Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: July 12, 2014

G. Chen
Z. Cao
China Mobile
C. Xie
China Telecom
D. Binet
France Telecom-Orange

January 8, 2014

NAT64 Operational Experience
draft-ietf-v6ops-nat64-experience-08

Abstract

This document summarizes NAT64 function deployment scenarios and operational experience. Both NAT64 Carrier Grade NAT (NAT64-CGN) and NAT64 server Front End (NAT64-FE) are considered in this document.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on July 12, 2014.

Table of Contents

1. Introduction 2
2. Terminology 3
3. NAT64 Networking Experience 4
3.1. NAT64-CGN Consideration 4
3.1.1. NAT64-CGN Usages 4
3.1.2. DNS64 Deployment 4
3.1.3. NAT64 Placement 5
3.1.4. Co-existence of NAT64 and NAT44 5
3.2. NAT64-FE Consideration 6
4. High Availability 7
4.1. Redundancy Design 7
4.2. Load Balancing 9
5. Source Address Transparency 9
5.1. Traceability 9
5.2. Geo-location 10
6. Quality of Experience 11
6.1. Service Reachability 11
6.2. Resource Reservation 12
7. MTU Considerations 13
8. ULA Usages 14
9. Security Considerations 15
10. IANA Considerations 15
11. Acknowledgements 15
12. Additional Author List 16
13. References 16
13.1. Normative References 16
13.2. Informative References 18
Appendix A. Testing Results of Application Behavior 20
Authors' Addresses 21

1. Introduction

IPv6 is the only sustainable solution for numbering nodes on Internet due to the IPv4 depletion. Network operators have to deploy IPv6-only networks in order to meet the needs of the expanding internet without available IPv4 addresses.

Single stack IPv6 network deployment can simplify networks provisioning. Some justifications have been described in 464xlat [RFC6877]. As an example, IPv6-only connectivity confers some benefits to mobile operators. In such mobile context, it enables the use of a single IPv6 Packet Data Protocol(PDP) context or Evolved Packet System (EPS) bearer if Long Term Evolution (LTE) network is

considered, which eliminates significant network costs caused by doubling the number of PDP contexts in some cases and the need of IPv4 addresses to be assigned to customers. In broadband networks overall, it can allow for the scaling of edge-network growth decoupled from IPv4 numbering limitations.

In a transition scenario, some existing networks are likely to be IPv4-only configured for quite a long time. IPv6 networks and hosts will need to coexist with IPv4 numbered resources. Widespread dual-stack deployments have not materialized at the anticipated rate over the last 10 years, one possible conclusion being that legacy networks will not make the jump quickly. The Internet will include nodes that are dual-stack, nodes that remain IPv4-only, and nodes that can be deployed as IPv6-only nodes. A translation mechanism based on a NAT64[RFC6146] [RFC6145]function is likely to be a key element of the Internet for IPv6-IPv4 interoperability.

[RFC6036] reports at least 30% of operators plan to run some kind of translator (presumably NAT64/DNS64). Advice on NAT64 deployment and operations are therefore of some importance. [RFC6586] documents the implications for IPv6 only networks. This document intends to be specific to NAT64 network planning.

2. Terminology

In regards to IPv4/IPv6 translation, [RFC6144] has described a framework of enabling networks to make interworking possible between IPv4 and IPv6 networks. This document has further categorized different NAT64 function locations and use cases. The principle distinction of location is if the NAT64 is located in a Carrier Grade NAT or server Front End. The terms of NAT-CGN/FE are understood to be a topological distinction indicating different features employed in a NAT64 deployment.

NAT64 Carrier Grade NAT (NAT64-CGN): A NAT64-CGN is placed in an ISP network. IPv6 subscribers leverage the NAT64-CGN to access existing IPv4 internet services. The ISP as an administrative entity takes full control on the IPv6 side, but has limited or no control on the IPv4 side. NAT64-CGN may have to consider the IPv4 Internet environment and services to make appropriate configurations.

NAT64 server Front End (NAT64-FE): A NAT64-FE is generally a device with NAT64 functionality in a content provider or data center network. It could be for example a traffic load balancer or a firewall. The operator of the NAT64-FE has full control over the IPv4 network within the data center, but only limited influence or control over the external IPv6 network.

3. NAT64 Networking Experience

3.1. NAT64-CGN Consideration

3.1.1. NAT64-CGN Usages

Fixed network operators and mobile operators may locate NAT64 in access networks or in mobile core networks. It can be built into various devices, including routers, gateways or firewalls in order to connect IPv6 users to the IPv4 Internet. With regard to the numbers of users and the shortage of public IPv4 addresses, stateful NAT64[RFC6146] is more adapted to perform some maximal sharing of public IPv4 addresses. The usage of stateless NAT64 can be seen with better transparency features [I-D.ietf-softwire-stateless-4v6-motivation], while it has to be coordinated with A+P[RFC6346] processes as specified in [I-D.ietf-softwire-map-t] in order to cope with IPv4 shortage.

3.1.2. DNS64 Deployment

DNS64[RFC6147] is recommended for use in combination with stateful NAT64, and will likely be an essential part of an IPv6 single-stack network that couples to the IPv4 Internet. 464xlat[RFC6877] is proposed to enable access of IPv4 only applications or applications that call IPv4 literal addresses. Using DNS64 will help 464xlat to automatically discover NAT64 prefix through [RFC7050]. Berkeley Internet Name Daemon (BIND) software supports the function. It's important to note that DNS64 generates the synthetic AAAA reply when services only register A records. Operators should not expect to access IPv4 parts of a dual-stack server using NAT64/DNS64. The traffic is forwarded on IPv6 paths if dual-stack servers are targeted. IPv6 traffic may be routed not going through NAT64. Only the traffic going to IPv4-only service would traverse NAT64. In some sense, it encourages IPv6 transmission and restrains NAT uses compared to NAT44(if used), on which all traffic flows have to be traversed and translated. In some cases, NAT64-CGN may serve double roles, i.e. a translator and IPv6 forwarder. In mobile networks, NAT64 is likely deployed as the default gateway serving for all the IPv6 traffic. The traffic heading to a dual-stack server is only forwarded on the NAT64. Therefore, both IPv6 and IPv4 are suggested to be configured on the Internet faced interfaces of NAT64. We tested on Top100 websites (referring to [Alexa] statistics). 43% of websites are connected and forwarded on the NAT64 since those websites have both AAAA and A records. With expansion of IPv6 supports, the translation process on NAT64 will likely be faded.

3.1.3. NAT64 Placement

All connections to IPv4 services from IPv6-only clients must traverse the NAT64-CGN. It can be advantageous from the vantage-point of troubleshooting and traffic engineering to carry the IPv6 traffic natively for as long as possible within an access network and translate packets only at or near the network egress. NAT64 can be considered as a feature of the Autonomous System (AS) border in fixed networks. And, it is likely to be deployed in an IP node beyond the Gateway GPRS Support Node (GGSN) or Public Data Network- Gateway (PDN-GW) in mobile networks or directly in the gateway itself in some situations. This allows consistent attribution and traceability within the service provider network. It has been observed that the process of correlating log information is problematic from multiple-vendor's equipment due to inconsistent formats of log records. Placing NAT64 in a centralized location may reduce diversity of log format and simplify the network provisioning. Moreover, since NAT64 is only targeted at serving traffic flows from IPv6 to IPv4-only services, the user traffic volume should not be as high as in a NAT44 scenario, and therefore, the gateway's capacity in such location may not be as much of a concern or a hurdle to deployment. On the other hand, the placement in a centralized way would require more strict high availability (HA) design. It would also make geo-location based on IPv4 addresses rather inaccurate as it is currently the case for NAT44 CGN already deployed in ISP networks. More considerations or workarounds on HA and traceability could be found at Section 4 and Section 5.

3.1.4. Co-existence of NAT64 and NAT44

NAT64 could likely co-exist with NAT44 in a dual-stack network mostly because IPv4 private addresses are allocated to customers. The coexistence has already appeared in mobile networks, in which dual stack mobile phones normally initiate some dual-stack PDN/PDP Type[RFC6459] to query both IPv4/IPv6 address and IPv4 allocated addresses are very often private ones. [RFC6724] always prioritizes IPv6 connections regardless of whether the end-to-end path is native IPv6 or IPv6 translated to IPv4 via NAT64/DNS64. Conversely, Happy Eyeballs[RFC6555] will direct some IP flows across IPv4 paths. The selection of IPv4/IPv6 paths may depend on particular implementation choices or settings on a host-by-host basis, and may differ from an operator's deterministic scheme. Our tests verified that hosts may find themselves switching between IPv4 and IPv6 paths as they access identical service, but at different times [I-D.kaliwoda-sunset4-dual-ipv6-coexist]. Since the topology on each path is different, it may cause unstable user experiences and some degradation of Quality of Experience (QoE) when fallback to the other protocol is not powerful enough. It's also difficult for operators

to find a solution to make a stable network with optimal resource utilization. In general, it's desirable to figure out the solution that will introduce IPv6/IPv4 translation service to IPv6-only hosts connecting to IPv4 servers while making sure dual-stack hosts to have at least one address family accessible via native service if it's possible. With the end-to-end native IPv6 environment is available, hosts should be upgraded aggressively to migrate to IPv6-only. There is an ongoing effort to detect host connectivity and propose new DHCPv6 option[I-D.wing-dhc-dns-reconfigure] to convey appropriate configuration information to the hosts.

3.2. NAT64-FE Consideration

Some Internet Content Providers (ICPs) may locate NAT64 in front of an Internet Data Center (IDC), for example co-located with load balancing function. Load balancers are employed to connect different IP family domains, meanwhile distribute workloads across multiple domains or internal servers actually. In some cases, IPv4 addresses exhaustion may not be a problem in some IDC's networks. IPv6 support for some applications may require some investments and workloads so IPv6 support may not be a priority. The use of NAT64 may be served to support widespread IPv6 adoption on the Internet while maintaining IPv4-only applications access.

Different strategy has been described in [RFC6883] referred to as "inside out" and "outside in". An IDC operator may implement the following practices in the NAT64-FE networking.

- Some ICPs who already have satisfactory operational experiences would adopt single stack IPv6 operations to build up their data center network, servers and applications since it allows new services delivery without having to integrate consideration of IPv4 NAT and address limitations of IPv4 networks. Stateless NAT64[RFC6145] is used to provide services for IPv4-only subscribers. [I-D.anderson-siit-dc] has provided further descriptions and guidelines.

- ICPs who attempt to offer customers IPv6 support in their application farms at an early stage may likely run some proxies, which are configured to handle incoming IPv6 flows and proxy them to IPv4 back-end systems. Many load balancers have already integrated some proxy functionality. IPv4 addresses configured in the proxy can be multiplexed like a stateful NAT64 performs. A similar challenge exists once increasingly numerous users in IPv6 Internet access an IPv4 network. High loads on load-balancers may be apt to cause additional latency, IPv4 pool exhaustion, etc. Therefore, this approach is only reasonable at an early stage. ICPs may learn from the experiences and move on to dual-stack or

IPv6 single stack in a further stage, since the native IPv6 is always more desirable than transition solutions.

[RFC6144] recommends that AAAA records of load-balancers or application servers can be directly registered in the authoritative DNS servers requiring to populate these servers with corresponding AAAA records. In this case, there is no need to deploy DNS64 servers. Those AAAA records can be some native IPv6 addresses or some IPv4-converted IPv6 addresses[RFC6052]. The type of IPv6 address does not give the possibility to nodes to get any information about NAT64 presence on communication path and the possibility to prefer IPv4 path or the IPv6 path in dual-stack networks. For the testing purpose, operators could use an independent sub domain e.g. ipv6exp.example.com to identify experimental ipv6 services to users. How to design the FQDN for the IPv6 service is out-of-scope of this document.

4. High Availability

4.1. Redundancy Design

High Availability (HA) is a major requirement for every service and network services. The deployment of redundancy mechanism is an essential approach to avoid single-point failure and significantly increase the network reliability. It's not only useful to stateful NAT64 cases, but also to stateless NAT64 gateways.

Three redundancy modes are mainly used hereafter: cold standby, warm standby and hot standby.

- Cold standby can't replicate the NAT64 states from the primary equipment to the backup. Administrators switch on the backup NAT64 only if the primary NAT64 fails. As the results, all the existing established sessions will be disconnected. The internal hosts are required to re-establish sessions to the external hosts. Since the backup NAT64 is manually configured to switch over to active NAT64, it may have unpredictable impacts to the ongoing services. Normally, the handover would take several minutes so as to wait for the whole process of NAT64 bootstrap loader.

- Warm standby is a flavor of the cold standby mode. Backup NAT64 would keep running once the primary NAT64 is working. This makes warm standby less time consuming during the traffic failover. Virtual Router Redundancy Protocol (VRRP)[RFC5798] can be a solution to enable automatic handover in the warm standby. It was tested that the handover takes as maximum as 1 minute if the backup NAT64 needs to take over routing and re-construct the Binding Information Bases (BIBs) for 30 million sessions. In

deployment phase, operators could balance loads on distinct NAT64s devices. Those NAT64s make a warm backup of each other.

- Hot standby must synchronize the BIBs between the primary NAT64 and backup. When the primary NAT64 fails, backup NAT64 would take over and maintain the state of all existing sessions. The internal hosts don't have to re-connect the external hosts. The handover time has been extremely reduced. Thanks to Bidirectional Forwarding Detection (BFD) [RFC5880] combining with VRRP, a delay of only 35ms for 30 million sessions handover was observed during testing. In some sense, it could guarantee the session continuity for every service. In order to timely transmit session states, operators may have to deploy extra transport links between primary NAT64 and distant backup. The scale of synchronization data instance is depending on the particular deployment. For example, If a NAT64-CGN is served for 200,000 users, the average amount of 800, 000 sessions per second is roughly estimated for new created and expired sessions. A physical 10Gbps transport link may have to be deployed for the sync data transmission considering the amount of sync sessions at the peak and capacity redundancy

In general, cold-standby and warm-standby is simpler and less resource intensive, but it requires clients to re-establish sessions when a fail-over occurs. Hot standby increases resource's consumption to synchronize the states, but it achieve seamless handover. The consideration of redundancy mode for stateless NAT64 is simple, because state synchronization is unnecessary. In regards to stateful NAT64, it may be useful to investigate performance tolerance of applications and the traffic characteristics in a particular network. Some testing results are shown in the Appendix A.

Our statistics in a mobile network shown that almost 91.21% of amount of traffic is accounted by browsing services. Those services don't require session continuity. The handover time of warm standby is qualified to the delay tolerance. Hot-standby does not offer much benefit for those sessions on this point. In a fixed network, HTTP streaming, p2p and online games would be the major traffic[Cisco-VNI]. Consideration should be given to the importance of maintaining bindings for those sessions across failover. Operators may also consider the Average Revenue Per User (ARPU) factors to deploy suitable redundancy mode. Warm standby may still be adopted to cover most services while hot standby could be used to upgrade Quality of Experience (QoE) using DNS64 with different synthetic responses for limited traffic. Further considerations are discussed at Section 6.

4.2. Load Balancing

Load balancing is used to accompany redundancy design so that better scalability and resiliency could be achieved. Stateless NAT64s allow asymmetric routing while anycast-based solutions are recommended in [I-D.ietf-softwire-map-deployment]. The deployment of load balancing may make more sense to stateful NAT64s for the sake of single-point failure avoidance. Since the NAT64-CGN and NAT64-FE have distinct facilities, the following lists the considerations for each case.

- NAT64-CGN equipment doesn't implement load balancer functions on a board card. Therefore, the gateways have to resort to DNS64 or internal host's behavior. Once DNS64 is deployed, the load balancing can be performed by synthesizing AAAA response with different IPv6 prefixes. For the applications not requiring DNS resolver, internal hosts could learn multiple IPv6 prefixes through the approaches defined in[RFC7050] and then select one based on a given prefix selection policy.

- A dedicated Load Balancer could be deployed at front of a NAT64-FE farm. Load Balancer uses proxy mode to redirect the flows to the appropriate NAT64 instance. Stateful NAT64s require a deterministic pattern to arrange the traffic in order to ensure outbound/inbound flows traverse the identical NAT64. Therefore, static scheduling algorithms, for example source-address based policy, is preferred. A dynamic algorithm, for example Round-Robin, may have impacts on applications seeking session continuity, which described in the Table 1.

5. Source Address Transparency

5.1. Traceability

Traceability is required in many cases such as identifying malicious attacks sources and accounting requirements. Operators are asked to record the NAT64 log information for specific periods of time. In our lab testing, the log information from 200,000 subscribers have been collected from a stateful NAT64 gateway for 60 days. Syslog[RFC5424] has been adopted to transmit log message from NAT64 to a log station. Each log message contains transport protocol, source IPv6 address:port, translated IPv4 address: port and timestamp. It takes almost 125 bytes long in ASCII format. It has been verified that the volume of recorded information reach up to 42.5 terabytes in the raw format and 29.07 terabytes in a compact format. Operators have to build up dedicated transport links, storage system and servers for the purpose.

There are also several implementations to mitigate the issue. For example, stateful NAT64 could configure with bulk port allocation method. Once a subscriber creates the first session, a number of ports are pre-allocated. A bulk allocation message is logged indicating this allocation. Subsequent session creations will use one of the pre-allocated port and hence does not require logging. The log volume in this case may be only one thousandth of dynamic port allocation. Some implementations may adopt static port-range allocations [I-D.donley-behave-deterministic-cgn] which eliminates the need for per-subscriber logging. As a side effect, the IPv4 multiplexing efficiency is decreased regarding to those methods. For example, the utilization ratio of public IPv4 address is dropped approximately to 75% when NAT64 gateway is configured with bulk port allocation (The lab testing allocates each subscriber with 400 ports). In addition, port-range based allocation should also consider port randomization described in [RFC6056] . A trade-off among address multiplexing efficiency, logging storage compression and port allocation complexity should be considered. More discussions could be found in [I-D.chen-sunset4-cgn-port-allocation].Basically, the decision depends on usable IPv4 resource and investments of log systems.

5.2. Geo-location

IP addresses are usually used as inputs to geo-location services. The use of address sharing will prevent these systems from resolving the location of a host based on IP address alone. Applications that assume such geographic information may not work as intended. The possible solutions listed in [RFC6967] are intended to bridge the gap. However, those solutions can only provide a sub-optimal substitution to solve the problem of host identification, in particular it may not today solve problems with source identification through translation. The following lists current practices to mitigate the issue.

- Operators who adopt NAT64-FE may leverage the application layer proxies, e.g. X-Forwarded-For (XFF) [I-D.ietf-appsawg-http-forwarded], to convey the IPv6 source address in HTTP headers. Those messages would be passed on to web-servers. The log parsing tools are required to be able to support IPv6 and may lookup Radius servers for the target subscribers based on IPv6 addresses included in XFF HTTP headers. XFF is the de facto standard which has been integrated in most Load Balancers. Therefore, it may be superior to use in a NAT-FE environment. In the downsides, XFF is specific to HTTP. It restricts the usages so that the solution can't be applied to requests made over HTTPs. This makes geo-location problematic for HTTPs based services.

- The NAT64-CGN equipment may not implement XFF. Geo-location based on shared IPv4 address is rather inaccurate in that case. Operators could subdivide the outside IPv4 address pool so an IPv6 address can be translated depending on their geographical locations. As consequence, location information can be identified from a certain IPv4 address range. [RFC6967] also enumerates several options to reveal the host identifier. Each solution likely has their-own specific usage. For the geo-location systems relying on a Radius database[RFC5580], we have investigated to deliver NAT64 BIBs and Session Table Entrys (STEs) to a Radius server[I-D.chen-behave-nat64-radius-extension]. This method could provide geo-location system with an internal IPv6 address to identify each user. It can get along with [RFC5580] to convey original source address through same message bus.

6. Quality of Experience

6.1. Service Reachability

NAT64 is providing a translation capability between IPv6 and IPv4 end-nodes. In order to provide the reachability between two IP address families, NAT64-CGN has to implement appropriate application aware functions, i.e. Application Layer Gateway (ALG), where address translation is not itself sufficient and security mechanisms do not render it infeasible. Most NAT64-CGNs mainly provide FTP-ALG[RFC6384]. NAT64-FEs may have functional richness on Load Balancer, for example HTTP-ALG, HTTPs-ALG, RTSP-ALG and SMTP-ALG have been supported. Those application protocols exchange IP address and port parameters within control session, for example the "Via" filed in a HTTP header, "Transport" field in a RTSP SETUP message and "Received: " header in a SMTP message. ALG functions will detect those fields and make IP address translations. It should be noted that ALGs may impact the performance on a NAT64 box to some extent. ISPs as well as content providers might choose to avoid situations where the imposition of an ALG might be required. At the same time, it is also important to remind customers and application developers that IPv6 end-to-end usage does not require ALG imposition and therefore results in a better overall user experience.

The service reachability is also subject to the IPv6 support in the client side. We tested several kinds of applications as shown in the below table to verify the IPv6 supports. The experiences of some applications are still align with [RFC6586]. For example, we have tested P2P file sharing and streaming applications including eMule v0.50a, Thunder v7.9 and PPS TV v3.2.0. It has been found there are some software issues to support IPv6 at this time. The application software would benefit from 464xlat[RFC6877] until the software adds IPv6 support.. A SIP based voice call has been tested in LTE mobile

environment as specified in [IR.92]. The voice call is failed due to the lack of NAT64 traversal when an IPv6 SIP user agent communicates with an IPv4 SIP user agent. In order to address the failure, Interactive Connectivity Establishment (ICE) described in [RFC5245] is recommended to be supported for the SIP IPv6 transition. [RFC6157] describes both signaling and media layer process, which should be followed. In addition, it may be worth to notice that ICE is not only useful for NAT traversal, but also firewall[RFC6092] traversal in native IPv6 deployment.

Different IPsec modes for VPN services have been tested, including IPsec-AH and IPsec-ESP. It has been testified IPsec-AH can't survive since the destination host detects the IP header changes and invalidate the packets. IPsec-ESP failed in our testing because the NAT64 does not translate IPsec ESP (i.e. protocol 50) packets. It has been suggested that IPsec ESP should succeed if the IPSec client supports NAT-Traversal in the IKE[RFC3947] and uses IPsec ESP over UDP[RFC3948].

Table 1: The tested applications

APPs	Results and Found Issues
Webservice	Mostly pass, some failure cases due to IPv4 Literals
Instant Message	Mostly fail, software can't support IPv6
Games	Mostly pass for web-based games; mostly fail for standalone games due to the lack of IPv6 support in software
SIP-VoIP	Fail, due to the lack of NAT64 traversal
IPsec-VPN	Fail, the translated IPsec packets are invalidated
P2P file sharing and streaming	Mostly fail, software can't support IPv6, e.g. eMule Thunder and PPS TV
FTP	Pass
Email	Pass

6.2. Resource Reservation

Session status normally is managed by a static timer. For example, the value of the "established connection idle-timeout" for TCP sessions must not be less than 2 hours 4 minutes[RFC5382] and 5

minutes for UDP sessions[RFC4787]. In some cases, NAT resource maybe significantly consumed by largely inactive users. The NAT translator and other customers would suffer from service degradation due to port consummation by other subscribers using the same NAT64 device. A flexible NAT session control is desirable to resolve the issues. PCP[RFC6887] could be a candidate to provide such capability. A NAT64-CGN should integrate with a PCP server, to allocate available IPv4 address/port resources. Resources could be assigned to PCP clients through PCP MAP/PEER mode. Such ability can be considered to upgrade user experiences, for example assigning different sizes of port ranges for different subscribers. Those mechanisms are also helpful to minimize terminal battery consumption and reduce the number of keep-alive messages to be sent by mobile terminal devices.

Subscribers can also benefit from network reliability. It has been discussed that hot-standby offers satisfactory experience once outage of primary NAT64 is occurred. Operators may rightly be concerned about the considerable investment required for NAT64 equipment relative to low ARPU income. For example, transport links may cost much, because primary NAT64 and backup are normally located at different locations, separated by a relatively large distance. Additional maintenance has to be spent to ensure the connectivity quality. However, that may be necessary to some applications, which are delay-sensitive and seek session continuity, for example on-line games and live-streaming. Operators may be able to get added-values from those services by offering first-class services. It can be pre-configured on the gateway to hot-standby modes depending on subscriber's profile. The rest of other sessions can be covered by cold/warm standby.

7. MTU Considerations

IPv6 requires that every link in the internet have an Maximum Transmission Unit (MTU) of 1280 octets or greater[RFC2460]. However, in case of NAT64 translation deployment, some IPv4 MTU constrained link will be used in some communication path and originating IPv6 nodes may therefore receive an ICMP Packet Too Big (PTB) message, reporting a Next-Hop MTU less than 1280 bytes. The result would be that IPv6 allows packets to contain a fragmentation header, without the packet being fragmented into multiple pieces. A NAT64 would receive IPv6 packets with fragmentation header in which "M" flag equal to 0 and "Fragment Offset" equal to 0. Those packets likely impact other fragments already queued with the same set of {IPv6 Source Address, IPv6 Destination Address, Fragment Identification}. If the NAT64 box is compliant with [RFC5722], there is risk that all the fragments have to be dropped.

[RFC6946] discusses how this situation could be exploited by an attacker to perform fragmentation-based attacks, and also proposes an improved handling of such packets. It required enhancements on NAT64 gateway implementations to isolate packet's processing. NAT64 should follow the recommendation and take steps to prevent the risks of fragmentation.

Another approach that potentially avoids this issue is to configure IPv4 MTU more than 1260 bytes. It would forbid the occurrence of PTB smaller than 1280 bytes. Such an operational consideration is hard to universally apply to the legacy "IPv4 Internet" NAT64-CGN bridged. However, it's a feasible approach in NAT64-FE cases, since a IPv4 network NAT64-FE connected is rather well-organized and operated by a IDC operator or content provider. Therefore, the MTU of IPv4 network in NAT64-FE case are strongly recommended to set to more than 1260 bytes.

8. ULA Usages

Unique Local Addresses (ULAs) are defined in [RFC4193] to be renumbered within a network site for local communications. Operators may use ULAs as NAT64 prefixes to provide site-local IPv6 connectivity. Those ULA prefixes are stripped when the packets going to the IPv4 Internet, therefore ULAs are only valid in the IPv6 site. The use of ULAs could help in identifying the translation traffic.[I-D.ietf-v6ops-ula-usage-recommendations] provides further guidance for the ULAs usages.

We configure ULAs as NAT64 prefixes on a NAT64-CGN. If a host is only assigned with an IPv6 address and connected to NAT64-CGN, when connect to an IPv4 service, it would receive AAAA record generated by the DNS64 with the ULA prefix. A Global Unicast Address (GUA) will be selected as the source address to the ULA destination address. When the host has both IPv4 and IPv6 address, it would initiate both A and AAAA record lookup, then both original A record and DNS64-generated AAAA record would be received. A host, which is compliant with [RFC6724], will never prefer ULA over IPv4. An IPv4 path will be always selected. It may be undesirable because the NAT64-CGN will never be used. Operators may consider to add additional site-specific rows into the default policy table for host address selection in order to steer traffic flows going through NAT64-CGN. However, it involves significant costs to change terminal's behavior. Therefore, operators are not suggested to configure ULAs on a NAT64-CGN.

ULAs can't work when hosts transit the Internet to connect with NAT64. Therefore, ULAs are inapplicable to the case of NAT64-FE.

9. Security Considerations

his document presents the deployment experiences of NAT64 in CGN and FE scenarios. In general, RFC 6146[RFC6146] provides TCP-tracking, address-dependent filtering mechanisms to protect NAT64 from Distributed Denial of Service (DDoS). In NAT64-CGN cases, operators also could adopt unicast Reverse Path Forwarding (uRPF)[RFC3704] and black/white-list to enhance the security by specifying access policies. For example, NAT64-CGN should forbid establish NAT64 BIB for incoming IPv6 packets if uRPF in Strict or Loose mode check does not pass or whose source IPv6 address is associated to black-lists.

The stateful NAT64-FE creates state and maps that connection to an internally-facing IPv4 address and port. An attacker can consume the resources of the NAT64-FE device by sending an excessive number of connection attempts. Without a DDoS limitation mechanism, the NAT64-FE is exposed to attacks. Load Balancer is recommended to enable the capabilities of line rate DDOS defense, such as the employment of SYN PROXY-COOKIE. Security domain division is necessary as well in this case. Therefore, Load Balancers could not only serve for optimization of traffic distribution, but also prevent service from quality deterioration due to security attacks.

10. IANA Considerations

This memo includes no request to IANA.

11. Acknowledgements

The authors would like to thank Jari Arkko, Dan Wing, Remi Despres, Fred Baker, Hui Deng, Iljitsch van Beijnum, Philip Matthews, Randy Bush, Mikael Abrahamsson, Lorenzo Colitti, Sheng Jiang, Nick Heatley, Tim Chown, Gert Doering and Simon Perreault for their helpful comments.

Many thanks to Wesley George, Lee Howard and Satoru Matsushima for their detailed reviews.

The authors especially thank Joel Jaeggli and Ray Hunter for his efforts and contributions on editing which substantially improves the legibility of the document.

Thanks to Cameron Byrne who was an active co-author of some earlier versions of this draft.

12. Additional Author List

The following are extended authors who contributed to the effort:

Qiong Sun
China Telecom
Room 708, No.118, Xizhimennei Street
Beijing 100035
P.R.China
Phone: +86-10-58552936
Email: sunqiong@ctbri.com.cn

QiBo Niu
ZTE
50,RuanJian Road.
YuHua District,
Nan Jing 210012
P.R.China
Email: niu.qibo@zte.com.cn

13. References

13.1. Normative References

[I-D.ietf-appsawg-http-forwarded]
Petersson, A. and M. Nilsson, "Forwarded HTTP Extension", draft-ietf-appsawg-http-forwarded-10 (work in progress), October 2012.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[RFC3704] Baker, F. and P. Savola, "Ingress Filtering for Multihomed Networks", BCP 84, RFC 3704, March 2004.

[RFC3947] Kivinen, T., Swander, B., Huttunen, A., and V. Volpe, "Negotiation of NAT-Traversal in the IKE", RFC 3947, January 2005.

[RFC3948] Huttunen, A., Swander, B., Volpe, V., DiBurro, L., and M. Stenberg, "UDP Encapsulation of IPsec ESP Packets", RFC 3948, January 2005.

[RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, October 2005.

[RFC4787] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", BCP 127, RFC 4787, January 2007.

[RFC5245] Rosenberg, J., "Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols", RFC 5245, April 2010.

[RFC5382] Guha, S., Biswas, K., Ford, B., Sivakumar, S., and P. Srisuresh, "NAT Behavioral Requirements for TCP", BCP 142, RFC 5382, October 2008.

[RFC5424] Gerhards, R., "The Syslog Protocol", RFC 5424, March 2009.

[RFC5580] Tschofenig, H., Adrangi, F., Jones, M., Lior, A., and B. Aboba, "Carrying Location Objects in RADIUS and Diameter", RFC 5580, August 2009.

[RFC5722] Krishnan, S., "Handling of Overlapping IPv6 Fragments", RFC 5722, December 2009.

[RFC5798] Nadas, S., "Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6", RFC 5798, March 2010.

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, June 2010.

[RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, October 2010.

[RFC6145] Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", RFC 6145, April 2011.

[RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, April 2011.

[RFC6147] Bagnulo, M., Sullivan, A., Matthews, P., and I. van Beijnum, "DNS64: DNS Extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", RFC 6147, April 2011.

[RFC6157] Camarillo, G., El Malki, K., and V. Gurbani, "IPv6 Transition in the Session Initiation Protocol (SIP)", RFC 6157, April 2011.

[RFC6384] van Beijnum, I., "An FTP Application Layer Gateway (ALG) for IPv6-to-IPv4 Translation", RFC 6384, October 2011.

[RFC6555] Wing, D. and A. Yourtchenko, "Happy Eyeballs: Success with Dual-Stack Hosts", RFC 6555, April 2012.

[RFC6724] Thaler, D., Draves, R., Matsumoto, A., and T. Chown, "Default Address Selection for Internet Protocol Version 6 (IPv6)", RFC 6724, September 2012.

[RFC6887] Wing, D., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", RFC 6887, April 2013.

[RFC6946] Gont, F., "Processing of IPv6 "Atomic" Fragments", RFC 6946, May 2013.

[RFC7050] Savolainen, T., Korhonen, J., and D. Wing, "Discovery of the IPv6 Prefix Used for IPv6 Address Synthesis", RFC 7050, November 2013.

13.2. Informative References

[Alexa] Alexa, "http://www.alexa.com/topsites", April 2013.

[Cisco-VNI] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2012-2017, http://ciscovni.com/forecast-widget/index.html", May 2013.

[I-D.anderson-siit-dc] Anderson, T., "Stateless IP/ICMP Translation in IPv6 Data Centre Environments", draft-anderson-siit-dc-00 (work in progress), November 2012.

[I-D.chen-behave-nat64-radius-extension] Chen, G. and D. Binet, "Radius Attributes for Stateful NAT64", draft-chen-behave-nat64-radius-extension-00 (work in progress), July 2013.

[I-D.chen-sunset4-cgn-port-allocation] Chen, G., "Analysis of NAT64 Port Allocation Method", draft-chen-sunset4-cgn-port-allocation-02 (work in progress), July 2013.

[I-D.donley-behave-deterministic-cgn]
Donley, C., Grundemann, C., Sarawat, V., Sundaresan, K., and O. Vautrin, "Deterministic Address Mapping to Reduce Logging in Carrier Grade NAT Deployments", draft-donley-behave-deterministic-cgn-06 (work in progress), July 2013.

[I-D.ietf-softwire-map-deployment]
Qiong, Q., Chen, M., Chen, G., Tsou, T., and S. Perreault, "Mapping of Address and Port (MAP) - Deployment Considerations", draft-ietf-softwire-map-deployment-03 (work in progress), October 2013.

[I-D.ietf-softwire-map-t]
Li, X., Bao, C., Dec, W., Troan, O., Matsushima, S., and T. Murakami, "Mapping of Address and Port using Translation (MAP-T)", draft-ietf-softwire-map-t-04 (work in progress), September 2013.

[I-D.ietf-softwire-stateless-4v6-motivation]
Boucadair, M., Matsushima, S., Lee, Y., Bonness, O., Borges, I., and G. Chen, "Motivations for Carrier-side Stateless IPv4 over IPv6 Migration Solutions", draft-ietf-softwire-stateless-4v6-motivation-05 (work in progress), November 2012.

[I-D.ietf-v6ops-ula-usage-recommendations]
Liu, B., Jiang, S., and C. Byrne, "Recommendations of Using Unique Local Addresses", draft-ietf-v6ops-ula-usage-recommendations-01 (work in progress), October 2013.

[I-D.kaliwoda-sunset4-dual-ipv6-coexist]
Kaliwoda, A. and D. Binet, "Co-existence of both dual-stack and IPv6-only hosts", draft-kaliwoda-sunset4-dual-ipv6-coexist-01 (work in progress), October 2012.

[I-D.wing-dhc-dns-reconfigure]
Patil, P., Boucadair, M., Wing, D., and T. Reddy, "DHCPv6 Dynamic Reconfiguration", draft-wing-dhc-dns-reconfigure-02 (work in progress), September 2013.

[IR.92] Global System for Mobile Communications Association (GSMA), , "IMS Profile for Voice and SMS Version 7.0", March 2013.

[RFC6036] Carpenter, B. and S. Jiang, "Emerging Service Provider Scenarios for IPv6 Deployment", RFC 6036, October 2010.

[RFC6056] Larsen, M. and F. Gont, "Recommendations for Transport-Protocol Port Randomization", BCP 156, RFC 6056, January 2011.

[RFC6092] Woodyatt, J., "Recommended Simple Security Capabilities in Customer Premises Equipment (CPE) for Providing Residential IPv6 Internet Service", RFC 6092, January 2011.

[RFC6144] Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation", RFC 6144, April 2011.

[RFC6346] Bush, R., "The Address plus Port (A+P) Approach to the IPv4 Address Shortage", RFC 6346, August 2011.

[RFC6459] Korhonen, J., Soininen, J., Patil, B., Savolainen, T., Bajko, G., and K. Iisakkila, "IPv6 in 3rd Generation Partnership Project (3GPP) Evolved Packet System (EPS)", RFC 6459, January 2012.

[RFC6586] Arkko, J. and A. Keranen, "Experiences from an IPv6-Only Network", RFC 6586, April 2012.

[RFC6877] Mawatari, M., Kawashima, M., and C. Byrne, "464XLAT: Combination of Stateful and Stateless Translation", RFC 6877, April 2013.

[RFC6883] Carpenter, B. and S. Jiang, "IPv6 Guidance for Internet Content Providers and Application Service Providers", RFC 6883, March 2013.

[RFC6967] Boucadair, M., Touch, J., Levis, P., and R. Penno, "Analysis of Potential Solutions for Revealing a Host Identifier (HOST_ID) in Shared Address Deployments", RFC 6967, June 2013.

Appendix A. Testing Results of Application Behavior

We test several application behaviors in a lab environment to evaluate the impact when a primary NAT64 is out of service. In this testing, participants are asked to connect a IPv6-only WiFi network using laptops, tablets or mobile phones. NAT64 is deployed as the gateway to connect Internet service. The tested applications are shown in the below table. Cold standby, warm standby and hot standby are taken turn to be tested. The participants may experience service interruption due to the NAT64 handover. Different interruption intervals are tested to gauge application behaviors. The results are illuminated as below.

Table 2: The acceptable delay of applications

APPs	Acceptable Interrupt Recovery	Session Continuity
Web Browse	As maximum as 6s	No
Http streaming	As maximum as 10s(cache)	Yes
Gaming	200ms~400ms	Yes
P2P streaming, file sharing	10~16s	Yes
Instant Message	1 minute	Yes
Mail	30 seconds	No
Downloading	1 minutes	No

Authors' Addresses

Gang Chen
China Mobile
53A,Xibianmennei Ave.,
Xuanwu District,
Beijing 100053
China

Email: phdgang@gmail.com

Zhen Cao
China Mobile
53A,Xibianmennei Ave.,
Xuanwu District,
Beijing 100053
China

Email: caozhen@chinamobile.com

Chongfeng Xie
China Telecom
Room 708 No.118, Xizhimenneidajie
Beijing 100035
P.R.China

Email: xiechf@ctbri.com.cn

David Binet
France Telecom-Orange
Rennes
35000
France

Email: david.binet@orange.com