

Address Administration in IPv6

Eric Hoffman (hoffman@epsilon.com),
k claffy (kc@nlanr.net)

October 17, 1996

Humans have not inhabited Cyberspace long enough or in sufficient diversity to have developed a Social Contract which conforms to the strange new conditions of that world. Laws developed prior to consensus usually serve the already established few who can get them passed and not society as a whole. —

John Perry Barlow

1 Introduction

As the Internet begins to enter the popular scope, portending large scale changes in how society will operate, important questions have emerged regarding issues of fairness and responsibility with respect to several aspects of the base Internet architecture.

The questions are many, and all have cumbersome legal, financial, and cultural ramifications. We focus here on only one: the addressing model. Addresses are the technical cornerstone of the Internet's ability to move data from sender to any connected receiver. The shape of the address space ultimately determines the effective scalability and constrains the financial model of the network. We contrast two models of address assignment, provider and geographic based, expanding on Tsuchiya's analysis [10], and explore their ramifications.

One difficulty in discussing address space is that although its effective use is essential for the technical feasibility of Internet operation, the ownership of address space and the responsibility for justifying its use remain ill-defined. There are several revealing analogies in other spheres, e.g., spectrum assignment and international telephony. Like spectrum bands, IP address space is a finite and contended resource that requires careful assignment and management. We note also that it has been, and continues to be, particularly challenging for regulatory bodies to create and enforce equitable and consistent policies for spectrum allocation.

Internet addressing policy must be constructed out of a careful balance among:

- service requirements of end users
- base technology of the network

- cost effectiveness of service provision
- ability of providers to regain infrastructure costs
- use of the Internet as a fundamental technology of society as a the phone system

Although the IETF has entertained discussion of addressing policy including all of the above issues, the Internet has reached a stage of maturity and breadth of scope that requires wider debate of these issues.

2 Addressing and Routing

Routing in the Internet requires network elements to proactively exchange information concerning reachability to sites on the Internet. Each of these sites is described by an *address*, much as a phone number or street address describes a location in their respective networks. Information about how to reach any destination in a given part of the network is referred to as a *route*. These routes are summarized into a table, which the switching element in the center of the network, a *router* uses to decide which output interface to use for each arriving packet.

The two most performance critical tasks for an Internet router are processing routing updates and consulting this forwarding table on a per-packet basis.

The costs of memory to store both the routing and forwarding tables, and the processing power needed to update and consult them, place economic constraints on table size. These processing costs are becoming dominant as the backbone routing table size grows and other components such as high speed line interfaces become cheaper. One other important design factor is that as interfaces become higher speed, routing systems have increasingly less time to make a forwarding decision for each packet.

Rather than maintaining information about each attached host in the forwarding tables of backbone routers, routers can summarize or *aggregate* reachability information. *Aggregation allows the Internet to scale*. The basic form of aggregation, collecting

4 Renumbering

Renumbering is a term used to describe the changing of addresses of the hosts in a network. Sites sometimes renumber in order to reorganize internal network topology, but it is most often the case that renumbering is caused by a change of providers or requesting a larger address block from the same provider. Because changing the network address of nodes involves reconfiguring nodes, services, and routing systems, sites generally prefer to avoid paying the administrative cost of renumbering. Since provider-based addressing assigns addresses out of larger provider blocks, sites have a strong disincentive to change providers. This situation can strongly hinder free competition, not only by making it less likely that a customer will continually seek better valued service, but by making providers that can offer relatively stable larger address blocks much more attractive to customers.

IPv6 efforts have focused substantial attention on making renumbering as automatic as possible [8]. However, renumbering equipment in the current Internet still imposes significant burden on even small organizations. Furthermore, many Internet applications still use host addresses as unique identifying keys; such applications range from transient sessions such as TCP connections, to globally cached information such as DNS mappings, to near permanent relationships such as tunnel endpoints and NNTP and AFS configurations.

Although such applications could use DNS records in place of IP addresses for these functions, software designers have preferred to avoid reliance on the DNS since transient DNS failures are quite common. Currently DNS itself requires manual configuration of server address information the forward direction, and an external registry to handle changes in the reverse direction.

Efforts to alleviate the renumbering burden have primarily focused on mechanisms to facilitate the assignment of arbitrary addresses to end systems, but another alternative, Network Address Translators (NATs), have also received attention. IPv6 itself has rules for dealing with multiple sets of addresses associated with an interface, primarily for phasing out old sets of addresses in deference to new prefixes. While these mechanisms can somewhat automate a transition, it is clear that without serious changes to hosts and application semantics, renumbering will never be fully transparent. Ultimately, the degree of transparency will determine the perceived customer cost of changing providers; if renumbering is sufficiently disruptive, provider-based addressing will seriously damage the purity of competition in the Internet service market.

Individual customer renumbering is not the worst

case. Singly homed resellers of Internet service, i.e., those fully dependent on a parent provider for transit service, bear a compounded risk. Current provider-based schemes, including RFC 1887[5], allow service providers their own address blocks, but this policy will be unsustainable as the number of leaf providers grows enough to inhibit routing scalability. Continued growth will inevitably involve *recursive aggregation*, resulting in singly homed smaller providers using address space allocated from the blocks of their parent providers. If such a provider needed to change transit providers for business or legal reasons, they would have to impose renumbering on every one of their customers.

5 Settlements For Route Propagation

In order to insure that the networks they serve will be universally reachable from the Internet, providers must arrange with one another for propagation of their routes through the system. Carrying transit routes incurs a load on provider infrastructure, and there is as yet no direct financial incentive to control the number of routes one announces. Unabated growth in routable entities with no feedback mechanism to control their proliferation has threatened the ability of current backbone routers to maintain the necessary routing tables. In order to limit routing table size and promote aggregation, at least one provider has already resorted to imposing a lower limit on the size of route blocks that they will announce.

Rekhter, Resnick, and Bellovin in PIARA [9] propose creating a market around route advertisements, so that closed loop economic feedback can balance between the global costs of route maintenance and selfish desire to consume large amounts of address space rather than renumber. In the limit, the PIARA scheme requires that settlements to carry individual routes occur on a contractual basis.

Internet reachability to prefixes increasingly involves a set of contractual and legal relationships that stretch far beyond the customer and immediate provider. Although providers need some mechanism to recover transit costs, whether usage based or flat rate, it is far less clear that their reachability should be subject to second-order business dynamics over which customers have no control.

Furthermore, although the economic ramifications of Internet access outside the first world is still slight, it is naive to assume they will remain so as countries and businesses rely more fully on electronic information exchange. Although any provider-based addressing scheme will likely involve allocating blocks to countries for local administra-

FP	country	metro	site	intra-site
bytes: 1	2	3		10

Figure 3: proposed metro address format

tion, control over route propagation will still likely fall under the control of a set of multinational contractual relationships. Considerable debate over this concern in the context of the current IPv4 provider-based addressing policy has already occurred [11].

6 Multihoming

Multihomed sites are those that attach to more than one provider, as shown in Figure 6. Sites multihome to enhance network reachability or to connect to task-specific (e.g., research) networks.

Multihoming has traditionally complicated provider-based aggregation, since by definition multihomed sites do not fit neatly underneath a single aggregate prefix of a parent provider. To multihome in the current Internet, a site must get its own *autonomous system* (AS) number in order to advertise its own reachability directly into a default-free, or core, routing table.

The extent that this is a problem depends on how many sites require this level of multi-provider connectivity. Multihoming is not supported well by the provider-based model, since it requires that customers peer directly with providers, something which providers are only willing to support for sites with enough experience and responsibility to manage wide area routing. It is not an option available to everyone.

7 Metro Addressing Scheme

One alternative approach to provider-based addressing uses network address prefixes that correspond to major metropolitan areas [1]. These area prefixes are allocated underneath country prefixes to facilitate aggregation at country boundaries when possible. Sites on the Internet are assigned addresses out of the metropolitan region to which they belong, and all such sites are aggregated outside of that region for purposes of routing abstraction.

Because routers outside a metro can by definition not distinguish among reachability to individual sites within a metro, the metro addressing scheme structures Internet backbone service around Metropolitan Exchange Points, or MIXs. These exchange points resemble the Network Access Points (NAPs) of today’s Internet but serve also as aggregation points for each of the defined metropolitan regions. The size of these metros is a tradeoff be-

tween the number of required MIXs and the number of destinations that need routing support within the Metro.

The essential design goals of Metro are: (1) the ability to easily change providers without renumbering; and (2) terse backbone routing tables. The first goal requires essentially flat addressing within a MIX, with no structure to exploit for aggregation. The proposed metro address scheme allocates 3 bytes within the IPv6 address field to represent this flat space. Providers permanently attached to the MIX receive a *site identifier*; more dynamic, address-on-demand customers can receive identifiers from the site through which they connect. Each site within the Metro area will receive a site identifier out of a pool of 10^7 sites within each Metro. Each site will have 80 bits, or 10^{24} addresses with which to number hosts and implement internal hierarchies.

The underlying assumption is that indefinite recursive aggregation is not necessary, rather only a high level of aggregation, based on short geographic prefixes, in backbone routers. Since the number of countries is small, and hierarchical aggregation can optionally occur across country boundaries, backbone (core) router forwarding tables will be much smaller than current ones, while serving a subscriber base several orders of magnitude larger.

This scheme bears a strong resemblance to the *stratum* addressing that Rekhter outlines in [12]. In both schemes addresses focus around interchange points and allow arbitrary movement within the exchange point. The major difference between the approaches is that the stratum approach does not impose any geographic context on the interexchange address space, instead choosing to number around interchange points convenient to providers. This allows less constrained interactions between the members of the stratum with a corresponding loss of permanence in the addresses assigned. Stratum addressing thus has the aggregation and multihoming properties of metros, but retains renumbering problems associated with the provider-based model.

8 Intra-MIX Routing

Metro addressing drastically simplifies the backbone routing problem, but requires careful engineering to solve the *intra-Metro* routing problem. Provider independence requires that any site identifier be reachable through any provider from the MIX. This routing space is completely flat and corresponds in size to the total number of sites active within the region. As this number grows, the traditional dynamic routing system for dealing with frequent changes in a small number of network prefixes will no longer be appropriate for exchanging reachability information

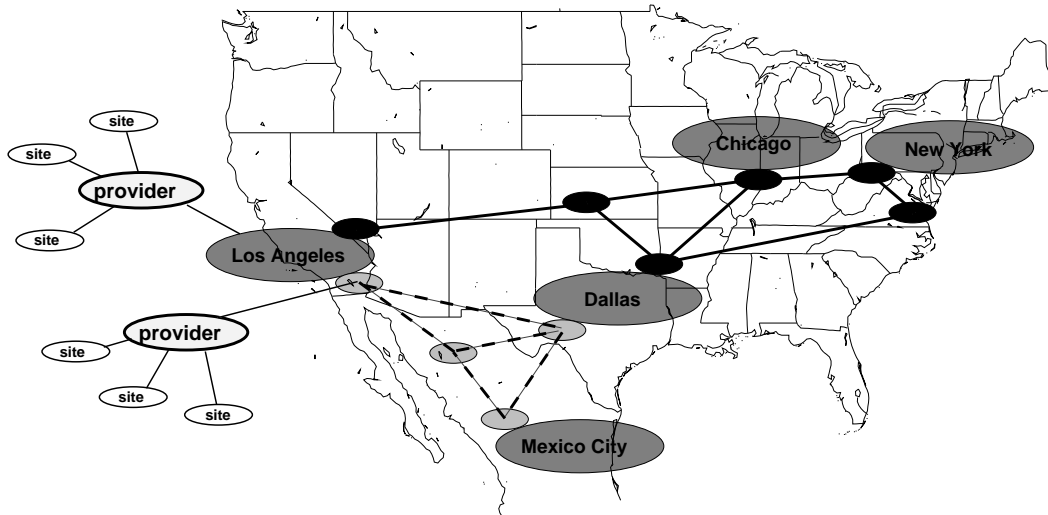


Figure 4: wide area metro routing

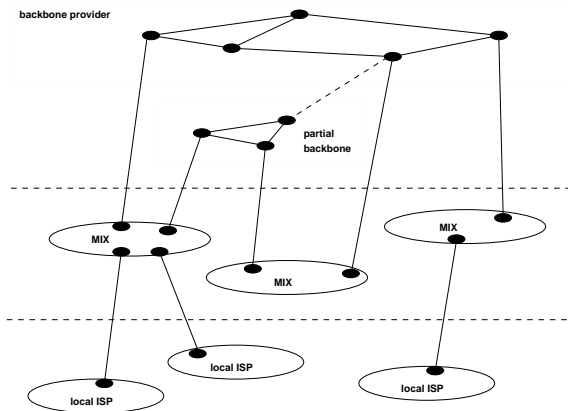


Figure 5: metro routing relationships

among sites. The size of this table also creates a special role for MIX routers, requiring them to maintain large routing table capacity and the ability to handle a large volume of routing information.

Deering [1] proposes a relatively static MIX-wide broadcast protocol that would result in the exchange of customer identifiers on daily basis. These routes would allow traffic between sites with different providers and traffic entering the metro from the wide area to travel to the correct destination provider for the destination customer site.

An alternative solution involves an on-demand mechanism: if a packet arrived at a provider's MIX router, and there was no destination provider, the router would send a message to a special broadcast group, and the provider serving that customer would respond with its router's address. Other routers would cache this information for some interval to avoid subsequent remote lookup delays for each packet they see for that customer.

A third possible solution would include a centralized server as part of the base MIX service. Providers would register customers with the server and could obtain partial or complete dumps of intra-MIX routing information. Although servers that maintain such mappings are single points of failure and often architecturally unnecessary, a single synchronization point for this information would enforce a consistent customer policy across the MIX.

Any of these solutions would be adequate for the simplified routing problem of distributing the information about which provider router to use to reach a customer. All of these techniques are similar to those used in another well-understood networking problem domain, that of address resolution.

9 Multihoming under Metro

Metro routing has as a beneficial side effect its simple support for multihoming within a metro region. Since aggregation occurs above the interchange point, multihoming within a metro area will not affect the routing table size outside the Metro.

Multihoming under Metro routing has good high-level aggregation properties, but it does require implementation changes. As the normal MIX routing system leverages off the fact that customer-provider bindings don't change more often than once a day, there is no way to support instantaneous failover within the basic intra-MIX routing model. However, application of a redirection technique is applicable if one only needs to redirect a small number of sites at one time.

Redirection can take the form of explicit messages, sent by the MIX exit router toward the entry point into the MIX, which create a transient routing

entry to direct future traffic to the backup router.

Redirection is not the only solution to providing failure robustness. Encapsulation could be used to create a temporary virtual link between the primary and failover router. This has the advantage of relying completely on the local information of the primary provider, but can result in undue traffic consumption as each packet has to enter and exit the primary provider on the path to the backup route.

10 Provider Constraints

Many providers feel strongly that metro addresses and the implied two-layer MIX routing is topologically constraining and thus imposes undue expense to supply the same service. This concern is due to the perception that Metro routing forces a fixed style of interconnection and routing without adequate means for expressing policy.

We submit that there is no architectural constraint preventing backbone providers from peering directly with each other using appropriate settlements for transit to metros that they do not serve directly. These peer relationships would be very similar to existing direct peerings, and could occur directly off the MIX or in some other context. Routing exchange across providers at such points traditionally benefits from provider based aggregation, but the Metro model treats such provider-based routes as exceptions rather than as the rule, and supporting non-default inbound Metro connectivity implies exchanging them as as full customer routes. Since involved providers typically have some kind of agreement between them, they should be able to arrange for mutual recovery of costs.

While the profit model for a provider that serves end customers is straightforward, it is less clear what business model will best serve providers acting solely in a backbone transport capacity. For current large scale backbone providers, subscriber fees from leaf customers cover much of the cost of maintaining the long haul resource. This cost is justified to end users by the end to end service they receive from transiting dedicated resources.

However, there is no reason that a backbone provider should not also serve as an intra-metro provider, in which case they would bypass the MIX for inter-Metro traffic to customers within the Metro. Although not strictly required, such a provider would likely use metro aggregation within its own routing system to prevent the insertion of non-scalable customer routes into its global routing system.

Direct second tier clients of a backbone provider can also arrange transit along dedicated links bypassing the MIX, as shown in figure 6. If this client

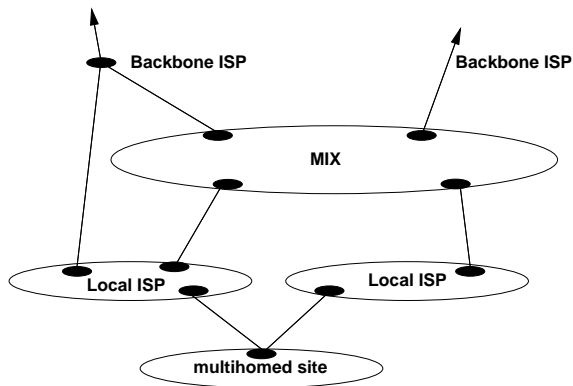


Figure 6: Routing relationships within MIX

is a provider, it would need to form an intra-MIX routing adjacency with the backbone provider and advertise its customers into the providers intra-MIX routing tables. These routes would allow traffic in both directions to use the dedicated link, but would not prevent the backbone provider from aggregating the entire Metro across the wide area, or force traffic from other sites within the MIX to traverse the backbone.

Singly connected smaller providers can still use regionals/backbone by arranging to get their customers routes carried to and advertised at the MIX by their parent provider. These small providers gain added leverage from Metro, since they have the ability to move their customers directly to another parent without disruptive renumbering.

Providers who must carry each other's customer routes when implementing direct peering relationships might see this as an inherent scalability problem with Metro routing. However, in this case Metro maintains a valuable property: scalability of policy. Although the providers involved must do extra work to peer directly, they do not impose any added complication in the routing system on entities outside the arrangement. Also, it is likely that if the provider spans multiple metros, providers can use metro aggregation used their own routing system to increase internal routing efficiency.

11 Establishment of Exchange Points

The MIXs play a central role in the proposed Metro architecture, and their establishment and maintenance merits careful consideration. The non-monopolistic nature of the exchange point is essential. Any second-tier provider capable of meeting some generally accepted criteria must be able to connect. Without this constraint the MIX itself would breed monopolistic behavior and encour-

age providers to violate the geographic locality of the address space. Possible models for insuring this property include mediation by a loose cooperative or government body.

Physically, a MIX would resemble either a centralized switched backbone or a physically dispersed shared media, analogous to the NAPs or MAEs today. Also similar to the NAP model, MIXes would serve a dual role, as a concentration point for traffic in the metropolitan area as well as a central point for the exchange of routing information. As noted above, it is not necessary that all traffic across the Metro cross this wire, but it will be the rule, especially for inbound traffic, rather than the exception.

In the San Francisco area there are currently several exchange points including MAE-West, the Pacific Bell NAP, PCH, FIX-west. Each of these has arisen out of differing needs and business models of providers that serve that region.

Metro routing does not preclude the creation of multiple exchanges in a Metropolitan area; accommodating all the service providers in a dense region might require more than one exchange point. Unless each regional provider attached to each of these exchanges, they would all need links among them. These links would carry traffic entering the Metro at one MIX, but destined to providers at some other MIX in the metro, as well as traffic between the exchange points within a geographic region.

These interexchange links are already difficult to manage in today's Internet, given that they are essentially public resources with no reasonable mechanism for cost recovery from traffic across large exchange points. Metro based routing however imposes one additional burden: the maintenance of a coherent fully qualified intra-MIX routing table consistent among all of the exchanges.

Proxied metros would relax the constraint of having backbone connections at each defined Metro, thus providing a crucial mechanism for any initial metro-based deployment. Assigned metro regions without an exchange would select a nearby exchange as a parent. Although they would number out of their assigned metro, each provider in the proxied metro would procure their own link to the nearby exchange.

As soon as there were enough providers in an area to justify an independent exchange point and attract a backbone provider, the network could incrementally rearchitect itself around the new MIX without any address reassignments.

12 Addressing Authority

Responsible use of allocated space in the provider-based model creates an interesting issue for re-

sellers of Internet service in a provider-based framework, who negotiate address space for their leaf customers. Although less critical in IPv6, scalability of provider-based addressing requires active management by addressing authorities and providers to insure conservative use of space. An organization capable of demonstrating that it serves a large number of end users and other providers can receive large blocks, with future allocations based on growth of the top level provider as well as effective use of previous allocations.

In contrast, the site identifiers used by Metro are neither scarce nor structured; assignments will not be highly dynamic and subject to much less policy consideration than a provider based scheme. Since the topmost bits are assigned to countries, it is perfectly natural to delegate addressing authority of each country to a national government or cooperative institution. Within each country, metropolitan regions themselves could use local registries or number out of the national registry. This natural match of the allocation problem to the region gives local governments and business cooperatives tools with which to shape the local networking landscape without having to contend with policy decisions elsewhere. Regardless of the scheme used locally for address assignment, it seems reasonable to permit providers to assign site IDs to new customers that need them as the purchase service, as long as these IDs are still fully transportable.

At the top level, some global organization will manage the country portion of the address space. Appropriate initial allocation should allow this address space to remain static over time scales that leave it amenable to management by a global treaty organization.

13 The Charging Problem

As the Internet continues its transition to a fully commercial framework, continued indeterminacy remains regarding viable underlying business models. Yet one cannot really design a routing framework without analyzing its affect on the ability of providers to charge for service.

Most differentiation among current backbones derives from their ability to effectively transit for large leaf customers and directly attached singly-homed providers. Since attachment points are a potential source of revenue, there is little economic incentive to provision a backbone to provide transit for default traffic.

Although some networks offer usage based billing at rough granularities, and other proposals for usage based pricing are emerging, the most common charging mechanism uses base connectivity fees with

implicit corresponding transit policy. ISPs typically provide service in two directions, by carrying packets from an attached customer to the backbone, and by providing routes at major exchange points to their customers. ISPs use explicit route filtering as well as route announcements to provide symmetric control over which routes to accept.

Since metro based addresses contain no provider information, and site IDs are flat and fully aggregated at the exchange boundary, advertising routes for directly attached customers is no longer possible. A provider could inject non-aggregated site routes to retain this policy flexibility, but it certainly would not scale to a large number of such routes. Service within the Metro model is thus necessarily constrained to providing settlements based on the sender, not the receiver. Ultimately, this unidirectional service model may turn out to be insufficient to express whatever charging policy may evolve. While it does offer providers the ability to instrument attachment and usage based policy for transmitters, it deliberately restricts the ability to filter based on receiver in the wide area.

An additional challenge to charging in the Metro model comes from the need to provision outbound service from second tier providers transiting the MIX. Since backbone providers currently only receive revenue from the destinations to which they explicitly route, they are unlikely to be willing to continue carrying traffic toward non-customers without some mechanism for revenue recovery. Unless each second tier provider is willing to negotiate separate interconnect agreements for outgoing traffic with each provider, the group of attached second tier providers will have to collectively subsidize default outbound connectivity. Enforcement in the first case, and in the second case if there are second tier providers that will not contribute to the subsidy fund, would require using layer 2 information to verify conformance to the agreement.

14 Conclusions

The IPv6 address space will host the Internet for some time to come. Careful consideration should precede initial numbering to insure that routing in the IPv6 will scale in usage as well as routing table size. Most discussion of addressing models to date has focused on the problem of allocation to support maximal aggregation. While aggregation is absolutely essential to maintaining a scalable infrastructure, it is not the only aspect of the wide area Internet that address assignment directly impacts.

Renumbering and route distribution policy are tools that can help improve the efficiency of the global routing system, but they also place the bur-

den of implementation on the end user, and could actually encourage a complete or partial monopoly over some segment of the market. Provider based addressing optimizes routing assuming a relatively static, hierarchical routing system that is unconnected at the edges. Metro based addressing provides an interesting alternative, although it presents a simplified costing model and requires further investigation into the details of intra-MIX routing.

We concede that excessive interference by regulatory can be harmful to technological development, but in this case the ramifications are too broad to be debated on technical merit alone. The deployment of an address model, whether provider-based, Metro, or some other alternative, will strongly determine the ultimate utility of the Internet to its users – society.

References

- [1] S. Deering, R. Hinden, “IPv6 Metro Addressing”, work in progress
- [2] V. Fuller, T. Li, J. Yu, K. Varadhan, RFC1519, “Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy”, 09/24/1993
- [3] Y. Rekhter, T. Li, RFC1618, “An Architecture for IP Address Allocation with CIDR”, 09/24/1993
- [4] S. Deering, R. Hinden, RFC1883, “Internet Protocol, Version 6 (IPv6) Specification”, 01/04/1996
- [5] Y. Rekhter, T. Li, RFC1887, “An Architecture for IPv6 Unicast Address Allocation”, 01/04/1996
- [6] R. Hinden, S. Deering, RFC1884, “IP Version 6 Addressing Architecture”, 01/04/1996
- [7] IESG, RFC1881, “IPv6 Address Allocation Management”, 12/26/1995. 01/04/1996
- [8] S. Thomson, T. Narten, RFC1971, “IPv6 Stateless Address Autoconfiguration”, 08/16/1996
- [9] Yakov Rekhter, Paul Resnick, Steven M. Bellovin “Financial Incentives for Route Aggregation and Efficient Address Utilization in the Internet: A Framework”, June 1996, Work in Progress
- [10] Paul Tsuchiya, “Comparison of Geographical and Provider-rooted Addressing”, INET '93
- [11] K. Hubbard, M. Koster, D. Conrad, D. Karrenberg, J. Postel, “Internet Registry Ip Allocation Guidelines”, August 1996, Work in Progress

- [12] Yakov Rekhter, “Stratum-Based Aggregation of Routing Information”, INET '95