causing the occurrence of packet collisions combined with a high packet loss.

All these factors contribute to the uncertainty of the sensor network behavior and function. These elements increase the probability of network functionality deviation from its normal operation and affects its' collected data accuracy.

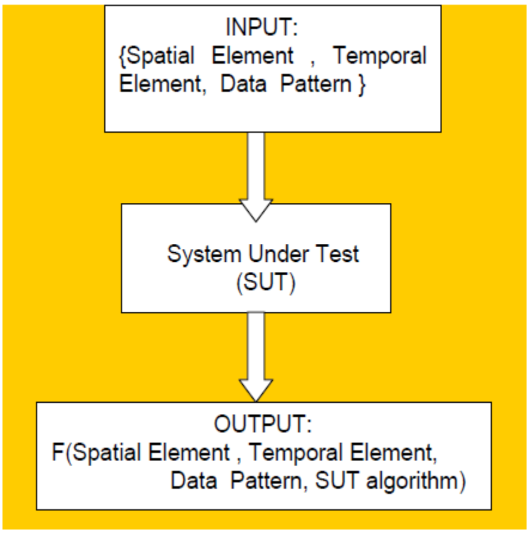


Fig. 1. Uncertainty modeling in WSN

In order to effectively develop parameters, we in [11] congregate hierarchical possible points of deviation into four groups.

1. Spatial elements uncertainty: include site specific characteristic: fading, signal attenuation, interferences, and network scale: topology, network size and density.
2. Temporal elements uncertainty: even on one particular spot, link state flips with time.
3. Data communication pattern uncertainty: include load burst pattern uncertainty (The volume, frequency of the data burst), communication interval difference (how often is the data communication happening, how long is the interval between two adjacent communications), and different communication modes (inquiry triggered, regular report, or event triggered communication).
4. Algorithm internal programming uncertainty: include malfunctioned models, assumption realization problem and other normal cooperation problem in programming,

We summarize above observation into Equation (1).

$$P = F\text{ (spatial, temporal, traffic load pattern, SUT)} \qquad (1)$$

The parameters in equation (1) as show in Figure 1 represent system level performance dynamics involving all four factors. As the input elements display statistical behaviors, output performance definitely will have a statistical distribution pattern with certain norm and deviations for specific scenarios. Since wireless performance is inherently statistical in nature, accurate performance testing must account for these random components [5]. More over, comparing performance curves produced by a number of metrics makes it difficult to evaluate how well a given protocol suits for the purpose of an application. It may also be difficult to estimate, which of the protocols at hand would perform the best with respect to that application [6].

2.2 Characterizing WSNs Uncertainty Performance with Statistical Concepts.

WSNs sense the targeted phenomenon, collect data and make decision to store, aggregate or send the data according to distributed local algorithm. The modulated electromagnetic waves propagate in free-space; interact with the environment through physical phenomenon such as reflection, refraction, fast fading, slow fading, attenuation and human activities. Even with the best wireless protocols, the best chipsets, the best RF design, the best software, wireless performance is going to vary. Wireless performance is inherently statistical in nature, and accurate performance evaluation must reflect this nature.

We observed that currently most ad hoc evaluations in wireless network field, especially in WSNs research, no matter in the form of test bed experiment or simulation, only focus on mean value of performance metrics, and do not pay much attention on performance deviation. For some applications, average performance is sufficient for data gathering and collective data analysis. However, average ‘throughput’, ‘lifetime’, ‘reliability’ or ‘delay response’ are not sufficient enough to predict performance on certain application scenarios. Any dip in performance, no matter how short, can result in dropped packets that cause visual artifacts or pixilation of the image of wireless video monitoring application. In extreme cases like in healthcare emergency monitoring application, any dropped packet may cause life or death difference. Consequently, the viewer/user experience is completely dependent on the wireless system "worst-case" performance [4].

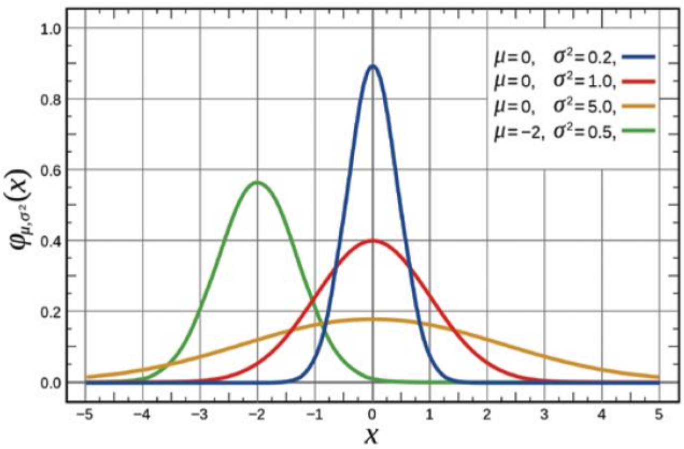


Fig. 2. An example PDF Bell curve graph depicting different average and standard deviation parameters from [5]

Figure 2 represents Probability Density Function" (PDF) of a sample performance metric with four different Systems Under Test (SUT), each with different average and standard deviation (variability) parameters. The graph illustrates „normal" probability distribution revealing statistical characteristics of metric X, representing at least approximately, any variable that tends to cluster around the mean as norm , shown as, but not necessary, the familiar „bell curve". It shows the relative probabilities of getting different values. It answers the question:

- What is the chance I will see a certain result?
- What is the mean value or norm of the respective SUT at this specific performance metric?
- How cohesive and stable is the performance for each SUT?

Examining the random process represented by the red curve in figure 3, we would expect outcomes with a value around „0" to be twice as prevalent as outcomes of around 1.25 (40% versus 20%). However, in some cases, we are more interested in a threshold performance value as benchmark value than individual probability point, what is the probability of having performance being less or greater than a threshold value? A transformed PDF ,Cumulative Distribution Function (CDF) (Figure 3.) helps answer this question.

When you use probability to express your uncertainty, the deterministic side has a probability of one (or zero), while the other end has a flat (all equally probable) probability. For example, if you are certain of the occurrence (or non-occurrence) of an event, you use the probability of one (or zero). If you are uncertain, and would use the expression “I really don't know,” the event may or may not occur with a probability of 50 percent. This is the Bayesian notion that probability assessment is always subjective. That is, the probability always depends upon how much the decision maker knows. Due to statistics science the quality of information and variation are inversely related. That is, larger variation in data implies lower quality data (i.e., information) [7].

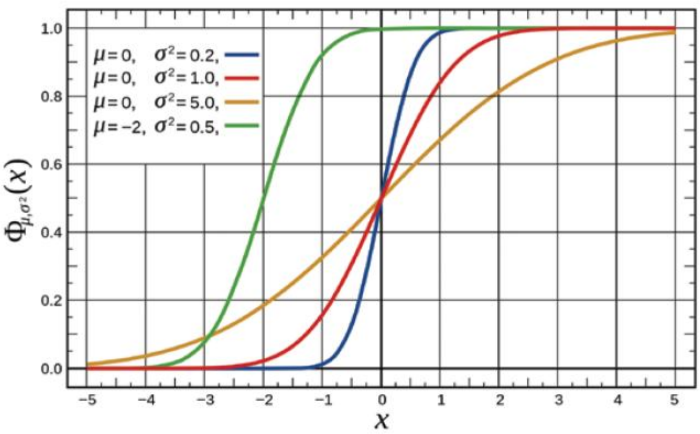


Fig. 3. A typical cumulative distribution function (CDF) graph from [5]

Examining the random performance metric represented by the red curve in figure 3,

- The probability of producing a result less than -0.75 is about 20%.
- The probability of producing a result less than 0 is 50% and
- The probability of producing a result less than 2 is about 95%.

To characterize wireless performance, CDF graphs can provide immensely useful information for system designers to compare performance of alternative solutions. Ideally, we prefer better mean value and smaller deviation, but if the ideal choice is unavailable, we have different optional solution to choose. All the choices should be put in application specific context to choose the right protocol for right application. The principle is:

- Drastic different mean value, if mean value represent positive metrics, like throughput or lifetime, bigger is better, ideally we prefer SUT with bigger mean value and less deviation.
- Same mean value, different deviation: long tail means performance not stable, we prefer smaller deviation.
- Slightly different mean value, but one with long tail, we prefer stable over slightly improved peak performance.
- But if the optimal solution is not available, we have choice over performance stability and higher norm performance according to different application scenario.
  - (option 1) Higher performance potential but less predictable performance
  - (option 2) Less performance potential but higher stable performance

Reference [5] presents practical guidelines on how to actually acquire the statistical performance PDF and CDF curve of a SUT; nevertheless sampling is the key to recover statistical performance and drawing the PDF and CDF curve of a wireless system. Furthermore, to predict real-life performance accurately, researchers ideally should conduct sampling tests across all relevant dimensions and variable if possible. However, in most cases, the design space is too big to exhaustively investigate all factors influencing the final performance. But planners must at least consider three rough dimensions, as we have mentioned above, to characterize wireless performance accurately: time, space, and data pattern. Under each category, there are vast known or unknown parameters that can affect the performance. Hence it is worthwhile to investigate the effect of parameter change to specific performance metric (sensitivity analysis). The effective way to deal with the vast design space is parameter reduction and inter-dependency analysis.

3 Application Specific Evaluation Based on AHP Method

WSNs energy-oriented research originates from conflict between application performance requirements and limited energy supply by battery. In foreseeable future, it will remain as a bottleneck for its widespread development unless a breakthrough at relevant material science field occurs. However, we can not overemphasize energy conservation while ignoring application specific requirements. To what extent we should emphasize the importance of energy aspect comparing other QoS objectives depends on application scenarios. Tradeoffs have to be made on per application basis.

Typically WSN lifetime (energy efficiency), response time (delay), reliability, and throughput are among the main concerns in the design and evaluation process. Under the constraint of wireless sensor node size, the energy budget and computing resources are unfeasible to afford any luxury algorithms. Under such constraint, there does not exist a perfect optimal solution satisfying all performance metrics in the problem (NP hard problem), rather, the question we sought is how to tradeoff multiple criteria explicitly leading to more informed and better decisions. The methodology should be general enough to contain different application scenario according to decision maker's preferences.

There have been few works on application-driven protocol stack evaluation for WSNs. Our evaluation methodology, similar to analytic hierarchy process (AHP) [9, 10], using a Single Performance Index (SPI) for each alternative solution or System Under Test (SUT), as the final quantified goodness measurement for alternative solutions comparison.

The end-user configures the relative importance of the individual design metrics in a function that maps the individual metric values to a single SPI value. Firstly we define the default overall objective function as a weighted sum of the individual design metric normalized values as other AHP methodologies normally do: separation of defining design metric, and weighing those functions' importance in an overall objective function.

$$SPI_{norm} = a * m(L) + b * m(R) + c * m(T); \qquad (2)$$

Here (a, b, c) represent corresponding weight of performance metrics such as lifetime (L), reliability(R) and timeliness (T). (m (L), m (R), m (T)) represents the mean value of multiple measurement of corresponding metric. A key feature of our approach is that, we introduce the statistical analysis of the resulting experiment data, not only using measurement mean value as supposed normalized value (which is not realistic representation of the dynamic truth of the wireless network nature), we introduce deviation of performance measurement PDF as a critical secondary performance metric to emphasis the importance of performance stability and cohesion. Even a higher mean performance metrics, if the performance spreads over a wide spectrum of measurement, not cohesive to so called norm performance (mean) value, it will be problematic for certain application scenarios which require consistent performance , such as health monitoring application and multimedia application. We introduce stability performance index, as:

$$SPI_{stability} = a' * (1/\delta^2 (L)) + b' * (1/\delta^2 (R)) + c' * (1/\delta^2 (T)) \qquad (3)$$

Here (*a', b', c'*) indicates the relative importance of the metrics cohesive characteristic of metrics (L, R, T) represented by deviation ($1/\delta^2$). So overall we have:

$$SPI = SPI_{norm} + SPI_{stability} = \sum_{i=1}^{n} (W_i * Metric_i(mean) + W_i' * (1/\delta_i^2)) \qquad (4)$$

Here 'n' represent the number of metrics considered, W_i = *(a, b, c…)* represents respectively the user specified relative importance of the performance metrics. And W_i' = (*a', b', c'* …) indicates the relative importance of the metrics cohesive characteristic. The relative importance of each design metric as weight is assigned by considered application specific scenarios, how important in your application is certain metric (network lifetime, reliability, throughput, delay, etc) respectively? How important is cohesive characteristic of performance to your application? Which metric is utmost important for you?

4 The Algorithm Proposed

System evaluation process starts with the end users as application experts who know very well what kind of performance they needs; they specify the most concerned QoS performance metrics, and the weights of each metric.

The WSNs designers decide the initial parameters according to the literature studies and previous experiment experience. Then for each performance metric start the iterative experiment process as such:

1. Parameters significance analysis for $metric_i$.

Repeat *l* experiment measurements, record each experiment the state of each parameter x_i as f_{ji} *f* (1<*i*<*n*, 1< *j*<*l*) and corresponding performance measurement ψ_j (*1<j<l)*. Then use linear aggregation and P-value to decide significant parameters to $metric_i$.

x_1	x_2	$\cdots$	x_n	Ψ
f_{11}	f_{12}	$\cdots$	f_{1n}	Ψ_1
f_{21}	f_{22}	$\cdots$	f_{2n}	Ψ_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
f_{l1}	f_{l2}	$\cdots$	f_{ln}	Ψ_l

2. Design space reduces from *n* parameters to *m* parameters for $metric_i$.

3. *m* parameters interaction analysis for $metric_i$.

Tune the parameters based on the reduced parameters set, Repeat l experiment measurements, record the state of each parameter x_i as f_{ji} (1<*i*<*n*, *1*< *j*<*l*) and corresponding performance measurement ψ_j (*1<j<l).* Then use the Choquet nonlinear aggregation model as described in later chapter to decide the most effective parameters set including interaction effect of individual parameter.

4. Now we have finally approach the effective parameters set. Tune the effective parameters set, repeat measurement and get the performance curve, get the ($metric_i(\delta^2)$) *and* $metric_i$ (mean) for $metric_i$.

5. Change another metric of interest, start over again from (1).
6. When all metrics of interest finish evaluation, calculate the Single Performance index as aforementioned formula.

$$\text{SPI}=\sum_{i=1}^{n}(W_i * Metric_i(mean) + W_i^{'} * (1/\delta_i^2))$$

7. For competing solution for pair wise comparison, repeat the above process and get SPI value and compare:

```
If system1 SPI > system2 SPI then system1 perform better
                      than system 2
```

Notice that we can setup threshold value as prerequisite filter for minimum requirement, any time if $Metric_i$ mean or deviation is less than the threshold value, the candidate solution is not qualified for further comparison due to unsatisfactory for minimum user specification.

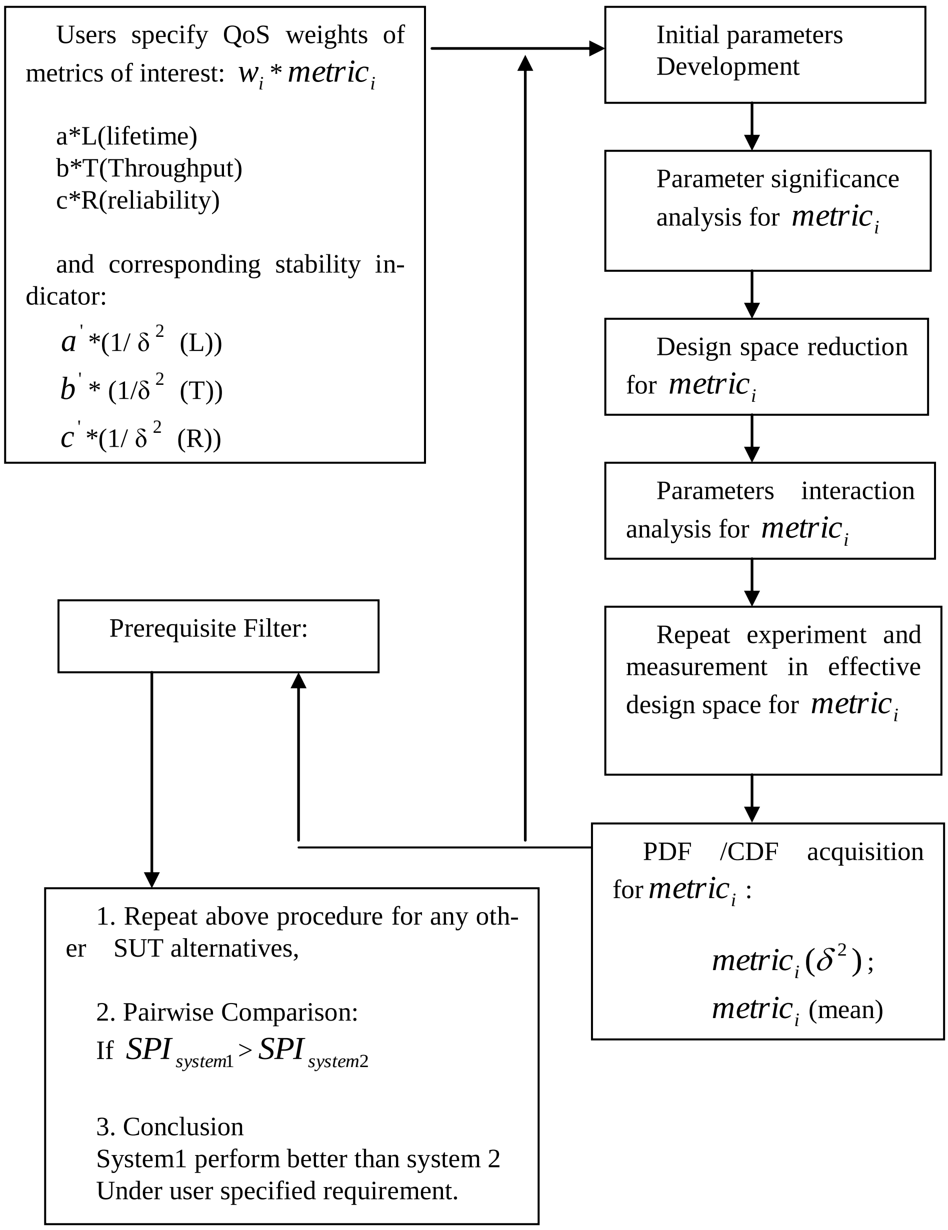
Prerequisite Filter:
If {
every $metric_i$ (mean)> threshold(mean)

And $metric_i(\delta^2)$< threshold(δ^2)
}
Then{
Single Performance Index: SPI=

$$\sum_{i=1}^{n}(W_i * Metric_i(mean) + W_i^{'} * (1/\delta_i^2))$$

)
Else
SPI=0 (not satisfy minimum requirement,
Not qualified for comparison)
Return

Fig. 4. Workflow of proposed benchmarking solution



5 Conclusion and Future Work

In this paper, we introduce a procedure to evaluate WSNs application performance according to application scenarios, in our approach, uncertainty attributes contribute to final performance index. Our future research direction is how to capitalize data mining technique to further dig deep experiment data and distil invaluable collective information from randomness. "Large-scale random phenomena in their collective action create strict, nonrandom regularity" [10].

References

1. Ko, B.J.G.; Lu, C.; Srivastva, M.B.; Stankovic, J.A.; Terzis, A.; Welsh, M. Wireless Sensor Network for Healthcare. *Proc. IEEE* 2010, *98*, 1947-1960.
2. M. Ali, T. Voigt, U. Saif, K. Rmer, A. Dunkels, K. Langendoen, J. Polastre, and Z. A. Uzmi, "Medium access control issues in sensor networks," SIGCOMM 2006, pp. 33–36, 2006.
3. G. Barrenetxea, F. Ingelrest, G. Schaefer, and M. Vetterli, "The Hitchhiker‟s Guide to Successful Wireless Sensor Network Deployments." in SenSys, 2008, pp. 43–56.
4. Yu, D., Nanda, P. He, X. "Performance Uncertainty Impact on WSNs Design Evaluation". IEEE Conference on Control Engineering and Communication Technology (*ICCECT2012*), 7-9 December 2012. Shenyang, China. Accepted.
5. The Ruckus white paper, available at: http://www.ruckuswireless.com/whitepapers/preview/wireless-network-performance
6. J. Haapola, F. Martelli, and C. Pomalaza-R´aez, "Application-driven analytic toolbox for wsns," in 8th International Conference on Ad-Hoc Networks and Wireless (ADHOC-NOW), LNCS 5793, pp. pp. 112 – 125, September 2009.
7. K. Khalili Damghani, M. T. Taghavifard, R. Tavakkoli Moghaddam, "Decision Making Under Uncertain and Risky Situations", Enterprise Risk Management Symposium Monograph Society of Actuaries - Schaumburg, Illinois, 2009.
8. J. Haapola, F. Martelli, and C. Pomalaza-R´aez, "Application-driven analytic toolbox for wsns," in 8th International Conference on Ad-Hoc Networks and Wireless (ADHOC-NOW), LNCS 5793, pp. pp. 112 – 125, September 2009.
9. T. Saaty, *The Analytic Hierarchy Process*. New York: McGraw-Hill, 1980.
10. B. V. Gnedenko and A. N. Kolmogorov *Limit distributions f or sums of independent random variables.* . Trans, and annotated by K. L. Chung. Cambridge, Addison-Wesley, 1954. 264+9 pp.
11. Yu, D., Nanda, P. He, X. "Performance Uncertainty Impact on WSNs Design Evaluation". IEEE Conference on Control Engineering and Communication Technology (*ICCECT2012*), 7-9 December 2012. Shenyang, China
12. Köksalan, M., Wallenius, J., and Zionts, S. (2011). *Multiple Criteria Decision Making: From Early History to the 21st Century*. Singapore: World Scientific