Fault Tolerant Scalable Support for Network Portability and Traffic Engineering

Marcelo Bagnulo[1], Alberto García-Martínez[2], Arturo Azcorra[2]

[1]Huawei Labs at UC3M [2]U. Carlos III de Madrid
Avda de la Universidad, 30, Leganés, 28911 Madrid
{marcelo, alberto, azcorra}@it.uc3m.es

Abstract. The P-SHIM6 architecture provides ISP independence to IPv6 sites without compromising scalability. This architecture is based on a middle-box, the P-SHIM6, which manages the SHIM6 protocol exchange on behalf of the nodes of a site, which are configured with provider independent addresses. Incoming and outgoing packets are processed by the P-SHIM6 box, which can assign different locators to a given communication, either when it is started, or dynamically after the communication has been established. As a consequence, changes required for provider portability are minimized, and fine-grained Traffic Engineering can be enforced at the P-SHIM6 box, in addition to the fault tolerance support provided by SHIM6.

1 Introduction[1]

The SHIM6 architecture [1] provides scalable support for IPv6 end site multihoming. As opposed to the BGP-style of multihoming, where the multihomed site injects its own prefix through the different providers, in the SHIM6 approach a multihomed site obtains a Provider Aggregatable (PA) prefix from each of its providers' address blocks. This fosters prefix aggregation in the global routing table, since the multihomed site prefixes do not need to be announced independently in the global routing table and only PA prefixes corresponding to the ISPs are announced. From the multihomed site perspective, this configuration results in the presence of multiple prefixes in the site (one per provider) and multiple global addresses configured in the hosts (again, one per provider).

The goal of the SHIM6 architecture is to preserve established communications through outages in the paths to a multihomed site with multiple addresses. The SHIM6 protocol [2] is an end-to-end protocol that is used between the peers of a communication to securely create SHIM6 contexts that contain the different addresses available for the communication. The SHIM6 architecture defines a SHIM6 sublayer located between the IP endpoint sublayer and the IP forwarding sublayer. This sublayer uses the SHIM6 context state to map the addresses used by the upper layers

[1] This work has been supported by the RiNG project IST-2005-035167 and by the IMPROVISA project TSI2005-07384-C03-02.

(known as *Upper Layer Identifiers*, *ULID*) and the actual addresses used for packet forwarding (called *locators*). In case that a failure is detected in the communication path, any of the alternative addresses stored in the SHIM6 context can be used as a new locator, while ULIDs are presented unchanged to the upper layers.

However, the SHIM6 protocol fails to provide some key features of the current BGP-based approach to multihoming. In particular, SHIM6 fails to provide the portability of the address block that is used by the multihomed site. This basically means that when a multihomed end-site changes one of its providers, the addresses, that were associated with this ISP, need to be changed in a process known as *renumbering*. Renumbering may be a costly and painful process, so imposing it when changing providers does increase provider lock-in. Moreover, another capability missing in the SHIM6 architecture is traffic engineering policy enforcement. In the BGP-based multihoming framework site administrators can deeply influence the links through which ingress and egress traffic is exchanged. In this way, objectives such as balancing the traffic proportionally to the capacities of the links with the neighbouring sites, or diverting the desired amount of traffic through the cheapest provider, can be fulfilled. While SHIM6 supports some forms of traffic engineering at the end nodes, because of its end-to-end nature it is hard to enforce traffic engineering policies at a site level. To end with missing features, it may be worthy to be able to off-load the SHIM6 context management from the end nodes to specialised middle boxes to ease the deployment in domains in which end-hosts cannot be upgraded to the SHIM6 protocol, and to distribute the performance penalty imposed by SHIM6 operation when required.

In this paper we present an architecture based on the functionality provided by a SHIM6 proxy (P-SHIM6) to achieve the following capabilities:

- Provide Upper Layer Identifier portability, in order to ease renumbering
- Provide Traffic Engineering policy enforcement
- Enable legacy IPv6 nodes located in the multihomed site to obtain full SHIM6 multihoming support, without the modification of the end nodes
- Off-load of the SHIM6 context management from the actual peers of the communication

The rest of the paper is structured as follows: First we introduce the SHIM6 protocol. Then we describe the P-SHIM6 architecture, first as an overview, and then detailing the configuration and data exchange phases. After this, support for multiple P-SHIM6 boxes in order to increase fault tolerance is discussed. Finally, we analyse related work, and draw the conclusions.

2. SHIM6 Overview

To provide fault tolerance to established communications, the SHIM6 architecture enables diverting a packet of a communication to an alternative address of the host, which may be delegated by an alternative ISP. Since current transport layers identify the endpoints of a communication through the IP addresses of the nodes involved, translation among ULIDs and locators must be performed in a transparent fashion with respect to transport and application layers. The SHIM6 architecture relies on the

SHIM6 protocol to allow both ends to exchange their alternative locators, and on a security framework based on addresses with cryptographic properties to ensure that only legitimate locators can be exchanged. Additionally, the *REAchability Protocol* (REAP) is used to detect communication failures and to explore new paths when required. The next paragraphs detail these components.

The security architecture proposed for the multihoming protocol is based in the use of cryptographic addresses such as CGA (Cryptographically Generated Addresses, [4]). CGA incorporate into the 64-bit interface identifier (II_{CGA}) a cryptographic one-way hash of a public key (K_{public_key}), a prefix owned by the node (P_{CGA}), and a Modifier, creating a binding between the public key and the resulting address. The Modifier is defined to enhance privacy by adding randomness to the resulting address.

II_{CGA}=hash$|_{64}$ (K_{public_key} | P_{CGA} | Modifier)

The CGA is built by appending the resulting CGA interface identifier to the CGA network prefix: $P_{CGA}::II_{CGA}$. The private key corresponding to the K_{public_key} can sign the alternative locators that are conveyed in the SHIM6 protocol exchange described later. The trust chain is as follows: the ULID used for the communication, that is a CGA, is securely bound to the key pair, because it contains the hash of the public key, and any alternative locator is bound to the public key through the signature.

The SHIM6 protocol [2] defines a 4-way handshake to create and manage the SHIM6 context associated with the communication between two end-points, so that data packets can be exchanged using different locators while preserving the established communication. After this handshake, the validity of the CGAs of both end-points are checked, along with the validity of the signature of the locators,

As it has been commented above, the SHIM6 layer performs the translation between the ULIDs and the locators used for a given communication. While a locator change is not required, the address included in the data packet assumes both identifier and locator roles, as it occurs in normal IP operation. However, if the locators are changed for an established communication, because of an outage, or the result of the application of a TE policy, the initial ULIDs have to be preserved when interfacing to upper layers. In this case, additional information is carried in the packets as a *context tag,* a number that is unique for each communication at the receiver. The context tag is conveyed into a SHIM6 Payload Extension Header in the packets for which the locators differ from the identifiers.

SHIM6 provide the means for recovering SHIM6 contexts that have been lost in one of the communication peers. This is achieved by repeating part of the initial 4-way handshake, using the context tag of a packet received in the end point that lost the context as a hint for the peer that still maintains the context. This may be needed, for example, in a heavily loaded server that uses aggressive strategies for releasing context state.

Additional protocols, as defined in [3], are used to detect failures affecting the currently used path, and to explore alternative paths and select among them the most appropriate one to divert the communication to.

3 P-SHIM6 Operation Overview

In the P-SHIM6 architecture (see fig. 1), a multihomed site obtains, apart from a PA prefix from each of its providers, non-routable globally unique addresses, i.e. independent of the location of the network to which they are assigned, that are permanently allocated to the end site. These addresses can be obtained from a central registry, such as it is specified in the Centrally Managed Unique Local Address (CMULA) [5] specification.

Consider that the hosts within the multihomed site are IPv6 hosts without SHIM6 support. These hosts are configured only with a single address with the CMULA prefix, so the hosts in the multihomed site do not depend on the ISPs, and a change in the ISP would not imply a renumbering of the hosts of the multihomed site. The multihomed site is served by one or more Proxy-SHIM6 (*P-SHIM6*) boxes, which execute the SHIM6 protocol functions on behalf of the hosts of the multihomed site.

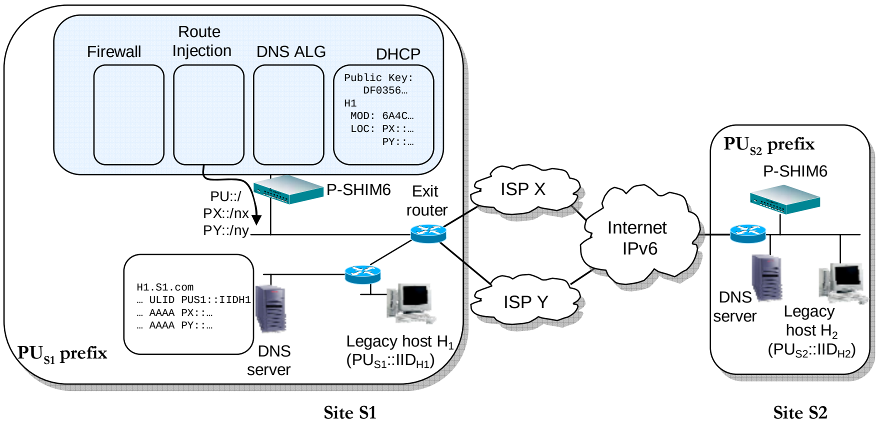

Fig. 1. P-SHIM6 architecture overview

An external communication can be established with a host located in another multihomed site or in a single-homed site. However, in order to enable the SHIM6 support for the communication, either the peer has to be SHIM6-capable or it has to be behind another P-SHIM6 that executed the SHIM6 protocol on its behalf. In the latter case, the result is that the SHIM6 protocol is executed between the P-SHIM6s that are serving each of the peers of the communication, as it is the case for figure 1.

For DNS operation when a P-SHIM6 is used, PA addresses are made public to the Internet in AAAA Resource Records (*RR*), while CMULAs are published in a newly defined *ULID RR*. The use of a new register prevents that an external non-SHIM6 aware node could try to use the CMULAs as a regular routable address.

When a (non-SHIM6 capable) host H1 located within the multihomed site S1 initiates a communication with a peer host H2, H1 normally performs a DNS query searching for `H2.foo.com` requesting an AAAA RR. Since the P-SHIM6 at S1 is configured as the DNS server for the hosts of the site, the query is sent to the P-

SHIM6. The P-SHIM6 behaves as a DNS ALG, and transforms the original request into a query for both AAAA and ULID records to the DNS of site S2. The reply from the DNS of S2 is processed by the DNS ALG of S1, so that only the CMULA is returned to the legacy host H1 in an AAAA RR. In addition to that, the P-SHIM6 at S1 stores the PA address information returned in the original DNS reply in the AAAA RR associated with the CMULA identifier obtained.

When H1 sends the first packet addressed to the CMULA of H2, the packet is intercepted and processed by the P-SHIM6 of the multihomed site S1. After forwarding the packet, the P-SHIM6 initiates the 4-way exchange to create a SHIM6 context with the P-SHIM6 of the peer network S2. This exchange conveys the PA addresses as locators and the CMULAs as ULIDs. Once the SHIM6 context is established between the local P-SHIM6 and the remote P-SHIM6, the local one can forward the first and subsequent data packets with a SHIM6 payload header referring to the established SHIM6 context. From now on, all packets belonging to the communication are intercepted by the P-SHIM6 and are processed so that the locators associated with the established context are included in the address fields of the packet and the negotiated context tag is included in all packets. Note this process only involves network-layer operations, as opposed to the application level rewriting that can be required by regular NAT operation, since in our case the applications of the communicating peers see the same identifiers at both sides.

The communication is now protected against failures by the SHIM6 protocol, in the sense that the reachability detection mechanisms of the REAP protocol will monitor the path availability of the communication. In case a failure is detected, alternative locator pairs are explored and the communication is diverted to an available path. Once the communication stops, heuristics are used at the P-SHIM6s to discard the associated SHIM6 state.

For intra-site communications, hosts can use CMULAs, which can be routed inside a domain in the same way as a regular address. So, in this case, the DNS should return the CMULAs of the internal hosts in AAAA records for internal queries. Again, the DNS ALG is responsible for processing the DNS reply of the actual DNS at S1, so that the CMULAs are returned in AAAA records.

The reverse tree of the DNS is used to store the locator set associated with the CMULAs in case that the communication does not start with a DNS query or there is no cached locator information available in the DNS ALG. This requires proper population of the reverse DNS tree of the CMULAs. Then, when a reverse DNS lookup is performed, the FQDN is returned and the locator information can be included in the Additional Information section of the DNS query.

Once a general overview of the mechanism has been presented, we next detail the configuration and the data exchange phases of the P-SHIM6 operation.

3.1 Detailed Configuration Phase

Consider the P-SHIM6 architecture depicted in figure 1, in which a multihomed site S1 is served by ISP X and ISP Y. Each of the ISPs delegates a Provider Aggregatable address block to the multihomed sites, with prefixes PX and PY respectively. Since these addresses are PA, the address block delegated by ISP X can only be reached through ISP X, and the address block delegated by ISP Y can only be reached through ISP Y. Besides, we assume that ISPs are performing ingress filtering, meaning that packets containing source addresses belonging to the address block delegated by a given ISP X can only exit through the same ISP. In addition, a CMULA block (prefix PU_{S1}) is assigned to the site so that CMULAs can be used as ULIDs for SHIM6 communications. So, each host within the multihomed site has conceptually three addresses: a CMULA from prefix PU and one address per PA prefix available in the site, prefixes PX and PY.

To enable SHIM6 operation, CMULAs have to be configured as cryptographic addresses, such as CGAs. Since the SHIM6 processing will be performed by the P-SHIM6, the CGA Parameter Data Structure and the associated private key must reside in the P-SHIM6 and not in the end host itself. Then, a DHCP component is required to generate CMULA CGAs on behalf of the hosts that are located behind the proxy, to store the associated parameters (CGA parameter Data structure and private key), and to assign the corresponding CMULA to each host when requested. As the hosts themselves are not involved in the SHIM6 protocol, the end hosts do not need to be aware that the address assigned is a CGA, neither they need to know the associated parameters. Note that all the different CMULA CGAs of the site can be generated using the same key pair, by only changing the Modifier field of the CGA Parameter Data Structure. This allows the P-SHIM6 to just maintain a single key pair for all its SHIM6 contexts.

In addition to the CMULA CGA, the P-SHIM6 internally assigns one address from each PA prefix available in the multihomed site to each host, although these addresses are not configured in the host itself. These addresses play the role of locators, and are permanently mapped in the P-SHIM6 to each corresponding host to allow external hosts to initiate a communication.

Regarding DNS configuration, the hosts inside S1 need to be configured to point to the P-SHIM6 as their DNS server, in order to assure that the DNS ALG is used. DHCP can be used to perform this configuration.

Finally, some configuration is required to assure that packets going from internal hosts to external ones, and vice versa are processed by the P-SHIM6. To do this, the P-SHIM6 injects an announce in the IGP (or either static routes are configured) to the root CMULA prefix, so that any packet generated from an internal host address to CMULA prefixes different from the ones assigned to the site are directed to the P-SHIM6. On the other hand, the P-SHIM6 announces internally reachability to PX and PY, so that the exit router(s) delivers to the P-SHIM6 any packet addressed to the locators assigned to the site.

Because of ingress filters, it may be necessary to route packets containing a given prefix in the source address through the ISP that has delegated this prefix. This can be achieved using tunnels from the P-SHIM6 to the exit routers, if many exist, and allowing the P-SHIM6 to route packets containing a given prefix in the source address through the corresponding ISP.

3.2 Data Exchange Phase

With the setup presented above, the behaviour of the P-SHIM6 architecture is the following:

1. A host H1 behind the P-SHIM6 at site S1 wants to initiate a communication with a host H2 located at site S2 with FQDN `H2.foo.com`. For that purpose, H1 performs a DNS query to its DNS server (the P-SHIM6) for `H2.foo.com`.
2. The P-SHIM6 performs a DNS query for `H2.foo.com`. If the query returns a ULID RR and one or more AAAA/A records, then the P-SHIM6 stores the information about the ULID and the associated locators and returns a single AAAA RR in the reply containing the CMULA (PU_{S2}:H2). At this point, the P-SHIM6 assumes that the host H1 will start sending data packets to the destination and it initiates the 4-way handshake defined in the SHIM6 protocol to establish a SHIM6 context.
3. When the host H1 receives the DNS reply containing a CMULA PU_{S2}:H2 in the AAAA record, it starts sending packets addressed to PU_{S2}:H2. Because of the longest prefix match of the address selection algorithm defined in RFC3484 [6], host H will choose the CMULA PU_{S1}:H1 as source address.
4. The intra-site routing will forward packets containing an external CMULA as destination address to the P-SHIM6. When a packet containing a CMULA as a destination address arrives, the P-SHIM6 performs the following processing:
   – If a SHIM6 context exists with the addresses contained in the packet as ULID pair, then it uses the existing SHIM6 context to process the packet (the context may be already in use, or may be just created when the DNS reply was received).
   – If no SHIM6 context exists, but there is locator information associated with the CMULA contained in the destination address (cached from the DNS reply), it uses that locator information to initiate the 4-way handshake to create a SHIM6 context for that ULID pair. Once the SHIM6 context is established, it is used to process the packet.
   – If no SHIM6 context exists and there is no locator information associated with the destination CMULA cached (for example, because the application used directly IP addresses to identify the peer, instead of a FQDN), the P-SHIM6 performs a DNS reverse lookup on the CMULA contained in the destination address field, and it obtains the locator set associated with the CMULA. Once the locator information is obtained, the 4-way handshake used to establish the SHIM6 context is performed. When the SHIM6 context is established, it is used to process the packet.
5. The packets addressed to any of the locators of site S2 are forwarded to the corresponding provider of site S2, then to S2, and finally to the P-SHIM6 at S2, since it internally propagates a route to those prefixes. Then this P-SHIM6 at S2,
   – If the packet is the first packet of the SHIM6 protocol exchange, it continues with the 4-way handshake for the establishment of the SHIM6 context.
   – If the packet is a payload packet and the P-SHIM6 has an existent context associated with it, it processes the data packet and replaces the locators by the associated identifiers, and forwards the packet to the final destination.

   - If the packet is a payload packet and the P-SHIM6 does not have an associated SHIM6 context, it initiates the SHIM6 Context Recovery Procedure, sending a *R1bis* packet [2] back, to the locator carried in the packet as source address, so that context can be restored.
6. After the SHIM6 context is established, the communication continues and both P-SHIM6s perform the translation between ULIDs and locator pairs as needed. In addition the REAP protocol for failure detection and alternative path exploration is used when needed, as defined in the SHIM6 protocol.
7. When the communication is finished, the P-SHIM6s use some heuristics to discard the SHIM6 context.

4 Multiple P-SHIM6s Support.

Since the main goal of multihoming is fault tolerance, it is critical to support multiple P-SHIM6s in a multihomed site, so that established communications could also be preserved in case of a failure in the P-SHIM6 that is being used for that communication. This can be done using the SHIM6 context recovery features.
We next consider the setup required to support multiple P-SHIM6s in a single site.

4.1 Configuration Phase

The described configuration uses one P-SHIM6 as the primary proxy for the multihomed site and the other P-SHIM6 as a backup in case the primary fails, as it is shown in figure 2.

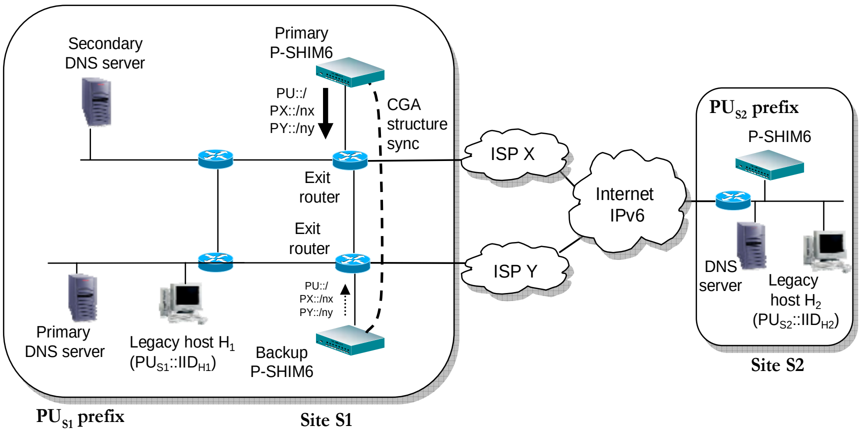

Fig. 2. Example of configuration with multiple P-SHIM6s within a site

In order to understand the implications of deploying multiple P-SHIM6s, we first summarize the interactions required between a single P-SHIM6 and the hosts being served by it.

1. DHCP address management: Delegation of CGA/HBA CMULA and storage of the associated parameters
2. DNS ALG service
3. Proxy function for egress packets: All packets generated by the internal hosts that are addresses to an external destination traverse the P-SHIM6, which establishes the correspondent SHIM6 context and then performs the appropriate ULID-locator translation.
4. Proxy function for ingress packets: All incoming packets are processed by the P-SHIM6, which restores the ULIDs.

It should be noted that operations 1, 3 and 4 require state in the P-SHIM6.

With respect to the CGA related information, to enable the use of multiple P-SHIM6s, all the P-SHIM6s within a site must have access to the CGA Parameter Data structure of each CMULA address assigned to a host within the site. Note that this state is per node, not per communication, so the overhead incurred in this replication may not be very high.

Outgoing data packets must be forwarded through the primary P-SHIM6 as long as it is working. This is achieved by configuring the primary P-SHIM6 to announce a route towards the generic CMULA prefix with a high priority, and configuring the backup P-SHIM6 to announce a route to the generic CMULA prefix with a low priority. In case of a failure of the primary P-SHIM6, the associated route would disappear and the alternative routes associated with the backup P-SHIM6 would be used. Similar considerations can be applied to incoming packets, so the primary P-SHIM6 will be configured to announce routes towards the prefixes assigned by the providers that are used to allocate the locators for end-hosts with high priority, while the backup P-SHIM6 announces the same routes with lower priority.

4.2 Data Exchange Phase in Case of Failures

In case the primary P-SHIM6 fails, the ongoing communications that have been established by the P-SHIM6 need to be preserved. This can be done by diverting the packets towards the secondary P-SHIM6 and allowing it to recover the SHIM6 context associated with the ongoing communication.

We assume that when a P-SHIM6 fails, the associated routes are no longer announced. This implies that the routes to the secondary P-SHIM6 will become the preferred ones. So, after the primary P-SHIM6 has failed, the following packets that belong to an ongoing communication can reach the secondary P-SHIM6:

- An incoming packet including a Payload Header with a context tag, that can be a data packet or a probe packet from the REAP protocol. The secondary P-SHIM6 will receive the packet and it will find that there is no existent context for that packet. Then the secondary P-SHIM6 will activate the recovery mechanism of the SHIM6 protocol by replying with a *R1bis* packet, and the remote P-SHIM6 or SHIM6 node will provide the missing context (identifiers being used, alternative locators for the remote node, context tag to use, etc.)
- An outgoing packet coming from one of the internal hosts is received. The secondary P-SHIM6 will unsuccessfully look for an existent SHIM6 context of for

cached locator information retrieved from the DNS query. Since there is no locator information associated with the destination identifier, it will perform a reverse DNS query using the CMULA included as destination in the packet, and it will obtain the locator information. At this point it will perform the 4-way handshake and the SHIM6 context will be re-established.

5. Related Work

IPv4 NATs are the reference middle-box architecture for IP networks. Compared to a P-SHIM6, both devices intercept packets exchanged between the site and the rest of the Internet, and process the IP header modifying the addresses using a per-communication state. Additionally, the P-SHIM6 model requires the middle-box to perform a 4-way handshake with external SHIM6-aware peers. However, P-SHIM6 provides many advantages compared to the deployment of NATs. Some derive from the fact that the identifiers are preserved in both end-points, avoiding the requirement for application inspection and processing at the middle-box, and allowing fully end-to-end operation, such as the one required by IPsec. Another advantage comes from the fact that IPv6 provides enough addresses so that stable mappings from PA addresses and CMULAs are possible, enabling externally initiated communications. Additionally, NATs are not able to preserve communications in case of failure, as P-SHIM6s do, even in case of failures in the P-SHIM6 itself. Protection against DoS attacks is provided by the use of the SHIM6 mechanism. The performance impact of deploying P-SHIM6 in a site is similar to deploying a NAT box in a site, since in both cases per-packet address rewriting and per-connection state maintenance are required, although for the P-SHIM6 case application processing is avoided.

An extended NAT architecture for IPv4 is proposed in IPnl [7]. Although this architecture may provide many benefits similar to P-SHIM6, such as network portability and fault tolerance against failures in data paths, it requires major changes not only in proxies but in hosts. More experience should be gained to determine the whole set of implications resulting form the deployment of this model.

GSE (Global, Site, End system) [8], is an IPng proposal in which a middle-box is used to rewrite addresses to gain provider independence, fault tolerance support, etc. However, this proposal raises some security vulnerabilities, such as the ones derived from the lack of tools to bind locators to an identifier.

In [9], a HIP proxy for 3G environments is described. The Host Identity Protocol (HIP) architecture [10] presents some commonalities with SHIM6, such as relying in an IP sublayer for performing a mapping between identifiers and locators. The fundamental difference between these two approaches is that a strict separation between locators and identifiers is proposed for HIP, so the identifiers are no longer valid locators, making difficult to manage application referrals and call-backs. Moreover, because of the non-hierarchical nature of the identifier name space, it is hard to deploy a directory service that stores the information about identifier to locator mapping. Apart from this, the proxy presented in [9] is specifically tailored to 3G environments, so 3G signaling is used to trigger state creation, and no hints for deployment in a full IP environment are described. In addition to this, the HIP approach imposes an extensive usage of public key cryptography, which is expensive in nature, and could be overkill for a proxy serving an IP site. Finally, proxy replication has not been considered for improving fault tolerance.

6. Conclusions

In this paper we have presented an architecture that relies on the configuration of provider independent addressed within a site and the deployment of SHIM6 proxies (P-SHIM6s) that intercept and process incoming and outgoing packets. In this way, non-SHIM6 aware hosts can benefit from SHIM6 when communicating with external SHIM6 hosts or hosts behind other P-SHIM6 proxies. The proxy uses the SHIM6 protocol to securely exchange the locators available for a communication, to detect communication failures, and to divert packets through an alternative path, in a transparent fashion to applications. The mechanism heavily relies on DNS to store the mapping between CMULA and the actual locators assigned by each of the actual providers of the site, and in some cases requires proper configuration of reverse DNS. The addressing and DNS specificities of this P-SHIM6 architecture affecting to legacy hosts are managed by the P-SHIM6 by means of a DHCP and a DNS-ALG component. Therefore, the P-SHIM6 architecture allows off-loading the SHIM6 protocol operation from the hosts inside the site, easing SHIM6 deployment since legacy hosts are not required to be migrated, and SHIM6 performance costs are not charged against existing nodes.

The resulting architecture enhances the SHIM6 multihoming model in several ways:

First, it enables multihomed sites to benefit from portability of address blocks when changing providers, freeing medium and small sites from the costs of a renumbering procedure, which *de facto* results in provider lock-in. In case a provider is changed, most of the configuration to be updated resides on the P-SHIM6, along with the DNS (both direct and reverse) and the site exit routers connected to the provider.

Next, P-SHIM6 determines the egress and ingress path for the packets of a given communication as a result of the selection of the locators. Therefore, Traffic Engineering policies can be easily enforced by properly configuring the selection of the locators. Since SHIM6 can enforce different ingress and egress paths for communications with different destinations, fine-grained Traffic Engineering can be achieved. Note that if the number of communications is high, the match with a target traffic profile can be achieved with very small deviations. If required, on-going communications could be reassigned to different locators to comply with Traffic Engineering objectives.

It should be highlighted that current BGP-based solution does not scale when applied to medium to small sites that require ISP independence and site traffic engineering capabilities when medium to small sites are involved.

Regarding to fault tolerance, the SHIM6 protocol executed between the P-SHIM6s uses the REAP protocol to detect failures along the communication path and to explore alternative paths. Once a failure is detected and an alternative path is discovered, the P-SHIM6 can divert the context affected by the failure through the new path, using the corresponding locator pair. It is also possible to feed the P-SHIM6 with additional information that can be used for failure detection. In particular, the P-SHIM6 can be fed with BGP information from the different ISPs – note that the site does not inject any information into BGP. In this case, the P-SHIM6 would have access to routing information and could divert the communication through an alternative ISP in case of a failure without requiring the use of REAP (or limiting it). Considering that multihoming, and therefore SHIM6, is aimed to enhance fault tolerance capabilities, special care has been devoted to describe configurations that preserve established communications in the case that the P-SHIM6 fails.

Communication with external legacy hosts that are not served by a P-SHIM6 is achieved by making the P-SHIM6 behave as a NATv6. In this case, the P-SHIM6 would simply translate the CMULA to one of the globally routable addresses. Of course this configuration presents some of the limitations of NATs in IPv4, including that the address of the host behind the P-SHIM6 is not restored end-to-end, so if addresses are included as application layer information, they will not match with the address actually contained in the header. However, since it is possible to perform a stateless one to many mapping between the CMULA and the global addresses, some of the limitations of NATs in IPv4, such as difficulties in allowing externally initiated communications, are lifted.

Finally, it should be noted that the architecture proposed does not require any modification neither in the hosts nor in the routers of the site.

References

1. G. Huston, "Architectural Commentary on Site Multi-homing using a Level 3 Shim", draft-ietf-shim6-arch-00 (work in progress), July 2005.
2. M. Bagnulo, E. Nordmark, "Level 3 multihoming shim protocol", draft-ietf-shim6-proto-07 (work in progress), November 2006.
3. J. Arkko, I. Beijnum, "Failure Detection and Locator Pair Exploration Protocol for IPv6 Multihoming", draft-ietf-shim6-failure-detection-07 (work in progress), December 2006.
4. T. Aura, "Cryptographically Generated Addresses (CGA)", RFC 3972, March 2005.
5. R. Hinden, B. Haberman, "Centrally Assigned Unique Local IPv6 Unicast Addresses", draft-ietf-ipv6-ula-central-01 (work in progress), February 2005.
6. R. Draves. “Default Address Selection for Internet Protocol version 6 (IPv6)”. Feb. 2003.
7. P. Francis, R. Gummadi. “IPNL: A NAT-extended Internet architecture”, Computer Communications Review 31 (4), pags 69-80, October 2001.
8. M. O’Dell, “GSE - An Alternate Addressing Architecture for IPv6”, draft-ietf-ipngwg-gseaddr-00.txt, February 1997.
9. P. Salmela. “Host Identity Protocol proxy in a 3G system”, Master Thesis, Helsinki University of Technology. February 2005.
10. R. Moskowitz, P. Nikander, P. Jokela, T. Henderson, “Host Identity Protocol”, draft-ietf-hip-base-07.txt, Internet Draft (work in progress), February 2007.