Influence of Geomarketing on the Rollout of New Telecom Network Infrastructure

Koen Casier, Jan Van Ooteghem, Manon Sikkema, Sofie Verbrugge, Didier Colle, Mario Pickavet, Piet Demeester
Internet Based Communication Networks and Services (IBCN)
Department of Information technology, Ghent University - IBBT
Gaston Crommenlaan 8 (Bus 201), B-9050 Gent, Belgium
{koen.casier, jan.vanooteghem, sofie.verbrugge, didier.colle, mario.pickavet, piet.demeester}@intec.ugent.be

Abstract—The rollout of new fixed network infrastructures requires significant capital investments and often a long payback period. To increase the net present value and to limit capital requirements a good rollout strategy is crucial. Cherry picking techniques will select the best areas, based on cost efficiency and highest average return per user, and will only install the new network technology in those areas, leading to higher return on investment. This paper investigates the impact of using geomarketing techniques to optimize a rollout strategy of a fiber to the home network in the Belgian city Ghent.

Keywords; geomarketing, Fiber to the Home, Techno-Economics

I. INTRODUCTION

Fiber to the Home, shortly FTTH, is most often seen as the next generation access network infrastructure. It offers much higher bandwidth and an almost unlimited bandwidth available with future equipment. Especially in Asian countries and to a lesser extent in the US, these networks are replacing the existing copper or coaxial based networks. Europe is clearly lagging in this trend. The current networks in Europe are good quality and already provide high bandwidth. This forms a high threshold to the installation of a fully new network. The high extra costs involved in the FTTH network installation should be compensated by the customers, and this extra cost could pose a competitive problem to the network operator.

Many European operators respond by upgrading their existing networks. Although so far providers have been able to keep up with demand, the question remains how long these existing networks will be able to fulfill the future demand. In order to be ready for this challenge, the operators should have a concrete planning for the possible future rollout of FTTH.

Planning the rollout of a whole new fixed infrastructure network for a whole country is a huge work which should of course be automated where possible. As the highest cost is on the physical infrastructure, most research and practical tools focus on calculating the lowest installation cost. This approach neglects the customer willingness to pay for the installation. In this paper we combine information of the customer with his location and find the best customers to connect first. Such a geomarketing approach is discussed in more detail in Section II. We implemented a calculation which takes the geographic information system (GIS) and returns a predefined number of separate clusters of customers. This calculation consists of two separate steps which are described in detail in Section III. In Section IV we build a full techno-economic model for the installation of a new FTTH network in these clusters using a phased installation. Finally in section V we show the results of combining the clustering with the techno-economic model. We vary the parameters of all calculations to find the best practices and trends in this geomarketing. We conclude the paper with a summary of the main findings.

II. GEOMARKETING MODEL

User segmentation is a widely used tool in the marketing world and leads to an adapted marketing and sales plan for each user segment to increase the net present value (NPV). In geomarketing we include spatial data in the segmentation and thus create geographical segments, or clusters, of users with similar characteristics. geomarketing has a wide set of applications such as branch network optimization, real estate expertise, customer localization, direct marketing, etc.

Spatial data is managed in a GIS. Each information element in a GIS has coordinates which can tie it back to a unique location on earth. Besides that, each element possesses attributes which contain additional information such as street names, number of inhabitants, etc. Typically GIS information consists of multiple layers, e.g. street layer, building layer, river layer. A multitude of tools exist which are capable of visualizing and/or working with the GIS information in a specific and tunable manner.

We used a two phased approach to perform this geomarketing. In the first phase we combine the detailed customer information to build a profile of the customer using data-mining tools. The second phase performs a clustering in which the customers are grouped according to their distance to connect and their profile.

After the clustering into geographically and profile-based customer groups, we build an economic model adapting existing investment models to the rollout order and different timings for the customer groups. In Figure 1 the developed model is shown. The focus of the research in this paper was on part 1, the clustering. Part 2 was based on [1] with some adaptations for the specific analysis in this paper. We will discuss the key parts of the model in the next paragraphs.

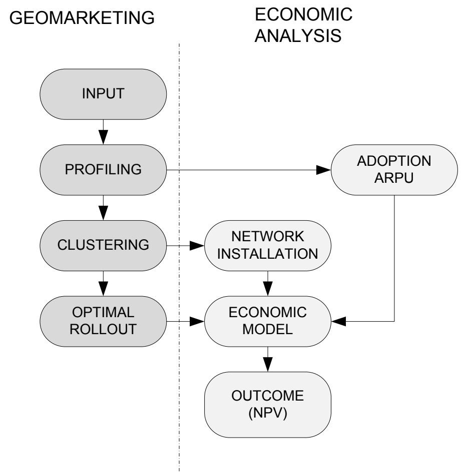

Figure 1: geomarketing model

This paper builds upon the approach suggested in [2], but has important differences increasing the applicability and usefulness of this approach. At first we use a profiling step classifying the families according to their expected eagerness to adopt the new technology. Secondly we cluster on a per customer base and try to find the optimal customer groupings, which we are sure will be able to connect to the central office. In a realistic situation, the central office might well be another free parameter and as such be chosen close to the most promising groups.

A. Input data for the inhabitants

A combination of data has been used for making the profiling of the customer. Most data was provided by the City of Ghent, resulting in detailed information regarding geographical location and demographics. The geographical information contains sector level info, streets, buildings and their function, location of residences, number of inhabitants per dwelling, etc. As geomarketing analysis also requires detailed information concerning the type of users and their potential, we included both age and income parameters. The latter parameter was not available per customer (due to privacy reasons) so a lognormal distribution over the population per sector was executed with standard deviation 1.2. This value is smaller than 1.45 proposed in [5], which is the value for a nationwide spread, where our model focuses on smaller clusters with more correlation of income levels. Other interesting parameters that would be very relevant are current Internet subscriptions and ARPU but this is information was not available.

B. Profiling the potential customers

In this first step we classify the customers according to their profile. A detailed profile of customers is typically output of dedicated marketing studies. As they fall outside the scope of this study, we profiled customers according to the aggregation of the GIS knowledge discussed above. This profiling step is used to reduce the amount of data available. As the aggregation of all customer information is the basis for making estimations of the willingness to adopt the new FTTH technology and new services offered by the operator, it is important to make sure that the profiles generated reflect this potential. As such the profiling step will either group customers with a dedicated adoption potential and estimation per group or put the customers in a linear scale linked to a scaling adoption potential. Both approaches have been investigated and are discussed below.

In the first approach, we used the Weka data-mining toolkit [4] for retrieving the different types of customers. Here we used an X-means clustering-algorithm and looked for a predefined number of customers-profiles. As each of these groups will be linked to a separate adoption model, it is key to limit the number of groups - for ease of use – between 5-10 customer profiles. Different tests with this X-means clustering algorithm indicated that the main parameters for the clustering are average income of the family and the average age of the family. The clustering is hard to tune and typically only takes one of both parameters as its main parameter, depending on the weight given to each of those two parameters. When asked for 10 profiles, the X-means clustering returns 10 groups, but 80% of the customers is assigned into one of four large groups and the remaining 20% of the customers are spread over 6 groups.

As both average income and average age are the most important information bases for the clustering, the second approach will start from this knowledge. Literature confirms this trend and when looking at market studies of the adoption of new technologies, they typically find both income and age to be of high importance [6]. We also expect the willingness to adopt to be higher for richer and younger families. This will reflect in a higher adoption potential – percentage of the potential customers actually buying the service, and a faster adoption. As such we can more easily construct a linear grading function for the customer profiles based on both parameters. Next to the fact that such a grading scale better reflects the impact of both income and age, it also has a much lower calculation effort as no clustering is involved, but a mathematical function is used. In the remainder of this paper we use the following grading function:

*Profile Customer = 2 * income rating + 1 * age rating*

The parameters for both income and ago rating can be found in Table I below.

TABLE I: GRADING OF AGE, INCOME AND FINAL CUSTOMER PROFILE

Rating	Age	Income (€)	Profile Customer
1	>65	<20,000	1-3
2	60-65	20,000-30,000	5-7
3	<25; 51-60	30,000-40,000	8-10
4	25-27; 41-50	40,000-55000	11-13
5	28-40	>55,000	14-15

Using the grading function indicated in the same table, we created a profile for each household in the range of 1-5 indicating their likely potential for FTTH adoption (based on outcome formula in the last column of the table above). In the following section this profile data is linked to the spatial data and the buildings are clustered based on both spatial and user profile data.

C. Clustering the potential customers

In the first step, all potential customers are profiled and in the second step, these customers should be optimally grouped.

The optimal grouping of the customers will certainly depend on their profiles but also on their geographic location. Rolling out an FTTH network requires the installation of a fully new network from the central office up to the customer and as such involves tremendous road works. One way to reduce costs in this installation is to connect only those customers which are close to each other and as such add little to the cost when coupled to the network.

We used a hierarchical clustering algorithm to find the optimal groupings. We start with each potential customer in a separate cluster and iteratively group the two closest clusters together. The distance between any two clusters is in this not just the geographical distance, but a function of the total estimated installation cost in the new cluster and the profiles of all customers in the cluster. As a very fast estimation of the installation length of the cluster, we use a minimal spanning tree over the street network.

The part of the estimated cost linked to the customer's profiles is based on the difference between the average profiles of both clusters. This will for instance prevent grouping two customers together if they have a very different profile, and rather look for very comparable customer profiles to group.

The implementation of this clustering algorithm has been done in Java. It takes in the GIS information map of the profiles of the and the GIS information of the area streets and customer location and gives as an output the mapping of the potential customers to their groups (number) again in a GIS information map. This way the outcome of this calculation can be visually inspected using any GIS visualization toolkit. The clustering implementation has an additional parameter to set the amount of clusters to look for.

The clustering implementation contains next to what has been described above, several additional tricks to increase the performance of the calculations. For instance for large full city areas, we split the area in different sub-areas, perform a clustering in each of the sub areas and cluster these. The combination of all clusters for the sub-areas into larger clusters over the full area is then again performed using the same clustering approach.

III. ECONOMIC MODEL

Once the customers are profiled and grouped in a limited set of clusters, a full economic model for the optimal FTTH rollout has been created.

Using the average user profile and user density of each cluster we rank the clusters on both potential and rollout cost per user. Based on the combined ranking of both we determine an optimal rollout strategy. This information is then fed to an Excel model [7] that models adoption, costs and revenues and calculates a 15-year NPV.

A. Adoption model

To predict the market potential of the newly installed network, we have to forecast the uptake of the service. We have chosen the Bass adoption model [8] for predicting the uptake of the five different profiles as discussed in the previous section. The Bass adoption model uses four parameters:

- The *market potential (m)* between the different groups varies between 65% and 95% for the lowest and highest profiles, respectively.
- The *year of introduction (t)* has been kept the same for all profiles due to the size of the city of Ghent.
- The *innovation coefficient (p)* indicates the influence of the innovators on the adoption. A higher value will lead to a faster adoption especially in the initial phase. Realistic values can be found in the range [0.005, 0.03] [9][10].
- The *imitation coefficient (q)* shows the influence of the imitators on the adoption. A higher value will lead to a faster adoption. We found realistic values for this parameter in the range [0.2, 0.5] for broadband access.

Both the innovation and imitation parameter are influenced by income of the customer [11]. The results of the adoption potential can be seen in Table II and Figure 2.

TABLE II: ADOPTION POTENTIAL

USER POTENTIAL	P	Q	M
Cat 1	0,008	0,20	65.0%
Cat 2	0,010	0,25	72.5%
Cat 3	0,020	0,30	80.0%
Cat 4	0,025	0,40	87.5%
Cat 5	0,030	0,50	95.0%

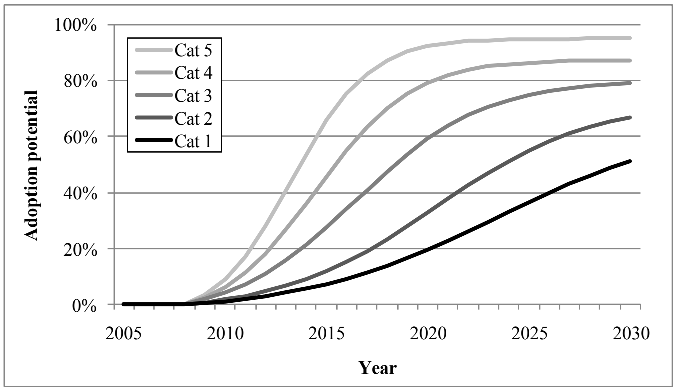

Figure 2: Adoption potential

B. Revenues

The ARPU (average revenue per user) is also determined on the type of user. High profile users (or families) will tend to subscribe to new technologies faster and thus to higher bandwidth services, leading in the end to higher revenues for the network operator. Based on this we chose to vary the market shares of the subscriptions based on the user profile.

In Table III an overview can be found of the subscription types and the forecasted adoption rates. Cat5, which has the highest potential, has a predicted \$66 ARPU per month. The lowest category (Cat 1) is assumed to have a \$45 ARPU per month. The other categories can be found in the range between these average revenues per month per user.

TABLE III: SUBSCRIPTION OVERVIEW

	ECO 35\$/MONT	STANDAARD 60\$/month	PREMIUM 80\$/month	ARPU \$/month
Cat 1	60%	40%	0%	\$45
Cat 2	40%	55%	5%	\$51
Cat 3	25%	65%	10%	\$56
Cat 4	20%	60%	20%	\$59
Cat 5	10%	50%	40%	\$66

C. *Cost model*

The costs are split into capital (CapEx) and operational expenditures (OpEx).

1) *Capital Expenditures (CapEx)*

To install a new optical fiber, there exist three major possibilities: (1) digging or trenching, (2) blowing or pulling in existing ducts and (3) an aerial rollout. Although aerial rollout (either on façades or poles) is the least expensive solution, often it cannot be used because of regulatory rules. Using available ducts (e.g. put in the ground during previous civil works) is a frequently used alternative. However, it can occur that these ducts are silted up after some years, and then it becomes very hard to reuse them. If no aerial rollout is permitted and no installed ducts can be used, which is very often the case, expensive digging works are required by which the connection to every home is by far the highest cost.

Digging is the most used method to rollout a new fiber network and represents the most important cost factor in a fiber rollout. Table IV shows the considered share of the different installation methods, together with the used costs including installation and duct costs, but without fibers. As we consider that a municipality has the advantage of combining this rollout with civil works or renovation of other city (utility) infrastructure, we take into account a 10% reduction of digging costs. The calculation of the digging cost and thus the distance of the new fiber plant is determined by the cluster analysis results.

TABLE IV: COSTS FOR THE CONSIDERED FIBER INSTALLATION METHODS

Installation Method	Cost (\$/m)	Share (%)
Digging	55	80
Pulling	3	10
Aerial	8	10

The costs of the fibers are not included in the above installation costs. We suppose that, on average, 40 fibers will be simultaneously installed, in which we expect a point-to-point rollout scenario. The cost for single mode fiber per meter is assumed to be 0.065\$/m exclusive installation (1.3\$/m) Further, the fiber length is over provisioned by a factor of 20% to provide branches to every house.

On the user side, an ONU (optical network unit) is required which is estimated at \$250. For the installation, a distinction is made between an early subscriber and customers who subscribe after the network is installed. For the early subscribers, the connection can be made at the time the network is rolled out, when the installation will be less time-consuming and thus cheaper. To promote an early subscription, the ONU will be paid by the operator, while the other customers have to pay themselves for the ONU. Finally, we want to stress that the fiber and electric costs will decrease with increasing production volumes. To model this cost erosion, we have used the extended learning curve model presented in [12], with a distinction between electrical and optical components (a faster cost erosion for optical components).

2) *Operational Expenditures (OpEx)*

While different models exist for calculating OpEx, the operational processes to feed these models with, are largely unknown especially when considering new technologies [13]. We have chosen to use a coarser modeling in which the different subtypes of OpEx are calculated as a fixed cost per subscriber. These OpEx then include, among other things, network operations, maintenance, marketing, billing, helpdesk, etc.

D. *Rollout strategy*

We now dispose of all information concerning the different sectors by combining their potential and appropriate costs and revenues. An important final step is determining the rollout sequence. The ratio between the total number of inhabitants in that section and the minimum spanning tree shows the population density; the potential of the area indicates the revenue potential. Clusters will thus be ranked according to the combination of both parameters, from most positive to most negative.

IV. RESULTS

We applied the explained model to the city of Ghent for which we have detailed demographic information as well as spatial data.

Figure 3: Overview of the City of Ghent

Ghent (Figure 3) is the third largest city of Belgium with 243.144 inhabitants (2009) and an area of 156.2 km² (corresponding to an average population density of 1.556 residents per km²) [14].

The city has been split up in 25 areas, which are subdivided in 223 sectors. In total this area consists of about 114.795 families and 96.892 buildings. Important to note is that Ghent has a very large student population consisting of about 65,000 students of which 19.480 have a residence in the city.

TABLE V: USER PROFILING RESULTS FOR THE INHABITANTS IN GHENT

	# OF HOUSEHOLDS	AVG. INCOME PER BUILDING	AVG. AGE	AVG. INCOME PER HOUSEHOLD
Cat 1	2,205	€ 26,798	67	€ 17,953
Cat 2	20,814	€ 43,586	56	€ 24,575
Cat 3	27,107	€ 59,482	48	€ 33,359
Cat 4	24,892	€ 74,096	44	€ 49,512
Cat 5	5,235	€ 78,146	38	€ 64,867

A. Base case: 40 clusters and 7 year rollout

Table VI: shows the results for the base case, where we assume a rollout of the city area in 7 years. The results are compared to the original rollout sequence with the same assumptions. A discount percentage of 10% is used.

TABLE VI: BASE CASE: 40 CLUSTERS, 7 YEAR ROLLOUT

	NPV 15y ($M)	Break-Even year	Minimal NPV ($M)
geomarketing (cherry picking)	74.5	2019	-31.4
geomarketing (full rollout)	69.3	2021	-36,5
Original sequence (full rollout)	60.4	2021	-40.5

Three important conclusions can be drawn from these results:

- The total NPV evaluated after 15 years is much higher (+23%) than for a non-geomarketing rollout. About a third of this profit ($ 5M) is thanks to a decision of not rolling out some areas (indicated in the cherry picking).
- The break-even point is 2 years earlier than for a full rollout (for either geomarketing or original sequence)
- The minimal NPV is lowest, meaning that the operator will have to invest less money than in the other scenarios.

An overview of the NPV curves can be observed in Figure 4.

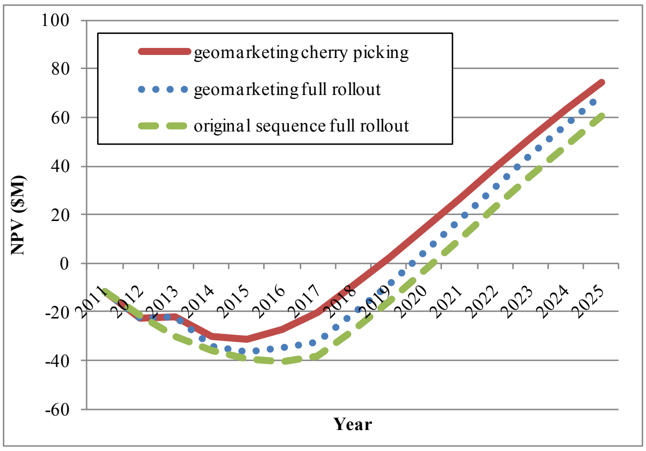

Figure 4: Base Case NPV results

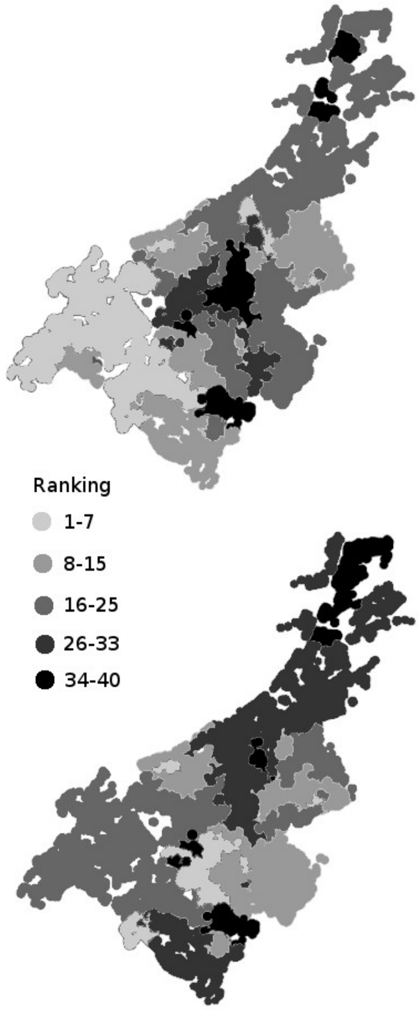

Figure 5: Rollout area sequence comparison

It is important to note that the cluster rollout strategy is based on both the user profiling as well the cluster density, which implies that we look both at revenue potential and expected cost when deciding the rollout sequence. This can be seen in Figure 5, showing on the left side the rollout sequence only based on the potential and on the right a combination of both parameters. The port area in the North has been clearly delayed in rollout (few inhabitants and large area); the city center has been put forward.

B. *Impact of varying rollout speed*

In the base case, we have worked with a rollout speed of 150 km per year. In this section we will show the impact of varying the rollout speed in the range of 1000 km/year (= a full rollout in 1 year) and 66 km/year (= 15 year rollout). In Figure 6 the results are illustrated.

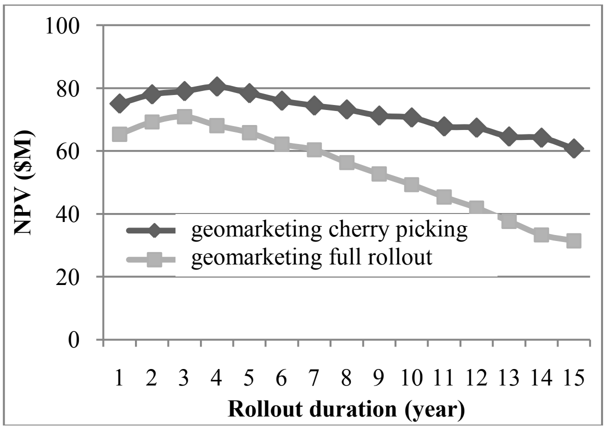

Figure 6: Impact of varying rollout speed

We clearly see that for all rollout speeds the geomarketing solutions result in better NPV's. The difference varies from ~10% at a one to three year rollout to over 100% at a 15 year rollout. The minimal NPV in each case remains ~10-20% better for geomarketing. The results of the geomarketing model also improve with longer project rollouts compared to the phased solution.

C. *Impact of varying number of clusters*

In Figure 7 we look at the impact of varying number of clusters. As expected, increasing the number of clusters results in an increased advantage of geomarketing over a randomized phased rollout. This advantages levels out at 25% around 40-50 clusters, implying that increasing the level of clusters beyond 40 is unnecessary (also the minimal NPV levels out at ~35% better).

Figure 7: Impact of varying number of clusters

V. CONCLUSION

The rollout of new telecom access networks is very costly, and thus requires a good planning. As such, geomarketing is definitely a valuable tool for network operators to optimize their network rollout by combining geographic and demographic information, either to decrease costs and increase revenues. Clearly a good approach should take into account the two most important aspects for the outcome of the project, namely the expectable customer revenue and the costs for connecting the customers.

This paper explores a two phased geomarketing approach in which the customers are profiled in the first step and are clustered based on this profile and their geographical distance during the second step. The customer profile is mainly based on the customer age and its average income. The distance is also taken into account, as in the rollout of an FTTH network the digging will constitute the main part of the costs, benefitting customers that are closer to each other.

Both phases are currently often still executed manually and are based on a lot of expert opinions, up-front knowledge and limitations imposed by the existing network structure (e.g. existing central offices and surrounding areas). This will certainly not lead to the best solution. This paper indicates how to automate both steps in the process using existing GIS data, existing tools and clustering algorithms. The solution of such automated process leads to a better (near optimal) solution. In this solution the areas are no longer based on known neighborhoods, telephone areas but are more organic smaller areas with people living closely together with the same profile. We complemented this solution with a full economic modeling. This allowed for the calculation of the outcome of a phased network rollout in those areas, rolling out first to the best areas. We found that the solution based on geomarketing as presented in this paper can lead to an improvement 23% in NPV, a faster break-even point with less initial investments. Finally, the solution we have proposed allows clustering into a predefined number of clusters, but we have shown in the paper that over 40 clusters the gain saturates.

ACKNOWLEDGMENT

This research was carried out as part of the IBBT TERRAIN project. This project is co-funded by IBBT, IWT and Acreo AB, Alcatel-Lucent, Comsof, Deutsche Telekom Laboratories, Digipolis, FTTH Council Europe, Geosparc, Stad Gent, TMVW, TE Connectivity, UNET and WCS Benelux BV. This work was also carried out in the framework of the COST ISO605 Econ@Tel project.

REFERENCES

[1] J. Yick, B. Mukherjee, D. Ghosal "Wireless sensor network survey. Computer Networks", 52 (12), pp. 2292--2330 (2008)

[2] B. Van Den Bossche, R. Meersman, J. Vanhaverbeke, A. Schoutteet, "Maximizing the Return on Investment for FTTx-rollout through the use of GIS Street Maps and Geomarketing data", CTTE 2010, 7-9 June 2010, Ghent

[3] V. Raghunathan, A. Kansai, J. Hsu, J. Friedman, M. Srivastava, "Design considerations for solar energy harvesting wireless embedded systems, 4th International Symposium on Information Processing in Sensor Networks (IPSN), pp. 457-462, Los Angeles, California, USA (2005)

[4] M. Hall, F. Eibe, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten (2009); "The WEKA Data Mining Software: An Update"; SIGKDD Explorations, Volume 11, Issue 1.

[5] C. Kleiber, S. Scheid, "The Generalized Lognormal Distribution as an Income Distribution", Bulletin of the International Statistical Institute 60:643-644.

[6] E. Rogers, "Diffusion of Innovations 4th ed.", New York: Free Press,1995

[7] B. Lannoo, K. Casier, J. Van Ooteghem, B. Wouters, S. Verbrugge, D. Colle, M. Pickavet, P. Demeester, "Economic Benefits of a Community Driven Fiber to the Home Rollout", Broadnets 2008, London, Sept UK, 8-11, 2008

[8] F. Bass, "A new product growth model for consumer durables", Management Science, 15(5), p215–227, 1969

[9] V. Mahajan, E. Muller, F.M. Bass, "Diffusion of New Products: Empirical Generalizations and Managerial uses", Marketing Science, Vol. 14, No. 3, Part 2, 1995

[10] I. Cluyse, "Diffusie van het Internetgebruik in Europa: Analyse door middel van het Bass Model", 2006 (dutch)

[11] S. Nam, "Empirical test of the Bass Diffusion Model using exogenous shocks on word of mouth effect", 2009

[12] B. T. Olsen and K. Stordahl, "Models for forecasting cost evolution of components and technologies", Telektronikk 4.04, pp. 138-148, 2004.

[13] S. Verbrugge, S. Pasqualini, F.-J. Westphal, M. Jäger, A. Iselt, A. Kirstädter, R. Chahine, D. Colle, M. Pickavet, P. Demeester, "Modeling Operational Expenditures for Telecom Operators", Proc. of ONDM 2005, pp. 455-466, Milan, Italy, Feb. 2005.

[14] City of Ghent, "Gent in cijfers", http://www.gent.be/gentincijfers/, (dutch)