.xxx Considered Dangerous

--- ---------------

<draft-eastlake-xxx-00.txt>

Status of This Document

Distribution of this draft is unlimited. Comments should be sent to the authors.

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC 2026. Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Copyright Notice

Copyright (C) The Internet Society (2001). All Rights Reserved.

Abstract

Periodically there are proposals to require the use of a special top level name or an IP address bit to flag "adult" or "safe" material or the like. This document explains why this is an ill considered idea.
Table of Contents

Status of This Document...........................................1
Copyright Notice...................................................1
Abstract......................................................................1

Table of Contents....................................................2

1. Background.......................................................3
2. Legal and Philosophical Problems............................4
3. Technical Difficulties..............................................6
4. Domain Name System (DNS) and Other Names...............7
4.1 Linguistic Problems.............................................7
4.1.2 The DNS Hierarchy and Use of TLDs.....................8
4.1.2 You Can’t Control Who Points At You...................8
4.1.3 Particular Protocol Considerations......................9
4.1.3.1 Electronic Mail (SMTP).................................9
4.1.3.2 Web Access (HTTP)....................................10
4.1.3.3 News (NNTP)........................................10
4.1.3.4 Internet Relay Chat....................................10
4.2 IP Addressing...................................................10
4.2.1 Hierarchical Routing.......................................11
4.2.2 IP Version 4 Addresses....................................12
4.2.3 IP Version 6 Addresses....................................12
5. Conclusions.......................................................13

References..........................................................14
Authors Addresses..................................................15

Full Copyright Statement..........................16
Expiration and File Name..........................16
1. Background

The concept of a .xxx, .sex, or similar top-level domain is periodically suggested by politicians and commentators. Other proposals have included a domain reserved exclusively for material viewed as appropriate for minors, or using IP address bits or ranges to segregate content.

In an October 1998 report accompanying the Child Online Protection Act, the House Commerce committee said "there are no technical barriers to creating an adult domain, and it would be very easy to block all websites within an adult domain." The report also said that the committee was wary of regulating the computer industry and that any decision by the U.S. government "will have international consequences." [HOUSEREPORT]

British Telecom has backed adult top-level domains, saying in a 1998 letter to the U.S. Department of Commerce that it "strongly supported" that plan. The reason: "Sexually explicit services could then be legally required to operate with domain names in this gTLD that] would make it much simpler and easier to control access to such sites.s..." [BT] One of ICANN's progenitors, the GTLD-MOU committee, suggested a "red-light -zone" top-level domain in a September 1997 request for comment. [GTLD-MOU]

Some adult industry executives have endorsed the concept. In 1998, Seth Warshavsky, president of the Internet Entertainment Group, told the U.S. Senate Commerce committee that he would like to see a .adult domain. "We're suggesting the creation of a new top-level domain called '.adult' where all sexually explicit material on the Net would reside," Warshavsky said in an interview at the time. [WARSHAVSKY]

More recently, other entrepreneurs in the industry have said that they do not necessarily object to the creation of an adult domain as long as they may continue to use .com.

Conservative groups in the U.S. say they are not eager for such a domain, and prefer criminal laws directed at publishers and distributors of sexually-explicit material. The National Law Center for Children and Families in Fairfax, Virginia, said in February 2001 that it did not favor any such proposal. For different reasons, the American Civil Liberties Union and civil liberties groups also oppose it.

Sen. Joseph Lieberman, the U.S. Democratic Party's vice presidential nominee, endorsed the idea at a June 2000 meeting of the federal Commission on Child Online Protection. Lieberman said in a prepared statement that "we would ask the arbiters of the Internet to simply abide by the same standard as the proprietor of an X-rated movie theater or the owner of a convenience store who sells sexually-explicit magazines." [LIEBERMAN]
In the 1998 law creating this commission, the U.S. Congress required the members to investigate "the establishment of a domain name for posting of any material that is harmful to minors." The commission devoted a section of its October 2000 report to that topic. It concluded that both a .xxx and a .kids domain are technically possible, but would require action by ICANN. The report said that an adult domain might be only "moderately effective" and raises privacy and free speech concerns. [COPAREPORT]

The commission also explored the creation of a so-called red zone or green zone for content by means of allocation of a new set of IP addresses under IPv6. Any material not in one of those two zones would be viewed as in a grey zone and not necessarily appropriate or inappropriate for minors. Comments from commissioners were largely negative: "Effectiveness would require substantial effort to attach content to specific IP numbers. This approach could potentially reduce flexibility and impede optimal network performance. It would not be effective at blocking access to chat, newsgroups, or instant messaging."

In October 2000, ICANN rejected a .xxx domain during its initial round of approving additional top-level domains. The reasons are not entirely clear, but former ICANN Chairwoman Esther Dyson said that the adult industry did not entirely agree that such a domain would be appropriate. One .xxx hopeful, ICM Registry of Ontario, Canada, in December 2000 asked ICANN to reconsider its decision. [ICM-REGISTRY]

2. Legal and Philosophical Problems

When it comes to sexually-explicit material, every person, court, and government has a different view of what’s acceptable and what is not. Attitudes change over time, and what is viewed as appropriate in one town may spark protests in the next. When faced with the slippery nature of what depictions of sexual activity should be illegal or not, one U.S. Supreme Court justice blithely defined obscenity as: "I know it when I see it."

In the U.S., obscenity is defined as explicit sexual material that, among other things, violates "contemporary community standards" -- in other words, even at the national level, there is no agreed-upon rule governing what is illegal and what is not. Making matters more knotty is that there are over 200 United Nations country codes, and in most of them political subdivisions can impose their own restrictions. Even for legal nude modeling, age restrictions differ. They’re commonly 18 years of age, but only 17 years of age in Sweden. A photographer in Oslo conducting what’s viewed as a legal and proper photo shoot there likely would be branded a felon and child pornographer in the U.S.
Saudi Arabia, Iran, and China are not likely to have the same liberal views as, say, the Netherlands. Saudi Arabia, like some other nations, filters its Internet connection and has created a government committee to protect its society from web sites that officials view as immoral. Their views on what should be included in a .xxx domain would hardly be identical to those in more liberal democracies.

Those wildly different opinions on sexual material make it improbable a global consensus can ever be reached on what is appropriate or inappropriate for a .xxx or .adult top-level domain. Moreover, the existence of such a domain would create an irresistible temptation on the part of conservative legislators to require controversial publishers to move to that domain.

Some conservative politicians already have complained that ICANN did not approve .xxx in its October 2000 meeting. During a February 2001 hearing in the U.S. House of Representatives, legislators warned that they "want to explore ICANN's rationale for not approving two particular top level domain names -- .kids and .xxx -- as a means to protect kids from the awful smut which is so widespread on the Internet."

It seems plausible that only a few adult publishers, and not those who have invested resources in building a brand around a .com site, would voluntarily abandon their current domain name. Instead, they'd likely add a propel legislators in the U.S. and other countries to require them to publish exclusively from an adult domain, a move that would invite ongoing political interference with Internet governance and raise concerns about forced speech and self-labeling.

In fact, the ultimate arbiter of generic top-level domain names -- at least currently -- is not ICANN, but the U.S. government. The U.S. Congress' General Accounting Office in July 2000 reported that the Commerce Department continues to be responsible for domain names allowed by the authoritative root. [GAO] The GAO’s auditors concluded it was unclear whether the Commerce Department has the "requisite authority" under current law to transfer that responsibility to ICANN.

The American Civil Liberties Union -- and other members of the international Global Internet Liberty Campaign -- caution that publishers speaking frankly about birth control, AIDS prevention, and gay and lesbian sex could be coerced into moving to an adult domain. Once there, they would be stigmatized and easily blocked by schools, libraries, companies, and other groups using filtering software. Publishers of such information who do not view themselves as pornographers and retain their existing addresses could be targeted for prosecution.

The existence of an adult top-level domain would likely open the door
for related efforts, either policy or legislative. There are many different axes through which offensive material can be defined: Sex, violence, hate, heresy, subversion, blasphemy, illegal drugs, profanity, political correctness, glorification of crime, incitement to break the law, and so on. Such suggestions invite the ongoing lobbying of ICANN, the U.S. government, or other policy-making bodies by special-interest groups that are not concerned with the technical feasibility or practicality of their advice.

An adult top-level domain could have negative legal repercussions by endangering free expression. U.S. Supreme Court Justice Sandra Day O’Connor has suggested that the presence of "adult zones" on the Internet would make a future Communications Decency Act (CDA) more likely to be viewed as constitutional. In her partial dissent to the Supreme Court’s rejection of the CDA in 1997 [CDA], O’Connor said that "the prospects for the eventual zoning of the Internet appear promising." (The Supreme Court ruled the CDA violated free speech rights by making it a crime to distribute "indecent" or "patently offensive" material online.)

Privacy could be harmed by such a proposal. It would become easier for repressive governments and other institutions to track visits to sites in a domain labeled as adult and record personally-identifiable information about the visitor. Repressive governments would instantly have more power to monitor naive users and prosecute them for their activities. It’s also not clear how effective a top-level domain would be when controlling access to chat, email, newsgroups and instant messaging.

4. Technical Difficulties

Even ignoring the philosophical and legal difficulties outlined above, there are substantial technical difficulties in attempting to impose content classification by domain names or IP addresses. Mandatory content labeling is usually advanced with the idea of using a top level domain name, discussed in section 4.1 below, but we also discuss the more fundamental possibility of using IP address bits or ranges in section 4.2 below.

In section 4.3 difficulties with a few particular higher level protocols are discussed. In some cases, these protocols use different name spaces.

We also discuss PICS labels [PICS] as an alternative technology in section 4.4.

Only a limited technical background is assumed so some basic information is included below and in some cases descriptions are
simplified.

4.1 Domain Name System (DNS) and Other Names

The most prominent user visible part of Internet naming and addressing is the domain name system [RFC 1034, 1035]. Domain Names are dotted sequences of labels such as aol.com, world.std.com, www.rosslynchapel.org.uk, or ftp.gnu.lcs.mit.edu [RFC 1035, 1591, 2606]. They form an important part of most World Wide Web addresses or URLs [RFC 2396], commonly appearing right after "/".

Actually, domain names just name nodes in a global distributed hierarchically delegated database. A wide variety of information can be stored at these nodes including IP addresses of machines on the network (see section 4.2 below), such things as mail delivery information, and many other types of information. Thus, the data stored at foo.example.com could be the numeric information for sending data to a particular machine, which would be used if you tried to browse <http://foo.example.com>, the name of a computer (say mailhost.example.com) to handle mail addressed to anyone @foo.example.com, and other information.

There are also other naming systems in use, such as news group names and Internet Relay Chat (IRC) channel names.

The usual labeling idea presented is to reserve a top level name, such as .xxx for "adult" material and/or .kids for "safe" material or the like. Ignoring the definitional and legal problems there are technical and linguistic problems with this are described in the subsections below.

4.1.1 Linguistic Problems

When using name labeling, the first problem is from whose language do you take the names to impose? Words and acronyms can have very different meanings if different languages and the probability of confusion is multiplied when phonetic collisions are considered.

As an example of possible problems, note that currently the government of Turkmenistan has suspended new registrations in ".tm", which had previously been a source of revenue, because some of the registered second level domain names may have been "legally obscene in Turkmenistan". <http://www.nic.tm>
4.1.2 The DNS Hierarchy and Use of TLDs

An important aspect of the design of the Domain Name System (DNS) is the hierarchical delegation of data maintenance. The DNS really only works, and has been able to scale the five orders of magnitude it has grown since its initial deployment, only due to this delegation.

The first minor problem is that one would expect most computers or web sites to have a mix of material only some of which should be specially classified. Using special TLDs multiples the number of DNS zones the site has to worry about. For example, assume the site has already sorted its material into "kids", "normal", and "adult" piles. Without special TLD labels, it can store them under kids.example.net, adult.example.net, and other.example.net, for instance, which requires only the maintenance of the single example.net zone of database entries. With special TLD labeling, at least example.net (for normal stuff), example.net.xxx, and example.net.kids would need to be maintained which are three separate zones in different parts of the DNS tree. As the number of categories expands and the number of category combinations explodes, this quickly becomes completely unmanageable.

4.1.2 You Can’t Control Who Points At You

The DNS system works as a database and associates certain data, called resource records, or RRs, with domain names. In particular, it can associate IP address resource records with domain names. For example, when you browse a URL, most commonly the domain name within that URL is looked up in the DNS and the resulting address (see Section 4.2) is used to address the packets sent from your web browser or other software to the server or peer.

Remember what we said in Section 4.1.1 about hierarchical delegation? Anyone controlling a DNS zone of data, say example.com, can insert data at that name or any deeper name (except to the extent they maintain delegations of some of the deeper namespace to yet others). So the controller of example.com can insert data so that purity.example.com has stored at it the same computer address which is at www.obscene.example.xxx. This directs any reference to purity.example.com to use the associated IP address which is the same as the www.obscene.example.xxx web site. The manager of that hypothetical web site, who controls the example.xxx zone, has no control over the example.com DNS zone and so is technically incapable of causing it to conform to any "xxx" labeling law. Or, in the alternative, someone could create a name conforming to an adult labeling requirement that actually pointed to someone else’s entirely unobjectionable site, perhaps for the purpose of polluting the labeling.
Thus, providers of data on the Internet cannot stop anyone from creating names pointing to their computer's IP address with misleading domain names.

4.1.3 Particular Protocol Considerations

There are additional considerations related to particular protocols. We consider only a few here. The first two, electronic mail and the World Wide Web, use domain name addressing. The second two, net news and IRC, actually use different name spaces and illustrate further technical problems with name based labeling.

4.1.3.1 Electronic Mail (SMTP)

The standard Internet electronic mail protocol separates "envelope" information from content [RFC 821, 822]. The envelope information indicates where a message claims to have originated and to whom it should be delivered. The content has fields starting with labels like "From:" and "To:" but these actually have no effect and can be arbitrarily forged using simple normally available software, such as telnetting to the SMTP port on a mail server. Content fields are not compared with envelope fields.

While different mail client display envelope information and headers from the content of email differently, generally the more common content fields are given prominence. Thus, while not exactly the same as content labeling, it should be noted that it is trivial to send mail to anyone with arbitrary domain names in the email addresses appearing in the From and To headers, etc.

It is also easy set up a host to forward mail to a mailing list. Mail sent with normal mail tools to this forwarder will automatically have content headers reflecting the forwarder’s name but the forwarder will change the envelope information and cause the mail to be actually sent to the original list. For example, (with names disguised) there is a social mailing list innocuous@foo.example.org and someone set up a forwarder at cat-torturers@other.example. Mail sent to the forwarder is forwarded and appears on the innocuous mailing list but with a "To: cat-torturers@other.example" header in its body. In some cases, similar things can be done using the "bcc" or blind courtesy copy feature of Internet mail.

Thus, standard Internet tools provide no way to control domain names appearing inside email headers.

There is work proceeding on securing email; however, such efforts at
present only allow you to verify whether or not a particular entity was the actual author of the mail. They do not generally relate to controlling or authenticating domain names in the content of the mail.

4.1.3.2 Web Access (HTTP)

At least with modern web servers and browsers supporting HTTP 1.1 [RFC 2616], the domain name used to access the site is available to access different web sites even though they are on the same machine at the same IP address.

(more to come)

4.1.3.3 News (NNTP)

Net news uses hierarchical structured newsgroup names that are similar in appearance to domain names except that the most significant label is on the left and the least on the right, the opposite of domain names. However, while the names are structured hierarchically, there is no central control. Instead, news servers periodically connect to other news servers that have agreed to exchange messages with them and then they update each other on messages only in those newsgroups in which they wish to exchange messages.

(more to come)

4.1.3.4 Internet Relay Chat

Internet Relay Chat is another example of a service which uses a different name space. (more to come)

4.2 IP Addressing

A key characteristic of the Internet Protocol (IP) on which the Internet is based is that it breaks data up into "packets". These packets are individually handled and routed from source to destination. Each packet has in it a numeric address for the destination point to which the Internet will try to deliver the packet.
(End users do not normally see these numeric addresses but instead deal with "domain names" as described in section 4.1 above.)

The numeric address system now primarily in use is called IPv4, or Internet Protocol Version 4, which provides for 32 bit addresses. There is a move to migrate to IPv6, which provides for 128 bit addresses.

One problem in using addressing for content filtering is that this is a very coarse technique. IP addresses address network interfaces which usually correspond to entire computer systems which could house multiple web pages, sets of files, etc., only a small part of which it was desired to block or enable. Increasingly, a single IP address may correspond to a NAT (Network Address Translation) box [NAT] which hides multiple computers behind it, although in that case these computers are usually not servers.

However, even beyond this problem of coarse granularity, the practical constraints of hierarchical routing make the allocation of even a single IPv4 address bit or any significant number of IPv6 address bits impossible.

4.2.1 Hierarchical Routing

As packets of data flow through the Internet, decisions must be made as to how to forward them "towards" their destination. This is normally done by comparing the initial bits of the packet destination address to entries in a "routing table" and forwarding the packets as indicated by the table entry with the longest prefix match.

While the Internet is actually a general mesh, if, for simplicity, we consider it to have a central backbone at the "top", a packet is typically routed as follows:

The local networking code looks at its routing table to determine if the packet should be sent directly to another computer on the "local" network, to a router to specially forward it to another nearby network, or routed "up" to a "default" router to forward it to a higher level service provider’s network. If the packet’s destination is "far enough away" it will eventually get forwarded up to a router on the backbone. Such a router can not sent the packet "up" since it is at the top or "default free" zone and must have a complete table of what other top level router to send the packet to. Currently, such top level routers are very large and expensive devices. They must be able to maintain tables of tens of thousands of routes. When the packet gets to the top level router of the part of the network within which its destination lies, it get forwarded "down" to successive routers which are more and more specific and local until
eventually its gets to a router on the local network where its destination address lies. This local router sends the packet directly to the destination computer.

Because all of these routing decisions are made on a longest prefix match basis, it can be seen that IP addresses are not general names or labels but are intimately associated with the actual topology and routing structure of the network. If there were assigned at random, routers would be required to remember so many specific routes for specific addresses that it would exceed the current technical capabilities for router design.

It should also be noted that there is some inefficiency in allocation at each level of hierarchy. Generally allocations are of a power of two addresses and as requirements grow and/or shrink, it is not practical to use every address for a computer.

(The above simplified description ignores multi-homing and many other details.)

4.2.2 IP Version 4 Addresses

There just isn’t any practical way to reallocate even one bit of IPv4 global Internet Addresses for content filtering use. Such addresses are in short supply and such an allocation would, in effect, cut the number of available addresses in half. There just aren’t enough addresses, given the efficiency of hierarchical allocation and routing, to do this. Even if there were, current numbers have not be allocated with this in mind so that a renumbering within every organization with hosts on the Internet would be required, a nightmarish and Herculean task costing in the billions of dollars. Even if these problems were overcome, the allocation of even a single bit would likely double the number of routes in the default free zone, exceeding the capacity of current routers and requiring the upgrade of thousands of them to new routers that do not exist yet.

And all this is for only a single bit, let alone more than one, is allocated to content labeling.

Basically, the idea is a non-starter.

4.2.3 IP Version 6 Addresses

IPv6 provides 128 bit address fields. (more to come)
4.3 PICS Labels

PICS Labels [PICS] have several modes. If content is required to have labels in it, it raises all the problems of categorization granularity and forced speech. But if used in a mode whereby a third party determines and provides labels for content and users are free to select whatever such third party or parties they wish to consult, it is a way to permit a myriad of categories, editors, and evaluators to exist in parallel.

(more to come)

5. Conclusions

TBD
References


[NAT] - ...

[PICS] - Platform for Internet Content Selection Service Descriptions <http://www.w3.org/TR/REC-PICS-services>
Label Format and Distribution <http://www.w3.org/TR/REC-PICS-labels>
PICS Rules <http://www.w3.org/TR/REC-PICSRules>
PICS Signed Labels (DSIG) 1.0 Specification <http://www.w3.org/TR/REC-DSig-label/>


Authors Addresses

Donald E. Eastlake 3rd
Motorola
155 Beaver Street
Milford, MA 01757 USA

Telephone: +1-508-634-2066 (h)
            +1-508-261-5434 (w)
FAX:       +1-508-261-4447 (w)
EMail:     Donald.Eastlake@motorola.com

Declan McCullagh
Wired News
1935 Calvert Street, NW #1
Washington, DC 20009

Telephone: +1-202-986-3455
FAX:       +1-202-986-3472
EMail:     declan@wired.com
Full Copyright Statement

Copyright (C) The Internet Society (2001). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Expiration and File Name

This draft expires August 2001.

Its file name is draft-eastlake-xxx-00.txt.