# Parametric distance distributions for fixed access network analysis and planning

Catherine Gloaguen*, Florian Voss† and Volker Schmidt†
*Orange Labs, 38–40, rue du Général Leclerc, 92794 Issy–les–Moulineaux, France.
catherine.gloaguen@orange–ftgroup.com
†Institute of Stochastics, Ulm University, Helmholtzstraße 18, 89069 Ulm, Germany.
{florian.voss,volker.schmidt}@uni–ulm.de

***Abstract*— The access network displays the important particularity that the locations of the network components strongly depend on geometrical features such as road systems and city's architecture. This paper presents a model for access networks arising from stochastic geometry, which takes this geometric structure into consideration, and thus offers a relevant view on location-dependent characteristics such as point-to-point connections as shortest paths along the road system. Closed analytical formulas are derived for point-to-point distance distributions, explicitly depending on the morphology of this road system. Theoretical predictions are successfully compared to real data from fixed access networks deployed in dense urban areas.**

## I. Introduction

No more pertinent introduction to this paper could be found than those sentences quoted from this very ITC21 Call for Papers "Now, the networking community is facing a rapid evolution and diversification of networks [...] As a matter of fact, advances in optical and wireless technologies will open new possibilities for networks in the near future with new performance problems in terms of traffic management. In addition, there is currently considerable activity worldwide on the design of new architecture principles and concepts for future networks [...] New architectural elements and business models are needed to finally meet user expectations for quality and security of their communications in a cost-effective way." Accordingly, the scope of the model presented here is to offer an easy to use, reliable and efficient tool to the network operator for the global analysis of huge networks that explicitly takes into account the geometry of the territory while being able to describe various technologies and architectures.

Although perhaps not perceived by modern customers immersed in a world of mobility, the fixed part of the access network is nevertheless an important and omnipresent entity, being based either on traditional (copper) or more recent (optical fibre) technologies. Due to its complexity in terms of variety of equipment types, geographical settings and history of successive amendments, this part of the network is for the historical operator a major cost element as well as an important source of incomes. A characteristic feature of the fixed access network is its strong dependency on the geography and the territory infrastructure, especially on the road system that is used as a natural guide for the physical telecommunication lines. Key components of the quality of service or the technical feasibility of architecture solutions, as for example the estimation of connection lengths, are thus very sensitive to the geometry of the network implantation and subject to regional and/or scale specificities.

Pertinent quantities relevant to the network performance evaluation or planning are usually estimated either by extraction from databases or by reconstruction. Each method has its own advantage and drawbacks. Databases are snapshots of the network and offer in principle a true picture of the reality. But they do not always describe all the network parts, are not available for non–existing networks, do not contain equally reliable data and are huge and not easy to handle. Moreover, they are only descriptive by nature and thus do not give any straightforward interpretation of the actual network state. Exhaustive reconstruction of a given network implementation scenario gives precise and local information. But it is often prohibitive in computation time and thus not fitted to global analysis at large scale. Then by necessity, the analysis is restricted to few cases that may not be representative for the total set of possible configurations. Finally, note that these two methods are often used to recover or reconstruct the same global information.

How to deal with these seemingly contradictory requirements: being able to provide rapid answers and estimation at global scale while being at the same time able to keep information on geographical features? A solution is to turn the intrinsic variability and the complexity of the whole network as an advantage. This is possible in the framework of stochastic modelling where the choice of random models and parameters allows to deal directly with the desired statistical information while taking into account the geographical features and offering an explicit relationship between the network and the underlying territory morphology. In the present paper, we propose the latest state of the art of the "Stochastic Subscriber Line Model" (SSLM) that was being developed in the past years [1], [2]. The mathematical tools for stochastic–geometric models have been greatly enriched since their first application to global network modelling [3], [4] and now offer reliable and instantaneous formulas for distance distributions involved in the global analysis of access networks.

The paper is organized as follows. Section II is a short description of the access network and the SSLM principles. Sections III and IV briefly introduce the mathematical aspects of spatial stochastic modelling and explain how one can obtain analytical formulas for point-to-point distance distributions,

with explicit dependence on the morphology of the underlying road system. In the last Section V, we show how the model fits with real analysis of networks in dense areas.

## II. Stochastic models for access networks

### A. Technical overview of the access network

The purpose of a network is to allow connections and data transfer between customers. The access network (or local loop) is the lower part of the network, connecting a subscriber (the phone or the computer at home) by a physical link to its corresponding Wire Center Station (WCS) via intermediate network components such as the Service Area Interface (SAI) and Network Node Device (ND). Connections between WCS nodes are ensured by the core network not considered here. The three types of access nodes play a concentration role, allowing to merge several cables of lower capacities to a single cable of higher capacity.

In a given area, the road system is built or planned in order to offer communication ways and/or access to the whole area. It is pre-existent to the telecommunication network: either in time or during the territory planning process. Since both kinds of networks have similar goals in connecting people, it is natural for the telecommunication network to use the road system that itself reaches the customers. Then access network nodes as well as connections strongly depend on the morphology of the road system. The cables run under pavements in trenches forming the civil engineering part of the whole telecommunication network: the fixed access network can be considered as the place, where the telecommunication network merges into civil engineering (Fig. 1). The set up of

Fig. 1. The access network merges in the urban road system.

a network with hierarchical architecture aims at decreasing its costs with economy of scale induced by the merging of cables. A delicate trade-off has to be found between the costs of individual nodes and the costs of the cables depending on their capacities, while ensuring the technical feasibility of the solution: node capacities vs. demand, length of connections vs. path loss requirements. It is then of extreme importance to be able to take into account the spatial structure of the road system in the process of network analysis or planning.

The underlying road system is itself a very complicated object. Its morphology depends on the scale of analysis, on the population density, the geographical constraints, the economic activity, etc. Since national roads and motorways are designed to connect major towns and since small streets or dead ends are designed to give access everywhere in inner cities, their overall shape and properties are very different (Fig. 2). Detailed road data are available in coordinate form, but as for network databases, they are only descriptive and need dedicated software to be handled with.

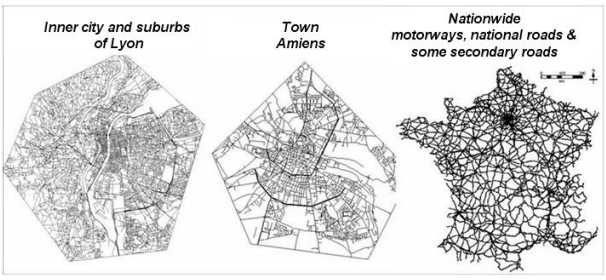



Fig. 2. The morphology of road systems depends on the scale of analysis.

In order to study such topics as the impact of the morphology of road systems on the performance of a given architecture and thus to propose a geographically based network analysis accounting for regional specificities in the analysis of fixed access networks, a natural way is to consider separately the geographical and purely telecommunication network components. This requires that reliable and fast models are available for each of the three building blocks: 1) geometrical support (i.e. the underlying road system), 2) locations of network nodes with respect to roads, and 3) connection topology (i.e. physical paths).

Such a description allows to deal with a variety of realistic situations by simply combining various models for these three parts of the network.

### B. Stochastic modelling principles

The power of stochastic modelling is to take advantage of the complexity of the system. It is particularly suited to our purpose which is to develop pertinent and reliable tools to analyse the network at global scale in order to compare various architectures and technologies.

Let us consider an example: knowing the exact distance from a given customer to its corresponding WCS may be of importance locally, but this kind of information, given for all the customers would be too huge to be handled with and also impossible to be reconstructed exactly for every scenario one should want to test. But if one considers the whole set of customers and their connection distances to WCS, one observes a variability that can be interpreted in a statistical way as a distance distribution. This distribution contains useful global information, say for example the percentage of customers that can be reached within a length threshold imposed by the technology.

At a macroscopic scale, global rules of connection or location are more relevant than minute details annihilating each other. Being able to directly link statistical distributions to some pertinent parameters of each of the three blocks (geometrical support, network nodes location and connection topology) is the key of stochastic modelling. Systems of randomly located lines or points are mathematical objects described in the framework of the theory of spatial processes.

They are defined by a small number of parameters while reproducing the geometrical and statistical features of real data and are ideal components for the SSLM blocks.

**Geometrical support.** Objects (lines, points) are thrown in a random way to generate a paving of the plane (a "tessellation") the edges of which can be used as a model for road systems. Several simple models (Fig. 3) are available: tessellations constructed from lines (Poisson Line Tessellation PLT), from links between points (Poisson-Delaunay Tessellation PDT) or from areas around points (Poisson-Voronoï Tessellation PVT). Iterated tessellations obtained by combinations, and the possibility to include a fraction of empty areas, provide even more realistic models. Under the Poisson assumption of maximum randomness (the objects are located independently from each other and their number follows a Poissonian law), a simple homogeneous random tessellation is fully defined by a single parameter: its intensity. Then, mathematical developments can lead to exact analytical formulas relating statistical features of the tessellation to its type and intensity [5], [6]. The vector

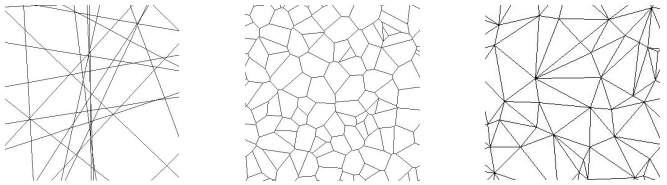


Fig. 3. Realizations of simple tessellations of the plane, PLT, PVT and PDT

$\mathcal{T} =$ (number of crossings, number of quarters, total length of streets, number of streets segments) of averaged quantities per unit area computed on the set of all possible realizations of the tessellation (or equivalently on an infinitely large realization) statistically describes the morphology of the road system. One can define the distance between a theoretical random model and the real road system by comparing the theoretical $\mathcal{T}$ to the vector of corresponding empirical quantities using, for example, the relative Euclidean norm. From the explicit dependence of $\mathcal{T} = \mathcal{T}(type, \gamma)$ on the type of model and its intensity $\gamma$, it is possible to choose the best model among a given set of candidates by minimizing the distance. This fitting procedure is described in [7]. The best model statistically reproduces the morphology of the real road system while being described only by its type and intensity, see Fig. 4.

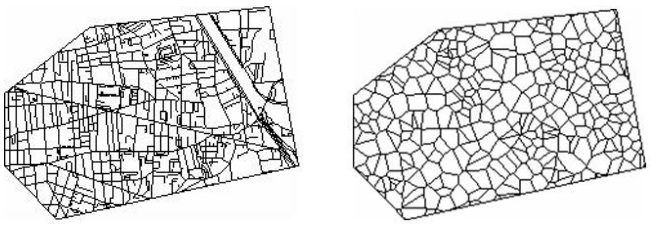


Fig. 4. A real road system (528 crossing, 324 quarters, 849 street segments and 97 km total street length) is replaced by a simple random PVT giving in the average (585 crossing, 293 quarters, 878 street segments and 82 km total street length)

**Locations of network nodes.** Since we are interested in point-to-point distances, it is sufficient to consider sub-

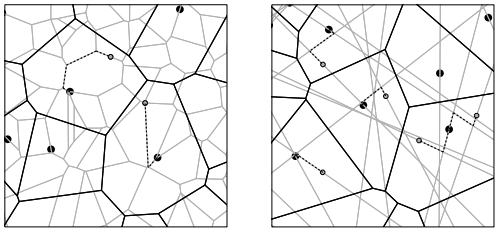


Fig. 5. HLC with their serving zones (black) and LLC (grey with black boundary) with shortest paths (dashed) along the edge set (grey). Street model PVT (left) or PLT (right).

networks involving two types of nodes: low-level components (LLC) and high-level components (HLC). Both types of nodes are independently and randomly located with respect to the road model. This reflects the variability of situations encountered in real networks, where the simplest model is also here a location of nodes under the homogeneous Poisson assumption. Then given a street, the intensity, i.e. the mean number of nodes per unit length of the street, is the only parameter required to fully determine the mathematical model for each node type. LLC are assumed to be logically connected to their nearest HLC, in the straight-line sense. This defines the serving zone of each HLC as the set of those points in the plane that are closest to it, see Fig. 5.

**Connection topology.** Physical connection is established as the shortest path from LLC to HLC following the streets. This appears to be a realistic assumption for the fixed access network. Then in SSLM, the point-to-point distances are computed as the lengths of shortest paths following the streets, linking two sets of nodes located on the streets, which themselves are modelled as random processes (Fig. 5). This framework is sufficiently well defined so that one can derive analytical formulas for the distribution of distances between nodes, with explicit dependence on the topology of the underlying road system. More detailed insight into mathematical methods is given below.

## III. Mathematical methods and results

### A. The Stochastic Subscriber Line Model

In this section we briefly describe the mathematical background of the stochastic network model which is considered in the present paper. More details can be found e.g. in [1], [7]–[9]. The model is based on (marked) point processes and random tessellations, see [5], [10], [11] for details.

**Random tessellations.** A random tessellation $T$ is a partition $\{\Xi_n\}$ of $\mathbb{R}^2$ into random (compact and convex) polygons $\Xi_n$ which are locally finite. The polygons $\Xi_n$ are called the cells of $T$. A random tessellation is called stationary if its distribution is invariant with respect to shifts of the origin $o$. We can identify $T$ with its edge set $T^{(1)} = \bigcup \partial \Xi_n$, i.e., the boundaries of the cells of $T$. Now suppose that $T$ is stationary. Then we define the intensity $\gamma$ of $T$ as $\gamma = \mathbb{E}\nu_1(T^{(1)} \cap [0,1]^2)$, i.e. the mean length of $T^{(1)}$ per unit area. In the following, we assume that $T$ is either a PLT, a PVT or a PDT, see Fig. 3.

**Typical shortest path length.** For any $T$ with intensity $\gamma$, we model the locations of HLC and LLC by (linear) Poisson

processes $X_H = \{X_{H,n}\}$ and $X_L = \{X_{L,n}\}$, respectively, on the edges $T^{(1)}$ of $T$ with (linear) intensities $\lambda_\ell$ and $\lambda'_\ell$, respectively. Then the planar intensities, i.e., the mean number of points per unit area, $\lambda$ and $\lambda'$ are given by $\lambda = \lambda_\ell \gamma$ and $\lambda' = \lambda'_\ell \gamma$. Now we associate to each HLC its Voronoi cell with respect to $X_H$ as its serving zone. All LLC are then connected to the HLC in whose serving zone they are located. In this way we can associate to each LLC a length $C_n$, namely the shortest path along $T^{(1)}$ to its closest HLC (Fig. 5). Thus we get a marked point process $X_C = \{(X_{L,n}, C_n)\}$. In the following we investigate the distribution of the typical shortest path length $C^*$ of $X_C$. Formally, the distribution of $C^*$ is defined as the Palm mark distribution ([10], Chapter 13.4) of $X_C$, but it can be regarded as the limit of empirical distributions of the shortest path lengths of all LLC in a sequence of unboundedly increasing sampling windows. Suppose e.g. that $W_n = [-n, n]^2$ and $h : \mathbb{R}^+ \longmapsto \mathbb{R}^+$ is some function, then

$$\mathbb{E}h(C^*) = \lim_{n\to\infty} \frac{1}{\#\{j : X_{L,j} \in W_n\}} \sum_{X_{L,j}\in W_n} h(C_j) \quad (1)$$

almost surely. Equation (1) motivates why we are interested in $C^*$, see also Section II-B. Furthermore, we can also regard $C^*$ as the shortest path length from the origin $o$ to its nearest HLC under the condition that there is a LLC at $o$.

**Typical serving zone.** In the next section we show how the density $f_{C^*}$ of $C^*$ can be estimated based on simulations of the typical serving zone $\Xi_H^*$ and the segment system $L_H^*$ inside $\Xi_H^*$. The distribution of the typical serving zone can be regarded again as the limit of empirical distributions of the serving zones in a sequence of unboundedly increasing sampling windows or as the (conditional) distribution of the Voronoi cell at $o$ given that there is a HLC located at $o$.

### *B. Density of shortest path length and its estimation*

We now state a representation formula for the density $f_{C^*}$ of $C^*$ which depends only on the typical segment system $L_H^*$ inside the typical serving zone $\Xi_H^*$. This formula is suitable to construct estimators for $f_{C^*}$ based on i.i.d. samples of $L_H^*$ which can be obtained from Monte Carlo simulation. If $T$ is a PLT, PVT and PDT, respectively, then simulation algorithms for $L_H^*$ are known, see [9], [12], [13].

**Density of shortest path length.** In the following, we derive a formula which allows us to compute the probability density of shortest path length $C^*$. But we first state a formula which represents the quantity $\mathbb{E}h(C^*)$ in terms of $L_H^*$.

*Lemma 1:* Let $c(y)$ denote the shortest path length from $y$ to $o$. Then, for any measurable $h : \mathbb{R}^+ \longmapsto \mathbb{R}^+$, it holds that

$$\mathbb{E}\, h(C^*) = \lambda_\ell\, \mathbb{E} \int_{L_H^*} h(c(y))\, \nu_1(dy)\,. \quad (2)$$

Eq. (2) follows from Neveu's exchange formula for stationary marked point processes, see [8].

An important fact is that $\mathbb{E}h(C^*)$ does not depend on $X_L$ and its linear intensity $\lambda'_\ell$. We can rewrite (2) as

$$\mathbb{E}\, h(C^*) = \lambda_\ell\, \mathbb{E} \sum_{i=1}^{M} \int_{c(A_i)}^{c(B_i)} h(c(u))\, du\,, \quad (3)$$

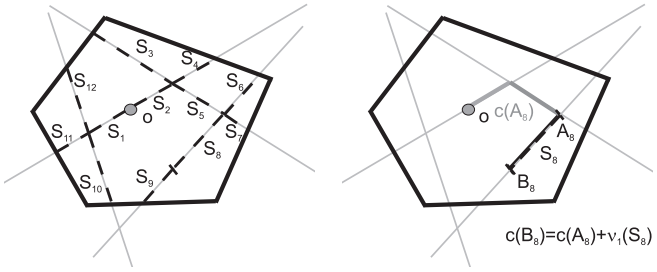



Fig. 6. $L_H^*$ split into segments $S_1, \ldots, S_M$ (left) and single segment with distance peak $B_8$ (right)

where the segment system $L_H^*$ is divided into line segments $S_1, \ldots, S_M$ with endpoints $A_1, B_1, \ldots, A_M, B_M$ such that $L_H^* = \bigcup_{i=1}^M S_i$ and $\nu_1(S_i \cap S_j) = 0$ for $i \neq j$, where $c(A_i) < c(B_i) = c(A_i) + \nu_1(S_i)$, see Fig. 6. Some segments of $L_H^*$ are split in this way at so–called distance peaks. A point $z$ on $L_H^*$ is called distance peak if there are two different shortest paths with the same length from $z$ to $o$. With the notation introduced above we can derive a representation for the probability density $f_{C^*}$ of $C^*$ which can be used in order to estimate this density.

*Corollary 1:* The density $f_{C^*}$ of the typical shortest path length $C^*$ is given by $f_{C^*}(0) = 2\lambda_\ell$ and

$$f_{C^*}(x) = \begin{cases} \lambda_\ell\, \mathbb{E} \sum_{i=1}^{M} \mathbb{1}_{[c(A_i), c(B_i))}(x) \text{ if } x \geq 0\,, \\ 0 \quad \text{otherwise}\,. \end{cases} \quad (4)$$

Formula (4) easily follows from (3), see [14].

**Estimation of the density of shortest path length.** In order to construct an estimator $\widehat{f}_{C^*}(x)$ for $f_{C^*}(x)$, we can use eq. (4). We are especially interested in the estimation of $f_{C^*}$ from synthetic data obtained by simulations. The concept is then to simulate the typical serving zone $\Xi_H^*$ together with the (typical) line segment system $L_H^*$ in $\Xi_H^*$. The shortest path lengths $c(A_i)$ and $c(B_i)$ to $o$ from all endpoints $A_1, B_1, \ldots, A_M, B_M$ in the cell are calculated using Dijkstra's algorithm. This procedure is then repeated $n$ times, so we obtain for each $j = 1, \ldots, n$ the shortest path lengths $c(A_1^{(j)}), c(B_1^{(j)}), \ldots, c(A_{M_j}^{(j)}), c(B_{M_j}^{(j)})$ from the endpoints of the line segments. Finally, we can construct the estimator

$$\widehat{f}_{C^*}(x; n) = \lambda_\ell\, \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{M_j} \mathbb{1}_{[c(A_i^{(j)}), c(B_i^{(j)}))}(x)\,. \quad (5)$$

Note that the estimated function $\widehat{f}_{C^*}(x; n)$ is a step function. For each pair $c(A_i^{(j)}), c(B_i^{(j)})$ we add $\lambda_\ell / n$ for all $x \in [c(A_i^{(j)}), c(B_i^{(j)}))$ to $\widehat{f}_{C^*}(x; n)$. Statistical properties of the estimator $\widehat{f}_{C^*}(x; n)$ are summarized in [14].

## IV. Empirical densities and parametric distance distributions

Now we present some numerical results obtained from a simulation study, where the road model $T^{(1)}$ is the edge set of a PDT, PLT and PVT, respectively, with intensity $\gamma$. Note that the considered models are scaling invariant, i.e., for all $\lambda_\ell, \gamma > 0$ with fixed quotient $\kappa = \gamma/\lambda_\ell$ we get the same model

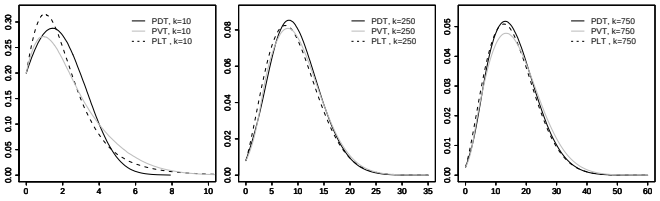



Fig. 7. Density for $\gamma = 1, \kappa = 10$ (left), 250 (middle), 750 (right) and PVT (grey), PDT (black), PLT (broken)

up to a scaling. Thus, for each $\kappa$, it is sufficient to compute numerical results for a single pair $(\gamma, \lambda_\ell)$ with $\gamma/\lambda_\ell = \kappa$. For further pairs $(\widetilde{\gamma}, \widetilde{\lambda}_\ell)$ with $\widetilde{\gamma}/\widetilde{\lambda}_\ell = \kappa$ the corresponding results can then be obtained by a suitable scaling, see [8]. Note that large $\kappa$ yield a dense network inside the serving zones, whereas for small $\kappa$ only a small number of segments intersect each serving zone. In the following we always consider the case that $\gamma = 1$ and values of $\kappa$ between 1 and 2000. With these parameter values we cover realistic network scenarios.

### A. Empirical densities

In order to estimate the density $f_{C^*}$ of the typical shortest path length $C^*$ we simulated $n = 50000$ cells for values of $\kappa$ between 1 and 2000 and computed the estimator $\widehat{f}_{C^*}(x; n)$ as explained in Section III-B for PDT, PLT and PVT. Some empirical densities obtained in this manner are displayed in Fig. 7. One can see that there is a clear difference between the shapes of the densities for small and large $\kappa$ as well as for the different considered models. The differences between the densities for different models seem to decrease for increasing $\kappa$, but it is still noticeable. In [15] it is shown that the typical shortest path length $C^*$ converges in distribution to $\xi X$ as $\kappa \to \infty$, where $X \sim Wei(\lambda\pi, 2)$ and $\xi \geq 1$ is some constant depending on the tessellation model. For PLT we have $\xi = 1$, but $\xi > 1$ for PDT and PVT. So there will always remain some difference between the densities. Based on the estimated densities we computed their means $\mathbb{E}C^*$, variances $\mathrm{Var}C^*$ and coefficients of variation $\mathrm{cv}C^* = 100\sqrt{\mathrm{Var}C^*}/\mathbb{E}C^*$. In Table I the means and cv's are displayed together with the corresponding results for the parametric densities fitted in Section IV-B to the estimated ones.

### B. Fitting of parametric densities

Especially for large $\kappa$ the estimation procedure for the densities $f_{C^*}$ is time-consuming and the means, variances and quantiles have to be calculated numerically. For applications it would be of great benefit if the densities were given as parametric functions, with parameters only depending on $\kappa$ and the type of the underlying road model $T$. Then, for real data, first an optimal tessellation $T$ could be fitted to the road system and in the next step the appropriate density could be chosen. Since the fitting procedure in [7] is done once for all, the distribution of $C^*$ for the given data would then be immediately available and time-consuming simulations could be avoided. Therefore, the aim of this section is the construction of a whole library of parametric distance distributions for PDT, PLT and PVT as road models and a large range of $\kappa$.

However, first we have to choose an appropriate parametric family of densities $\{f(x;\theta), \theta = (\theta_1, \ldots, \theta_k) \in \Theta\}$, where $\Theta \subset \mathbb{R}^k$ for some $k \geq 1$. In [15] it was shown that $C^*$ converges in distribution to the parametric limit distributions $Wei(\lambda\pi/\xi^2, 2)$ and $Exp(2\lambda_\ell)$ for $\kappa \to \infty$ and $\kappa \to 0$, respectively, where $\xi \geq 1$ is some constant depending on the road model $T$. So it is reasonable to choose a parametric family which contains the exponential- and Weibull distribution as limiting cases. Furthermore, the family $\{f(x;\theta), \theta = (\theta_1, \ldots, \theta_k) \in \Theta\}$ should possess the following properties.

1) The dimension $k$ of $\theta = (\theta_1, \ldots, \theta_k)$ is small.
2) The parametric density $f(x;\theta)$ fits well for PDT, PLT and PVT and for a large range of $\kappa$, especially with respect to expectation and variance.
3) For each $\theta \in \Theta$, it holds that $f(0;\theta) = 2\lambda_\ell = 2/\kappa$.
4) The densities of $Wei(\alpha, 2), \alpha > 0$ and $Exp(\lambda), \lambda > 0$ are contained in $\{f(x;\theta), \theta \in \Theta\}$ as limiting cases.

It is not easy to choose a family of densities which fulfills all these conditions. The limit distributions $Wei(\lambda\pi/\xi^2, 2)$ and $Exp(2\lambda_\ell)$ are both special cases of a $Wei(\alpha, \beta)$-distribution. Since condition 3 can not be fulfilled in general by the $Wei(\alpha, \beta)$-distributions, we shift their densities to the left and truncate them at zero such that condition 3 is fulfilled. In this way we get as one possible type of candidates the truncated Weibull distribution with density

$$f(x;\alpha,\beta) = C\Big(x+\big(\frac{2}{\alpha\beta\kappa}\big)^{\frac{1}{\beta-1}}\Big)^{\beta-1} e^{-\alpha\left(x+\left(\frac{2}{\alpha\beta\kappa}\right)^{\frac{1}{\beta-1}}\right)^{\beta}} \qquad (6)$$

for $x \geq 0$, where $C = \alpha\beta \exp((\alpha^{-1}(2/(\beta\kappa))^{\beta})^{1/(\beta-1)})$. This density has two parameters. Another candidate is a mixture $p\, f_1(x) + (1-p) f_2(x), p \in (0,1)$ of the densities $f_1$ of $Exp(\lambda)$ and $f_2$ of $Wei(\alpha, \beta), \beta > 1$. Again, condition 3 should be fulfilled, so we get

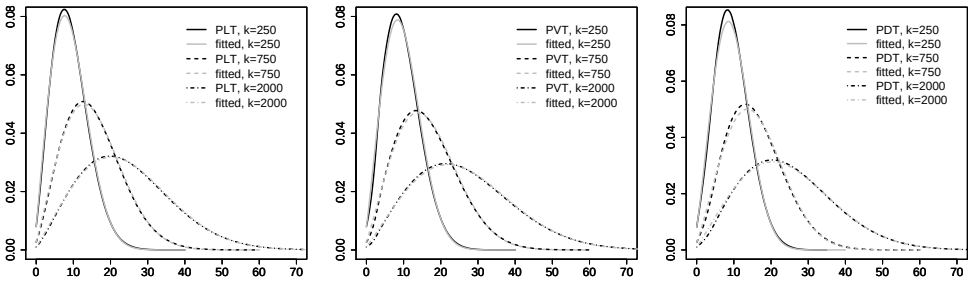
$$f(x;\alpha,\beta,\lambda) = 2\kappa^{-1} e^{-\lambda x} + (1 - 2(\lambda\kappa)^{-1})\alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}, \qquad (7)$$

which has three parameters. We used Matlab to perform a weighted least squares fit of these parametric densities to the data $(\widehat{f}_{C^*}(x_1), \ldots, \widehat{f}_{C^*}(x_n))$ obtained from the empirical densities for a vector $(x_1, \ldots, x_n)$ with equidistant components. As weights we chose the reciprocals $1/\widehat{f}_{C^*}(x_1), \ldots, 1/\widehat{f}_{C^*}(x_n)$ in order to get a better fit at the tails of the densities. Then the optical fit of the densities is worse than without weighting, but the means and variances fit much better. Both regarded types of parametric densities fit optically quite well for all models and a large range of $\kappa$. If we look at the expectations and variances of the fitted truncated Weibull distribution compared to the empirical ones, we can see that they match almost perfectly for all models and a large range of $\kappa$, see Table I. Moreover, it can be seen that for the mixture of the exponential and Weibull distributions the expectations fit quite good, but the variances differ clearly from the variances obtained from the empirical densities. The reason might be that the exponential term in the mixture dominates the tails of this distribution yielding too large variances. So the truncated Weibull distribution was chosen for the library. Some estimated densities together with the fitted ones are displayed in Fig. 8. We calculated the

| $\kappa$ | $\mathbb{E}C^*$ | | | cv$C^*$ | | | $\mathbb{E}C^*$ | | | cv$C^*$ | | | $\mathbb{E}C^*$ | | | cv$C^*$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PVT | trunc. | mix | PVT | trunc. | mix | PLT | trunc. | mix | PLT | trunc. | mix | PDT | trunc. | mix | PDT | trunc. | mix |
| 5 | 1.397 | 1.391 | 1.387 | 73.5 | 73.7 | 83.7 | 1.510 | 1.450 | 1.511 | 95.4 | 93.8 | 120.8 | 1.744 | 1.723 | 1.712 | 92.5 | 91.8 | 80.5 |
| 10 | 2.054 | 2.055 | 2.059 | 63.5 | 63.3 | 80.3 | 2.181 | 2.111 | 2.175 | 83.9 | 78.0 | 103.6 | 2.367 | 2.378 | 2.373 | 76.6 | 76.1 | 77.9 |
| 50 | 4.552 | 4.511 | 4.540 | 52.0 | 52.9 | 69.6 | 4.505 | 4.469 | 4.450 | 60.2 | 59.7 | 70.8 | 4.780 | 4.757 | 4.768 | 54.1 | 54.6 | 69.0 |
| 500 | 13.46 | 13.49 | 13.51 | 50.9 | 50.6 | 61.2 | 12.69 | 12.71 | 12.72 | 51.8 | 51.4 | 57.5 | 13.05 | 13.06 | 13.08 | 49.3 | 49.1 | 57.7 |
| 1000 | 18.89 | 18.86 | 18.87 | 50.7 | 50.7 | 56.0 | 17.68 | 17.60 | 17.61 | 51.2 | 50.8 | 55.7 | 17.89 | 17.95 | 17.97 | 49.9 | 49.4 | 55.5 |
| 2000 | 26.34 | 26.34 | 26.36 | 51.0 | 50.9 | 54.7 | 24.24 | 24.30 | 24.31 | 51.1 | 50.7 | 54.3 | 24.72 | 24.64 | 24.85 | 50.4 | 50.8 | 54.0 |

TABLE I

MEAN AND CV OF $C^*$ TOGETHER WITH THESE VALUES FOR FITTED TRUNCATED WEIBULL AND MIXED EXPONENTIAL-WEIBULL DISTRIBUTION




(a) PLT (b) PVT (c) PDT

Fig. 8. Density for $\gamma = 1, \kappa = 250, 750, 2000$ and: (a) PLT, (b) PVT, (c) PDT with fitted truncated Weibull distribution

parameters $\alpha$ and $\beta$ for all three considered models and a large range of $\kappa$, where functions $\alpha(\kappa), \beta(\kappa)$ depending on $\kappa$ were fitted to the estimated parameters for $\gamma = 1$. So, for all three model types, the distribution of $C^*$ is now directly available up to a scaling via

$$f_{type}(x;\kappa) = f(x;\alpha_{type}(\kappa), \beta_{type}(\kappa))\,. \tag{8}$$

## V. APPLICATION TO REAL NETWORK ANALYSIS

In [2] we have shown that real road systems can be replaced by best fitted models as support for nodes by successfully comparing the histograms of shortest path connections in the two settings of nodes: randomly located on the real road system or on the best fitted model. This validated the first building block of the SSLM. In the present paper, we go further by comparing real distance distributions obtained from data bases for fixed access networks in dense urban areas with the fitted parametric distributions that we developed above in the framework of the SSLM model.

### A. Dealing with real data sets

The first step is to extract from the whole complexity of the database (being the result of a long history of modifications and changes) a synthetic view of the network not taking into account marginal situations. Again, the philosophy of stochastic models is to provide a global vision of a very complex situation that involves a huge number of equipments of various types, where one has to extract the main features of the current network state from the amount of real data. The difficulty is to think "stochastic geometry" while analyzing data so that they can be matched by the SSLM building blocks. Equipment or connections are not to be sorted only according to their usual names or obvious functions, but more by their geographical range and connection principles. Let us briefly review these steps that have been done for the area of Paris.

**Geometrical support and node location.** In this paper, we shall consider only three simple tessellation models (PVT, PDT, PLT) that provide a sufficiently good description of the road system. The fitting procedure applied to the road data, restricted to a possible choice from these simple tessellations, proposes a PVT model of intensity $\gamma_{PVT} = 18km^{-1}$. The theoretical values of the following four characteristics of the fitted PVT model are very close to the measured ones:

| | crossings | quarters | street segments | total length km | required input |
|---|---|---|---|---|---|
| real data | 15462 | 10613 | 26056 | 2146 | database |
| model | 17692 | 8846 | 26537 | 1931 | 2 parameters |

All the network equipments referred in the databases to a common address are considered as geographical sites for node equipments located along the streets.

**Structure in sub-networks.** Data analysis shows that the total area $A$ covered by the network can be divided in a set of non-overlapping (macro-) zones on which distinct two-levels sub-networks are deployed. Each of these zones contains only one highest WCS node. This division process of the total area $A$ is then repeated a second and a third time, producing more and more disjoint two-level sub-networks, deployed in smaller and smaller (micro-) areas, still non-overlapping, where the union of them covers $A$. The sets of zones arising at each of these three steps show some variability in shape and size, but the ranges of their average areas are well distinct.

This structure has been visualized by plotting the locations of the nodes using their geographical coordinates, where it must be kept in mind that the "frontiers" of the zones become more and more fuzzy from steps 1 to 3. Still, it remains possible to assume that the LLC considered at each step are logically connected to their closest HLC in flybird sense. Apart from $A$ and the road-system characteristics, the only parameter required to describe a typical serving zone at a given step $i$

is the number of HLC-nodes $n_i$ from which the parameter $\kappa_i$ can be computed; $i = 1, 2, 3$.

**Histograms for the connection lengths.** The histograms for the connection lengths are recovered from the databases with respect to the above division into sub-networks. They concern the overall lengths of the telecommunication lines from any LLC node to its respective HLC node, but no information is available about their real physical path.

The SSLM model proposes a global analysis in the most simple (idealized) setting that is compatible with the conclusion inferred from the analysis of data bases: homogeneous random street system which captures the most important structural properties of the real street system, homogeneous random Poisson repartition of nodes along the streets, logical connection to the closest HLC node in straight-line sense, physical connection as shortest path along the street system. All the geometrical objects are idealized points or line segments, whereas real cables do possess a reality that induces physical constraints as a minimum curvature radius as well as necessary cable-joining in chambers, for example. Then the theoretical prediction for connection lengths from SSLM should naturally be lower than the observed ones.

### *B. Results*

In order to describe the global statistical behavior of point-to-point connections in the real access network of Paris, structured in three distinct two-level sub-networks, the SSLM only requires the knowledge of six global parameters:

- the size $|A|$ of network area $A$ and the type and intensity of the road system it contains, i.e. $(type, \gamma)$,
- the numbers $n_1$, $n_2$ and $n_3$ of higher-level nodes for each sub-network, where for steps 2 and 3, a HLC node may be part of the set of corresponding LLC nodes.

In all three examples considered below, the same family of parametric distance distributions was used, where eq. (8) was used for the same model type (PVT), but for different values $\kappa_1$, $\kappa_2$ and $\kappa_3$ directly deduced from the above parameters. The whole model is proposed under the form of an Excel sheet with functions encoded in Visual Basic, thus giving an instantaneous answer and also avoiding the need of specialized computing languages not easily accessible for managers.

Note that a few features of the real network were not mentioned above in order to simplify the presentation. Nevertheless, the figures presented here integrate all the reality of the network including some connections that arise in a non-purely hierarchical architecture as well as exclusion areas. This does not change the number of required parameters and analytical formulas and only asks for some right combination.

The typical serving zone at step 1 is relatively large compared to the scale of the road system since it contains in the average up to 200 quarters. This corresponds to a large value of $\kappa_1 \simeq 1000$ as illustrated in the left part of Fig. 9. The theoretical probability density fits extremely well with the histogram computed from the real network data (see Fig. 9, right), in shape as well as length scale, where $D_1$ denotes a reference distance to enable comparisons between results at different scales.

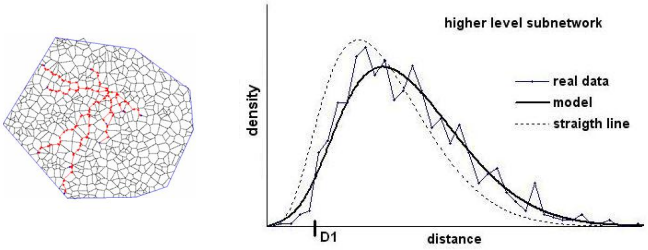



Fig. 9. Typical serving zone of the larger scale sub-network ($\kappa_1 \simeq 1000$); theoretical distance densities compared with real data, showing that the assumption of physical connections as straight-line shortest paths is incorrect.

The typical serving zone at step 2 is much smaller since it contains roughly only 10 quarters in the average, where $\kappa_2 \simeq 35$ (see Fig. 10, left). The networks results concern all the point-to-point connections involved at that scale, i.e., they are a mixture of links rising from various types of lower-level sub-network nodes provided that they are all connected to the same HLC node. In this case, SSLM underestimates the reality by a factor $c = 1.15$ on the averaged connection length. Thus, the theoretical distance density $f$ has been rescaled in order to recover the measured average, i.e. $f(x) \to f(x/c)/c$, where it is assumed that the underestimation is due to an effect proportional to the distance. Model and real data are still close (see Fig. 10, right), whereas the distance scale (with unit distance $D_2$) is very different from the previous one (given by $D_1$). The typical serving zone at step 3 is comparable in

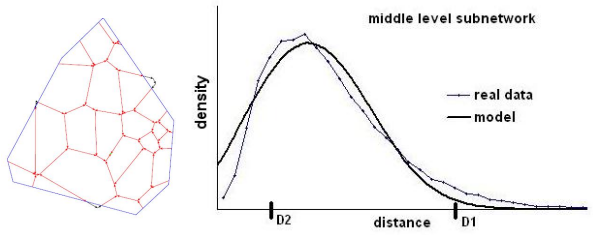



Fig. 10. Typical serving zone of the middle-scale sub-network ($\kappa_2 \simeq 35$) and (rescaled) theoretical connection-length density compared with real data.

size with the area covered by a group of buildings; it contains only a few street segments, where $\kappa_3 \simeq 4$ (see Fig. 11, left). The theoretical average connection length underestimates the measured one by some value $l > 0$. This can be explained by the fact that real lengths include a part that goes from the building to the end customer, not taken into account by SSLM. Since this distance simply adds to the predicted length, the theoretical distance density $f$ has been shifted in order to recover the measured average, i.e. $f(x) \to f(x - l)$ for $x > l$. This leads to a very good fit between model and real data (see Fig. 11, right). Note that $l$ thus determined from the comparison of model and real data is compatible with the value currently admitted in the network community. Again, the distance scale is rather different from the previous ones (given by $D_1$ and $D_2$, respectively).

We also remark that Fig. 9 illustrates the importance to take into account the underlying road system. The theoretical distance distribution computed under the assumption that the

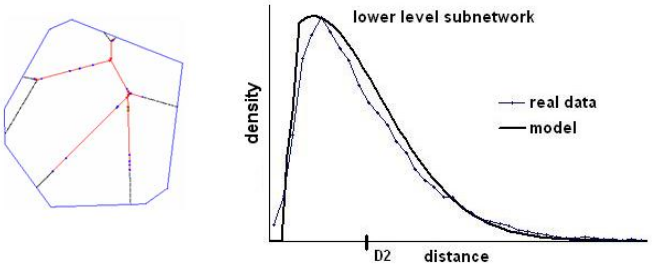



Fig. 11. Typical serving zone of the lower-scale sub-network ($\kappa_3 \simeq 4$) and (shifted) theoretical connection-length density compared with real data.

physical connections are straight lines (dashed graph) does not fit to the histogram for real data and clearly underestimates the real average connection length. Classical studies introduce an estimated correction factor for the average length, but SSLM explicitly relates it to the geometry of the underlying road system. Single WCS sub-networks have also been satisfactory addressed by the SSLM model provided that one deals with a sufficiently large or diversified set of connections.

## VI. Conclusions and Outlook

We have shown that SSLM based on methods from stochastic geometry is quite a good model to address huge access networks in a global way. SSLM relies on a small number of parameters while taking into account the morphological features of the underlying road system that is a key issue of fixed access networks. The explicit separation of three model components (geometrical support, location of network nodes, connection topology) offers the possibility to consider a wide range of situations. Thus, SSLM fulfills the requirements of the telecommunication operator who needs simple and global tools to cope with major changes and new technologies, to reconstruct networks of competing operators or foreign countries and to negotiate with regulatory authorities.

Probability theory and stochastic processes provide analytical formulas that are formally equivalent to statistical results computed on some reconstruction scheme, considering all the possibilities compatible with the geometrical setting. The hard work of listing and analyzing all those possibilities, with appropriate coefficients depending on their probability of occurrence is done by considering stochastic integrals, Palm probabilities, and the notion of the typical cell. Any reconstruction by hand, e.g. regarding the locations of nodes, would produce a configuration already considered.

The implementation of the SSLM model has helped to improve the synthetic view of existing networks that are very complicated due to their long history, where the main focus of the present study lies in that it quantitatively validates the use of random road models and shortest paths connections. Furthermore, the SSLM model can be used as a sound basis for the construction of random cabling trees (point to multipoint connections) for which no detailed data bases are available. Note that the explicit description of the underlying road system cannot be avoided in this case since it is impossible to build trees from straight-line connections that never share a common path. Also, the morphology of the road system, especially the number of streets incoming in crossings, greatly influences the capacities of the cables to be installed.

The analytical formulas for distance distributions depend on the type of the road system and its parameter(s). They are obtained as explained in Section IV using simulation procedures for corresponding typical Cox–Voronoï cells, which can be extended to iterated tessellations constructed from PDT, PVT and PLT. On the other hand, the fitting procedure also applies to road systems at a regional or territory scale. Thus, we are able to adapt the SSLM model to regional scales and to address such applications as in the process of development of optical fibre networks in a given region. In particular, one can directly relate a specified network architecture to the percentage of customer lines that will exceed a given length threshold. This may be used to decide on which network scenarios are possible and thus greatly restrict the domain of parameters to be analyzed by detailed optimization procedures before implementation.

## Acknowledgment

This research was supported by Orange Labs through Research agreement 46143714 to the Institute of Stochastics at Ulm University.

## References

[1] C. Gloaguen, P. Coupé, R. Maier, and V. Schmidt, "Stochastic modelling of urban access networks," in *Proc. 10th Internat. Telecommun. Network Strategy Planning Symp.* Munich: VDE, Berlin, June 2002, pp. 99–104.

[2] C. Gloaguen, H. Schmidt, R. Thiedmann, J. P. Lanquetin, and V. Schmidt, "Comparison of network trees in deterministic and random settings using different connection rules," in *Proceedings of the 3rd Workshop on Spatial Stochastic Models for Wireless Networks (SpaSWiN)*, Limassol, April 2007.

[3] F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev, "Géométrie aléatoire et architecture de réseaux," *Annales des Télécommunication*, vol. 51, pp. 158–179, 1996.

[4] F. Baccelli and S. Zuyev, "Poisson-Voronoi spanning trees with applications to the optimization of communication networks," *Operations Research*, vol. 47, pp. 619–631, 1996.

[5] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, 2nd ed. Chichester: J. Wiley & Sons, 1995.

[6] R. Maier and V. Schmidt, "Stationary iterated random tessellations," *Advances in Applied Probability*, vol. 35, pp. 337–353, 2003.

[7] C. Gloaguen, F. Fleischer, H. Schmidt, and V. Schmidt, "Fitting of stochastic telecommunication network models, via distance measures and Monte-Carlo tests," *Telecommunication Systems*, vol. 31, pp. 353–377, 2006.

[8] C. Gloaguen, F. Fleischer, H. Schmidt, and V. Schmidt, "Analysis of shortest paths and subscriber line lengths in telecommunication access networks," *Networks and Spatial Economics (to appear)*, 2008.

[9] F. Voss, C. Gloaguen, F. Fleischer, and V. Schmidt, "Distributional properties of typical Euclidean distances in telecommunication networks involving road systems," *Preprint*, 2009.

[10] D. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, 2nd ed. New York: Springer, 2003/2008, vol. I and II.

[11] R. Schneider and W. Weil, *Stochastic and Integral Geometry*. Berlin: Springer, 2008.

[12] F. Fleischer, C. Gloaguen, V. Schmidt, and F. Voss, "Simulation of the typical Poisson-Voronoi-Cox-Voronoi cell," *Journal of Statistical Computation and Simulation (to appear)*, 2009.

[13] C. Gloaguen, F. Fleischer, H. Schmidt, and V. Schmidt, "Simulation of typical Cox-Voronoi cells, with a special regard to implementation tests," *Mathematical Methods of Operations Research*, vol. 62, pp. 357–373, 2005.

[14] F. Voss, C. Gloaguen, F. Fleischer, and V. Schmidt, "Density estimation of typical shortest path lengths in telecommunication networks," *Preprint*, 2009.

[15] F. Voss, C. Gloaguen, and V. Schmidt, "Scaling limits for shortest path lengths along the edges of stationary tessellations," *Preprint*, 2009.