Proc. 15th International Conference on Advanced Computing and Communication (ADCOM2007), Guwahati, India, December 2007

Agent-Based VPN Architecture: A Framework and The Optimal User Connectivity Problem (Static Case)

D. Medhi[1], B.-Y. Choi[1], C. Scoglio[2], S. Song[3†], S. Dispensa[4]
[1]University of Missouri–Kansas City, USA
[2]Kansas State University, USA
[3]Wichita State University, USA
[4]Positive Networks, Inc., USA

***Abstract*—Remote access VPN services have gained significant popularity in recent years. In this paper, we present the basic framework of an agent-based VPN architecture (ABVA) in which organizations do not need to maintain VPN servers. We then consider the problem of the user connectivity decision in an ABVA, which consists of the selection of the best Rendezvous Point (RP) for each user in a heterogeneous access environment. Specifically, we consider the optimal user connectivity problem from the perspective of an ABVA provider. For the static case of known users, we present an optimization formulation. An advantage of our formulation is that it can incorporate users with heterogeneous access technology, along with delay and penalty factors. Through our computational study, we observe that the delay factor is not necessarily dominant in all cases, while the bandwidth access requirement plays a critical factor in an overloaded situation. Finally, we briefly comment on how the results from the static case are useful with regard to the random arrival case.**

I. INTRODUCTION

Virtual Private Network (VPN) services have seen significant growth over the years; they are becoming critical for many organizations for their day-to-day operations. For example, an employee of an organization can use the VPN service to access the intranet of the organization from an off-location through the public Internet by establishing a VPN tunnel. A norm for deploying such VPN services is that the organization installs a VPN server on its premise, which acts as the gateway for accessing intranet services by its users. While this model serves well for moderate to large organizations, this model does not work well for small organizations. For instance, for organizations with a hundred employees or less, the cost of having a VPN server and skilled IT personnel is cost-prohibitive. Furthermore, even for a moderate size organization, maintaining a VPN server in its own location can be a major cost in terms of personnel expertise and day-to-day maintenance issues. This is because of the diversity of their user base, and especially its satellite locations, for which it may not be cost effective to have IT personnel. These organizations would want to have the remote access VPN service for their users if this service can be provided and maintained by a third party in a cost-effective manner. Such a service would be particularly appealing if it is flexibly adaptable to changing organizational and technological requirements.

To allow such an environment, we consider a VPN architectural concept called an *Agent-Based VPN Architecture* (ABVA) that suites the third party offering. In our approach, one or more VPN agents (or brokers) are located outside the organizations that serve as *Rendezvous Points* (RP), where users and their associated organizations meet. Here, the Rendezvous Point (RP) service is provided by a third party (ABVA provider) for multiple organizations while maintaining separate virtual tunnels for each organization. An agent-based VPN service is currently available; see [7].

In this paper, we address the problem of optimal user connectivity in an ABVA from the perspective of an ABVA provider. Note that users belong to different organizations; they use heterogeneous network access mechanisms over the public Internet for access to such a service. Furthermore, each organization may have a different service level agreement with the ABVA provider. We factor in a number of issues and present a generic optimization formulation when the users are given to be static. We then consider a set of scenarios to shed light into the optimal user connectivity decision problem.

Undoubtedly, there is a close relationship between an ABVA and overlay networks. There have been significant research on overlay networks of various types in recent years (for example, see [1], [5]; an ABVA poses its unique challenges. There have been many works on various aspects about VPN (for example, see [2], [3], [4], [6], [8]), which primarily address the core design of overlay networks; however, how to efficiently provide user connectivity for agent-based VPN services over IP-based networks remains an important problem that has not been addressed in the literature.

The remainder of this paper proceeds as follows. In Section II, we present the basic framework of an ABVA. The connectivity problem is described and the integer linear programming formulation is presented in Section III. In Section IV, numerical results are presented and analyzed showing how the proposed architecture provides the required flexibility to

---

([†]work done while at Positive Networks, Inc., USA)

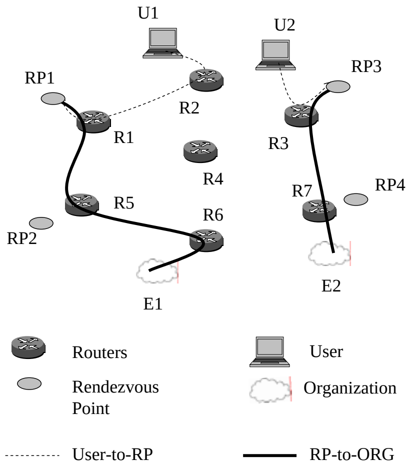


Fig. 1. ABVA: Basic conceptual framework

adapt the user connectivity for variable network and traffic conditions. We conclude our work in Section V.

II. ABVA: FRAMEWORK

We begin with a high-level conceptual view of the ABVA framework where users, Rendezvous Points, and organizations are considered (see Figure 1). For example, User U1 is associated with organization ORG1, and user U2 with organization ORG2. The Rendezvous points are marked as RP1, RP2, RP3, and RP4. The Internet serves as the native network that has IP routers identified as R1, R2, and so on. When user U1 wants to connect to organization ORG1 for VPN services, it is actually connected to Rendezvous Point RP1, which, in turn, connects to organization ORG1; similarly, for user U2 connecting to organization ORG2. The underlying path from the user to the Rendezvous point is over the Internet, which a user can access from anywhere on the Internet. Clearly, the Rendezvous Points must have presence on the Internet. However, the path from a Rendezvous point to an organization can be over the public Internet or it can be over a private network; for ease of illustration, the figure shown illustrates only the former case.

To illustrate further, consider again user U1 wanting to connect to organization ORG1, and user U2 to organization ORG2. We will now elaborate it through a layered approach, as shown in Figure 2. Note that each organization can have one or more gateway nodes as VPN entrance points to its intranet. In this case, ORG1 has gateway node G11, and ORG2 has gateway nodes G21 and G22. From a layered point of view, we consider the top layer to consist of users and organizations, the middle layer to consist of Rendezvous Points in the ABVA framework, and the bottom layer to be the Internet consisting of IP routers and links; these relationships are depicted in Figure 2. Note that we do not impose any restrictions on location of IP routers; i.e., they can be located in multiple different administrative domains; this is important to consider from a flexible IP-based service access point of view. While conceptually Rendezvous Points can be located anywhere, from an efficient service delivery perspective it can be arranged to be physically collocated with an Internet service provider's facility (e.g., a PoP) that gives bandwidth guarantee between the Rendezvous Point and the router of the ISP, as agreed upon between the ABVA provider and the ISP. An important advantage of collocating at an ISP's PoP is that such facilities have a redundant power supply and thus avoid typical power failure problems. In turn, the Rendezvous Point are remotely accessible to ABVA network administrators for performing any system reconfiguration and updates. For ease of reference, the direct link between a Rendezvous Point and the ISP's router will be referred to as an *ABVA-link*. For example, RP1–R1 is an ABVA-link.

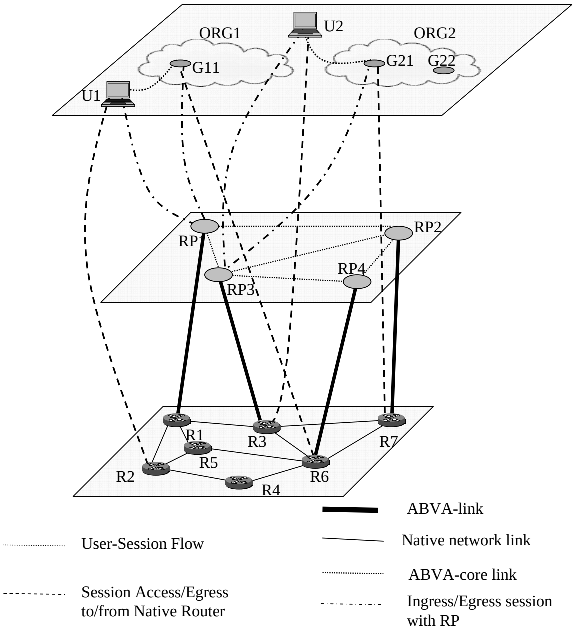


Fig. 2. ABVA: Service Layering View, with user-Organization connectivity (with sessions from Figure 1)

Consider Figure 2 in relation to Figure 1. User session flow from U1 to ORG1 is shown to enter at gateway node G11 that belongs to ORG1. First, this user session flow is shown at the top layer from U1 to G11. To provide this session, U1 connects to Rendezvous Point RP1 (first segment) which is then routed to G11 on RP1–G11 (second segment); this is shown by connecting from the top layer to the middle layer and then back to the top layer. If we now take the IP network ("the bottom layer") view into account for user session flow U1–G11, we see that segment U1–RP1 is actually routed on path U1–R2–R1–RP1 that includes ABVA-link R1–RP1, and that segment RP1–G11 is routed on path RP1–R1–R5–R6–G11, which again visits ABVA-link R1–RP1. Thus, ABVA-link R1–RP1 is visited twice for a session flow. While this double-visit appears to be an extra cost in the agent-based VPN approach, in terms of overall benefit this service provides, this cost is nominal. Note that RP1–G11 can be set up as a semi-permanent tunnel that might take path RP1–R1–R5–R6–G11 on the public Internet. Similarly, we can describe session U2–G21. Both sessions U1–G11 and U2–G21 are summarized in Table I. An important variation is to consider the case in which

TABLE I
CONNECTIVITY/PATH FOR USER SESSION FLOWS SHOWN IN FIGURE 2

U1 to G11	Consists of segments U1–RP1 & RP1–G11
U1–RP1	U1–R2–R1–RP1
RP1–G11	RP1–R1–R5–R6–G11
U2 to G21	Consist of segments U2–RP3 & RP3–G21
U2–RP3	U2–R3–RP3
RP3–G21	RP3–R3–R7–G21

organizations prefer to have dedicated private-line connectivity to one or more RPs (not shown in figure).

III. USER CONNECTIVITY DECISION

Simply stated, the user connectivity decision is to select the optimal Rendezvous Point to which a user connects. In Figure 2, we show that user U1 connects to RP1. In general, this may not be the optimal Rendezvous Point from the ABVA design point of view; for example, the quality of experience (QoE) as measured through user-perceived performance, such as delay, might be higher than desired when the user connects. This means that ideally, each user should not be statically assigned to a particular Rendezvous Point. Connecting to an optimal Rendezvous Point from a user is critical for an ABVA provider to support and maintain the desired performance while efficiently using its resources. Excessive overloads of particular Rendezvous Points would degrade network performance of the users, and moreover, may lead to dropping user access while other Rendezvous Points are available.

Recall that there are two types of connectivity decisions in regard to a user session flow: one is related to the user connectivity to a Rendezvous Point (ABVA ingress link, e.g., U1-N1) and the other is related to connectivity from an organization to the Rendezvous Point. In this work, we assume that the latter is established on a semi-permanent basis by setting up a tunnel based on service level agreements. Thus, the connectivity decision problem we consider is with regard to user connectivity to a Rendezvous Point. It should be noted in a practical environment, the user would arrive at random; hence, a mechanism is needed to select which RP to connect to. In this work, however, we consider the static case, i.e., the users are known and static. While this may not be practical, the static model allows us to determine the ultimate (albeit utopian) optimal decision; yet, through the understanding of the optimal decision, our goal is to understand the overall system behavior for future development of heuristic schemes that can be used in practice.

The optimal user connectivity problem for the static case can be stated as follows: upon a user requesting access from variable locations we need to connect the user to the optimal Rendezvous Point in the network so that an overall objective is minimized. A user connects to at most one Rendezvous Point at a particular time. Once connected, the user session flow impacts the ABVA-link on a Rendezvous Point through which it is connected to in two ways: 1) for the incoming flow from the user side *and* its connectivity flow continuing on the ABVA tunnel to the organization that also uses the same hose-link (see Figure 3), and 2) for the reverse direction of flow (due to asymmetric flow) from the organization in response to user's requests, which uses the ABVA tunnel to enter the same Rendezvous Point and then returns out on to the user ingress link. In order to address this effect, we consider two different flow rates: one for the forward direction, R^f (from user), and the other for the reverse direction, R^r (from organization for this user); note that each of these traverses *twice* on the ABVA-link and the actual flow can be directional depending on the type of layer-2 transport leased (see Figure 3 for an example of directional links). While, in general, such flow rates are not known beforehand, for agent-based VPN services, the rates can be derived or estimated, as discussed later.

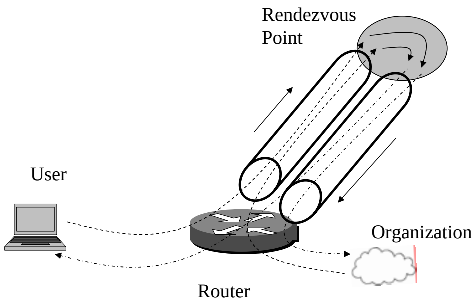


Fig. 3. ABVA: Service Access View.

Formally, assume that a set $\mathcal{E}$ of organizations uses the agent-based VPN service; each organization $e \in \mathcal{E}$ has a set of users $\mathcal{U}_e$ where each user is identified by $u \in \mathcal{U}_e$, and the user-organization pair is denoted by $\langle u, e \rangle$. The set of Rendezvous Points in the ABVA framework is denoted by $\mathcal{N}$. Since there is an one-to-one mapping between Rendezvous Points and the ABVA-links, we can use the set of RPs as the identifier for ABVA-links. Let the bandwidth allocated on the ABVA-link $n \in \mathcal{N}$ be $c_n, n \in \mathcal{N}$.

We denote the decision variable for connecting user $u \in \mathcal{U}_e$ for organization $e \in \mathcal{E}$ to the Rendezvous Point $n \in \mathcal{N}$ by y_{eun}, i.e., $y_{eun} = 1$ if $\langle u, e \rangle$ connects to Rendezvous Point n, and 0, otherwise. However, in an overloaded environment, not all users trying to gain access might be connected. We use indicator variable $\hat{y}_{eu}$ for this purpose, which takes the value of either 1 or 0. It indicates whether user u for organization e is not connected by setting $\hat{y}_{en}$ to 1. If this user is connected, then a corresponding y_{eun} will be 1 and artificial variable $\hat{y}_{eu}$ will take the value 0. Since each user is to be connected to at most one Rendezvous Point (or not connected in an overloaded situation), the following must be satisfied for each user:

$$\sum_{n \in \mathcal{N}} y_{eun} + \hat{y}_{eu} = 1, \quad u \in \mathcal{U}_e, e \in \mathcal{E} \tag{1}$$

If we distinguish the forward and reverse flow rate for user u in organization e by R^f_{eu} and R^r_{eu} respectively, then the following capacity constraint on each ABVA-link n must be satisfied due to forward and reverse traversal as well as due to double visits from ABVA ingress and egress sessions that use this ABVA-link:

$$\sum_{e\in\mathcal{E}}\sum_{u\in\mathcal{U}_e} 2(R^f_{eu}+R^r_{eu})y_{eun} \le c_n, \quad n\in\mathcal{N}. \tag{2}$$

We now define a generic latency function $\ell_{eun}(y_{eun})$ for connecting user u for organization e to Rendezvous Point n and a high penalty cost $P_e(\hat{y}_{eu})$ for user u for organization e when not connected. Note that both the penalty cost and generic latency can be organization-dependent and will be discussed in more detail later. We define the following objective that captures the total latency induced along with the penalty cost:

$$F=\sum_{e\in\mathcal{E}}\sum_{u\in\mathcal{U}_e}\sum_{n\in\mathcal{N}}\ell_{eun}(y_{eun})+\sum_{e\in\mathcal{E}}\sum_{u\in\mathcal{U}_e}P_e(\hat{y}_{eu}). \tag{3}$$

Thus, the optimal connectivity decision problem can be stated as the minimization of the overall latency and penalty cost given by (3) subject to connectivity constraints (4) and ABVA-link capacity constraints (2). The total latency function is advantageous to consider since (a) it represents the average latency if divided by the total number of users, i.e., by a fixed factor, (b) the functional component can be modeled further to capture user requirements. In general, this problem falls into the class of assignment problems; however, this is a heterogenous assignment problem due to different service requirements by different users and organizations.

A limitation of the above formulation is that it does not address the minimal user connectivity guarantee that might be expected by an organization due to SLAs. This can be addressed by introducing the following constraints:

$$\sum_{u\in\mathcal{U}_e}\sum_{n\in\mathcal{N}} y_{eun} \ge K_e, \quad e\in\mathcal{E} \tag{4}$$

where K_e denotes the minimum number of users that are to be connected for a specific enterprise $e\in\mathcal{E}$. This constraint is, therefore, a form of fairness constraint, which can be incorporated in the above model if necessary.

We now comment on the choice of the generic latency function. From initial measurement data available to us from service offerings by [7], we have observed that the latency cost can be approximately modeled with a fixed cost part and a variable cost part using weights, α,β (≥ 0), as given by

$$\ell_{eun}(y_{eun})=\alpha+\beta v_{eun} \tag{5}$$

where v_{eun} is an estimate on the delay. Here, α signifies a fixed latency cost incurred by every user. A variation of this cost model considers weights to be organization-dependent, i.e.,

$$\ell_{eun}(y_{eun})=\alpha_e+\beta_e v_{eun} \tag{6}$$

An advantage of the organization-dependent latency model is that α_e and β_e can be appropriately selected for different organizations, for example, to differentiate services; thus, any priority to be given to certain organizations based on service level agreements can be reflected. In this sense, (6) reflects a generic latency function that incorporates such needs.

The second part of the objective function concerning penalty is also important to consider. First, this allows the users of different organizations to be assigned a different penalty, based on service level agreements; second, it serves as a revenue loss factor which the ABVA provider incurs if some users are not connected. Thus, the composite objective function of latency and penalty together provides a generic abstraction for capturing a wide variety of variations that can be considered for the user connectivity problem.

IV. NUMERICAL RESULTS AND ANALYSIS

In this section, we use the problem formulation presented in the previous section to study a variety of scenarios. We first start with our study environment.

A. Study Environment

Working with Positive Networks [7], an ABVA provider, we obtained some measurement data and types of users, and access modes used. For instance, the user access bandwidth requirement has certain patterns. Typically, a user accesses either from his/her home or from a hotel room; thus, often a user has either a cable-modem or DSL connectivity. In order to satisfy such needs, it is desirable that the ABVA-link is not a bottleneck. Thus, for such access, the download rate would need to meet approximately 1.5 Mbps and the upload link would need to meet about a third of this rate; due to the dual direction we discussed earlier, approximately 2 Mbps bandwidth is sufficient for each user of this class of users. While there are several others rates noted from actual data, another class that is noticeable is due to the access mode of ISDN lines and the number of users; due to dual direction issues, this bandwidth is typically at 128 Kbps. For our study, we consider two types of users: ”Type-A” users with 2 Mbps and ”Type-B” users with 128 Kbps which correspond to $R^f_{eu}+R^r_{eu}$ in (2). We considered three different user ratios: all Type-A users, 80% Type-A users with 20% Type-B users, and 50% Type-A users with 50% Type-B users. For brevity, the case of all Type-A users will be called ’homogeneous’ users, and the other two as ’mixed’ users (with appropriate percentage).

To understand the user dynamics and connectivity, we consider a 100×100 grid in which we randomly selected the location of RPs and the users. Without loss of generality, we assume that a single enterprise is considered in our study since our interest is in the basic connectivity decision problem and also since the two type of users can be classified as two different enterprises. In our study, we considered the number of RPs, $N(=\#(\mathcal{N}))$, to be 3, 7, and 10 and assumed that the capacity c_n to be the same for all RPs. We varied the number of simultaneous users to be 50, 100, 200, 500, and 2000. Due to space limitation, we discuss results mainly for $N=3$ and 50 users.

Based on the bandwidth requirement of users, we selected ABVA-link capacity, c_n. In order to understand congestion, we considered four provisioning factors: 2.0, 1.0, 0.9, 0.8. At provisioning factor 1.0 all users are essentially accommodated since the sum of ABVA links has just enough capacity; this serves as the normalized value. With this normalization, provisioning factor 2.0 means that links are essentially twice over-provisioned. On the other end, provisioning factors 0.9 or 0.8 reflect underprovisioned situations, which allow us to determine the user connectivity layout in congested/overloaded situations in which some users would not be connected at all. Since we already allow an artificial variable in our model, any users not-connected in an underprovisioned situation would incur penalty as indicated through the penalty cost function. From the measurement data available to us from [7], an ABVA provider, we have found that we can approximately model the latency function $\ell_{eun}(y_{eun})$ using (5).

B. Observations

We first conducted a set of experiments to understand the computational time. Note that due to the two part definition of the latency function, the objective function in (3) is linear, while parameters α and β allow us to study different variations. The resulting optimization problem is then an integer linear programming (ILP) problem. We solve this ILP problem for all cases using CPLEX, a well-known optimization solver. In many instances, CPLEX solved the problem with 10 RPs ($N = 10$) and 200 users in less than a second to optimality. The general observation about the computational time is that when the provisioning factor is 0.8 or 0.9, the problem is capacity 'tight' and requires more computing time to solve. The second observation is that when there is mixed traffic (Type-A and Type-B users), the problems tend to take more time for 'tight' problems; this is because the system tries to pack heterogeneous group of users in a resource limited environment. While an ILP based computation may not be usable in real-time, this approach helps in the case of *forecasted* users and in computing a layout profile ahead of time very quickly so that heuristic rules can be used for actual connectivity in real-time.

Next, we discuss the assignment of users to different RPs for the homogeneous case. Recall that for the homogeneous case, all users are Type-A users with 2 Mbps requirement for bandwidth. We observe that when the provisioning factor is 2.0, then due to the ABVA link bandwidth being abundant, most users are attracted to the closest location, as long as the available bandwidth can allow them to do so. This is irrespective of the values of the scaling factors α and β in the latency function. Intuitively, this is not surprising. When the provisioning factor is reduced to the normalized value 1.0, then the users are uniformly distributed among the RPs, and continue to do so when the provisioning factor is reduced to 0.8, irrespective of all the different scaling values α and β we tested. This says that in the homogeneous case, the bandwidth constraint is the most stringent requirement in an overloaded/congested scenario, and on average, assignment is about the same. This is also fairly obvious. However, when we keep the number of users same, and increase the number of RPs to 10, then we do observe some imbalance in user allocation to different RPs, even for the provisioning factor of 1.0. As illustrated, column 'Homogeneous' in each of Table II and III shows the number of users accepted and total bandwidth used per RP for provisioning factors of 2.0 and 0.8, respectively.

We next discuss the case when there is a mixture of Type-A and Type-B users. We started by assuming that the penalty cost is the same for both users. While the general observation is often similar to the homogeneous case, there is a distinctive difference when it comes to a tight situation, e.g., at provisioning factor 0.8. When the fixed component of the cost function is neglected (α = 0), we found that at this provisioning factor (provisioning factor = 0.8) all Type-B users are accommodated by two RPs as shown in Figure 4(d). The disconnected users are 6 Type-A users (Table IV). The accommodated Type-A users are nearly equally distributed among the three RPs. This means that a lower bandwidth service (Type-B) gets preference over a higher bandwidth service regardless of the delay they might incur. When the provisioning factor is increased to 0.9, some Type-A and Type-B users are redistributed to allow connecting some of the left out users as shown in Figure 4(c). After the redistribution of users, the remaining bandwidth of each RP is not sufficient to include all the disconnected users. All Type-B users are accommodated while 5 Type-A users are still left out. Again, the accommodated Type-A users are equally distributed. When the provisioning factor is increased to 1, we expect that all the users would be connected. However, due to modularity constraints and equal partition of the total bandwidth, we found that 2 Type-A users are still disconnected as shown in Figure 4(b). As the provisioning factor is increased to 2, the bandwidth constraint is no longer a bottleneck and each user is connected to the closest RP as shown in Figure 4(a). Table IV summarizes the above numerical results. When α is small, same topologies are obtained, except when $\alpha = 1000$ and the provisioning factor is equal to 0.9; in this case, we observe different user connectivity to different RPs (compare $\alpha = 1$ and $\alpha = 1000$ in Table IV in which the differing results for $\alpha = 1000$ from $\alpha = 1$ are indicated in parenthesis), a form of permutation is taking place. This is significant from a stability point of view since the optimal behavior can be different as the provisioning factor is varied.

In summary, we can say that the delay factor is not necessarily dominant while the bandwidth access requirement plays a critical factor in that users with lower bandwidth access requirements get priority in getting connected over those with higher bandwidth access requirements when the penalty cost is the same for both services. To see how the result would change, we lowered the penalty cost for Type-B users and re-ran the case in which the provisioning factor is still set at 0.8. We found that indeed more Type-A users get connected while some Type-B users are now left out; the actual number of Type-B users not connected depends on the value of the penalty cost associated with this user class compared to the penalty cost associated with the Type-A user class. In any case, not all Type-B users are connected at provisioning factor 0.8 since there is not enough bandwidth.

V. CONCLUSION

The Agent-based VPN framework is an emerging service type for providing VPN services to users and enterprises. In

TABLE II
NUMBER OF ACCEPTED USERS AND TOTAL BANDWIDTH SERVED PER RENDEZVOUS POINT (N=3, TOTAL USERS=50, PROVISIONING FACTOR=2.0)

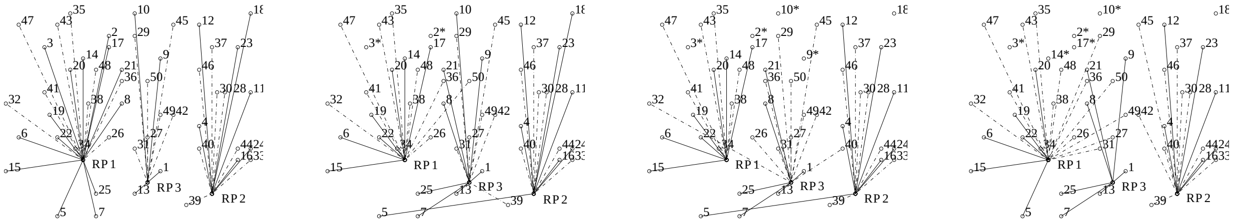
	Homogeneous		Mixed with Type-A users=80%		Mixed with Type-A users=50%	
RP	Users Acc.	TotBW Served	Users Acc.	TotBW Served	Users Acc.	TotBW Served
1	24	48	24	40.5	24	29.3
2	11	22	15	26.2	15	15
3	15	30	11	14.5	11	8.9

TABLE III
NUMBER OF ACCEPTED USERS AND TOTAL BANDWIDTH SERVED PER RENDEZVOUS POINT (N=3, TOTAL USERS=50, PROVISIONING FACTOR=0.8)

	Homogeneous		Mixed with Type-A users=80%		Mixed with Type-A users=50%	
RP	Users Acc.	TotBW Served	Users Acc.	TotBW Served	Users Acc.	TotBW Served
1	13	26	14	20.5	21	13.9
2	13	26	12	20.2	16	13.2
3	13	26	14	20.5	7	14.0

TABLE IV
α=1 (1000), β=5, RPS=3, USERS=50 AND 50% TYPE-A USERS

	Provisioning Factor=0.8		Provisioning Factor=0.9		Provisioning Factor=1.0		Provisioning Factor=2.0	
RP	Type-B	Type-A	Type-B	Type-A	Type-B	Type-A	Type-B	Type-A
1	15	6	7 (11)	7 (6)	11	7	10	14
2	10	6	7	7	7	8	8	7
3	0	7	11 (7)	6 (7)	7	8	7	4
# users not connected	0	6	0	5	0	2	0	0



(a) Provisioning factor =2.0 (b) Provisioning factor=1.0 (c) Provisioning factor =0.9 (d) Provisioning factor=0.8

Fig. 4. User Connectivity Layout, $\alpha = 1$, $\beta = 5$ (solid and dotted lines present Type-A and Type-B users, respectively; Type-A users not connected are marked with *)

this paper, we present the basic framework for an ABVA. In particular, we addressed the issue of the user connectivity decision taking practical factors into consideration. Our formulation generically encompasses the bandwidth access requirement of different users as well as the latency and penalty for different user classes. We have shown that the latency function can incorporate different types of users through two different parameters.

Our computational study considered both homogeneous users and mixed users, when the provisioning factor is varied. The results for the homogeneous user case is straight forward. The mixed user case was particularly interesting since we observed that users requiring higher bandwidth access do get turned away more so than the users requiring lower bandwidth access, almost irrespective of the latency cost while the penalty cost is kept the same. Note that this observation is for the static case; however, this observation seems to suggest that the system behavior is similar to a loss model for the stochastic case, i.e., perhaps a loss system can be used for modeling the stochastic arrival case. On the other hand, this does not consider the delay in connectivity cost, i.e., the system is not exactly like a loss model. Currently, we are investing the stochastic case by developing a number of heuristics on how user connectivity decisions can be considered; the results will be reported in a subsequent paper.

REFERENCES

[1] Z. Duan, Z. Zhang, and Y. Hou, "Service overlay networks: SLAs, QoS, and bandwidth provisioning," *IEEE/ACM Trans. Networking*, vol. 11, pp. 870–883, 2003.
[2] N. G. Duffield, P. Goyal, and A. Greenberg, "A flexible model for resource management in virtual private networks," in *Proc. ACM SIGCOMM*, 1999.
[3] N. G. Duffield, P. Goyal, A. Greenberg, P. P. Mishra, K. K. Ramakrishnan, and J. E. van der Merwe, "Resource management with hoses: point-to-cloud services for virtual private networks." *IEEE/ACM Trans. Networking*, vol. 10, no. 5, pp. 679–692, 2002.
[4] B. Gleeson, A. Lin, J. Heinanen, G. Armitage, and A. Malis, "A framework for IP based virtual private networks," *Internet RFC 2764*, February 2000.
[5] K. Liu, J. Lui, and Z. Zhang, "Distributed Algorithm for Service Replication in Service Overlay Network," in *Proc. IFIP TC6 Networking Conference*, May 2004.
[6] A. Nagarajan, Ed., "Generic requirements for provider provisioned virtual private networks (PPVPN)," *Internet RFC 3809*, June 2004.
[7] Positive Networks, http://www.PositiveNetworks.com/
[8] E. Rosen and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)," *Internet RFC 4364*, February 2006.