A New Service Overlays Dimensioning Approach Based on Stochastic Geometry

Anthony Busson[a] Daniel Kofman, Jean-Louis Rougier[b]

[a]*Institut d'Electronique Fondamental, Université Paris Sud Orsay, F91405 Orsay Cedex*

[b]*Ecole Nationale Supérieure des Télécommunications, 46 rue Barrault, F-75634 Paris Cedex 13*

Abstract

The global provisioning of some types of evolved services over the Internet requires the deployment of servers; the approach is usually called service overlays. An optimal design of the number of servers and of their location is critical in order to minimize the deployment cost while providing the quality of service expected by users.

In this paper, we propose an original modelling, analysis and optimization approach based on stochastic geometry to deal with this design issue. Two particular cases of service overlays are emphasized: servers deployed for content delivery (i.e. to deploy a Content Delivery Network or CDN) and servers deployed for reliable multicast communications.

In our approach, the network topology, the subscribers' location and the servers' location are represented by point processes in the plane. Point process theory and Palm Calculus are used in order to obtain closed form expressions for the expected value of realistic cost functions. These closed forms allow us to evaluate the optimal cost of a future deployment.

Key words: network planning, stochastic geometry, content delivery network, multicast transport

1 Introduction

Service overlays became very popular for the deployment of some types of evolved services over the Internet. In this paper, we propose a new modelling and analysis method, based on stochastic geometry, to address several topology optimization problems related to the cost-effective deployment of service

overlays. An optimal design of the number of servers and of their location is critical in order to minimize the deployment cost while providing the quality of service expected by users.

Two particular cases of service overlays are emphasized: servers deployed for content delivery (i.e. to deploy a Content Delivery Network or CDN) and servers deployed for reliable multicast communications.

Our work is motivated by the difficulty of solving optimal server deployment problems using traditional methods, due to the fact that we usually deal with large systems made up of a huge number of entities (e.g. routers, links, servers, users, etc.). As a consequence, it is not possible to model the detailed behaviour of each node, as is done when using a classical queuing theory approach. A macroscopic modelling approach is therefore required. Tractability issues are also critical, as the complexity of the problem itself leads to NP complete problems when using classical optimization approaches.

In the stochastic-geometry based approach we propose in this paper, the network topology, the subscribers' location and the servers' location are represented by point processes in the plane. Point process theory and Palm Calculus are used in order to obtain closed form expressions for the expected value of key quantities, such as the expected value of realistic cost functions. The expressions obtained depend only on the point process intensities; this is why we say that we propose a macroscopic modelling approach. In particular, the number of parameters involved does not depend on the size of the system. The main advantages of this approach can be summarized as follows:

- A very complex system can be represented with a very limited number of parameters, i.e. point process intensities.
- Networks of arbitrary size can be considered
- Closed form expressions of objective cost functions can be computed and optimal structures easily deduced.

The counterpart of the simplicity is that, since the results are the average of well-chosen objective functions (i.e. average over all point process realizations), the approach is useful for obtaining global dimensioning rules, but cannot be used for optimal placement of individual equipment.

Our analysis is thus aimed at global technico-economical studies, strategic planning and cost analysis for deploying service overlays.

Classical optimization tools are necessary to refine the proposed optima for specific topologies (usually for a local topology for tractability reasons).

Our approach was motivated by former successful applications of stochastic geometry for network optimization. For instance, similar tools have been used

to optimize multicast routing protocols (5) and mobile networks (6; 7; 8; 9). Other examples of the use of stochastic geometry in modelling telecommunication systems can be found in (10). The application of stochastic geometry to the problems considered in this paper have already been addressed by the authors in several international conferences' papers (12; 13; 14). However, the emphasis in these previous papers was on networking aspects and mathematics analysis was, for the most part, omitted. In the present paper, the mathematical model and analysis is detailed and we refer to the papers cited above for details on numerical results.

The paper is organized as follows: in Section 2 we introduce the Content Delivery Network concept and present our Stochastic Geometry model and the cost functions we want to optimize (2.2, 2.3 and 2.4). The closed form expressions of the cost functions are obtained in 2.5 and a numerical example is studied in 2.6.

In Section 3, we apply a similar approach to the planning of an overlay for reliable multicast communications. The model and cost functions are given in 3.3. The mathematical analysis is presented in 3.4 and a numerical example is studied in 3.5.

Finally, conclusions are drawn and some future works suggested in Section 4.

2 Optimal Content Delivery Networks Deployment

2.1 Context and Motivations

The aim of Content Delivery Networks (CDN) is to enable content providers to improve the quality of delivery of their content to potential customers/users. The basic idea to enhance the quality of content delivery is to bring the content closer to potential consumers. This implies the deployment of a large number of servers and the corresponding cost can be prohibitive for content providers. A CDN allows for virtualization of a server infrastructure. A content provider will have the feeling of having a proprietary cache system, with the required number of servers, but in reality he is sharing the server infrastructure with other content providers. CDNs allow a content provider to replicate content (web pages, pictures, audio and video files, media, etc.) in servers called *surrogates* deployed over the Internet. Contrary to usual caching, CDN guarantee the content providers (i.e. the customers of the CDN) the presence and consistency of their content in the surrogates. (This is sometimes referred to the "push" approach, as content is placed in advance. In the "pull" approach, content is placed in cache following a user request).

3

When a given user requests a piece of content stored in the CDN, the request is redirected to one of the surrogates, which should be located close to the user in order to provide the required Quality of Service (QoS). Thus, a client may download a copy of a piece of content from the nearby CDN rather than having to fetch it from the server of the content provider, which may be far away. This allows the user to perceive a good QoS in terms of delay and download rate. Moreover, it reduces the load on the network and the content provider servers.

On the one hand, the QoS perceived by users clearly depends on the number and placement of the surrogates. On the other hand, the deployment and management cost of the CDN depends on the number of servers and placement (number of sites on which servers are deployed). Therefore, the main obejective of CDN design is to minimize the cost of the CDN while providing the required QoS to content consumers.

The problem of CDN server placement has already been addressed in related works. Authors using a graph theory approach usually define the problem as finding an optimal placement of servers in a certain set of predefined potential sites. The deployment of a server in a site k has a cost c_k and a user i assigned to this server incurs a cost $b_i.d_{ik}$, where b_i is the demand from the user and d_{ik} is the distance between the user and the server. The optimal placement is the solution which minimizes the global cost. There are several variants of this problem. In the *facility location problem* the number of servers is not predefined and is unlimited (26). In the *K-median problem*, the number of servers to be deployed is limited (at most K) and the cost of a server is nil ($c_k = 0$) (27). The same problems have been considered for servers of limited capacity: servers cannot respond to infinite demand (28; 29). We can also cite (30; 31), where the cost of writing and storing content and the consumed bandwidth are considered in the optimization model.

In all the cited cases, the approach assumes that certain parameters are known in advance. Unfortunately, the exact topology, user location, bandwidth and demand (which is linked to the popularity of the content) are usually unknown, or not available to the designer, and vary in time. Moreover, the static approach leads to very complex problems (NP-hard for the cited formulations), which are difficult to manage when large networks such as the Internet are considered.

Stochastic geometry allowed us to develop a scalable modelling, analysis and optimization approach that encompasses the main characteristics of the system with just a few parameters. This approach provides results that are easy interpret and suited to economic analysis and strategic planning. The modelling approach is presented in the following subsection.

We model the topology of the Internet by a Poisson point process distributed in the plane, where each point represents an Autonomous System [1] (AS) or a cluster of users as defined in (32).

Note that the representation proposed here is somewhat arbitrary. We have chosen a Poisson point process for its simplicity. It should also be noted at this point that, no assumption has been made on the connectivity between the ASes (the points of the Poisson process). In the following sections, we shall see that the distance between any points is mapped to a realistic "distance" function (related to the number of ASes in the paths between clients and surrogates).

Potential consumers of the content available in the CDN may lie in each AS or cluster (we shall use *site* to be more generic). Similarly, some *sites* may contain a surrogate. We model these properties of the sites by marking the Poisson process with two marks.

The first mark indicates the presence of a surrogate in a site. These marks form a set of independent random Bernoulli variables of parameter p (in other words, for each point of N, the mark takes the value 1 with probability p and 0 with probability $1 - p$), with $0 < p < 1$, independent of the Poisson process and of the second marks. The mark has the value 1 if there is a surrogate in the site and 0 otherwise.

The second mark represents user demand from the site, such as the number of requests per time unit. These also form a set of independent identically-distributed random variables, independent of the Poisson process.

In this paragraph, we introduce the notation and the mathematic formalism.

- Let N be a Poisson point process in $I\!\!R^2$ of intensity λ ($\lambda \in I\!\!R^+$), whose points represent the sites, with independent marking (s, m), where s indicates the presence of a surrogate in the corresponding site ($s = 1$ if there is a surrogate in the corresponding site and $s = 0$ otherwise), with $P(s = 1) = p$, and m is a random variable representing the demand from the corresponding site. Independent marking means here that the marks are independent of each other and independent of the point process N. In particular, the s marks are i.i.d. and so are the m marks.

---

[1] An Autonomous System is a set of routers under a single technical administration which has a single routing policy as seen by the others Autonomous Systems. An IP external routing protocol (today BGP4) allows for the interconnection of the various ASes to build up the Internet

- Let N_0 and N_1 be the two Poisson point processes obtained from the thinning of N. The points of the process N_0 (resp. N_1) correspond to the points of N where the marks s are 0 (resp. 1). Giving the independence hypothesis, N_0 and N_1 are two independent Poisson processes.
- For a point $x \in N$, we denote by m_x the mark representing the demand from the site located at x.

In order to keep finite quantities without loss of generality, the point processes N, N_0 and N_1 are observed in the open ball of radius R and centred at the origin of the plane. This ball is denoted by $B(o, R)$, where R can be seen as the "radius" of the Internet.

We suppose that the users will fetch documents from the nearest site with regard to the Euclidian distance (the nearest site is the local one when a surrogate has been deployed in the site; in this case the distance equals 0).

From the previous hypothesis, a surrogate deployed in a site represented by point x of N_1 is responsible of all the clients of all the sites represented by the points of N which are in the Voronoï cell of x. For a point y of N_1, the Voronoï cell of y denoted by $V_y(N_1)$ is the set of points of $\mathbb{R}^2$ which are closer to y than any other point of N_1. More precisely:

$$V_y(N_1) = \{x \in \mathbb{R}^2 | \|x - y\| \leq \|x - y_j\| \forall y_j \neq y, y_j \in supp(N_1)\},$$

where $\|x - y\|$ is the Euclidian distance between x and y. The set of the Voronoï cells of a Poisson point process forms a tesselation of the plane (3). A Poisson point process and its Voronoï tesselation are represented in Figure 1(a). Figure 1(b) shows the process N_0 and the process N_1 with its tessellation (the crosses represent the points of N_1 and the dots the points of N_0).
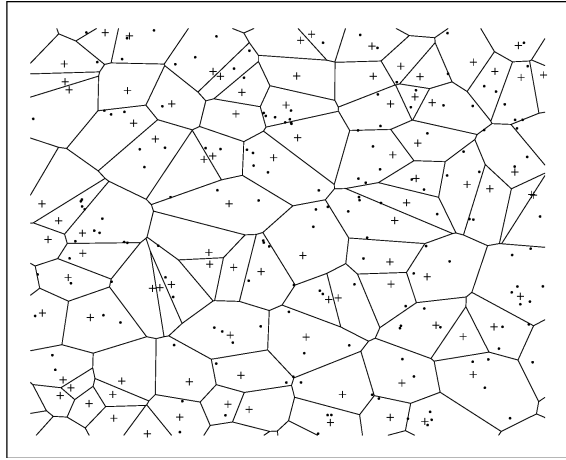
2.3   Cost function

A tradeoff must be found between the cost of the surrogates and the quality of service perceived by users. Therefore, our cost function has two components.

- We allocate each surrogate a cost α which encompasses the server, installation and maintenance costs.
- We allocate each site a metric representing the "annoyance" perceived by users.

For the users of a given site, annoyance is measured as a function of the distance between the site and the nearest surrogate. This function, that we denote by f, is an increasing function of the distance: the further the surrogate
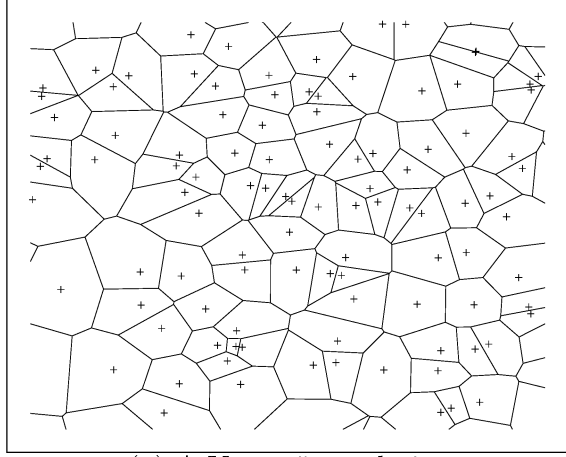
(a) A Voronoï tesselation

(b) The N_0 process and the N_1 process with its tesselation

Fig. 1. Example of a Voronoï tesselation

is, the greater the annoyance perceived by the user. We consider the following cost function:

$$Cost = \mathbb{E}\left[\sum_{x \in N_1 \cap B(o,R)} \alpha + \sum_{y \in N \cap B(o,R)} m_y f(z_y)\right], \qquad (1)$$

where z_y is the distance between the point y of N and the nearest point of N_1.

2.4   Extended model

The previous model and cost function may easily be extended to consider a system with different classes of servers. In fact, content elements hosted by

7

the CDN are not usually placed on all servers. Therefore, different classes of surrogates, hosting different content elements, can be introduced. Considering heterogeneous servers grouped into classes (a class being composed of all the servers of a given type) allows us to evaluate different content placement policies among the classes, which can depend, for example, on content popularity.

We assume that a document is stored in all the surrogates of a given class. A class of surrogates is thus in charge of a set of content elements. We consider n ($n \in I\!N^*$) different classes of surrogates. The location of the class i surrogates ($1 \geq i \geq n$) is obtained by an independent thinning of parameter p_i ($0 < p_i < 1$) of the process N. Let us denote by N_i this process, which has intensity $\lambda_i = \lambda.p_i$. Let q_i be the probability that a randomly chosen content element is stored in the class i surrogates and let α_i be the cost of a surrogate of this class. As we will see in the numerical example, the parameters q_i are used to model the content placement policy. For this extended model, we define the following objective cost function :

$$Cost = \sum_{i=1}^{n} \mathbb{E}\left[\sum_{x \in N_i \cap B(o,R)} \alpha_i\right] + \sum_{i=1}^{n} q_i \mathbb{E}\left[\sum_{y \in N \cap B(o,R)} m_y f(z_y^i)\right], \qquad (2)$$

where z_y^i is the Euclidian distance between y, a point of N, and the closest point of N_i. From the analytical expression of this cost function, we will show, for example, that the cost does not depend on the distribution of m, but solely on its average. Another interesting result we will present is that we can easily deduce the optimal proportion of sites where class i surrogates must be deployed i.e. the optimal values for p_i when the other parameters are fixed.

2.5   Computation

In this section, we present the complete computation of the cost function for the case of a single class of servers. The computation of the cost function for the heterogeneous case is analogue, so here we just present the final result. In books (1) and (2), the reader can find an introduction to Stochastic Geometry and a presentation of the different formulae used in this paper.

The cost function

$$Cost = \mathbb{E}\left[\sum_{x \in N_1 \cap B(o,R)} \alpha\right] + \mathbb{E}\left[\sum_{y \in N \cap B(o,R)} m_y f(z_y)\right],$$

can be rewritten as

$$Cost = \mathbb{E}\left[\int_{B(o,R)} \alpha N_1(dx)\right] + \mathbb{E}\left[\int_{B(o,R)} \int_{V_x(N_1)} m_y f\left(\|y - x\|\right) N(dy) N_1(dx)\right].$$

$$(3)$$

Since α is a constant, we have, for the first term of the cost :

$$\mathbb{E}\left[\int_{B(o,R)} \alpha N_1(dx)\right] = \lambda p \pi R^2 \alpha \tag{4}$$

For the second term of the cost, since $N = N_0 + N_1$, we can write:

$$\mathbb{E}\left[\int_{B(o,R)} \int_{V_x(N_1)} m_y f\left(\|y - x\|\right) N(dy) N_1(dx)\right] = \mathbb{E}\left[\int_{B(o,R)} m_y f(0) N_1(dy)\right]$$

$$+ \mathbb{E}\left[\int_{B(o,R)} \int_{V_x(N_1)} m_y f\left(\|y - x\|\right) N_0(dy) N_1(dx)\right] \tag{5}$$

The marks $(m_y)_{y \in \mathbb{R}^2}$ are i.i.d., thus, if we note $\overline{m} = \mathbb{E}\left[m_y\right]$, we have, for the first term:

$$\mathbb{E}\left[\int_{B(o,R)} m_y f(0) N_1(dy)\right] = \lambda p \pi R^2 \overline{m} f(0) \tag{6}$$

For the second term, we have:

$$\mathbb{E}\left[\int_{B(o,R)} \int_{V_x(N_1)} m_y f\left(\|y - x\|\right) N_0(dy) N_1(dx)\right] \tag{7}$$

$$= \lambda p \int_{B(o,R)} \mathbb{E}\left[\int_{V_0(N_1)} m_y f\left(\|y\|\right) N_0(dy)\right] dx \tag{8}$$

$$= \lambda(1 - p) \int_{B(o,R)} \mathbb{E}^o_{N_0}\left[m_{Z_0} f\left(\|Z_0\|\right)\right] dx \tag{9}$$

$$= \lambda(1 - p) \pi R^2 \overline{m} \mathbb{E}^o_{N_0}\left[f\left(\|Z_0\|\right)\right] \tag{10}$$

$$\tag{11}$$

where: $\mathbb{E}^o_{N_0}[.]$ is the expectation under the Palm measure with respect to the point process N_0 and Z_0 is the point of N_1 which is closest to the origin of the plane. The first equality is obtained by Campbell's formula; the second is obtained by the application of Neveu's exchange formula.

The point processes N_1 and N_0 being independent, we have:

$$\mathbb{E}_{N_0}^o \left[f(\|Z_0\|) \right] = \mathbb{E} \left[f(\|Z_0\|) \right] \tag{12}$$

$$= \mathbb{E} \left[\int_{I\!R^2} f(\|z\|) \, \mathbb{1}_{\{N_1(B(o,\|z\|))=0\}} N_1(dz) \right] \tag{13}$$

$$= \lambda p \int_{I\!R^2} f(\|z\|) \, \mathbb{P}_{N_1}^o (N_1(B(-z,\|z\|)) = 0) \, dz \tag{14}$$

$$= \lambda p \int_{I\!R^2} f(\|z\|) \exp\{-\lambda p \pi \|z\|^2\} dz \tag{15}$$

$$= \lambda p 2\pi \int_0^{+\infty} r f(r) \exp\{-\lambda p \pi r^2\} dr \tag{16}$$

$$\tag{17}$$

These equalities are obtained from Campbell's Formula and from Slyvniak's theorem. We note that the expression of the cost does not depend on the distribution of the marks $(m_x)_{x \in I\!R^2}$.

In the case where $f(0) = 0$ (case considered in the following numerical example), we obtain :

$$Cost = \lambda \pi R^2 \left[p\alpha + (1-p)\overline{m}\lambda p 2\pi \int_0^{+\infty} r f(r) \exp\{-\lambda p \pi r^2\} dr \right]. \tag{18}$$

In the case of heterogeneous servers and with the notations presented in Section 2.4, an analogue analysis leads to the following result:

$$Cost = \lambda \pi R^2 \sum_{i=1}^n \left[p_i \alpha_i + (1-p_i)\overline{m}\lambda q_i p_i 2\pi \int_0^{+\infty} r f(r) \exp\{-\lambda p_i \pi r^2\} dr \right]. \tag{19}$$

2.6 Example

In this subsection we present some numerical results for the case of heterogeneous servers with two classes of surrogates. In order to highlight the breadth of the approach, we consider the whole Internet as the network. Here a site represents here an Autonomous System (AS).

We set $\lambda \pi R^2 = 10\,000$; $10\,000$ being the approximate number of ASes in the Internet.

The annoyance metric that we consider is the number of ASes crossed to fetch the document. Indeed, bottlenecks on the Internet occur most often in the interconnecting links between ASes. Relating the number of AS hops in a path between a user and a surrogate with the Euclidian distance separating the user from the surrogate, the function f is obtained from BGP routing

10

tables (see (33) for more details on BGP). The computed function f is a step function (more details are given in (14)).

For class 1 (resp. 2) surrogates, we allocate a constant cost γ_1 (resp. γ_2) for the installation and maintenance costs, and we allocate a second cost $\beta_1 l_1$ (resp. $\beta_2 l_2$) where β_1 (resp. β_2) is the storage cost of a document and l_1 (resp. l_2) is the number of hosted documents. The server costs are then given by $\alpha_1 = \gamma_1 + \beta_1 l_1$ and $\alpha_2 = \gamma_2 + \beta_2 l_2$ We assume that the most popular documents are stored on class 1 surrogates. The popularity of a web document is the ratio between the number of requests for this object and the total number of requests for all documents. For the Web, it has been shown that if we order the documents by popularity, document 1 being the most popular, the ordering distribution follows a Zipf-law of parameter 1 (34; 35). More precisely, let $K > 0$ be the total number of documents stored by the CDN and let v_i be the probability that a document is the i^{st} ($0 < i \leq K$) most popular document, we have $v_i = \frac{C}{i}$, where C is a non-negative constant. Therefore, the probability that a user requests a document stored on class 1 is $q_1 = \sum_{i=1}^{l_1} v_i$, where l_1 is the number of documents hosted on class 1 surrogates. Unfortunately, the optimal intensity values (λp_1 and λp_2) which minimize the global cost cannot be found analytically. However, a numerical computation which is scalable since its complexity depends only on the number of involved processes and not on the size of the systems, allows us to evaluate these optimal values when they exist. In Figure 2(a), we have plotted the parameters p_1 and p_2, which represent the optimal proportion of ASes where surrogates should be deployed, as a function of the mean number of requests per site ($\overline{m}$). The values of the parameters were chosen as follows: $\gamma_1 = 300$, $\gamma_2 = 600$, $\beta_1 = 0.5$, $\beta_2 = 0.5$, $l_1 = 150$ and $l_2 = 850$. The aim is to have a large number of low storage capacity (cheap, class 1) servers that will store the most popular content close to users and a few high-storage-capacity (expensive, class 2) servers that will store the rest of the content. Due to the nature of Zipf-law, the $l_1 = 150$ documents stored on class 1 servers will correspond approximately to 75% of the requests. As expected, we observed that in the optimal configuration, p_1 is significantly larger than p_2. Therefore, class 1 surrogates will respond to the majority of requests (75%) with a good quality of service. The fewer, more expensive class 2 surrogates will handle 25% of the requests and host 85% of the documents.

In Figure 2(b), we have plotted the cost as a function of the number of documents stored on class 1 servers, for the optimal values of p_1 and p_2, for $\overline{m} = 300$ and for a total number of 1000 documents. It appears that there is an optimal storage policy which minimizes the global cost. We see here that the modelling approach has the benefit of allowing us to identify the optimal storage policy, which has a significant impact on the cost of the CDN.

With this example, we have shown that the approach can take into account elaborate considerations in the structure of the CDN (like storage policies

and document popularity) and allows us to deduce optimal values for the parameters characterizing the CDN infrastructure.
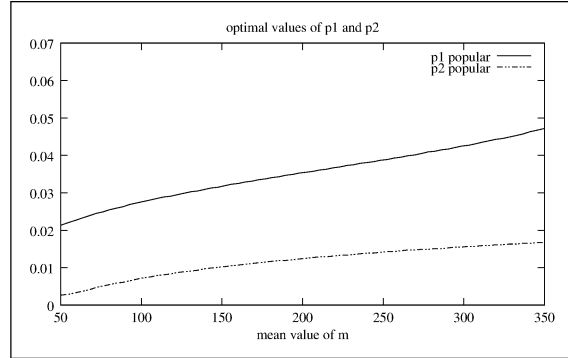
(a) Optimal values of p_1 and p_2 : Proportion of ASes where surrogates should be deployed
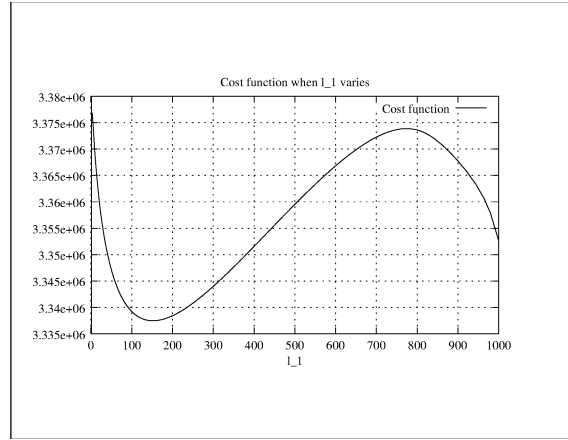

(b) Cost function when l_1 varies

Fig. 2. Optimal parameters for surrogate planning and storage policy

## 3   Optimization of overlays for reliable Multicast transport

3.1   Technical Presentation: multicast

A multicast session is composed of several sources and several receivers. Each packet sent by a source has to be received by all the receivers. Without loss of generality concerning the reliable multicast issues we deal with in this paper, we consider here the case of one source and several receivers.

The multicast session is identified by a multicast address. The source sets the destination address of each packet of the session (that has to reach all the
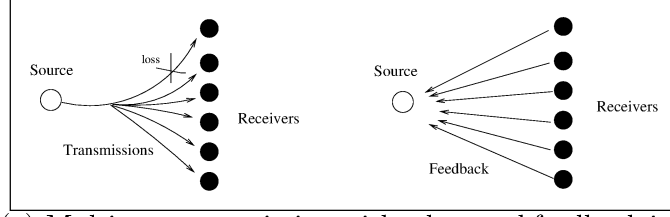
(a) Multicast transmission with a loss and feedback implosion
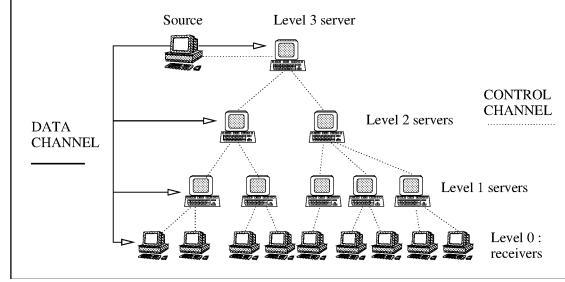
(b) Control tree

Fig. 3. Multicast transmission, feedback implosion and the control tree

destinations) with this multicast address. We suppose here that the network implements a multicast routing protocol. Therefore, the source sends each packet only once and the network is in charge of splitting the packet in such a way that it will reach all the destinations. We suppose that the multicast routing is based on a shortest path tree approach (although the analysis can be applied to other approaches, such as to the Core Based Tree approach). In this approach, based on knowledge of multicast membership, the routers build a shortest path tree from the router to which the source is attached to all routers to which destination users are attached.

Multicast routing may significantly reduce the network load as each packet is transmitted at most once on each link of the network. There are different protocols and mechanisms that allow the routers to build multicast trees in a distributed manner. Examples of protocols used in the Internet and other IP networks are DVMRP, PIM and PIM-SSM which are described in (18) and (15; 16; 17). In the following we assume that a multicast routing protocol is used in the network, without specifying which one. As explained in the following subsection, we are interested here in a protocol of a higher layer, which is used to increase the reliability of the multicast transmission and does not require the use of a particular multicast protocol.

3.2 Technical Presentation: reliable multicast transport protocols

IP multicast is now mature and widely available. (The interest in IP multicast increased significantly during the past few years with the generalization of

13

telecommunications services like "triple play", which includes the distribution of TV programs). However, the lack of reliable transport protocols limits its application for many services. The Internet is actually an unreliable "best effort" network, meaning that it does its best but does not guarantee reliability. In particular, packets can be lost in the network and the IP protocol is not able to recover these losses.

For reliable point to point communications, TCP (4) protocol is run at source and destination in order to recover from potential losses (the protocol implements the detection of losses and the corresponding retransmissions). The approach is based on the fact that the destination sends acknowledgement packets to the source to indicate which packets have been correctly received. In the multicast context this approach is not scalable. Indeed, if all the receivers indicated which packets had been received, the source would be flooded with feedback as represented in Figure 3(a).

Moreover, even if a packet has not been received by just one of the receivers, as represented in Figure 3(a), in a classical approach the packet would be retransmitted to all destinations.

One of the most efficient approaches to solve these scalability problems consists in the deployment of servers in the network to take care of reliability, instead of dealing with the reliability issue end to end. Such an architecture is described in (19) and (20). The deployed servers are in charge of detecting losses and retransmitting packets for a subset of receivers, thus relieving the source of most of its reliability work.

When a large number of receivers are involved in the multicast session, a hierarchy of reliability servers can be used in order to keep the architecture scalable: a server of the lowest level (level 1) is responsible for a limited number of receivers, a server of level 2 is responsible for a set of level 1 servers, and so on.

In such a way, each server is in charge of a limited number of receivers or servers of a lower level.

The architecture distinguishes two different channels: the data channel used by the source to transfer data to receivers and servers (the multicast tree), and the control channel used by the servers to recover lost packets. It is important to highlight that in the data channel the servers are considered as leaves of the

routing tree (the impact of the servers in the cost of the data channel is therefore nil or negligible). Therefore, we are only interested in the dimensioning of the control channel. Figure 3(b) represents the control tree as the hierarchy of servers used to implement reliable multicast transport. The servers infrastructure will support several multicast sessions from, usually, different applications

Little research has been done on the dimensioning aspects of this architecture, and the optimal configuration of the control tree remains an open issue. For a given multicast environment (multicast applications and users), important questions remais unanswered: How many levels are required? How many servers should be deployed at each level?

The multicast architecture has to support several multicast sessions to be cost efficient. In addition, the number and location of the receivers change in time and may differ between sessions. . This means that classical optimization approaches, which consider static topologies, such as in (22) and (23) and simulation techniques, which can only consider one topology at a time, are both inadequate.

Here, we propose a new approach to design the control channel. We use point processes in the plane to represent the location of both the receivers and the servers. The efficiency of the architecture is captured by a wide family of cost functions whose optimization gives the optimal intensities of the point processes describing the placement of reliable multicast transport servers.

The model and the computations for the proposed cost functions are presented in the following subsections.

3.3   Model

3.3.1   Modelling users and servers location

Two different point processes are used to represent, respectively, the location of users and the location of the servers.

Users are gathered in local regions: Autonomous Systems, companies, cities, etc. In order to model these concentration regions, we represent the location of subscribers by the points of a Neyman-Scott process (presented with more details in (1)). The process is built from a stationary Poisson point process Π_p of intensity λ_p which represents the location of concentration regions. This Poisson point process is called the parent point process. To build the Neyman-

15

Scott process, each point $x \in I\!R^2$ of the parent point process is replaced by a cluster S_x^0 which is itself a point process with a finite number of points. The clusters $(S_x^0)_{x \in \Pi_p}$ are independent and identically distributed. The resulting Neyman-Scott process, denoted here by S_0 and defined as

$$S_0 = \bigcup_{x \in \Pi_p} S_x^0, \tag{1}$$

is stationary.

As explained at the beginning of this section, there may be several hierarchical levels of servers. We denote by H the number of levels, with $H \geq 1$.

We define a point process for each level. The servers of level 1 are responsible for receivers, the servers of level 2 are responsible for servers of level 1 and so on.

Level 1 to $H-2$ servers are modelled as Neyman-Scott processes with parent point process Π_p. Π_p is the same parent point process as for S_0. Indeed, we suppose that the servers of level i for $1 \leq i \leq H-2$ are located in the same concentration regions as the receivers. Each server of level i of a given concentration region is responsible for a subset of children of level $i-1$ of this concentration region.

Let S_i be the point process representing the location of the servers of level i for $1 \leq i \leq H-2$. The clusters composing S_i are i.i.d. The cluster of the different levels generated by the same point x of Π_p are also independent. In other words, the different point processes S_i are generated by the same point process Π_p, but the clusters at the points of Π_p are independent. If we denote by S_x^i $x \in \Pi_p$, the cluster of S_i at x, we have:

$$S_i = \bigcup_{x \in \Pi_p} S_x^i, \text{ for } i = 1, 2, .., H-2. \tag{2}$$

All of these point processes are stationary. The intensity of the point process of level i ($i = 0, 1, .., H-2$) is $\lambda_i = \lambda_p \mathbb{E}_{\Pi_p}^0 [S_0^i (I\!R^2)]$.

A server of level $H-1$ may be responsible for servers (of level $H-2$) of several concentration regions. The servers of level $H-1$ are modelled by a Poisson point process S_{H-1} of intensity λ_{H-1}. This process is independent of the other point processes. All the points in the cluster S_x^{H-2} ($x \in \Pi_p$) are bound to the point of S_{H-1} which is the closest to x. A point y of S_{H-1} is thus responsible for the points of the clusters which have their generating points (the point

of Π_p they replaced in the construction of the process) in the Voronoï cell $V_y(S_{H-1})$ (see Subsection 2.2 for the definition of the Voronoï cells generated by a point process).

The highest level (level H) is a single point located at the origin of the plane. This server is supposed to be very close to the source. Therefore, we assume that the source is located at the same location as this server (at the origin). The hierarchy of point processes in a Voronoï cell of S_{H-1}, for $H = 3$, is represented in Figure 4(a).

All of these processes are observed in the open ball $B(o, R)$ of radius R centered in the origin of the plane. In the numerical evaluation, R represents the radius of the Internet.

3.3.2 Cost Functions

The servers are in charge of managing the Acknowledgements from participants (receivers or servers, depending on the considered level) and retransmitting lost data.

A large number of children (participants or servers) bound to a given server will degrade the performance of the multicast session for two main reasons: the server will have to manage a large number of acknowledgements and the probability of performing a repair (retransmission) will be high (a drawback given that the retransmission will be sent to all children). The number of repairs carried out by a given packet is a measure of its performance.

As a consequence, we define our cost function as a linear combination of the mean number of retransmissions at the different levels and of the server cost (acquisition, installation and maintenance costs). More precisely, we define:

$$\mathbb{C}ost = \sum_{i=1}^{H} \mathbb{C}_i, \qquad (3)$$

with

$$\mathbb{C}_i = \mathbb{E}\left[\sum_{x \in S_i \bigcap B(o,R)} \left(a_i \overline{m}_x^i + \alpha_i \right) \right] = \lambda_i \int_{B(o,R)} \overline{m}_x^i dx + \lambda_i \alpha_i \pi R^2 \qquad (4)$$

where $\overline{m}_x^i$ is the mean number of retransmissions from a server of level i located at x. Only active servers are taken into account. Active level 1 servers are those

17

with receivers bound to them and active level i servers are those with active servers of level $i - 1$ bound to them. Finally, α_i represents the cost of a level i server and a_i is a constant $(a_i \in I\!\!R^+)$ for each i.

The equality is obtained by applying the Campbell formula (1). The use of the Campbell formula allows us to reduce the computation of the cost function to the mean number of retransmissions for a point located at x. This is very interesting. Indeed, the random variable describing the number of retransmissions may have been considered as a mark bound to the point, but due to the non-homogeneity of its distribution introduced by the loss model (Subsection 3.3.3), the marked point process would not have been stationary and the computations would not have been possible.

The choice of the a_i is very important, it allows us to favour the retransmissions from one level rather than another (the impact on the network of retransmissions at different levels are usually different). For instance, the cost of a server at each level can be weighted by the mean distance between the servers and their children.
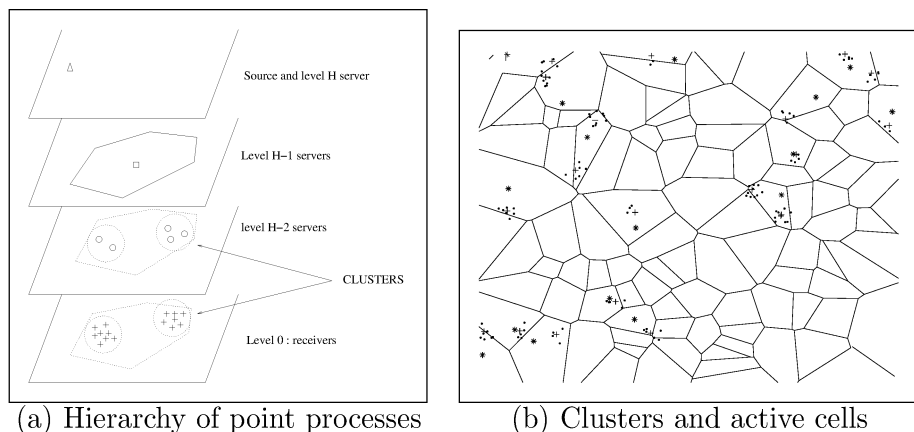

(a) Hierarchy of point processes   (b) Clusters and active cells

Fig. 4. The hierarchy of point processes and the active point of level $H - 1$

3.3.3 Loss model

In this section, we present the loss model we used, which was inspired by the analysis of losses measured during a multicast session (presented in (21)). There are two kinds of transmissions: the original transmission from the source to all the receivers and servers (on the data channel), and the retransmissions from the servers to their children on the control channel. Therefore, we distinguish two loss models: the first one describes the loss distribution for the transmission on the data channel and the second one describes the loss distribution of retransmissions on the control channel. The models are based on a

function $p : I\!R \to [0,1]$.

3.3.3.1 Loss model on the data channel The original transmission is sent from the source to all receivers and servers. This transmission reaches a point of level $H-1$ located at x with probability $1 - p(\|x\|)$ independently of the other points. The probability of reaching any cluster within the Voronoï cell generated by x is given as $1 - p(\|x\|)$. This assumption is a simplification which allows us to avoid introducing a heterogeneous distribution within a Voronoï cell. We assume that the losses are independent between clusters. For a packet that reaches a cluster, there is a probability q $(0 < q < 1)$ of the packet being lost within the cluster for a given receiver or server, independently of the losses of other receivers and servers. This model is depicted in Figure 5(a).

3.3.3.2 Loss model on the control channel The level H server is a single server located at the origin. It retransmits packets to level $H-1$ servers. A retransmission reaches a level $H-1$ server located at x with probability $1 - p(\|x\|)$ independently of the other points. For a retransmission from level $H-1$, the probability of the packet being lost before reaching the cluster is $p(l)$ where l is the mean distance between a cluster (the point of Π_p which generates the cluster) and the nearest point of S_{H-1}. We find $l = \frac{1}{2\sqrt{\lambda_{H-1}}}$. As for the original transmission, within a cluster the probability of losing a packet for a receiver or a server is q, independently of the other points. This loss model is represented in Figure 5(b).

For a retransmission from a server of level $i < H-1$ (within a cluster), the probability of losing the retransmission for a given child is q independently of the other children.

3.4 Computation

We shall now present the computation of the mean number of retransmissions at the different levels, which is required, according to equation 4, for the computation of the cost function. First, we compute the mean number of retransmissions between a server and a set of children for a simpler model.

3.4.1 Preliminary results

In this paragraph we compute, for a simple model, the number of retransmissions between a server and a number of receivers described by a discrete
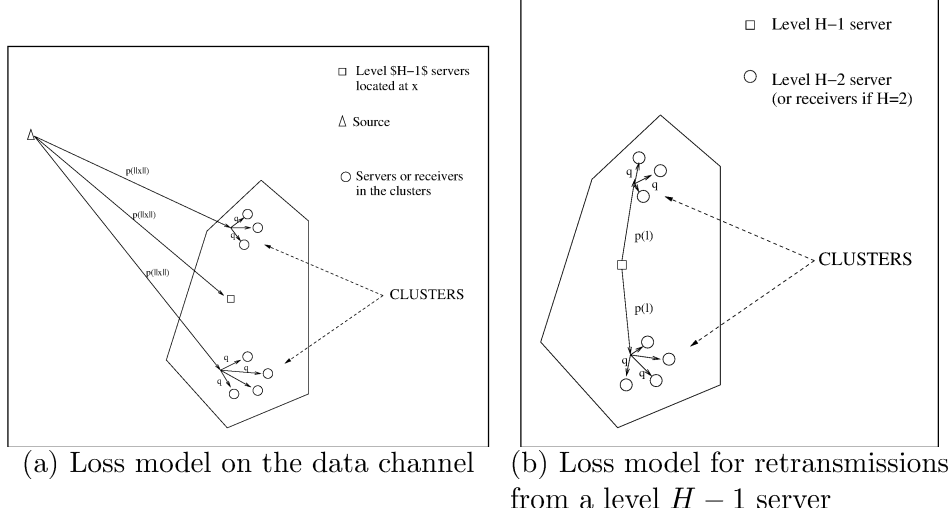
(a) Loss model on the data channel   (b) Loss model for retransmissions from a level $H-1$ server

Fig. 5. Loss model

random variable. The results will be used in the computations for the more realistic loss model presented earlier.

Let N be a discrete r.v. and q the probability of losing a packet between the server and a receiver. We assume that the losses are independent between the receivers. If, after a transmission, there is a receiver that has not received the packet, it is retransmitted to all receivers. We compute the mean total number of retransmissions m. We do not take into account the first transmission (only retransmissions are considered). We have,

$$\mathbb{E}[m] = \sum_{k \geq 0} \left(1 - \mathbb{P}\left(m \leq k\right)\right) = \sum_{k \geq 0} \left(1 - \sum_{i=0}^{+\infty} \mathbb{P}\left(m \leq k | N = i\right) \mathbb{P}\left(N = i\right)\right) \quad (5)$$

$$= \sum_{k \geq 0} \left(1 - \sum_{i=0}^{+\infty} \left(1 - q^{k+1}\right)^i \mathbb{P}\left(N = i\right)\right) = \sum_{k > 0} \left(1 - g_N(1 - q^k)\right), \quad (6)$$

where g_N is the generating function of N.

3.4.2   Mean number of retransmissions from the higher level

Here, the higher level (H) is a single point located at the origin of the plane. The level $H-1$ is a Poisson point process. We have to consider only active points (as explained in 3.3.2), i.e points of level $H-1$ with active servers in their Voronoï cell. If there is a cluster in a Voronoï cell, there is inevitably an active server of level $H-2$. The process of active points of level $H-1$ is stationary but it is not a Poisson point process. Figure 4(b) shows such a process with $H = 3$; the stars represent the active points of level 2, the crosses represent the centres of the clusters and the points represent the receivers. We

20

denote the intensity of the *active point process* by λ^*_{H-1}, we have

$$\lambda^*_{H-1} = \lambda_{H-1}.\mathbb{P}^0_{S_{H-1}}\left(S_{H-1}\left(V_0\left(\Pi_p\right)\right) > 0\right) \qquad (7)$$

The second factor is the probability that the Voronoï cell of a typical point of S_{H-1} contains at least one cluster (a point of Π_p). $\mathbb{P}^0_{S_{H-1}}$ is the Palm probability with respect to the process S_{H-1}. We have,

$$\lambda^*_{H-1} = \lambda_{H-1}\left(1 - \mathbb{E}^0_{S_{H-1}}\left[\exp\left\{-\lambda_{H-1}|V_0\left(\Pi_p\right)|\right\}\right]\right) \qquad (8)$$

where $|V_0\left(\Pi_p\right)|$ is the size of the Voronoï cell. To the best of our knowledge, the size distribution of a typical Voronoï cell is not known. A very accurate approximation of this distribution is given in (3). Based on these results, we use a gamma law of parameter $\alpha = 3.571$, $\beta = 3.571\lambda_{H-1}$.

We shall consider the active point process as a Poisson point process of intensity λ^*_{H-1}. Simulation has shown that this approximation is not penalizing for the considered quantities. Moreover, in practical cases, the intensity of the point process S_{H-1} is much smaller than the one of the point process which generates the clusters (so almost all the cells contain clusters and therefore the approximation is very accurate).

According to our notation, the mean number of retransmissions between the point at the origin and the set of active points of the Poisson point process S_{H-1} is denoted by $\overline{m}^H_0$. We have $\overline{m}^H_0 = \sum_{k>0}\mathbb{P}\left(m^H_0 \geq k\right)$. Since we have assumed that the active point process is still Poisson, the set of points which have not received the information at the k^{st} retransmission (k retransmissions plus the original transmission) is still a Poisson point process of intensity $\Lambda_k(B) = \lambda^*_{H_1}\int_{B(o,R)}\left(p\left(\|x\|\right)\right)^{k+1}dx$. Thus,

$$\mathbb{P}\left(m^H_0 < k\right) = \exp(-\Lambda_k(B)), \qquad (9)$$

and,

$$\overline{m}^H_0 = \lambda^*_{H-1}\sum_{k\geq 0}\left(1 - \exp(-\Lambda_k(B))\right). \qquad (10)$$

3.4.3   Mean number of retransmissions between the levels $H-1$ and $H-2$

Let us compute $\overline{m}^{H-1}_x$, the mean number of retransmissions between the point of level $H-1$ located at x and the set of points of the clusters of level $H-2$ having their generating point in the Voronoï cell of x. We have :

$$\overline{m}^{H-1}_x = \sum_{k\geq 0}\mathbb{E}[m^{H-1}_x \mid \Pi_p(V_x(S_{H-1})) = k]\mathbb{P}^0_{S_{H-1}}\left(\Pi_p\left(V_0\left(S_{H-1}\right)\right) = k\right) \qquad (11)$$

By definition, $\mathbb{E}[m_x^{H-1}|\,\Pi_p(V_x(S_{H-1})) = k]$ is the mean number of retransmissions from the point x given that there are k clusters in the Voronoï cell of x. It depends solely on k random variables which describe the number of points within each of the k clusters. Therefore, this quantity does not depend on the geographical component of the model. The distribution of the number of clusters in a cell does not depend on the location of the cell since the process is stationary.

Thus, we have,

$$\mathbb{E}\left[m_x^{H-1}|\,\Pi_p\left(V_x(S_{H-1})\right) = k\right] = \sum_{i>0}\left(1 - \mathbb{P}\left(m_x^{H-1} < i|\Pi_p\left(V_x(S_{H-1})\right) = k\right)\right)$$

$$(12)$$

Since the number of points in the different clusters are independent and the event $\{m_x^{H-1} < i\}$ holds if and only if the number of retransmissions is less than i in each of all the clusters of the Voronoï cell, we can write :

$$\mathbb{P}\left(m_x^{H-1} < i|\Pi_p(V_x(S_{H-1})) = k\right) = [B(i,x)]^k \qquad (13)$$

where $B(i,x)$ is the probability that the number of retransmissions in a cluster located at x is less than i.

The following expression of $B(i,x)$ is derived from the preliminary results (Subsection 3.4.1) and by conditioning in the number of retransmissions (resp. the transmission) from the level $H-1$ server (resp. from the source) which reached the cluster.

$$B(i,x) = \sum_{u=0}^{i}\left(C_i^u p(l)^{i-u}\left(1 - p(l)\right)^u\left[p\left(\|x\|\right)g_{S*}\left(1 - q^u\right) + \left(1 - p\left(\|x\|\right)\right)g_{S*}\left(1 - q^{u+1}\right)\right]\right)$$

$$(14)$$

where g_{S*} is the generating function of the number of active points in a cluster of level $H-2$.

For the second factor in 11 we have:

$$\mathbb{P}_{S_{H-1}}^0\left(\Pi_p\left(V_0\left(S_{H-1}\right)\right) = k\right) = \mathbb{E}_{S_{H-1}}^0\left[\frac{V_0\left(S_{H-1}\right)^k}{k!}\exp\left(-\lambda_p V_0\left(S_{H-1}\right)\right)\right] \quad (15)$$

Using the same statistical evaluation of the size of a Voronoï cell used in subsection 3.4.2, we find:

$$\mathbb{P}^0_{S_{H-1}}\left(\Pi_p\left(V_0\left(S_{H-1}\right)\right) = k\right) = \frac{(\lambda_{H-1})^k (\alpha\lambda_p)^\alpha}{k!(\lambda_{H-1} + \alpha\lambda_p)^{k+\alpha}} \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)} \qquad (16)$$

Finally, after a simple but heavy computation that we shall not explicit to lighten the presentation, we have:

$$\overline{m}_x^{H-1} = \sum_{i>0}\left[1 - \left(\frac{\alpha\lambda_p}{\lambda_{H-1} + \alpha\lambda_p - B(i,x)\lambda_{H-1}}\right)^\alpha\right] \qquad (17)$$

3.4.4   Mean number of retransmissions in the inferior levels (within the clusters)

Finally, let us compute the last terms of 4 (for $i = 1$ to $i = H - 2$). For the original transmission from the source to a given destination, the loss probability depends on the centre of the Voronoï cell of S_{H-1} to which the cluster of the destination is associated (Figure 5(a)). Therefore, to calculate the total number of retransmissions at level i for $i = 1..H - 2$, we have to take into account for each cluster the Voronoï cell to which it belongs. We denote by τ_x^i the mean number of retransmissions performed by a point of level i within a cluster for which the generating point is in the cell centred at x. We have:

$$\lambda_i \int_{B(o,R)} \overline{m}_x^i dx = \mathbb{E}\left[\int_{B(o,R)} \int_{V_x(S_{H-1})} \int_{I\!R^2} \tau_x^i S_y^i(dz) \Pi_p(dy) S_{H-1}(dx)\right] \qquad (18)$$

$$= \lambda_{H-1} \int_{B(o,R)} \tau_x^i \mathbb{E}^0_{S_{H-1}}\left[\int_{V_0(S_{H-1})} \int_{I\!R^2} S_y^i(dz) \Pi_p(dy)\right] dx \qquad (19)$$

$$= \lambda_p \mathbb{E}^0_{S_{H-1}}\left[S_0^i(I\!R^2)\right] \int_{B(o,R)} \tau_x^i dx \qquad (20)$$

$$\qquad (21)$$

The second equality is obtained from the Campbell formula and the last one from the exchange formula (11). Since the process S_{H-1} and S_i for $i < H - 1$ are independent, we have $\mathbb{E}^0_{S_{H-1}}\left[S_0^i(I\!R^2)\right] = \mathbb{E}\left[S_0^i(I\!R^2)\right]$.

τ_x^i is obtained from the preliminary results in subsection 3.4.1. We express τ_x^i assuming that the original transmission has reached the cluster.

$$\tau_x^i = (1 - p(\|x\|)) \sum_{k>0}\left(1 - g_{N_{i-1}}(1 - q^{k+1})\right) + p\left(\|x\|\right) \sum_{k>0}\left(1 - g_{N_{i-1}}(1 - q^k)\right) \qquad (22)$$

Here, N_{i-1} is the random variable describing the number of active points of

the inferior level $(i-1)$ bound to a typical point of level i.

3.5 Example

In this example, we present the optimal tree configuration for the case of two levels of servers. We ignore the cost of the servers ($\alpha_i = 0, \forall i$), which leads to a simple expression of the cost function allowing us to highlight the efficiency of the approach through the analysis of a reduce number of parameters.

Here, we have four point processes:

- Π_p of intensity λ_p which represents the center of the clusters (the concentration regions or the ASes),
- S_0 which represents the receivers within the clusters,
- S_1 which represents the servers within the clusters,
- S_2 which consists in a single point located at the origin representing the upper level server.

From 4, the cost function is given by:

$$\mathbb{C}ost = \mathbb{E}\left[\sum_{x \in S_1 \cap B(o,R)} a_0 \overline{m_x^1}\right] + \mathbb{E}[a_1 \overline{m_x^2}] \tag{23}$$

where a_0 (resp. a_1) is the mean number of receivers (resp. servers) reached by a retransmission from a server of level 1 (resp. 2) and $\overline{m_x^1}$ (resp. $\overline{m_x^2}$) is the mean number of retransmissions performed by a server of level 1 (resp. 2).

When the number of level 1 servers increases, we find that

- the mean number of receivers bound to these servers decreases and so does the probability of carrying out a retransmission from any of these servers to its receivers.
- the probability of a retransmission from level 2 servers to level 1 servers increases.

Therefore, the chosen cost function has a global minimum with regard to the number of level 1 servers. The function $p(x)$ has been chosen as a function of the number of concentration regions crossed. Most losses in the Internet are localized at exchange points and rarely within a domain. The function $p(x)$ depends on the probability, denoted by b, of losing the packet between two concentration regions. The exact definition of $p(x)$ is given in (25). The radius R has been chosen equal to 1 and $\pi R^2 \lambda_p = 10\,000$ (10 000 concentration regions are considered, corresponding to the 10 000 ASes of the Internet).

Figure 6 shows the mean number of level 1 servers per cluster which minimizes the cost function when $\nu_0 = \mathbb{E}[S_0^0(I\!\!R^2)]$ (the mean number of receivers per cluster) increases and for three different values of b (we vary the probability of losing the data between the level 2 server and the servers/receivers within the clusters). We observe that the optimal number of servers is close to a straight line with regard to ν_0. It can thus be expressed approximately as $d.\nu_0$ where d is a function of b (the inter-domain loss probability) and q (the intra domain loss probability). This observation can allow us to derive simple engineering rules in order to dimension the number of servers.

As it may be easier to dimension the network with regard to the expected number of receivers per cluster, a simple dimensioning rule may consist in planning the number of servers in such a way that each server is responsible for an average of $\frac{1}{d}$ receivers.
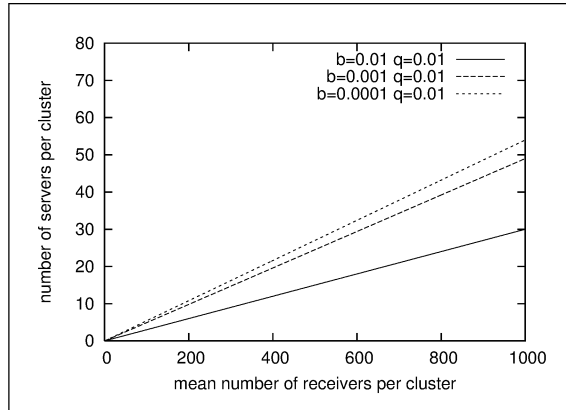

Fig. 6. Optimal number of servers

4   Conclusion

In this paper, we have presented an innovative approach to model service overlays based on Stochastic Geometry.

Although the approach is quite generic, we have concentrated on two particular cases: content delivery networks and reliable multicast transport. In both cases, we have focused on the optimization of the deployment cost under realistic cost functions. Since the distribution over the networks of the users of a given application is not known in advance, Stochastic Geometry proved to be a very well adapted modelling tool. Indeed, the use of point processes allows us to represent the geographical locations of the users and servers in the network. We presented the detailed computation of the cost functions which lead to closed expressions. From this analysis, an a priori evaluation of minimum-cost deployments is done. As we indicated, the approach is quite

25

generic: specific point processes and cost functions have to be designed to model a given service architecture.

References

[1] D.Stoyan, W.Kendall, J.Mecke. *Stochastic Geometry and its Applications.* 2nd Edition, J.Wiley & Sons Chichester, 1995.

[2] D.Daley, D.Vere-Jones. *An Introduction to the Theory of Point Processes.* Springer Series in Statistics, Springer-Verlag New-York, 1988.

[3] A. Okabe *et al. Spatial tesselations.* 2nd Edition, Wiley.

[4] Richard Stevens. *TCP/IP Illustrated, Volume 1: The Protocols.* Addison-Wesley, 1994.

[5] F.Baccelli, D.Kofman, J.L.Rougier.*Self-Organizing Hierarchical Multicast Trees and their Optimization.* IEEE Infocom'99, New-York (USA), March 1999.

[6] F.Baccelli, S.Zuyev. *Poisson-Voronoi Spanning Trees with applications to the Optimization of Communication Networks.* INRIA Research Report No. 3040. Nov.1996.

[7] F. Baccelli, M. Klein, M. Lebourges and S. Zuyev, *Stochastic Geometry and Architecture of Communication Networks*, Telecommunication Systems, 7, pp. 209-227, 1997.

[8] F. Baccelli and B. Blaszczyszyn.On a Coverage Process Ranging from the Boolean Model to the Poisson Voronoi Tessellation, with Applications to Wireless Communications INRIA. Rapport de recherche 4019, Oct. 2000, to appear in Adv. Appl. Prob.

[9] F. Baccelli, B. Blaszczyszyn and F. Tournois.Spatial Averages of Coverage Characteristics in Large CDMA Networks INRIA. Rapport de recherche 4196, June 2001.

[10] Stochastic Geometry: a Tool for Modeling Telecommunication. *http://www.di.ens.fr/~ mistral/sg/*

[11] J. Neveu. *Sur les mesures de Palm de deux processus ponctuels stationnaires.* Z. Wahrsch. verw. Gebiete, 34:199-203, 1976.

[12] A. Busson, J-L Rougier, D. Kofman *Analysis and optimization of hierarchical reliable protocols* ITC 17, Salvador (Brazil), December 2001.

[13] A. Busson, J-L Rougier, D. Kofman *Impact of tree structure on retransmission efficiency for TRACK* NGC 2001, London, November 2001.

[14] A. Busson, D. Kofman, J-L Rougier *Optimization of content delivery networks server placement* ITC 18, Berlin, September 2003.

[15] A. Adams *et al. Protocol Independent Multicast-Dense Mode (PIM-DM): Protocol Specification.* RFC 3973.

[16] D. Estrin *et al. Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification.* RFC 2362.

[17] S. Bhattacharyya *An Overview of Source-Specific Multicast (SSM).* RFC

3569.

[18] D. Waitzman *et al. Distance Vector Multicast Routing Protocol.* RFC 1075.

[19] B. Whetten *et al. TRACK Architecture, A Scalable Real-Time Reliable Multicast Protocol.* Internet draft, IETF, November 2000.

[20] B. Whetten *et al. Reliable Multicast Transport Building Block for TRACK.* IETF working document.

[21] M. Yajnik, J. Kurose, and D. Towsley. *Packet Loss Correlation in the MBone Multicast Network.* IEEE Global Internet Conference 1996.

[22] Levine, B., D. Lavo, and J. J. Garcia-Luna-Aceves, *The Case for Reliable Concurrent Multicasting Using Shared ACK Trees,* In Proceedings of ACM Multimedia (Boston, Massechusettes, USA, Nov. 18-22, 1996). ACM New York, N. Y., pp.365-376.

[23] Christian Maihfer. *Bandwidth Analysis and Simulation of Reliable Multicast Transport Protocols.* Telecommunication Systems 23(1-2): 201-237 (2003).

[24] B. Halabi. *Internet Routing Architectures.* Cisco Press, 1997.

[25] Anthony Busson, Jean-Louis Rougier and Daniel Kofman. *Analysis and optimization of hierarchical reliable transport protocols.* ENST-Research report 2001 D 009.

[26] M. Charikar, S. Guha. *Improved Combinatorial Algorithms for the Facility Location and K-Median Problems.* In Proc. of the 40th Annual IEEE Conference on Foundations of Computer Science, 1999.

[27] L. Qiu, V. Padmanabhan, G. Voelker *On the Placement of Web Server Replicas,* In Proc. of 20 th IEEE INFOCOM, Anchorage, USA, April 2001.

[28] M. Charikar, S. Guha, E. Tardos, and D. B. Shinoys. *A constant-factor approximation algorithm for the K-median problem.* In Proc. of the 31st Annual ACM Symposium on theory of Computing (1999).

[29] F.A. Chudak and D.B. Shmoys. *Improved approximation algorithms for the capacitated facility location problem.* In Proc. of the 10st Annual ACM Symposium on Discrete Algorithms.

[30] Konstantinos Kalpakis, Koustuv Dasgupta, Ouri Wolfson *Optimal Placement of Replicas in Trees with Read, Write, and Storage Costs.* IEEE Transactions on Parallel and Distributed Systems 12(6): 628-637 (2001).

[31] A. Venkataramani, M. Dahlin, and P. Weidmann. *Bandwidth constrained placement in a WAN.* In Proceedings of the 20th International Conference on Distributed Computing Systems, Aug 2001.

[32] B. Krishnamurthy and J. Wang. *On network-aware clustering of web clients.* Proc. of the ACM SIGCOMM'00, 2000.

[33] BGP - the Border Gateway Protocol Advanced Internet Routing Resources *http://www.bgp4.as/*

[34] L. Breslau et al., *Web caching and Zipf like distribution : Evidence and implications,* In procceedings of IEEE Infocom , March 1999, New York.

[35] C. Cunha, A. Bestavros and M. Crovella. *Characteristics of WWW Client*

based traces. Technical Report TR-95-010, Boston University, Computer Science Dept., USA, 1995.