

Survivable Free Space Optical Mesh Network using High-Altitude Platforms

Dieu Linh Truong^a, Xuan Vuong Dang^a, The Ngoc Dang^b

^a*School of Information and Communication Technology, Hanoi University of Science and Technology, Vietnam*

^b*Department of Wireless Communications, Posts and Telecommunication Institute of Technology, Vietnam*

Abstract

Free space optical (FSO) communication refers to the information transmission technology based on the propagation of optical signals in space. FSO communication requires that the transmitter and receiver directly see each other. High-altitude platforms (HAPs) have been proposed for carrying FSO transceivers in the stratosphere. A multihop HAP network with FSO links can relay traffic between ground FSO nodes. In this study, we propose an end-to-end switching model for forwarding traffic between massive pairs of ground FSO nodes over a HAP network. A protection mechanism is employed for improving the communication survivability in the presence of clouds, which may break the line of sight (LoS) between HAPs and ground nodes. We propose an algorithm to identify a set of necessary HAPs and to design the HAP network topology in integrating the protection mechanism. The design aims to a network with the minimal equipment cost. The results demonstrate that, even though networks with survivable capacity use more resources, they are not necessary much more expensive than those without survivability in terms of equipment, i.e., HAPs and FSO devices, and in terms of wavelength resource utilization.

Keywords: Free Space Optics, Survivable networks, Topology design, Routing

1. Introduction

Free space optics (FSO) communication technology uses a laser beam to transmit information through air or vacuum between a pair of transceivers at high speed. Commercial FSO transceivers available in the market usually operate at 1.25 Gbps, 2.5 Gbps or even higher speeds. However, FSO communication requires line of sight (LoS) between transceivers, which is sometimes difficult to satisfy, mostly at long distances on the ground. In addition, terrestrial FSO links are sensitive to weather conditions such as rain, fog, or air turbulence leading to link failures.

Some studies have addressed terrestrial FSO link failures by building terrestrial FSO networks with spare capacity, for example, the works in [1–3]. These works propose network dimensioning solutions, taking into account single link failures, multiple link failures, and partial link failures.

Recent research has proposed the use of high-altitude platforms (HAP) for wireless communications, including FSO communication at different scales [4–6]. HAPs are high-flying aircraft or airships, which operate at altitudes of 17 to 24 km in the stratosphere. HAPs are used to avoid physical obstacles and weather impact present on the ground. In stratosphere, HAPs suffers the least wind impact thus less power is required to stabilize the platform. HAPs carry FSO transceivers that play the role of

intermediate nodes forwarding traffic in the stratosphere or between the stratosphere and the ground. Figure 1 illustrates the communication between the ground FSO transceivers (also called ground FSO nodes) using HAPs. The traffic between a pair of ground FSO transceivers follows a path comprising three parts: i) *uplink*: from the source-ground FSO node up to an FSO transceiver on a HAP, ii) *inter-HAP links*: between FSO transceivers on different HAPs, iii) *downlink*: from the last HAP down to the destination-ground FSO transceiver. A HAP can communicate with several ground FSO transceivers. Although inter-HAP links are not effected by weather conditions, uplinks and downlinks suffer interference from clouds floating at an altitude of approximately 10km.

Several research projects on designing HAPs to deploy broadband networks such as Halo, HeliNet, CAPANINA, SkyNet [7], Loon project [8] of Google, UAV project of Facebook [9] and Stratobus project of Thales Alenia Space [10] have been conducted worldwide. HAPs are classified in two categories based on the underlying physical principle that provides the lifting force for the HAPs: aerodynamic (the HAP is heavier than air) and aerostatic (the HAP is lighter than air) [11]. Aerostatic platforms use buoyancy to float in the air, their locations are relatively stable with the ground. Aerodynamic platforms use dynamic forces created by movement through the air. In general, HAP is not only costly but also limited in capacity in terms of payload weight and available energy [12]. Therefore, the utilization of HAPs should be well planned.

Other studies focus on the utilization of HAPs. In [5],

*Corresponding author

Email address: linhtd@soict.hust.edu.vn (Dieu Linh Truong)

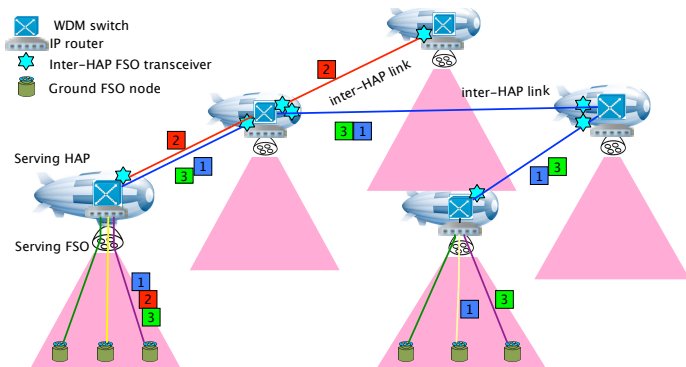


Figure 1: Data communication in a FSO network using HAPs without protection. Uplinks and downlinks can be compromised by clouds.

each HAP carries a mobile based station that serves a cluster of mobile ground nodes. The work proposed a solution to divide ground nodes into clusters but the connectivity between HAPs has been omitted from the scope of this research. In [13], the authors proposed an optical transport network based on mesh HAP systems with inter-HAP links using FSO technology, whereas radio frequency (RF) was used between HAPs and the ground. All-optical data transmission with wavelength routing was used between HAPs. Nodes on HAPs are capable of MUX/DEMUX data received from the ground using a time division multiplexing (TDM) technique. The authors agreed that the number of required wavelengths on the inter-HAP links and the wavelength routing scheme depend on the actual physical HAP topology. However, they only evaluated the maximum number of required wavelengths on two very particular network topologies: bus and full mesh with five HAPs where each HAP only serves one ground station. The possibility that a HAP serves more than one ground station has not been clearly discussed.

European studies on the availability of optical links between HAPs and ground in [14] show that single optical link availability varies from under 20% during winter in Northern Europe to over 70% during summer in Southern Europe. Research has focused on improving the availability of single optical downlink to ground stations by using a ground-station diversity scheme, i.e., ground stations at geographically separate locations. Availability increases with the number of available stations. The scheme requires a HAP network to carry traffic toward available ground links. Four ground stations in Southern Europe can achieved 99% availability. Nevertheless, ground-station diversity is applicable only if the location of the transmission destination is not important.

In [4], data from Earth observation satellites were relayed to some HAPs at 20 km altitude through FSO links before sending them down to ground stations via microwave point-to-point links. Since there are no clouds at an altitude of 20 km, the availability of the FSO links is almost 100%. In cloud-free conditions, FSO links can also be used

in place of microwave links for a higher data rate.

Research in [6] proposed multi-HAP networks for carrying data between ground FSO nodes and optical slant links for transmitting data up to and down from HAPs when primary connections fail owing to air turbulence. This research focused only on analyzing the bit error rate (BER) and the availability of a multihop FSO link. The choice of HAP locations, slant links, connectivity between HAPs and routing from ground nodes to ground nodes were left outside of the study.

Other studies either focused on analyzing the BER of multihop HAP-based FSO connections [15] or HAP network dimensioning without survivability against LoS loss [16].

In summary, to the best of our knowledge, there is no work on the dimensioning of HAP-based FSO networks that are survivable against optical HAP-ground link failures. We address this problem in this study. Given a set of ground FSO nodes with communication demands between each other, we will try to identify a set of necessary HAPs carrying FSO transceivers in the stratosphere for forwarding data between ground FSO nodes using FSO technology in all links including uplinks and downlinks. Similar to [6] we use FSO slant links for enhancing end-to-end communication survivability against clouds. Although RF may be used in alternative links in case of clouds, we do not elect the RF technology for slant links because RF links offer much lower data rate than optical links leading to connection quality down-grading when failures occur. Unlike previous studies, we focus on the dimensioning of the HAP-ground network. In addition, we assume that all HAPs are aerostatic and their locations are relatively stable with the ground during their operation. There are two main contributions in this paper: i) proposing an end-to-end data switching mechanism between massive pairs of ground FSO nodes through a HAP network with a 1+1 protection mechanism against uplink and downlink failures, ii) a dimensioning solution for the survivable HAP-ground network with the minimum cost of equipment including HAPs and FSO transceivers costs.

The remainder of this paper is organized as follows. Section 2 presents the switching mechanism from end-to-end and the protection mechanism against clouds. Section 3 describes the network dimensioning problem. Section 4 presents the proposed dimensioning algorithm. Section 5 presents the simulation results and discussions. Finally, Section 6 concludes the paper.

2. Switching and Protection mechanisms

2.1. WDM communication between HAP and ground

Each HAP carries one down-facing FSO device to communicate with a cluster of ground FSO transceivers. For the ground FSO transceivers, the HAP is called the *servicing* HAP and the FSO device on the HAP is called the *servicing* FSO transceiver. On the same HAP, there may

be other FSO transceivers for communicating with other HAPs. These FSO transceivers are called *inter-HAP* FSO transceivers. The *servicing* FSO transceiver and *inter-HAP* FSO transceivers on the same HAP should be wired together back-to-back using optical fibers (see Figure 1).

We divide the ground FSO transceivers into clusters, each having a *servicing* HAP. Each ground FSO transceiver communicates with its *servicing* FSO bidirectionally by using two fixed wavelengths for the up and down directions. The wavelengths in the two directions are managed separately because each FSO transceiver has a light source and a light receiver for sending and receiving data independently. The ground FSO devices of the same cluster share access to the common *servicing* FSO transceiver using wavelength division multiplexing (WDM) technology. Figure 1 illustrates this multiple-access method, where each color represents a wavelength. For simplification, we show only the wavelengths for the up or down direction. Depending on wavelength density of the WDM technique to be used, 32, 64 or more ground FSO transceivers can be served by a single FSO transceiver on HAP.

On *inter-HAP links*, WDM technique will also be used. Between HAPs data are forwarded on a wavelength-switching basis without wavelength conversion. Data sent from one HAP to the other may travel through multiple HAPs using a continuous wavelength, the so-called lightpath. When the volume of traffic between a pair of HAPs is larger than the capacity of a wavelength, multiple lightpaths may be necessary. These lightpaths do not need to follow the same route.

2.2. Data switching on HAP

A ground FSO transceiver may have data to send to different remote ground FSO transceivers. This section explains how the data are switched from end to end. We propose to employ one of two models: IP over WDM and SONET over WDM.

In *IP over WDM* model, the ground FSO nodes in the same cluster must be assigned IP addresses of the same network. An IP router is installed on each HAP to groom traffic received from the *servicing* FSO transceiver based on destination clusters. Then, the groomed traffic follows different lightpaths between HAPs to reach destination clusters. The end-to-end data are switched as follows:

- At a source ground FSO node, data are encapsulated in IP packets, of which the destination IP addresses are those of the destination ground FSO nodes.
- The packets follow an uplink and arrive at the *servicing* FSO of the source. Here, the IP packets pass through an IP router that directs the packets to different ports dedicated to different destination cluster networks. If the amount of data addressed to a cluster network is larger than the capacity of a wavelength then multiple ports should be used for that cluster.

The flow of IP packets going out from each router port is regenerated in optical domain using a wavelength and then forwarded to a WDM switch to take a lightpath to get to the *servicing* HAP of the destination cluster.

- At the *servicing* HAP of the destination cluster, IP packets inside each lightpath are DEMUX by the IP router of the HAP according to their destination IP addresses. IP packets heading to the same destination ground FSO node are regenerated in optical domain, using the wavelength assigned to the destination ground FSO node, and transmitted to the ground.

In this switching mechanism, the set of lightpaths between HAPs should be pre-configured beforehand. These lightpaths can be considered as physical links between IP routers. Figure 2 illustrates data switching on a HAP. Port p-HAP-1 is dedicated to receiving traffic for HAP-1 to get to network cluster-1. Port p-HAP-2 receives traffic for HAP-2 to get to network cluster-2. Table 1 shows the IP routing table of the router in Figure 2.

Network	Interface
Network address of cluster-1	p-HAP-1
Network address of cluster-2	p-HAP-2
...	...

Table 1: Routing table of the IP router in Figure 2

In *SONET over WDM* model, SONET switches replace IP routers. While IP routers forward data by packet switching basis, SONET switches work on TDM circuit switching basis. Therefore, a circuit should be established for each pair of ground FSO nodes by taking a fixed time slot inside the lightpath between the associated source and destination *servicing* HAPs. The data rate of the time slot corresponds to the volume of requested traffic between the two ground FSOs.

Once one of the above models is employed, the remaining question is setting up lightpaths between HAPs and configuring the IP routers/SONET switches for grooming traffic to lightpaths. In Section 4, we propose algorithms for building the HAP topology and the lightpaths between HAPs, i.e., their paths and wavelengths. Subsequently, IP routers/SONET switches can be easily configured.

2.3. Protection model

On cloudy days, clouds may cut the LoS between a ground FSO node and its *servicing* FSO device. Clouds generally float at an altitude of 10 km. To maintain continuous communication in the network, we propose the following protection mechanism against clouds.

When the LoS between ground FSO nodes and their *servicing* FSO are cut, these ground nodes should forward all their traffic to a nearby alternative HAP, for which the LoS is still clear. Let *primary* HAP be the *servicing*

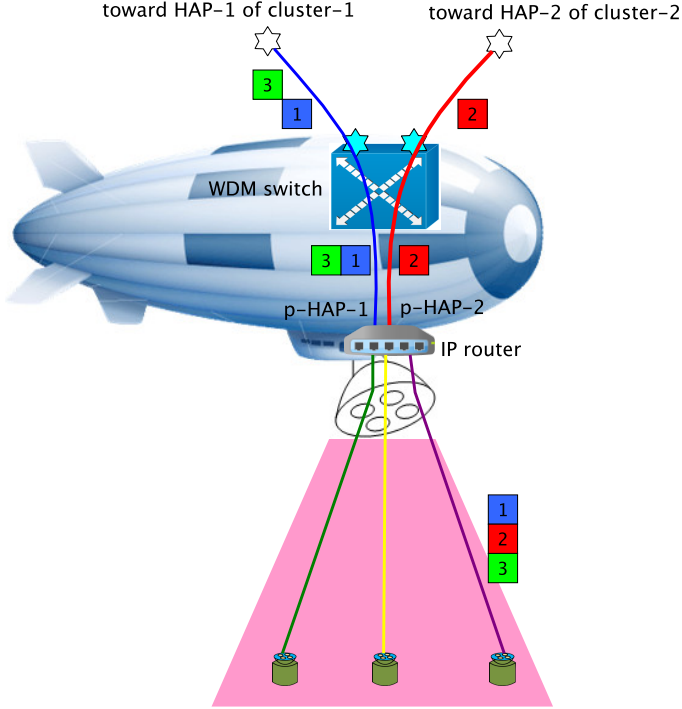


Figure 2: Data switching on HAP.

HAP where data travel in case of no failure whereas the alternative HAP is *backup* HAP. The *backup* HAP, after receiving data from a ground FSO node on a wavelength, will simply forward that wavelength to the *primary* HAP without any processing. At the *primary* HAP, traffic is handled exactly as it is received directly from the ground, i.e. passing through the IP router and continuing the original lightpaths to the destinations. To receive the backup traffic from the ground, each *backup* HAP should have a dedicated FSO device that looks at its backup zone. This FSO device is called *backup-serving FSO*.

Figure 3 illustrates the protection model where zone A is the serving zone of HAP H_A and HAP H_B backs up for H_A . Every ground node in the serving zone of H_A should forward its traffic to H_B whenever its LoS to H_A is cut. H_B has a *backup-serving FSO* device for covering zone A.

To implement this model, each ground node should have two light sources and two light detectors oriented toward H_A and H_B . Such an FSO device has been mentioned in [6]. We suggest that the ground nodes transmit two copies of data to both H_A and H_B , and H_B always forwards the copies it receives to H_A . H_A chooses to receive the traffic directly from ground nodes or from H_B depending on the availability of LoS between H_A and its serving zone. This is a 1+1 backup mechanism.

For downstream flows, the backup flows travel along the same paths with the upstream flows but in the reverse direction. When H_A has to forward data to its ground FSO nodes, it transmits a copy to H_B to relay the copy to the ground FSO nodes. It is the ground node that decides

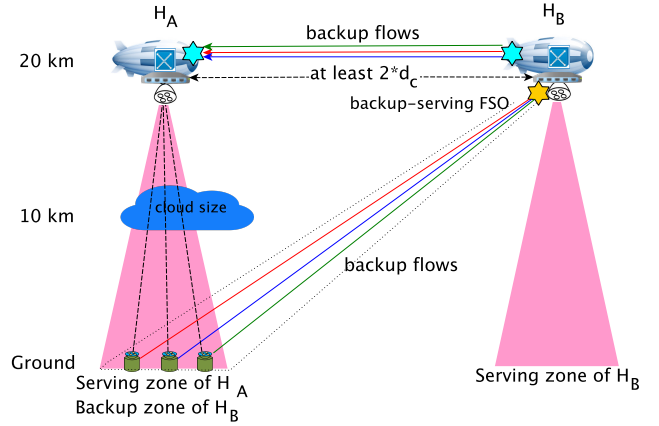


Figure 3: Protection model: when LoS between H_A and its serving zone is cut, every ground node in serving zone of H_A should forward its traffic to H_B . *Backup* HAP should be at a distance of at least 2 times the size of cloud from its *primary* HAP.

whether to receive from H_A or H_B .

A *backup* HAP H_B must be sufficiently far from its *primary* HAP H_A to avoid that a cloud cuts both links to *primary* HAP and *backup* HAP. Based on the altitude of HAPs and clouds, we deduce that the minimum distance between H_A and H_B should be twice the maximum size of clouds. However, a *backup* HAP should not be too far from its backup ground FSOs because greater distance causes greater attenuation and air turbulence impact leading to high BER. Therefore, the distance $d(H_A, H_B)$ between H_A and H_B is restricted by:

$$2 \times d_c < d(H_A, H_B) < L^{HH} \quad (1)$$

where d_c is the maximum cloud size, L^{HH} is the maximum length of inter-HAP links with BER under the requirement threshold of the current Forward Error Correction techniques.

In brief, the condition that the communication between H_A and its serving zone is protected against a single cloud is that there exists a HAP H_B at a distance from H_A that satisfies the constraint (1). This constraint guarantees that either H_A or H_B has LoS to the serving zone of H_A making the communication available against a single cloud.

If there are many clouds but they are scattered in the sky, LoS from the serving zone of H_A to H_A and H_B may still not be cut simultaneously. However, if there are dense clouds in a large region, it is possible that both LoS are cut. Section 5 analyzes further link availability that the proposed protection scheme offers.

It is worth noting that if H_B backs up H_A , then topologically, H_A could also backup H_B .

3. Survivable ground-HAP network dimensioning problem statement

In this section, we formulate the problem of dimensioning a survivable ground-HAP network. The dimensioning aims to minimize the investment cost in terms of HAPs and FSO devices on HAPs while ensuring that all traffic demands between ground nodes are accommodated and backed up against clouds.

In order to model the impacts of air turbulence, we use Gamma-Gamma distribution model proposed in [15] for both ground-HAP and inter-HAP channels. The impacts are represented as BERs of ground-HAP and inter-HAP links. Then, while dimensioning the ground-HAP network, we require that all lightpaths between HAPs must have end-to-end BER under a threshold δ such that bit errors could be corrected by current FEC techniques.

In the formulation, following notations are used as parameters, values of some parameters are listed in Table 2:

- \mathbb{N} : represents the set of ground FSO nodes that has communication demands to be carried by HAPs.
- \mathbb{M} : represents the matrix of traffic demands between ground FSO nodes in \mathbb{N} .
- \mathbb{D} : is the ground coverage diameter of an FSO transceiver on HAP. We assume that light beam sizes of *servicing* FSO transceivers are identical, thus their coverages on the ground are fixed by \mathbb{D} .
- \mathbb{C} : is the payload capacity of a HAP in terms of the number of FSO transceivers that a HAP can carry.
- \mathbb{W} : is the number of wavelengths available on an FSO link according to WDM technique.
- r^{max} : is the data rate of a wavelength.
- d_c : is the maximum size of a cloud.
- δ : is the maximum allowable BER of a connection.
- L^{HH} : is the maximal allowable length of an inter-HAP link such that BER of the link does not exceed threshold δ
- c_{HAP} : is the investment cost for a HAP
- c_{FSO} : is the investment cost for an FSO on HAP.

Following notations are used for variables:

- n_{HAP} : represents the number of HAPs will be used in the HAP network to be designed.
- n_{FSO} : represents the total number of *servicing* FSO transceivers to be installed on HAPs.

The dimensioning problem is stated as follows.

Given:

Param.	Description	Sim. value
\mathbb{D}	Ground coverage diameter of an FSO transceiver on HAP	15 km
\mathbb{C}	Payload capacity of a HAP in terms of FSO devices	10 devices
\mathbb{W}	Number of wavelengths in WDM technique	128
r^{max}	Data rate of a wavelength	1 Gbps
d_c	Maximum cloud size	10 km
L^{HH}	Maximal distance between HAPs	60 km
δ	BER threshold	10^{-3}

Table 2: Parameters of the dimensioning problem and their simulation values.

- a set of ground FSO nodes \mathbb{N} ,
- a matrix of traffic demands between ground FSO nodes \mathbb{M} .

The objective is to find a set of HAPs and a HAP network topology that minimize the total investment cost

$$\min(n_{HAP} * c_{HAP} + n_{FSO} * c_{FSO}) \quad (2)$$

and subject to following constraints:

- all traffic demands in \mathbb{M} are routed,
- all traffic demands in \mathbb{M} are backed up using the proposed protection scheme,
- data traffic between HAPs are carried by continuous lightpaths with end-to-end BER under threshold δ .

The constraint that all traffic demands in \mathbb{M} are routed means that: for each demand, there exists a serving HAP receiving the demand from its source ground node, an inter-HAP lightpath carrying the demand over the HAP network, and another serving HAP forwarding the demand to the destination ground node.

4. Proposed dimensioning solution

The dimensioning problem as stated in Section 3 is complex due to multiple constraints related on working and backup traffic. It is difficult to find an optimal HAP topology with the minimum total cost of HAPs and FSO transceivers in a single step. Therefore, we divide the design problem into smaller sub-problems and attempt to solve them by heuristic algorithms in independent steps.

For carrying a given demand matrix \mathbb{M} , a HAP network with few HAPs should require high network connectivity in order to be able to carry the required traffic and thus uses many FSO devices on HAPs. If the HAP network uses more HAPs, it may need fewer inter-HAP links thus fewer FSO devices. In other words, n_{HAP} and n_{FSO} may vary in opposite direction meanwhile they both contribute in the total investment cost in (2).

The unity cost of an FSO device c_{FSO} , for example the SONABeam of fsona [17], is about thousands of dollars. The investment cost of a HAP c_{HAP} is much higher. Consequently, the total investment cost increases faster when the number of HAPs increases than when the number of FSO devices increases. The number of HAPs should be prioritized to minimize than the number of FSO devices while minimizing the total cost. From this observation, we try to find a network configuration with the minimal number of HAPs first then minimize the number of FSO devices on HAPs later by linking HAPs with the smallest number of inter-HAP links.

The design problem is divided into three steps: i) *minimum clustering*: identifying the minimum set of HAPs for serving all ground FSO nodes, ii) *backup matching*: identifying a *backup* HAP for each *primary* HAP due to survivability purpose, iii) *topology design*: linking HAPs together with the minimum number of inter-HAP links in such a way that both the traffic demands of matrix \mathbf{M} and their corresponding backup traffic are accommodated.

4.1. Minimum clustering

To identify a minimum set of HAPs for serving all ground nodes, the ground nodes are organized into clusters, each of which is assigned a *servicing* HAP. There are several constraints to be considered:

- *Limited coverage of FSO transceivers on HAPs*: ground nodes in a cluster should be inside the coverage diameter \mathbb{D} of its *servicing* FSO. Larger \mathbb{D} allows a *servicing* FSO communicating with more ground nodes leading to fewer HAPs, but requires larger *servicing* FSO apertures thus higher transmit power to maintain the link quality. As a result, HAPs need to recharge more often. The optimal diameter \mathbb{D} should be considered in taking count the trade off between HAP recharging cost and HAP cost. We reserve this optimization problem for a future work as fix \mathbb{D} as a parameter in this research.
- *Limited number of wavelengths*: The number of ground FSO nodes per cluster is limited by \mathbb{W} , the number of wavelengths available according to the WDM technique.

Let us describe the location of a ground FSO node by two coordinates x and y in a 2-dimension Cartesian coordinate system. Assume that all ground FSO nodes are distributed in the quarter where the x and y coordinates are all positive. We propose the following clustering algorithm:

- Step 1: Divide the ground FSO region into non-overlapping bars parallel to the x-axis such that bar indexed k contains the FSO devices with coordinates $\forall y \in [k\mathbb{D}\sqrt{2} \dots (k+1)\mathbb{D}\sqrt{2}]$. See Figure 4.
- Step 2: We consider bars consecutively, from the index $k = 0$ onward. For each bar, the FSO nodes are

screened from the smallest to the largest x coordinates:

- Step 2.1: Find the FSO node, called F_1 , with the smallest coordinate x that has not been clustered.
- Step 2.2: Find the FSO node, called F_2 , with the largest coordinate x that has not been clustered such that: i) its distance to F_1 does not exceed $\mathbb{D}\sqrt{2}$, and ii) the number of ground FSO nodes located between F_2 and F_1 in the bar does not exceed \mathbb{W} .
- Step 2.3: Create a new cluster including all FSO nodes in the bar between F_1 and F_2 inclusively. The *servicing* HAP for this cluster is placed at the coordinates $x = (x_{F_1} + x_{F_2})/2$ and $y = (k + 1/2)\mathbb{D}\sqrt{2}$.
- Step 2.4: If the cluster is not full, i.e., the number of FSO nodes is still under \mathbb{W} , then put into the cluster the FSO nodes that are in the coverage diameter \mathbb{D} of the newly identified *servicing* HAP, even if those nodes do not belong to the current bar. The process is repeated until the cluster is full.
- Step 2.5: If there are still FSO nodes unclustered in the bar, repeat step 2.1 to create other clusters.

The results of the algorithm are the projected locations of all *servicing* HAPs on the ground and the set of ground FSO nodes served by each HAP. Figure 5 shows the clustering result for the set of ground FSO nodes shown in Figure 4. Each cluster is represented by a circle in which the centroid is the projected location of its *servicing* HAP. Ground FSO nodes are represented by green + symbols in the figure. We can remark that more clusters are formed where the density of the ground FSO nodes is high, while fewer or even no clusters are resulted where the density of ground FSO nodes is low.

4.2. Backup matching

In this step, we find a backup HAP for each *servicing* HAP. Assume the maximum cloud size is d_c , the *backup* HAP should be at a distance of at least $2 \times d_c$ from the primary one.

Our idea is that, starting from the set of HAPs found from the clustering step, we match them in pairs so that HAPs in a pair backup each other. The matching problem can be solved using Edmonds' Blossom algorithm [18] which finds a maximum cardinality matching in a graph. Applying to our problem, the graph contains all HAPs as vertices and edges are inter-HAP links whose lengths are greater than $2 \times d_c$ but smaller than L^{HH} , the maximum allowable length for an inter-HAP link. The maximum matching provides the largest number of pairs of primary

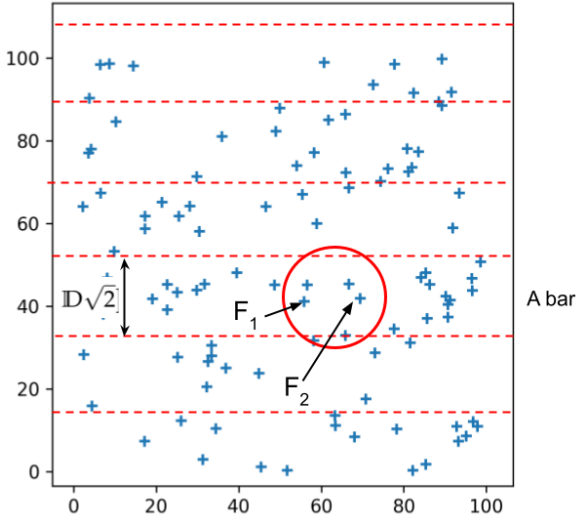


Figure 4: Clustering process with the ground region divided into bars.

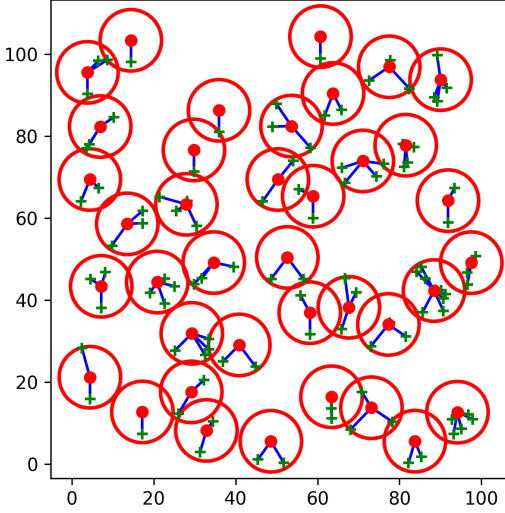


Figure 5: Example of the clustering result.

and backup HAPs. If the algorithm fails to find a matching for a HAP, we add a new HAP as the backup HAP for it. The newly added backup HAP must be at $2 \times d_c$ distance from the primary HAP.

After matching backup, a number of wavelengths need to be reserved on the link between a *primary* HAP and its *backup* HAP for carrying backup upstream and downstream traffic. In Figure 3, when H_B backs up H_A , the number of wavelengths to be reserved on link $H_B \rightarrow H_A$ is equivalent to the number of ground FSO nodes served by H_A because the upstream traffic from a ground FSO node travels on a single wavelength over link $H_B \rightarrow H_A$. Similarly, the same number of wavelengths must be reserved on the opposite link $H_A \rightarrow H_B$ for downstream traffic.

Since H_A also backs up H_B , a number additional wavelengths equivalent to the number of ground FSO nodes

served by H_B should be reserved on the links $H_A \rightarrow H_B$ and $H_B \rightarrow H_A$. As a result, the link (H_A, H_B) can accommodate backup traffic for both H_A and H_B only if the total number of ground FSO nodes of H_A and H_B does not exceed W .

4.3. Topology design and wavelength routing

Once the locations of primary HAPs are identified and backup HAPs have been assigned to them, the remaining issue is to determine the inter-HAP links to be deployed and the lightpaths to be setup between HAPs in order to accommodate all traffic requests. Since the wavelengths dedicated to backup traffic are reserved during the backup matching step, the remaining traffic to be accommodated is the primary traffic. The amount of primary traffic to be sent between two HAPs is the sum of the demanded bandwidth between their ground FSO nodes. Therefore, the number of lightpaths to be setup between a pair of HAPs is the ratio of the demanded bandwidth between the HAPs and the bit rate r^{max} of a wavelength. The topology design problem is stated as follows:

Given:

- Locations of HAPs obtained from the clustering step and those added in the backup matching step.
- \mathbb{M}^{HAP} : set of traffic demands between pairs of HAPs. Each demand is represented by a tuple (s, d, n) , where s, d and n are, respectively, the source HAP, destination HAP and number of lightpaths needed to carry traffic from s to d .

Constraints:

- All end-to-end lightpaths should have BER under δ .
- Wavelength continuity constraint along each lightpath.

Objective:

- We need to identify the necessary inter-HAP links and lightpaths for carrying all traffic in \mathbb{M}^{HAP} such that the total number of FSO transceivers to be carried on HAPs is minimized.

The idea of the proposed algorithm is that: from a set of possible inter-HAP links, we find routes for demands in \mathbb{M}^{HAP} one by one, and incorporate the links on the routes in the topology. The links already included in the topology are assigned small weights so that they will be prioritized in subsequent demands, thus new links are involved only when there is no possible route within the current topology.

Wavelengths are assigned to routes based on the *least used* strategy. The least used wavelength is defined as the wavelength that carries traffic on the smallest number of inter-HAP links.

Below are some notations used in the algorithm:

- G_w denotes the directed graph, where vertices represent HAPs, and arcs represent all possible directional inter-HAP links with wavelength w available.
- $nFreeW(e)$ denotes the number of free wavelengths of directional link e .
- c_e denotes the weight assigned to directional link e ,
$$c_e = \begin{cases} 100 & \text{if } e \text{ has never been used} \\ 1 - \frac{nFreeW(e)}{\mathbb{W}} & \text{otherwise} \end{cases} \quad (3)$$
- T denotes the topology of the HAP network during construction. Initially, T contains all bidirectional inter-HAP links connecting *primary* HAPs and their *backup* HAPs. On these links, wavelengths for backup purpose were marked as used.
- $C(u)$ denotes the number of FSO devices to be on HAP u according to topology T . $C(u)$ includes all serving FSO devices of HAP u and one FSO device as the endpoint for each bidirectional inter-HAP link adjacent to u .

Algorithm 1 shows the pseudo-code of the proposed algorithm. For each demand $(s, d, n) \in \mathbb{M}^{HAP}$, we find n lightpaths from s to d , each one uses a single wavelength and possibly take a individual route. The main steps are explained bellow:

- Step 1 (lines 13–23 of Algorithm 1): Assume that the current *least used* wavelength is w , we find the weighted shortest path from s to d in G_w using Algorithm 2. Algorithm 2, proposed in [16], is a modified Dijkstra algorithm, which finds the shortest path having end-to-end BER under threshold δ . The found path is called p . Each arc of p , that is not in T yet, is checked to see if both HAPs adjacent to it having free payload capacity for installing an additional FSO device for the arc. If at least one HAP violates the constraint, the arc is removed from G_w , and another path will be sought using Algorithm 2 again. This step could be repeated until no violation is found, or no path is found.
- Step 2: If no path is found on G_w for the current demand, then try to find a path on G_j , where j is the *next least used wavelength* and so on (line 24). When all wavelengths have been screened but no path is found, this demand is rejected (lines 46–47).
- Step 3: When no violation is found, the path is good for the current demand, wavelengths are assigned to it to make lightpaths:
 - The new arcs of p are added to T and the first lightpath for the demand (s, d, n) is recorded as using wavelength w along path p (lines 26–30).

Algorithm 1 Topology design and wavelength routing algorithm

```

1: function MAIN
2:   for all  $r(s, d, n) \in \mathbb{M}^{HAP}$  do
3:     ROUTING-ONE-DEMAND( $s, d, n$ )
4:   end for
5: end function
6: function ROUTING-ONE-DEMAND( $s, d, n$ )
7:    $T \leftarrow \emptyset$ 
8:   LP-set  $\leftarrow$  NULL  $\triangleright$  set of lightpaths to be found
9:   nb-trial-w  $\leftarrow$  0  $\triangleright$  number of trial wavelengths
10:  repeat
11:     $w \leftarrow$  the least used wavelength
12:    nb-trial-w ++
13:    repeat
14:       $p \leftarrow$  SP-CONSTRAINT( $G_w, s, d$ )
15:      valid-path  $\leftarrow$  true
16:      for all  $(u, v) \in p$  and  $(u, v) \notin T$  do
17:        if  $C(u) \geq \mathbb{C}$  or  $C(v) \geq \mathbb{C}$  then
18:          valid-path  $\leftarrow$  false
19:           $G_w \leftarrow G_w \setminus (u, v)$ 
20:          break;
21:        end if
22:      end for
23:      until  $p$  not found or valid-path = true
24:      if  $p$  not found then next  $\triangleright$  try the next
wavelength
25:      end if
26:      if valid-path then  $\triangleright$  Assign wavelengths
27:        for all  $(u, v) \in p$  do
28:           $T \leftarrow T \cup (u, v)$   $\triangleright$  add new arcs to  $T$ 
29:        end for
30:        LP-set  $\cup$  path  $p$  wavelength  $w$ .
31:        remain-LP  $\leftarrow n - 1$ 
32:        while remain-LP > 0 and nb-trial-w <  $\mathbb{W}$ 
do
33:           $w_k \leftarrow$  the next least used wavelength
34:          if  $w_k$  is available along  $p$  then
35:            LP-set  $\cup$  path  $p$  wavelength  $w_k$ 
36:            remain-LP  $\leftarrow$  remain-LP - 1
37:          end if
38:        end while
39:        if remain-LP > 0 then
40:           $\mathbb{M}^{HAP} \leftarrow \mathbb{M}^{HAP} \cup (s, d, \text{remain-LP})$ 
41:        else
42:          break
43:        end if
44:      end if
45:      until nb-trial-w =  $\mathbb{W}$ 
46:      if remain-LP > 0 then
47:        return false  $\triangleright$  Reject demand
48:      end if
49:      return LP-set
50: end function

```

- If the number of requested lightpaths $n > 1$, the remaining number of lightpaths to be created is $remain - LP = n - 1$. We try to make the remaining lightpaths by using always path p but other wavelengths starting from the *least used* wavelength available first (lines 31–38).
- If there are still not enough available wavelengths to route the entire demand over p , the un-routed lightpaths should follow a different path. We consider the remaining part of the current demand as a new one with parameters $(s, d, remain - LP)$ and add it to \mathbb{M}^{HAP} for later handling (lines 39–40).

At the end of the algorithm, we obtain the topology T for the HAP network and all lightpaths for the accepted demands in \mathbb{M}^{HAP} .

Algorithm 2 Shortest path with BER constraint

```

1: function SP-CONSTRAINT( $\mathbb{G}, s, d$ )
2:    $Q \leftarrow$  vertex set of  $\mathbb{G}$ 
3:   for all vertex  $v \in Q$  do
4:      $dist[v] \leftarrow$  INFINITY
5:      $prev[v] \leftarrow$  UNDEFINED
6:   end for
7:    $dist[s] \leftarrow 0$ 
8:   while  $Q \neq \emptyset$  do
9:      $u \leftarrow$  vertex in  $Q$  with min  $dist[u]$ 
10:    for all neighbor  $v$  of  $u$  do
11:       $alt \leftarrow dist[u] + length(u, v)$ 
12:      if  $alt < dist[v]$  and  $BER-e2e(s,u,v)$  then
13:         $dist[v] \leftarrow alt$ 
14:         $prev[v] \leftarrow u$ 
15:      end if
16:    end for
17:  end while return  $dist[d], prev[d]$ 
18: end function
19: function BER-E2E( $s,u,v$ )
20:    $prod \leftarrow (1 - BER_{(u,v)})$ 
21:   for  $\ell \in$  current path from  $s$  to  $v$  do
22:      $prod \leftarrow prod \times (1 - BER_{\ell})$ 
23:   end for
24:   if  $prod > 1 - \delta$  then
25:     return true
26:   else
27:     return false
28:   end if
29: end function

```

Figure 6 shows the topology of the HAP network designed by the proposed algorithm for the test case shown in Figure 5. Blue dots are HAPs providing backup function, and blue dashed lines are inter-HAP links between a pair of primary and backup HAPs.

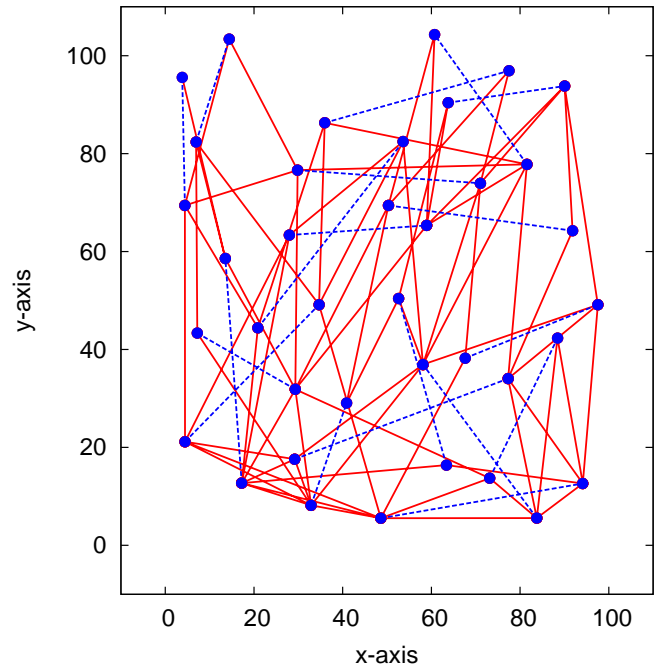


Figure 6: HAP network designed by the proposed algorithms. Blue dots are HAPs acting as backup HAPs, blue dashed lines are inter-HAP links between a pair of primary and backup HAPs.

5. Numerical results

We implemented the algorithms in Section 4, including FSO clustering, backup matching and topology design steps. Since there is no previous work on dimensioning problem for survivable FSO networks using HAPs against optical uplink and downlink failures, thus we can not evaluate our dimensioning solution against any previous one. We will thus concentrate on analyzing the impact of protection on topologies by comparing the topologies with protection against the topologies without protection. Topologies without protection were obtained by running only the clustering and topology design steps.

The comparisons were made on different test cases. The ground FSO node locations were generated randomly on a square surface of $100\text{km} \times 100\text{km}$, which is the size of a metropolitan area. The total number of ground FSO nodes was generated randomly in the range of 100 to 4000 nodes. The following parameters were set based on the technological and theoretical review in [7] on optical communications for HAP. The transmission rate per wavelength r^{max} was set to 1 Gbps. The aperture of *servicing* FSO devices was set approximately 40° ; thus, the ground coverage diameter of a *servicing* FSO was $\mathbb{D} \sim 15$ km. Traffic requirement was generated randomly between ground FSO nodes, but the total incoming or outgoing traffic of a ground FSO node does not exceed 1 Gbps, which is the capacity of a single wavelength.

The BER of an inter-HAP link was calculated according to air turbulence model using Gamma-Gamma distri-

bution in [15] with moderate turbulence condition. BER threshold for an end-to-end lightpath as well as for a single inter-HAP link was set as $\delta = 10^{-3}$, so that the errors could be recover by current FEC techniques. According to the BER computation model, inter-HAP link longer than 60 km having BER greater than $\delta = 10^{-3}$, therefore, L^{HH} was set to 60 km. The wavelength density of the WDM technique was $\mathbb{W} = 128$ wavelengths. The recapitulation of the parameter values is shown in Table 2.

In all test cases, we observed no demand rejection either with or without protection. We will discuss the impact of protection on the number of HAPs and the number of FSO devices on HAPs to be invested. We also analyze the resource occupation with and without protection.

5.1. Availability of end-to-end connections

Although the links from a ground node to its primary and backup HAPs are not influenced simultaneously by a cloud, they can still be interrupted by two independent clouds.

Let the probability that a link between a ground node and a HAP is cut by clouds be p_{cut} . A connection between two ground nodes through intermediate HAPs contains three legs: from the source ground node to the source HAP, from the source HAP to the destination HAP, and from the destination HAP to the destination ground node. The links between HAPs are not impacted by clouds thus the availability of the second leg is considered as 1. The availability of the end-to-end connection without protection between two ground nodes is

$$A_{e2e} = (1 - p_{cut})^2 \quad (4)$$

The probability that two links from a ground node to its *primary* HAP and *backup* HAP are cut by two independent clouds is p_{cut}^2 . Consequently, the probability that one of the two links is available for transmission is $1 - p_{cut}^2$. The availability of the end-to-end connection with the proposed protection scheme is

$$A_{e2e}^p = (1 - p_{cut}^2)^2 \quad (5)$$

Figure 7 illustrates the availability attained with and without the proposed protection mechanism when the probability of cloud cut on a single ground-HAP link increase from 0 to 1. It is clear that the protection offers better link availability.

In the Southern of Europe, the availability of a single ground-HAP link varies between 50% – 85% over a year [14] which is equivalent to cloud cut probability $p_{cut} = 0.15 \sim 0.5$. Without protection, the availability of an end-to-end connection is between 25% and 72.25%. The proposed protection mechanism allows increasing the availability to 56.25% – 95.55%.

In the Northern of Europe, the mean availability of a single link is lower than 40% that means $p_{cut} > 0.6$. The end-to-end availability without protection is up to 16%, the protection helps to increase the availability to 40.96%.

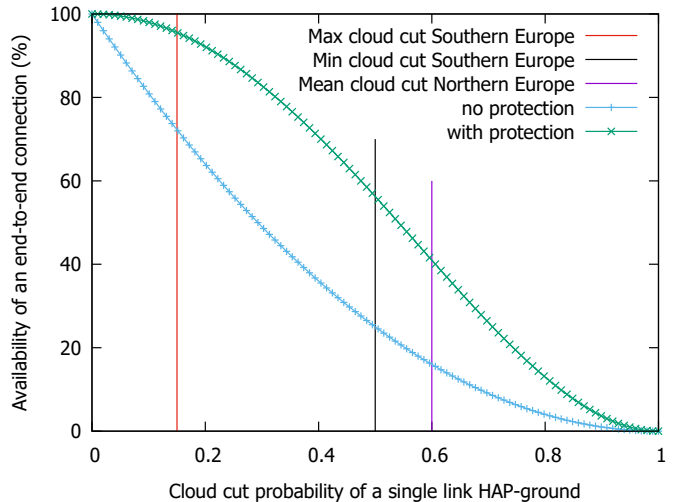


Figure 7: Availability of an end-to-end connection between two ground nodes

More than one slant link might help to improve the availability. Indeed, if m *backup* HAPs are used for each *primary* HAP the end-to-end availability of a connection becomes:

$$A_{e2e}^{mp} = (1 - p_{cut}^m)^2 \quad (6)$$

With $m = 4$, the end-to-end availability of a connection under the lowest cloud cut probability of Southern of Europe ($p_{cut} = 0.15$) reaches to 99.89%. This number is equivalent to what reported by [14] when a HAP sent data to 4 diverse ground stations to avoid clouds. The improvement is much better in regions with high cloud cut probability. For example, when $p_{cut} = 0.6$ the availability is 75.75% with $m = 4$ in comparison with 40.96% when $m = 2$.

Formula (6) shows also that when the single link availability is already good, for example $p_{cut} < 0.1$, adding more backup HAPs does not help to increase much the availability. However, such a small improvement trades off in cost of backup HAPs and backup FSO transceivers since many backup HAPs and inter-HAP links are needed. Therefore, multiple backup HAPs should be considered only in cloudy regions or cloudy seasons. In this research, we focus on single backup HAP only.

5.2. Number of additional HAPs due to protection

Figure 8 shows the number of HAPs as a function of the number of ground nodes. This shows that protection requires *at most one more HAP* in all test cases. The reason is that: if HAPs still have free payload for installing backup serving FSO devices, *serving* HAPs of different clusters backup each other. The additional HAP is only involved when the number of clusters is impaired. Thus, *the protection does not lead to a higher cost of HAPs if the HAPs are not too loaded prior protection.*

5.3. Number of FSO devices on HAPs

Figure 9 shows the total number of FSO devices to be installed on HAPs with and without protection. Evidently, cases with protection require additional FSO devices to serve ground nodes in the protected zones and to connect the primary and backup HAPs. Nonetheless, the total number of FSO devices to be deployed on HAPs increases only by 5%–12% owing to protection.

Figure 10 shows the average number of FSO devices that a HAP carries with and without protection. We observe that, on average, each HAP has to carry from 0.45–0.95 more FSO devices for protection purpose.

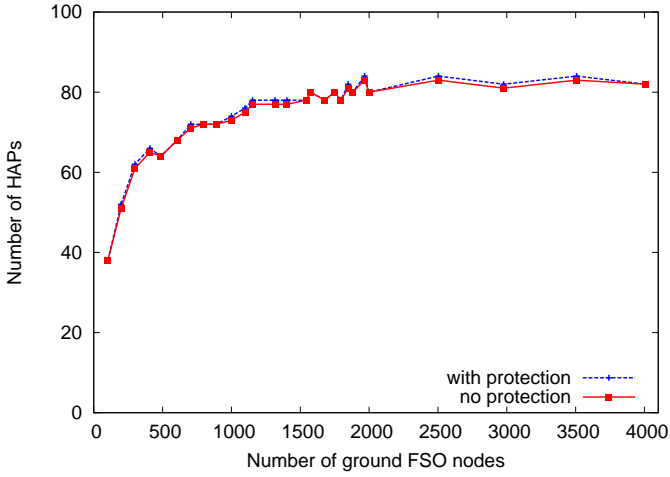


Figure 8: Number of HAPs

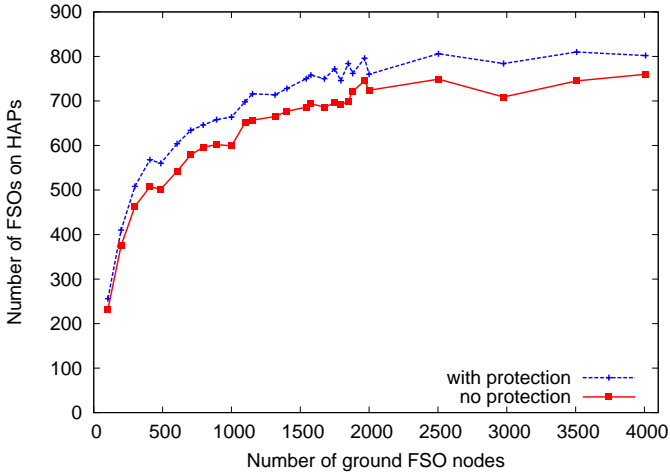


Figure 9: Total number of FSO devices carrying by all HAPs

5.4. Number of additional inter-HAP links and occupancy due to protection

Figure 11 shows the number of inter-HAP links when the number of ground nodes increases. Since traffic demands were generated between all ground nodes, the volume of traffic increases proportionally with the number

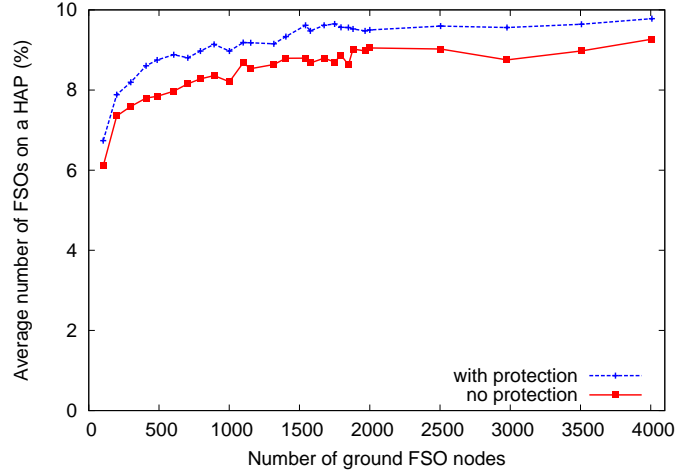


Figure 10: Average number of FSO devices carrying by one HAP

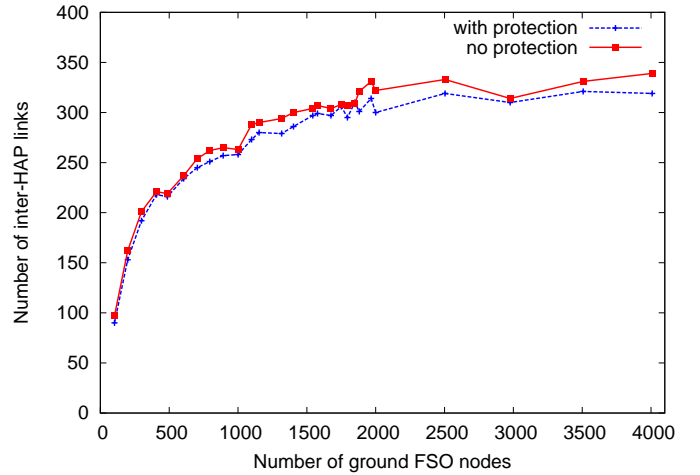


Figure 11: Number of inter-HAP links

of ground nodes. As a result, more inter-HAP links are involved in networks with larger number of ground nodes.

Counter intuition, the statistics demonstrate that the topology with protection generally uses fewer links than the topology without protection. The difference becomes clearer when the number of ground nodes increases. We believe that this phenomenon is due to the characteristics of the proposed algorithm. In the backup matching step, a number of inter-HAP links connecting pairs of primary and backup HAPs are included in the initial topology T . Consequently, the topology design algorithm frequently uses these links, resulting in fewer added links than in the no-protection scenario. We believe that by modifying the topology design algorithm, it is possible to achieve a topology without protection with fewer links.

The analysis of link occupation confirms that links are exploited more intensively in topologies with protection. We define link occupation as the ratio between the number of wavelengths used and the total number of wavelengths

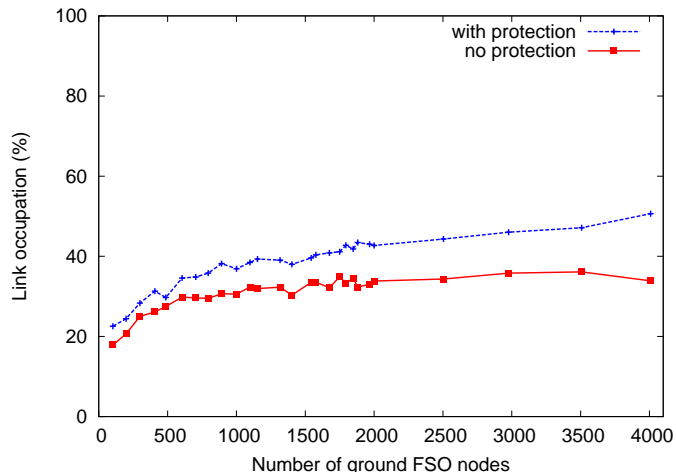


Figure 12: Link occupation

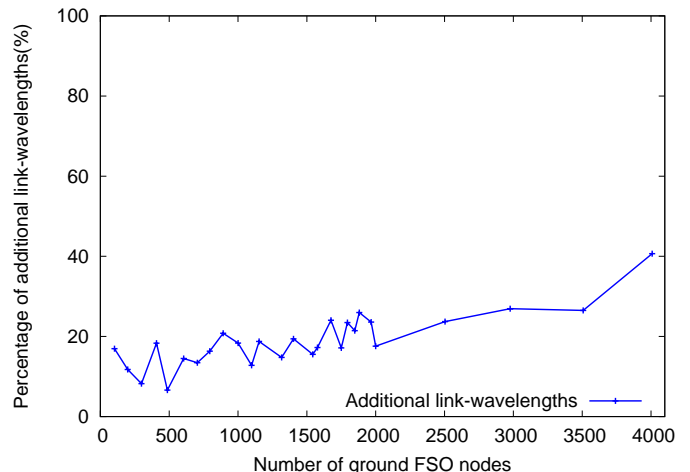


Figure 13: Percentage of additional wavelength resources for protection in comparison with no protection

of the entire network:

$$\text{link occupancy} = \frac{\text{Number of wavelengths used on all links}}{\text{Number of inter-HAP links} \times 2W}. \quad (7)$$

Coefficient 2 in the denominator is introduced because a link is bidirectional, and wavelengths are counted separately in opposite directions.

Figure 12 shows that link occupation in topology with protection is clearly higher than link occupation in topology without protection mostly when the number of ground FSO nodes increases.

5.5. Wavelength cost for protection purpose

We evaluated the impact of protection on the quantity of wavelength resources in the network topology. Let *link-wavelength* be a wavelength of a link. The number of link-wavelengths increases when more wavelengths are used in many links. We calculated the total number of link-wavelengths used in all inter-HAP links of the topologies with and without protection for each test case. Figure 13 shows the percentage of additional link-wavelengths used for protection. The topologies with protection use from 8% to 40% more link-wavelengths than those without protection. The amount of additional link-wavelengths is modest because in the case with protection, demands travel over only one more link (between the backup and primary HAPs) compared to the case without protection.

6. Conclusions

HAPs as intermediate systems for long-distance FSO communication are promising since they provide broadband communication. The presence of clouds in the middle of the uplinks and downlinks between HAPs and the ground is a major obstacle. In this research, we proposed a switching and protection mechanisms as well as a network dimensioning solution for survivable end-to-end communication against clouds. Data travel on both primary and

backup paths using uniquely FSO communication technology. The simulation results show that a network with survivable capability is not necessary much more expensive than that without protection in terms of HAP and FSO investment costs, meanwhile connection availability is significantly improved and can reach to 95.55% based to Europe cloud coverage data.

The network dimensioning would be better optimized if the apertures of different *servicing* FSOs could be adjusted independently leading to various cluster sizes. In a future work, we will address the *servicing* FSO apertures optimization problem.

Acknowledgements

This research was funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.02-2018.305.

References

- [1] L. D. Truong, H. T. T. Pham, N. T. Dang, T. V. Doan, Topology design and cross-layer optimization for FSO mesh networks impaired by atmospheric turbulence and misalignment fading, *IEEE/OSA Journal of Optical Communications and Networking* 9 (12) (2017) 1097–1107. doi:10.1364/JOCN.9.001097.
- [2] F. D’Andreagiovanni, D. Nace, M. Pióro, M. Foss, M. Shehaj, A. Tomaszewski, On robust FSO network dimensioning, in: 2017 9th International Workshop on Resilient Networks Design and Modeling (RNDM), 2017, pp. 1–8. doi:10.1109/RNDM.2017.8093027.
- [3] D. Nace, M. Pióro, M. Poss, F. D’Andreagiovanni, I. Kalesnikau, M. Shehaj, A. Tomaszewski, An optimization model for robust fso network dimensioning, *Optical Switching and Networking* 32 (2019) 25 – 40. doi:https://doi.org/10.1016/j.osn.2018.11.004.
- [4] M. Knapek, J. Horwath, F. Moll, B. Epple, N. Courville, H. Bischl, D. Giggenbach, Optical high-capacity satellite downlinks via high-altitude platform relays, in: A. K. Majumdar, C. C. Davis (Eds.), *Free-Space Laser Communications VII*, Vol. 6709, International Society for Optics and Photonics, SPIE, 2007, pp. 107 – 118. doi:10.1117/12.734053.

- [5] H. Y. Song, A method of mobile base station placement for High Altitude Platform based network with geographical clustering of mobile ground nodes, in: 2008 International Multiconference on Computer Science and Information Technology, 2008, pp. 869–876. doi:10.1109/IMCSIT.2008.4747344.
- [6] M. Sharma, D. Chadha, V. Chandra, High-altitude platform for free-space optical communication: Performance evaluation and reliability analysis, *IEEE/OSA Journal of Optical Communications and Networking* 8 (8) (2016) 600–609. doi:10.1364/JOCN.8.000600.
- [7] F. Fidler, M. Knappek, J. Horwath, W. R. Leeb, Optical Communications for High-Altitude Platforms, *IEEE Journal of Selected Topics in Quantum Electronics* 16 (5) (2010) 1058–1070. doi:10.1109/JSTQE.2010.2047382.
- [8] B. Moision, B. Erkmen, E. Keyes, T. Belt, O. Bowen, D. Brinkley, P. Csonka, M. Eglington, A. Kazmierski, N. hyong Kim, J. Moody, T. Tu, W. Vermeer, Demonstration of free-space optical communication for long-range data links between balloons on Project Loon, in: H. Hemmati, D. M. Boroson (Eds.), *Free-Space Laser Communication and Atmospheric Propagation XXIX*, Vol. 10096, International Society for Optics and Photonics, SPIE, 2017, pp. 259 – 272. doi:10.1117/12.2253099.
- [9] C. Chen, A. Grier, M. Malfa, E. Booen, H. Harding, C. Xia, M. Hunwardsen, J. Demers, K. Kudinov, G. Mak, B. Smith, A. Sahasrabudhe, F. Patawaran, T. Wang, A. Wang, C. Zhao, D. Leang, J. Gin, M. Lewis, D. Nguyen, K. Quirk, High-speed optical links for UAV applications, in: H. Hemmati, D. M. Boroson (Eds.), *Free-Space Laser Communication and Atmospheric Propagation XXIX*, Vol. 10096, International Society for Optics and Photonics, SPIE, 2017, pp. 316 – 324. doi:10.1117/12.2256248.
- [10] Thales group, What’s up with stratobus, <https://www.thalesgroup.com/en/worldwide/space/news/whats-stratobus>, accessed Jan. 2022.
- [11] G. Karabulut Kurt, M. G. Khoshkholgh, S. Alfattani, A. Ibrahim, T. S. J. Darwish, M. S. Alam, H. Yanikomeroglu, A. Yongacoglu, A Vision and Framework for the High Altitude Platform Station (HAPS) Networks of the Future, *IEEE Communications Surveys & Tutorials* 23 (2) (2021) 729–779. doi:10.1109/COMST.2021.3066905.
- [12] S. C. Arum, D. Grace, P. D. Mitchell, M. D. Zakaria, N. Morozs, Energy management of solar-powered aircraft-based high altitude platform for wireless communications, *Electronics* 9 (1) (2020). doi:10.3390/electronics9010179. URL <https://www.mdpi.com/2079-9292/9/1/179>
- [13] M. Mohorčić, A. Vilhar, M. Berioli, A. Werner, A. Donner, Optical transport network based on a meshed HAP system with interplatform links, in: *Proc. Advanced Satellite Mobile Systems (ASMS) Conference*, 2005, pp. 29–31.
- [14] F. Moll, M. Knappek, Wavelength selection criteria and link availability due to cloud coverage statistics and attenuation affecting satellite, aerial, and downlink scenarios, in: A. K. Majumdar, C. C. Davis (Eds.), *Free-Space Laser Communications VII*, Vol. 6709, International Society for Optics and Photonics, SPIE, 2007, pp. 347 – 358. doi:10.1117/12.734269.
- [15] N. T. T. Nguyen, M. B. Vu, H. T. Le, V. V. Mai, N. T. Dang, Hap-based multi-hop fso systems using all-optical relaying and coherent receiver, in: 2019 6th NAFOSTED Conference on Information and Computer Science (NICS), 2019, pp. 119–124. doi:10.1109/NICS48868.2019.9023866.
- [16] D. L. Truong, N. T. Dang, P. M. Hien Vu, Dimensioning of free-space optical mesh network using high-altitude platforms, in: 2019 6th NAFOSTED Conference on Information and Computer Science (NICS), 2019, pp. 125–129.
- [17] fSONA, SONABeam 2500-E+ model specifications, <http://fsona.com>, accessed Jan. 2022.
- [18] J. Edmonds, Paths, trees, and flowers, *Canadian Journal of Mathematics* 17 (1965) 449–467. doi:10.4153/CJM-1965-045-4.