Network Architectures for Load Balancing in Multi-HAP Networks

MIHAEL MOHORCIC and ALES SVIGELJ
Department of Communication Systems
Jozef Stefan Institute
Jamova 39, SI-1000 Ljubljana
SLOVENIA
http://www-e6.ijs.si

Abstract: - In this paper the network architectures are investigated in a network of high altitude platforms (HAPs) with overlapping coverage areas, focusing on the additional network elements/functionalities required to make use of load balancing mechanisms. According to the location within the network and the complexity of additional equipment/functionality the paper proposes two different architectures. In the case of the basic utilisation of multiple HAPs a new equipment/functionality is installed only at the user premises, supporting load balancing only for the connections initialised from the HAP user network to the Internet (i.e. the outbound connections). In the case of advanced utilisation of multiple HAPs a new equipment/functionality is needed both at the user premises and in the HAP network, thus supporting load balancing in both directions.

Key-Words: - High Altitude Platforms, network architectures, load balancing

1 Introduction

In the last decade high altitude platforms (HAPs) have been extensively investigated as an alternative infrastructure for provision of different applications, with the main focus on provision of wireless communications. Equipped with communications payload HAPs will be typically operating in lower stratosphere at altitudes around 20 km above the ground, thus providing broadband wireless access (BWA) particularly to remote and sparsely populated areas [1, 2]. Station-keeping at/around the nominal position, while in operation, depends on the type of aeronautical platform chosen for the mission (e.g. balloon, airship, manned/unmanned airplane). However, regardless of the choice of the platform, from the perspective of the user on the ground HAPs are considered quasi stationary objects in the sky providing notably larger coverage areas compared to terrestrial BWA systems. HAPs are seen as particularly attractive solution for on-demand establishment of a wireless communication access for a specific geographical region, for instance for disaster relief and special event servicing, as well as for the gradual service roll-out with incremental deployment of the network as the need emerges for larger coverage area and/or capacity. In particular, multiple HAPs can be deployed by the same network operator to serve a common coverage area in order to increase the capacity provided and/or to improve resilience. However, multiple HAPs can be installed also if more than one HAP network operator exists in the same geographical area. In general, there are three different cases with regard to coverage areas of multiple platforms:

- The coverage areas of HAPs in the system are not overlapping. This case is not interesting for further investigation as the user can only be in the coverage area of a single HAP, so it cannot exploit the benefits of multiple HAP constellations.
- The coverage areas of HAPs in the system are partially overlapping. This is the most likely case in the network deployment as in the real HAP network the coverage areas will have to be partially overlapped to support seamless handover between different HAPs.
- The coverage areas of HAPs in the system are fully overlapping. This situation is expected in the latter deployments of HAP systems where multiple HAPs are serving the same area to increase the overall capacity or improve resilience, or simply because of additional HAP network provider entering the market.

Constellations of multiple HAPs with partially/completely overlapping coverage areas have been shown (i) to enhance the overall system capacity by exploiting highly directional fixed user antennas used to discriminate spatially between different HAPs [3], essentially increasing the spectral efficiency of the system, or (ii) to increase the total link availability between the HAP and a user by exploiting the diversity gain in mobile user environment [4].

In this paper the network architecture implications of using multiple platforms with partially/completely overlapping coverage areas are investigated. In particular, we investigate the load balancing in the multiple platform constellations by introducing additional network elements/functionalities. In Section 2 we present a general HAP system architecture and provide background assumptions for this study. The basic utilisation of multiple HAP constellations for load balancing in outbound direction (i.e. connections initialised from the HAP user network to the Internet) is discussed in Section 3, essentially distinguishing between and comparing per-packet, per-destination and per-flow load balancing mechanisms. Section 4 extends the capabilities of load balancing to both inbound and outbound directions, while Section 5 concludes the paper.

2 System Architecture

From the system architecture point of view HAPs can be used in different configurations. The most general HAP system architecture including all partial network scenarios discussed in the following is depicted in Fig. 1.

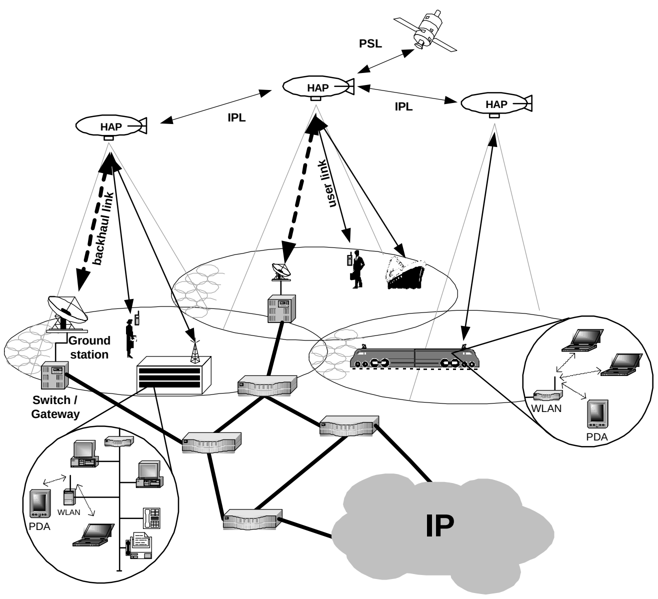

Fig. 1 General HAP System Architecture

The simplest HAP system configuration consists of standalone platform where the system coverage is limited to a single HAP cellular coverage. Only communication between fixed, portable and mobile user terminals within this coverage is enabled. Additionally, connection to other public and/or private networks via Gateway Station (GS) is foreseen. This scenario can be further divided into two distinct topologies taking into account where the switching is taking place [5]:

- Bent-pipe standalone platform (scenario with on ground switching), where the path between two users encompasses uplink from the user to the platform, feeder downlink to GS, where the switching is performed, feeder uplink from GS to the platform and downlink to the target user.
- Standalone platform with onboard switching, where the path between two users takes only uplink from the user to the platform, where switching is performed; and downlink from the platform to the target user. Standalone platform scenario with onboard switching is particularly suitable for temporary provision of basic or additional capacity required for the short term events with a large number of participants, in case of natural disasters, or in areas where the fixed infrastructure has suffered a major failure.

In the case of multiple platform constellation HAPs can be interconnected via ground stations or by the interplatform links (IPL) forming a network of HAPs, thus arbitrarily extending the system coverage.

The most extensive scenario, as depicted in Fig. 1, can also include platform to satellite links (PSL), which are particular useful if HAPs are placed above the areas with deficient (rural and remote areas) or non-existent terrestrial infrastructure. Using PSL HAPs can be connected to other remote public or private networks. Furthermore, PSLs could also be used as a backup solution in the case when the connection with the rest of the network via IPLs or GSs is disabled due to a failure or extreme rain fading on up/down link segment.

The scenario used in this paper represents a subset of the general HAP system architecture shown in Fig. 1 and comprises the fixed users which are in the coverage area of at least two HAPs. In addition, we distinguish if the available HAPs are operated by the same network provider or they belong to different providers. However, we do not distinguish between fully and partially overlaid coverage areas as from the fixed user perspective it is not important as long as it is within the overlapping area. We are only focusing on solutions which in general do not need changes at the application level, thus from the users perspective the usage of more than one HAP is transparent resulting only in better performance and/or better reliability/availability of the services. A typical network architecture which is investigated in this paper is depicted in Fig. 2.

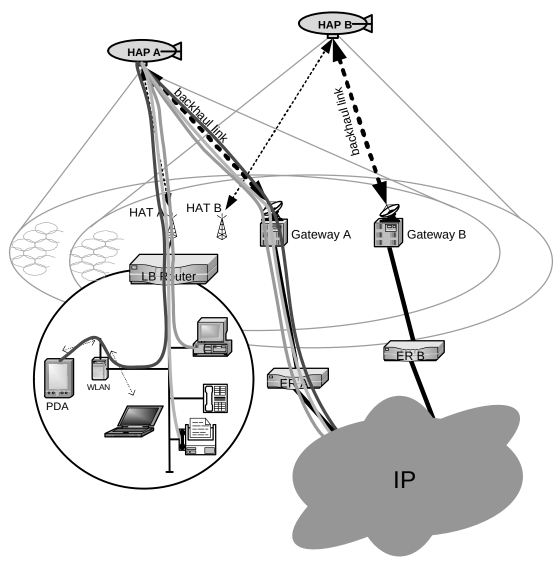

Fig. 2. Traffic flows if only one HAP is utilized.

We assume that there are one or more users which are connected to the HAP network via router (i.e. LB Router) using wired or wireless access (e.g. WLAN). In the original scenario depicted in Fig. 2, with different shades of gray representing different flows, these users are utilizing only one HAP (i.e. HAP A), although they are also within the coverage area of HAP B. All users (and applications) are accessing internet via the same link, which can cause the congestion in the access segment and/or backhaul link.

In order to utilize also the second HAP (i.e. HAP B) we can make use of different scenarios. They can be divided upon the position and complexity of additional equipment/functionality in the network:

- Basic utilisation of multiple HAPs – in this scenario the new equipment/functionality is installed only at the user premises.
- Advanced utilisation of multiple HAPs – in this scenario the new equipment/functionality is needed at the user premises and also in the HAP network (i.e. between the Gateway and Edge Router (ER).

3 Basic Utilisation of Multiple HAPs for Load Balancing

This solution is based on installing a router with load balancing capabilities at the user premises. The router can combine two or more broadband connections, at least summing up the amount of bandwidth which is available to users and at the same time creating a more resilient solution. In this solution the load balancing works only when connections are initialised from the HAP user network to the Internet (i.e. outbound direction). Typically, the load balancing can work per-packet, per-destination or per-flow.

Per-packet load-balancing means that the router sends one packet for destination1 over the first path, the second packet for (the same) destination1 over the second path, and so on [6]. Typically the round robin scheduling policy is applied. Per-packet load balancing guarantees equal load across all links. However, packets may arrive at the destination out of order because different delays may exist within the network. For per-packet load balancing, the forwarding process determines the outgoing interface for each packet by looking up the route table and picking the least used interface. This ensures equal utilization of the links, but is a processor intensive task and impacts the overall forwarding performance. This form of per-packet load balancing is not well suited for high speed interfaces. Graphical representation of the per-packet load balancing is depicted in Fig. 3, where different shades of gray represent different flows. It is clearly seen that packets belonging to the same flow are splitting at LB router thus utilizing both HAPs and increasing the throughput. From the user perspective the overall capacity a single application can achieve is the sum of the capacities of each particular connection.

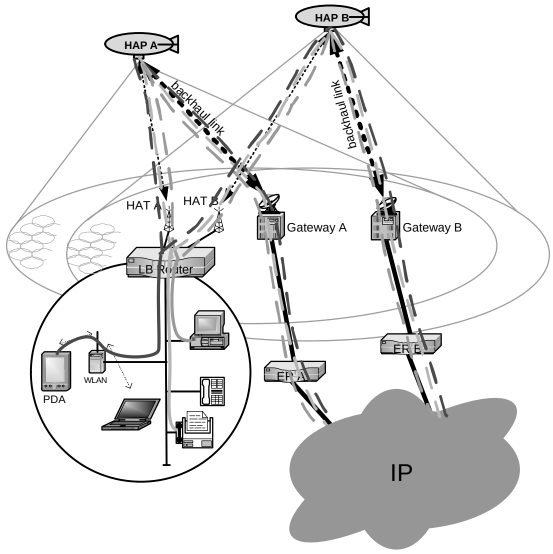

Fig. 3. HAP architecture for basic utilization of multiple HAPs: per-packet load balancing.

Per-destination load balancing means that the router distributes the packets based on the destination address [7]. Given two paths to the same network, all packets for destination_1 on that network go over the first path, while all packets for destination_2 on that network go over the second path. Also in this case the round robin scheduling policy is typically applied. This solution preserves the packet order, yet potentially yields unequal utilisation of the links. If one host receives the majority of the traffic all packets use a single link, leaving bandwidth on other links unused. A larger number of destination addresses leads to more equally used links. To achieve a more equal link utilisation a route-cache entry has to be built for every destination address, instead of every destination network, as is the case when only a single path exists. Therefore traffic for different hosts on the same destination network can use different paths. The downside of this approach is that for core backbone routers carrying traffic for thousands of destination hosts, memory and processing requirements for maintaining the cache become very demanding. From the user perspective the overall capacity a single application can achieve is the same as the capacity of each particular connection. However the overall capacity for all applications/users is increased.

Per-flow load balancing means that connections or flows, are shared between users [7]. The result is that, whilst a single flow cannot use more bandwidth that provided by a single link, multiple users/applications balance across multiple links. Graphical representation is depicted in Fig. 4. Two flows are using the first broadband connection via HAT A, while the third flow is utilizing the second HAP (HAP B), thus not utilising the first HAP at all.

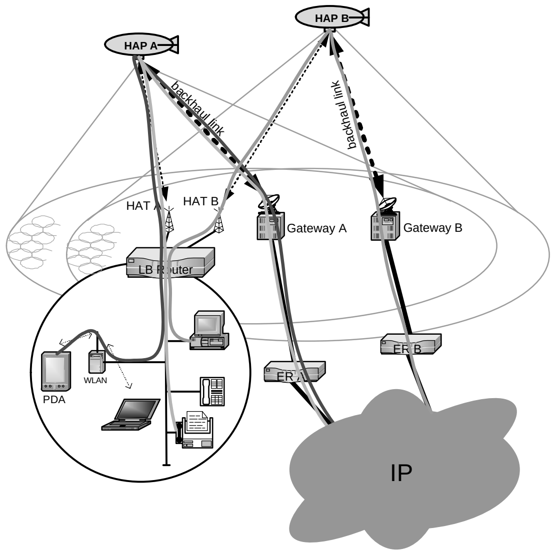

Fig. 4. HAP architecture for basic utilization of multiple HAPs: per-flow load balancing.

Routers with per-flow load balancing capabilities have a range of methods by which they balance traffic between the broadband links. Typical criteria, which are used by the router when deciding which connection to use, include:

- Round robin – every new flow is using different link from the previous one.
- Least connections – the new flow is established via the broadband connection with less active connections.
- Traffic load – the new flow is established via the broadband connection with less traffic load.
- Best quality – the new flow is established via the broadband connection with better quality.
- Least hops – the new flow is established via the broadband connection with less hops to the final destination.
- Scheduled – the new flow is established according to the predefined time schedule.
- Traffic type – particular traffic classes are utilizing always the same HAP.

Advantages and disadvantages of Per-packet, Per-destination and Per-Flow load balancing are summarised in Table 1. It is worth noting that presented solution is independent of HAP network operator (i.e. each HAP can belong to different network domain / ISP). The capacity can be increased also if HAPs are run by two different providers. The main disadvantage of the basic utilisation of multiple HAPs is that it is outbound only, which means that inbound traffic is forced to traverse the dedicated connection (link).

4 Advanced Utilization of Multiple HAPs for Load Balancing

In order to provide load balancing in both directions (i.e. outbound and inbound) additional router with load-balancing capability should be added to the network architecture. The proposed architecture is depicted in Fig. 5. The main difference comparing to the basic utilization of multiple HAPs is that the flows can be split also in the inbound direction in LB router 2. It is worth pointing out that in this case both HAPs should be run with the same network operator.

Table 1. Advantages and disadvantages of different load balancing mechanisms.

	Load Balancing Mechanism		
	Per-packet	Per-destination	Per-flow
Advantages	Per-packet load balancing allows the router to send successive data packets over paths without regard to individual hosts or user sessions. Allows more evenly loaded links.	Packets for a given destination / source-destination host pair are guaranteed to take the same path, even if multiple paths are available. Traffic destined for different pairs tend to take different paths.	Packets for a given flow are guaranteed to take the same path, even if multiple paths are available. Traffic destined for different flows tend to take different paths.
Disadvantages	Packets for a given source-destination host pair take different paths, which could introduce reordering of packets. This is not recommended for Voice over IP (VoIP) and other flows that require in-sequence delivery.	It may result in unequal distribution with a small number of destination / source-destination pairs. Per-destination load balancing depends on the statistical distribution of traffic; load sharing becomes more effective as the number of destinations / source-destination pairs increases.	It may result in unequal distribution with a small number of flows. Per-flow load balancing depends on the statistical distribution of traffic; load sharing becomes more effective as the number of flows increases.

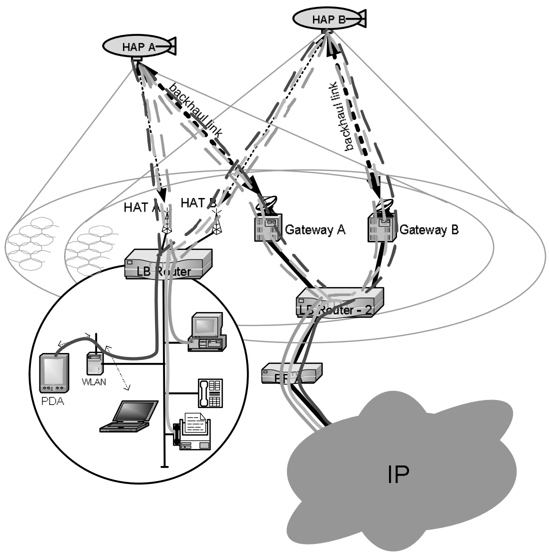

Fig. 5. HAP architecture for advanced utilization of multiple HAPs: per-packet load balancing.

The load balancing in the advanced utilisation of multiple HAPs can be performed with per-packet, as shown in Fig. 5, per-flow or per-destination load balancing mechanism, by using the mechanism of equal cost routes in router LB and router LB-2 for the routes over both HAPs. The same advantages/disadvantages for per-packet and per destination load balancing apply as described in Table 1.

5 Conclusion

In this paper we investigated the network architecture implications of introducing additional network elements/ functionality for the support of load balancing in the network with multiple platforms with partially or completely overlaid coverage areas. We proposed two different architectures with regard to complexity of additional elements/functionality, which should be installed in the system. For the basic utilisation of multiple HAPs an additional network element is added at user's premises, allowing the capacity increase on outbound connections only. The advanced utilisation of multiple HAPs is more complex and requires also the load balancing router at the gateway, allowing the load balancing in both directions, outbound and inbound.

Acknowledgement:

This work has been partially funded by the European Community through the 6th Framework Programme IST project SatNEx (FP6-IST-027393).

References:

[1] G. M. Djuknic, J. Freidenfelds, Y. Okunev, Establishing Wireless Communications Services via High-Altitude Aeronautical Platforms: A Concept Whose Time Has Come?, *IEEE Communications Magazine*, September 1997, pp. 128-135.

[2] T. C. Tozer, D. Grace, High-altitude platforms for wireless communications, *Electron. & Commun. Engineering Journal*, June 2001, pp.127-137.

[3] D. Grace, J. Thornton, G. P. White, T. C. Tozer, Improving spectrum utilisation for broadband services in mm-wave bands using multiple High Altitude Platforms, *IEE Conference – 'Getting the most of the Radio Spectrum*,' London, UK, 24-25 October 2002.

[4] T. Celcer, G. Kandus, T. Javornik, M. Mohorcic, S. Plevel, Evaluation of diversity gain and system capacity increase in a multiple HAP system, *International Workshop on Satellite and Space Communications (IWSSC2006)*, Madrid, Spain, 14-15. September, 2006.

[5] M. Mohorcic, D. Grace, G. Kandus, T. C. Tozer, Broadband Communications from Aerial Platform Networks, *13th IST Mobile and Wireless Communications Summit 2004*, Lyon, France, June 27-30, 2004, pp. 257-261.

[6] http://www.cisco.com

[7] http://broadband.nildram.net