Combate ao spam em IPv6

novos problemas e propostas de solução

Danton Nunes, InterNexo Ltda.
danton.nunes@inexo.com.br

O mesmo velho problema de sempre...

```
Date: Sat, 11 Mar 2017 06:55:50
From: CBN <godwinnewcbngovernor6@gmail.com>
Reply-To: godwinnewcbngovernor5@gmail.com
Subject: GOOD NEWS

Attn: Fund Recipients

 Based on the meeting we had with United Nations and IMF
 officials on your payment DID YOU PERMIT MR. COLE WANDA TO
 RECEIVE YOUR FUND AS YOUR NEXT OF KIN We wait to hear from
 you soon.
 Thanks
 Prof.Godwin Emefiele
 Special Adviser to Mr.
 President
```

...agora em IPv6!

```
Return-Path: <godwinnewcbngovernor6@gmail.com>
Received: (qmail 2323 invoked from network); 11 Mar 2017 21:01:49 -0000
Received: from unknown (HELO rs000039.fastrootserver.de)
    (2001:4ba0:babe:0039:0000:0000:0000:0000)
  by 2001:12c4:b010:c0c0:0000:0000:0000:000a with SMTP; 11 Mar 2017 21:01:49
    -0000
Received: from User (unknown [154.118.33.157])
    by rs000039.fastrootserver.de (Postfix) with ESMTPA id BAB6E222540;
    Sat, 11 Mar 2017 10:55:36 +0100 (CET)
Reply-To: <godwinnewcbngovernor5@gmail.com>
From: "CBN"<godwinnewcbngovernor6@gmail.com>
Subject: GOOD NEWS
Date: Sat, 11 Mar 2017 10:55:50 +0100
MIME-Version: 1.0
```

IPv6 trouxe novos problemas na mitigação de spam!

O problema:

- Identificar usuários finais (ADSL, xPON, ...) para bloqueio ou pontuação.
- Reconhecer servidores legítimos de email.

Especificidades do IPv6

- Esqueça o reverso!
- SLAAC e extensões de privacidade podem ajudar.

Identificando usuário final

HIPÓTESE: o usuário final tem endereços IP obtidos por SLAAC ou gerados pelo algoritmo de extensão de privacidade (RFC–4941)

Por outro lado servidores legítimos normalmente tem endereços IP estáticos, atribuídos por seres humanos.

Resumindo

provavelmente é:

SLAAC
RFC–4941 } → **usuário final**

criação humana → **servidor válido**

SLAAC (StateLess Address AutoConfiguration)

- Definido pela RFC–4862 e RFC–4291 App. A

```
# ip addr sh dev eth0
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP group default qlen 1000
    link/ether 00:23:7d:43:e7:fb brd ff:ff:ff:ff:ff:ff
    inet6 2001:12c4:5afe:b175:223:7dff:fe43:e7fb/64 scope global mngtmpaddr dynamic
       valid_lft 86400sec preferred_lft 14400sec
```

construido a partir do endereço MAC
ff:fe constante em todos os endereços SLAAC de ethernet
prefixo da rede, vindo do roteador.

O padrão ff:fe é um indicativo de endereço auto configurado em rede tipo ethernet.

É claro, mas muito improvável, que um endereço estático seja construído com esse padrão.

A presença desse padrão pode ser um critério para suspeitar da origem. E sem DNS!

Endereços dinâmicos segundo a RFC–4941

- O identificador da interface é aleatório e não tem um padrão facilmente reconhecível.
- Aleatório a grosso modo significa:
  - alta entropia
  - estacionário

Entropia: $$H = -\sum_i P_i \log_2 P_i$$ Claude E. Shannon, 1948 (!)

Para "mensagens" binárias, *i=0 ou i=1:*

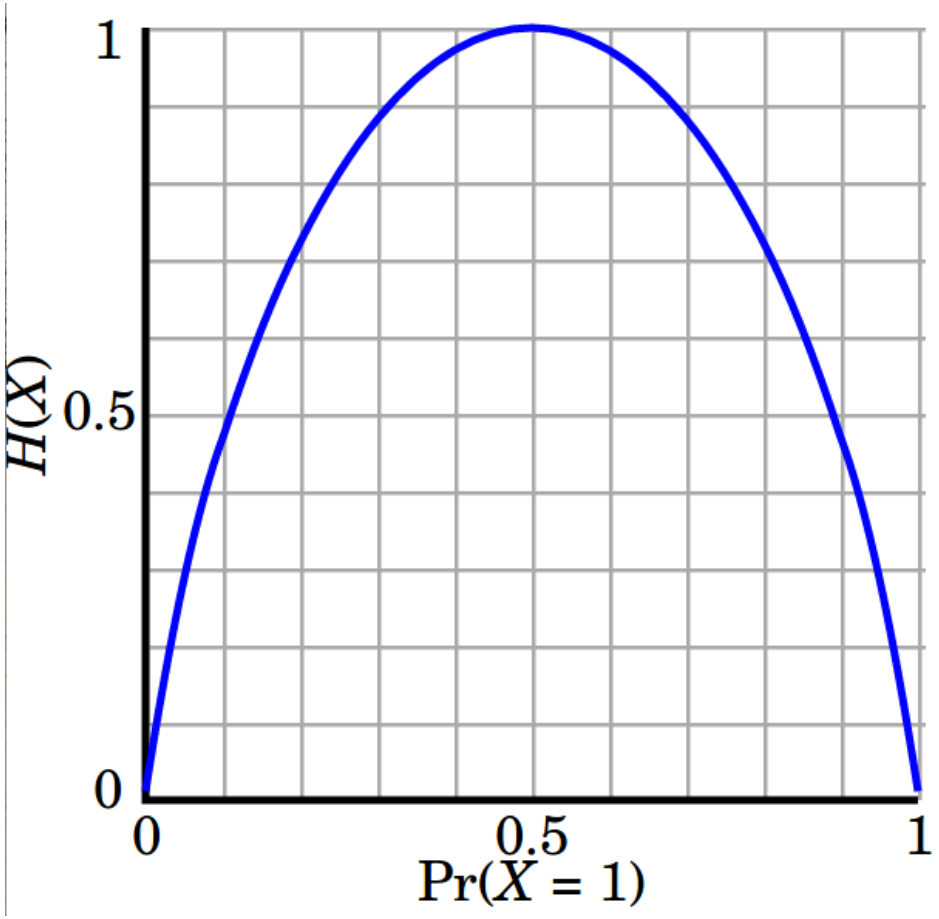
$$-H = P_0 \log_2 P_0 + P_1 \log_2 P_1$$

probabilidade estimada pela frequência relativa.

Endereços dinâmicos segundo a RFC–4941

A entropia é máxima quando há igual número de 0s e 1s e é zero quando há somente 0s ou 1s.

Não é necessário avaliar a entropia, pois ela fica perfeitamente definida se contarmos quantos 0s há nos últimos 64 bits.



Fonte: Wikipedia

Então, um indício de aleatoriedade é a quantidade de 0s e 1s serem da mesma ordem. Endereços feitos por humanos tendem a ter muitos 0s e poucos 1s.

Endereços dinâmicos segundo a RFC–4941

Alguns endereços de servidores reais

origem	endereço IP	0s	1s	entropia
gmail	2800:3f0:4003:c00::1a	61	3	0.272
internexo	2001:12c4:b010:c0c0::a	62	2	0.200
ix.br	2001:12ff:0:6192::201	62	2	0.200
he.net	2001:470:0:76::2	63	1	0.116
nasa	2001:4d0:a302:1100::152	60	4	0.337
unesp	2801:88:1000:bf::80	63	1	0.116
bmw	2620:101:200a:d100::f	60	4	0.337

E alguns endereços dinâmicos

Ahá!

endereço IP	0s	1s	entropia
2001:12c4:5afe:b175:878:9b0:4b5c:9dda	36	28	0.988
2001:12c4:5afe:0a:68b3:29da:9893:e340	35	29	0.993
2001:12c4:5afe:0a:d538:7556:b20e:eb77	28	36	0.988
2001:12c4:5afe:0a:2236:32c4:2878:4fec	36	28	0.988
2001:12c4:5afe:0a:4715:6ddb:404b:893c	34	30	0.997

Entropia, porém, não é suficiente!

Endereços dinâmicos segundo a RFC–4941

Além da entropia temos que ver se a sequência de bits e estacionária.

`0000:ffff:0000:ffff` tem entropia 1 (metade dos bits de cada sabor) mas cada pacote de 16bits tem entropia 0!

Numa sequência estacionária as propriedades estatísticas, entre elas a entropia, variam pouco ao longo do índice.

Assim consideraremos "suficientemente aleatória" a sequencia que tiver alta entropia e que cada segmento de 16bits também tenha alta entropia.

Endereços dinâmicos segundo a RFC–4941

Exemplos (só os 64 bits finais dos endereços IP)

4059	0598	5989	989e	8b1f	b1ff	1ffd	ffd6
0.896	0.896	0.988	0.999	0.988	0.811	0.811	0.696
f30d	30de	0de5	de56	5277	277c	77c0	7c0b
0.988	0.999	0.999	0.954	0.988	0.988	0.999	0.999

Mas alguns casos nitidamente artificiais passam no teste. Falsos positivos à vista:

beba	c0ca	c01a	b0b0	face	b00c	0000	25de
0.896	0.954	0.896	0.954	0.896	0.896	0.000	0.988

ou quase!

Endereços dinâmicos segundo a RFC–4941

Finalmente temos critérios para determinar com alguma chance de acerto os endereços dinâmicos:

1. **Presença do padrão ff:fe**
2. **Alta entropia e estacionariedade porcamente avaliadas.**

Pontos positivos: não requer consulta ao DNS e é fácil de implementar (máscara e contar bits).

Pontos negativos: podem ocorrer falsos positivos.

Conclusão: usar como critério complementar ou para pontuação.

Legitimação da MTA remetente

Quem está transmitindo uma mensagem de um domínio tem autoridade para isso?

O SPF (Sender Policy Framework, RFC–7208) ataca este problema. O "dono" do domínio publica uma lista de controle de acesso dizendo quem pode enviar email em nome do dado domínio.

SPF suporta IPv6 corretamente, mas tem problemas:

- **– domínios criados especificamente para enviar spam tem listas de controle de acesso muito permissivas;**
- **– a interpretação do 'softfail' é ambígua. Foi criado para transição, mas até hoje é muito usado.**

Legitimação da MTA remetente

Uma maneira de evitar SPF muito permissivo é testar um endereço sabidamente absurdo, por exemplo, ::1. Se passar, então a lista é podre e provavelmente de spammer.

No caso de 'softfail' você pode:

- **considerar como 'pass', mas é arriscado;**
- **considerar como 'fail', mas também é arriscado por conta de falso positivo;**
- **aplicar critérios complementares, como a análise do endereço IP sugerida anteriormente.**

Legitimação da MTA remetente

Há uma tendência em IPv6 a listas SPF muito amplas, como neste exemplo do Google/Gmail:

```
$ host -t txt _netblocks2.google.com
_netblocks2.google.com descriptive text "v=spf1 ip6:2001:4860:4000::/36
ip6:2404:6800:4000::/36 ip6:2607:f8b0:4000::/36 ip6:2800:3f0:4000::/36
ip6:2a00:1450:4000::/36 ip6:2c0f:fb50:4000::/36 ~all"
```

Será que o Gmail tem tantos endereços autorizados a enviar email em seu nome assim?

Listas assim amplas fazem com que o próprio SPF perca seu valor como fator de legitimação do MTA remetente.

Felizmente mensagens legítimas do Gmail são assinadas pelo DKIM, mas esse é um método que se aplica depois que a mensagem já entrou.

Legitimação da MTA remetente

E se não existir registro SPF?

Uma abordagem de algum sucesso é ter uma lista padrão para esses casos: `+mx -all`

Se a MTA remetente for também um receptor válido para o domínio, a consideramos válida.

Outra idéia é validar o certificado usado na abertura de sessão TLS, se for o caso, se contém o domínio do remetente em seu 'dn', como se faz em https.

É uma idéia incipiente mas que pode ser explorada.

Conclusões

– É fácil distinguir usuários finais em IPv6 sem recorrer a DNS a partir de padrões do SLAAC e de outros endereços dinâmicos, com razoável chance de acerto.

– SPF funciona para IPv4, e também para IPv6, mas alguns cuidados são necessários para evitar armadilhas de listas muito promíscuas e ter uma política razoável para o caso 'softfail'.

– Outros objetos da transação do envio de uma mensagem podem ser escrutinados, por exemplo, o certificado usado para estabelecer sessão cifrada.

Agradecimentos

às inspiradoras discussões que rolam nas listas

gter@eng.registro.br

masoch-l@eng.registro.br

aos organizadores deste evento.