2020

IDN-домены:
правила безопасного использования

Материал подготовлен **Координационным центром доменов .RU/.РФ**
при поддержке проекта **Поддерживаю.РФ**

IDN-домены: правила безопасного использования

Сегодня аудитория интернета оценивается в 4,5 миллиарда человек, и для большинства интернет-пользователей английский язык не является родным. Однако именно латинская кодировка ASCII (American standard code for information interchange) — основной «язык» интернета, и долгие годы именно на ней основывалось большинство сервисов и технологий глобальной сети, включая и систему доменных имен DNS.

язык-

это система звуковых обозначений, позволяющая описывать окружающую действительность и формулировать мысли при помощи устной речи.

алфавит или азбука-

форма письменности, основанная на наборе знаков, соответствующих фонемам (звукам языка). Алфавит используется, например, в русском и английском языке, но китайские иероглифы им не являются, так как обозначают отдельные понятия, а не звуки.

кодировка-

способ записи знаков азбуки или иероглифического письма в виде числовых значений (кодов символов) на основе «кодовых таблиц» или «таблиц символов». Самыми распространенными кодировками сегодня являются ASCII и Unicode. В информационных технологиях под кодировкой также часто подразумевают способ хранения кода символа. Например, популярная кодировка UTF-8 означает кодирование символа при помощи Unicode с хранением кода символа в специальной структуре длиной от 1 до 4 байт.

В доменных именах изначально разрешалось использовать только основное подмножество символов ASCII **(буквы a–z, цифры 0–9 и дефис «-»).** После того как **в 1996 году Мартин Дюрст** впервые предложил идею создания доменных имен с символами национальных алфавитов, была начата техническая проработка этой идеи. Чтобы способствовать интернационализации глобальной сети, Инженерная проектная группа интернета (IETF[1]), начиная с 2003 года, выпустила ряд стандартов, описывающих технические принципы использования интернационализированных доменных имен (IDN-доменов, Internationalized Domain Names). Использование IDN-доменов было реализовано через специальный **механизм преобразования Punycode,** позволяющий обеспечить корректную работу в интернете доменных имен с использованием символов, которые поддерживаются в стандарте кодирования Unicode, например普遍接受- 测试.世界, ua-test, .كاثو لـــيك, тестовая-зона.рф и т. д.

После длительных обсуждений Правление ICANN одобрило ускоренный процесс создания новых национальных нелатинских доменов верхнего уровня (Fast Track Process) в октябре 2009 года, а первые IDN ccTLD (country code Top Level Domain), включая и российский кириллический домен .РФ, были добавлены в корневую зону уже в мае 2010 года. В июне 2011 года Правление ICANN утвердило и санкционировало запуск программы New gTLD, которая предоставила возможность для создания новых общих доменов верхнего уровня (generic Top Level Domain), включая новые латинские в кодировке ASCII – например: **.ONLINE, .SPACE, .SITE, .FUN, .HOST, .PRESS** и интернационализированные TLD – например, **.МОСКВА, .ДЕТИ, .РУС, .САЙТ** и другие.

Первая партия общих доменов верхнего уровня (более 1200) в рамках этой программы была добавлена в корневую зону в 2013 году и с тех пор широко используется пользователями интернета.

В 2020-м году кириллическому домену .РФ исполнилось 10 лет.

Он был делегирован России 12 мая 2010 года, а первыми доменными именами в новой доменной зоне стали президент.рф и правительство.рф. Открытая регистрация доменных имен в зоне .РФ стартовала 11 ноября 2010 года. Сегодня в домене .РФ насчитывается около 730 тысяч доменных имен, и каждый месяц это число увеличивается. Причем доменные имена в .РФ всё чаще регистрируются как основной адрес для запуска новых проектов.

1 https://www.ietf.org/

Новая технология – новая опасность

Доменные имена на родном языке легче запоминаются, произносятся и записываются с меньшим количеством ошибок, именно это делает IDN-домены такими удобными для пользователей и является причиной их широкого использования. Однако, как известно, появление любых новых технологий привлекает внимание и сетевых мошенников, которые стремятся использовать их в своих нелегальных схемах. Такие ситуации случаются и с IDN-доменами.

Основной способ мошенничества с IDN-доменами – омоглифические атаки, когда символы одного алфавита подменяются символами другого, **похожими внешне, но по сути являющимися совершенно разными кодами Unicode.** В простейшем случае заменяется только одна буква в доменном имени, в более сложных - весь домен может состоять из символов другого алфавита, но выглядеть как настоящий.

Простейший пример **омоглифа** – буква «о», которая одинаково пишется во многих алфавитах. Например: U+004F – O латинское, U+039F – O (омикрон) греческое, U+041E – O кириллическое. Соответственно и доменные имена с ними будут выглядеть одинаково, например google.com (используется «о» кириллическое) и google.com (используется «о» латинское), но являться разными доменными именами для программного обеспечения.

Самый распространенный метод борьбы с омоглифическими атаками – ограничение возможности одновременного использования нескольких алфавитов в доменных именах и в локальной части адресов электронной почты (до знака @). Для защиты от омоглифов регистратуры доменов верхнего уровня разрабатывают собственные правила регистрации доменных имен, которые чаще всего определяют ограниченный список символов одного алфавита (в редких случаях - нескольких совместимых алфавитов), разрешенных для использования при регистрации доменных имен.

В российском кириллическом домене .РФ проблема омоглифов решена простым образом: для регистрации доменных имен доступны только буквы русского алфавита, арабские цифры и дефис. Смешение нескольких алфавитов в доменных именах в зоне .РФ не допускается. То есть можно зарегистрировать доменное имя «рао.рф», где «рао» составлено из трех кириллических символов, но нельзя зарегистрировать имя «рао.рф», где «рао» состоит из ASCII-символов.

Поэтому российский домен .РФ можно назвать одним из самых безопасных с точки зрения омоглифических атак.

Еще один вариант противоправного использования схожести IDN-доменов встречается значительно реже и только в отдельных языках. Речь идет о **наличии в самом алфавите так называемых лигатур** - составных символов, образованных путём соединения двух и более графем (например, в арабском, французском, норвежском, нидерландском, армянском и многих других). Такие составные символы могут выглядеть для пользователя одинаково при последовательном написании составляющих их символов. Например, строчная буква хорватского алфавита **«lj» (код U+01c9)** и две отдельные буквы **«l» и «j» (два отдельных кода U+006c и U+006a),** которые при написании вместе выглядят так же. Кроме того, во множестве алфавитов используются символы с модификаторами, называемыми диакритическими знаками. Причем в Unicode-представлении как лигатуры, так и символы с диакритическими знаками могут кодироваться двумя способами: либо одним уникальным кодом, либо сочетанием соответствующих кодов.

К примеру, строчная русская буква «ё» может иметь код U+0451, а может быть представлена как сочетание кодов U+0435 и U+0308, т.е. базового символа «е» и комбинируемого символа надстрочного двоеточия «¨» соответственно. Также и кириллическая буква «й» с кодом U+0439 может быть представлена как сочетание базового символа «и» U+0438 и комбинируемого бреве «˘» U+0306. В связи с этим возникает опасность, что одно и то же доменное имя на программном уровне может восприниматься как два разных.

Для борьбы с этой опасностью регистратуры доменов верхнего уровня, в которых разрешена регистрация доменных имен на таких языках, при проверке уникальности и допустимости для регистрации доменного имени рекомендуют **использовать определенные в стандарте Unicode алгоритмы нормализации строк, приводящие их к каноническому представлению.** Приведение осуществляется путем замены символов на эквивалентные с использованием определенных таблиц и правил. **Декомпозицией** называется замена (разложение) одного символа на несколько составляющих символов, а **композицией**, наоборот, — замена (соединение) нескольких составляющих символов на один.

Рекомендуемой практикой при работе с IDN доменами является использование **формы нормализации C**[2] (NFC - алгоритм, согласно которому последовательно выполняются каноническая декомпозиция и каноническая композиция). Кроме того, в процессе нормализации нельзя забывать о приведении буквенных символов к нижнему регистру, т.к., например, символы «Й» и «й» имеют в стандарте Unicode разные кодовые точки: U+0419 и U+0439 соответственно. В доменных именах и адресах электронной почты используется **только нижний регистр.**

Чтобы избежать различного рода ошибок, проверка уникальности доменных имен при регистрации обычно производится после процедуры нормализации. В ходе такой проверки аккредитованным регистратором или регистратурой домена верхнего уровня может быть отказано в регистрации доменного имени, потому что в нормализованном представлении оно может совпадать с другим, уже зарегистрированным.

2 http://unicode.org/reports/tr15/

Как распознать поддельное доменное имя?

Главная опасность для пользователей, связанная с подделкой IDN-доменов, – это **фишинг,** то есть создание сайтов или почтовых сообщений, копирующих известные сервисы, где предусмотрено введение критически важных данных. При этом доменное имя сайта или почтовый адрес, которые используют мошенники, внешне выглядят так же, как и официальные адреса сервиса. Например, может быть создан фальшивый веб-портал банка, где вас попросят ввести номер банковской карты или SMS-код подтверждения платежа, а также сайт почтового сервиса, где потребуется пароль от вашего ящика. Если вы введете эти данные на фальшивом интернет-ресурсе, они будут украдены, а с их помощью злоумышленники получат доступ к вашей электронной почте или банковскому счету. При этом ссылки с поддельными IDN-доменами на фальшивый сайт распространяются мошенниками в письмах или сообщениях в мессенджерах якобы от имени настоящего сервиса и визуально могут не отличаться от ссылок на подлинные веб-ресурсы.

Большинство браузеров помогают пользователю бороться с фишинговыми атаками и распознают доменные имена, в которых присутствуют символы сразу нескольких алфавитов. Такое имя они показывают в адресной строке в Punycode, то есть в его ASCII-представлении с префиксом «xn--» (первые четыре символа). А если в IDN-домене используется один алфавит, то такие имена отображаются символами этого алфавита без преобразования, то есть вы будете видеть доменное имя, например, на русском языке. Так что если вы открыли в браузере ссылку на сайт с IDN-доменным именем, а его название отображается в адресной строке латиницей, **начинающейся с символов xn--, то с этой ссылкой что-то не в порядке и, вероятнее всего, это фишинговый сайт.**

Однако важно помнить, что отдельное программное обеспечение и сервисы, например, почтовые клиенты и онлайн почтовые сервисы, до сих пор не умеют распознавать доменные имена с омоглифами или все еще не до конца поддерживают IDN-технологию, поэтому в таких программных продуктах все IDN-домены будут показываться только в Punycode - и корректные, и поддельные. В результате, пользуясь такими сервисами, вы можете перейти из письма по фальшивой ссылке, либо, наоборот, увидев корректный IDN-домен в ASCII-представлении, посчитать его мошенническим.

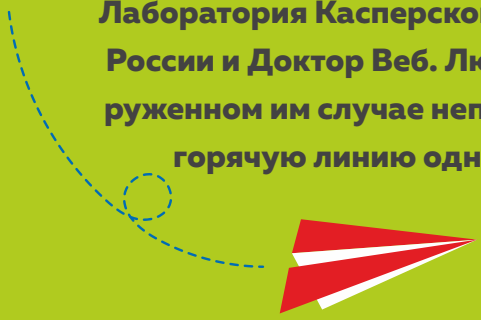
ПОЭТОМУ ОБЯЗАТЕЛЬНО ВНИМАТЕЛЬНО ПРОВЕРЯЙТЕ, КАК ВЫГЛЯДИТ ДОМЕННОЕ ИМЯ В БРАУЗЕРЕ ПЕРЕД ТЕМ, КАК ВВОДИТЬ НА САЙТЕ СВОИ ДАННЫЕ.

В доменных именах на русском языке стоит также внимательно относиться к окончаниям, определяющим форму слова. Например, домен go.com может иметь не только прямой аналог идти.рф, но и иду.рф, идем.рф, идут.рф, которые легко перепутать, если не запоминать специально.

Как же поступить, если все усилия разработчиков браузеров и других программных продуктов оказались недостаточно успешными, и вы все-таки столкнулись с фактом неправомерного использования доменного имени?

Если речь идет о фишинге, несанкционированном доступе к информационным системам или о распространении вредоносных программ с использованием доменных имен в зонах .РФ и .RU, то вы можете **обратиться на горячую линию компетентной организации.**

Сегодня с Координационным центром доменов .RU/.РФ сотрудничают десять компетентных организаций — Национальный координационный центр по компьютерным инцидентам, Лига безопасного интернета, Group-IB, Лаборатория Касперского, RU-CERT, РОЦИТ, Роскомнадзор, БИЗон, Банк России и Доктор Веб. Любой пользователь сети может сообщить об обнаруженном им случае неподобающего использования доменного имени на горячую линию одной из этих компаний — и меры будут приняты незамедлительно.



СОВЕТЫ

#1

Внимательно относитесь к ссылкам, которые вам присылают по электронной почте. Большинство почтовых клиентов до сих пор не полностью поддерживают IDN-домены и могут как показать корректный интернационализированный домен в Punycode, так и, наоборот, представить в виде Unicode-символов фишинговую ссылку, содержащую омоглифы.

#2

Внимательно проверяйте, как ссылка выглядит в браузере. Если IDN-домен отображается в Punycode – с ним точно что-то не так, и по такой ссылке лучше не переходить, а если вы уже перешли, то на этой странице не стоит вводить никаких данных о себе.

#3

Набирая доменное имя по памяти или на слух, помните, что в домене .РФ используются символы только русского алфавита.

Материал подготовлен **Координационным центром доменов .RU/.РФ**
при поддержке проекта **Поддерживаю.РФ**

тел. +7 495 7302971

cctld.ru | кц.рф
поддерживаю.рф

КООРДИНАЦИОННЫЙ ЦЕНТР
ДОМЕНОВ .RU/.РФ

Поддерживаю.РФ