De onde vem o spam?

Seis meses de funcionamento de um ’spamtrap’

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

Objetivo do experimento

1. **Fornecer subsídios para melhorar nosso arsenal anti−spam;**

2. **Confirmar ou refutar alguns mitos sobre a origem dessa poluição digital.**

Limitações

1. **A armadilha fica atrás de um mata−burro com duas listas de bloqueio:**
   zen.spamhaus.org + bl.spamcop.net

2. **Também são rejeitados os IPs/envelopes com SPF fail.**

Material coletado

Período de operação: maio a outubro de 2010

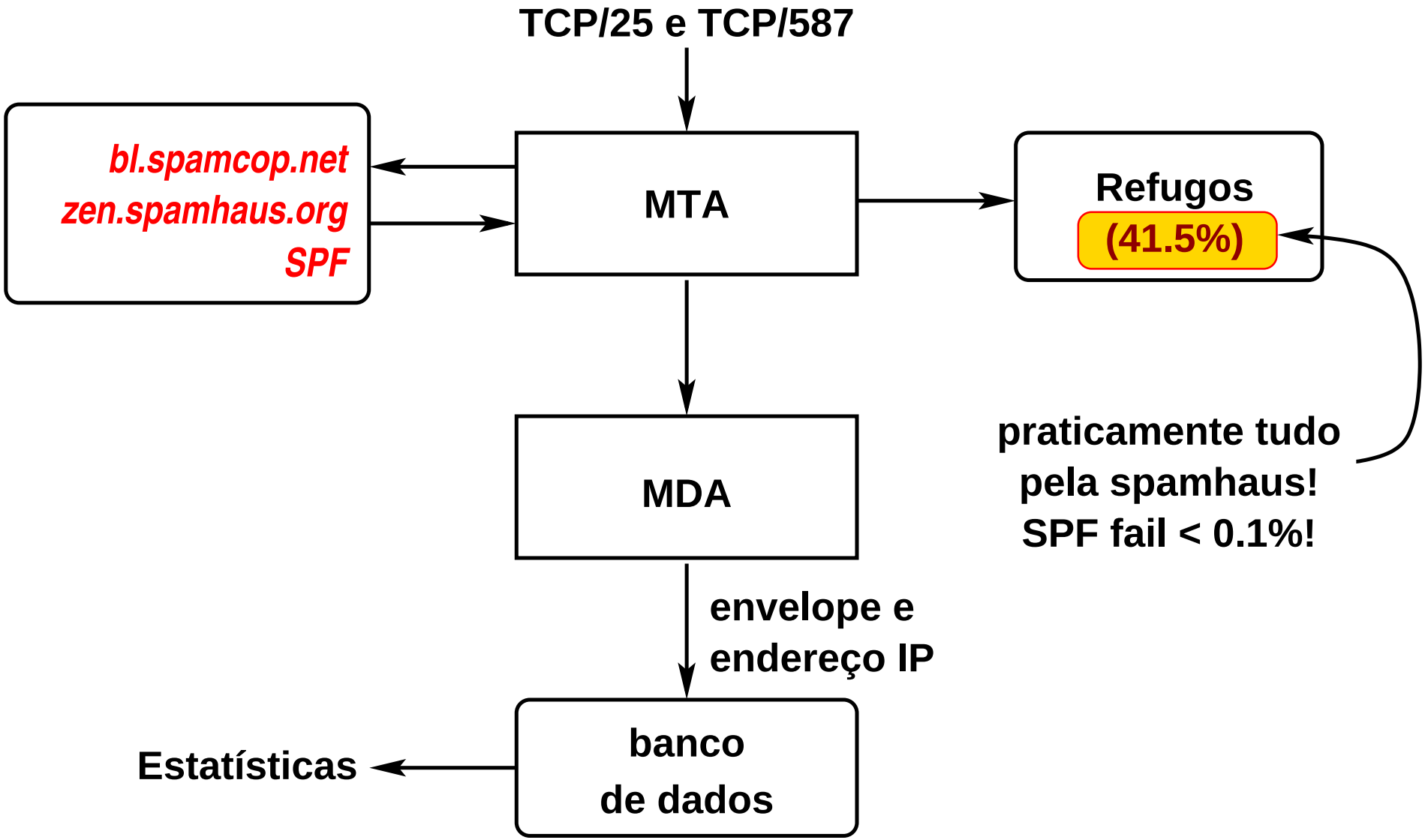
Aproximadamente 750 mil mensagens/mês

Apenas 8098 endereços IPv4 únicos (e um IPv6!)

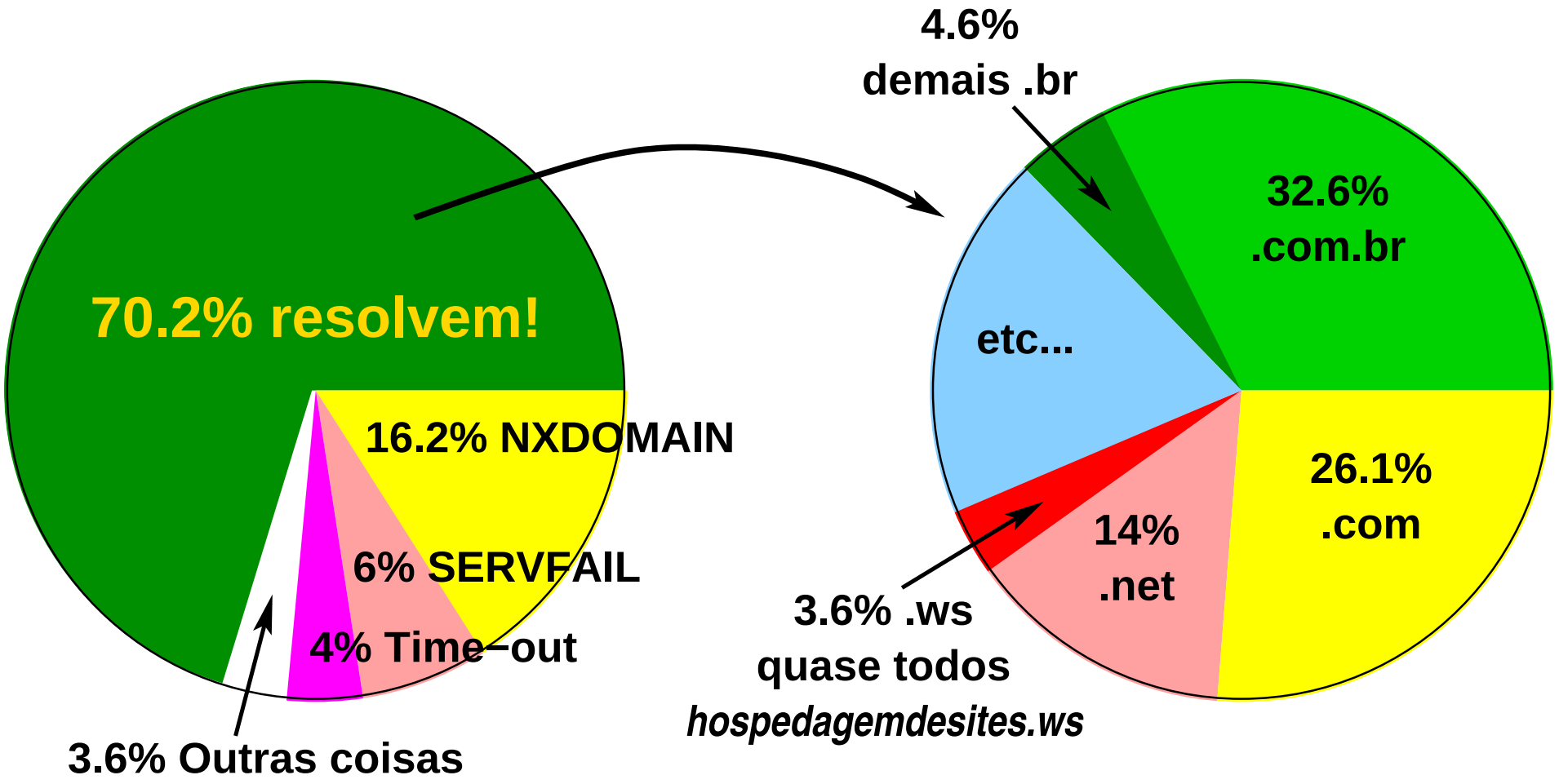
Constatação:

Relativamente poucas origens para muitas mensagens!

Arranjo da arapuca



O famoso REVERSO!



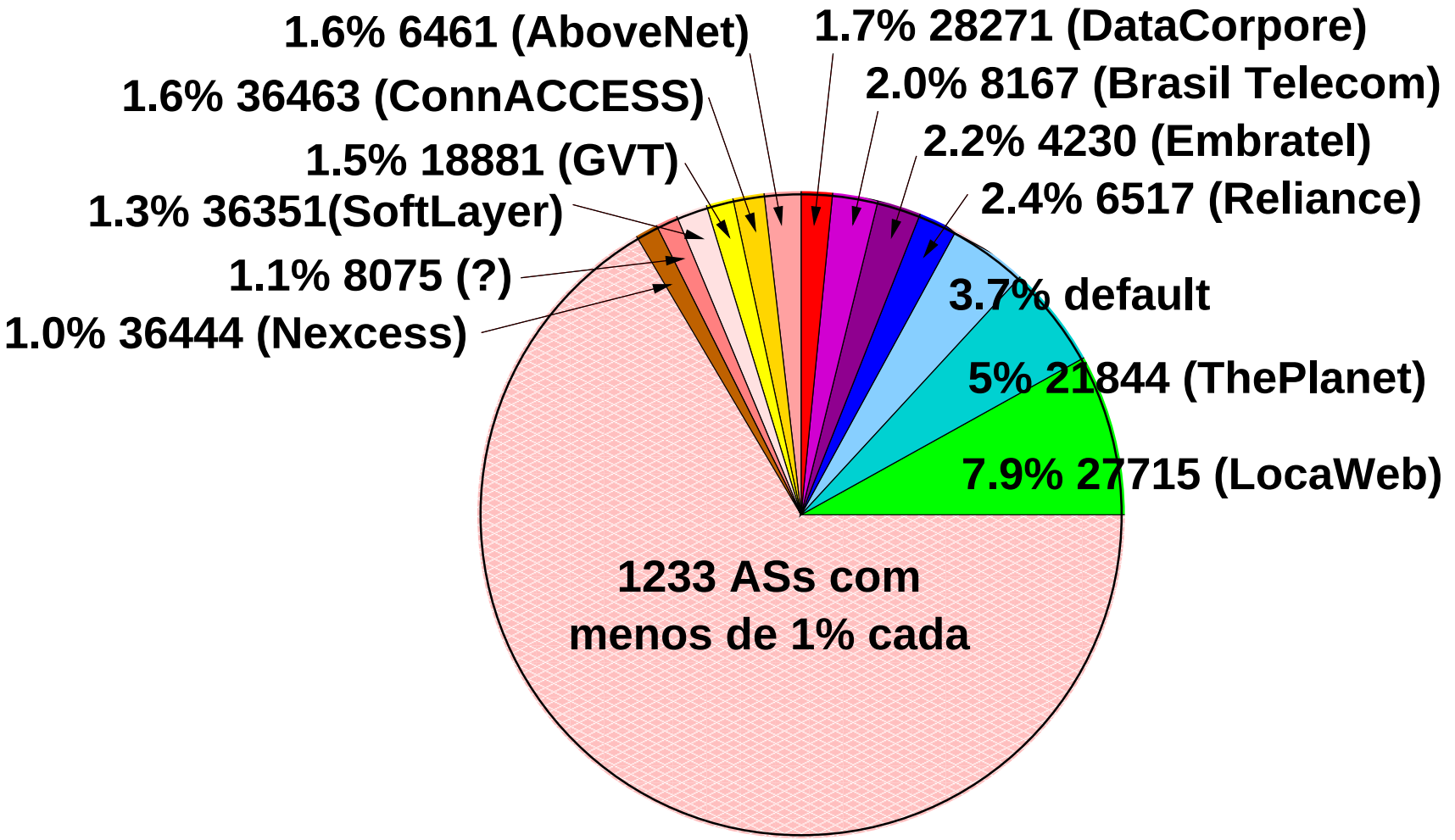
Ainda o reverso, os TOP−10, exceto TLD/SLDs.

pos.	(%)	domínio
1	5.9	locaweb.com.br
2	3.7	hospedagemdesites.ws
3	1.7	available.above.net
4	1.6	yahoo.com
5	1.2	hotmail.com
6	0.9	telesp.net.br (quase tudo dsl.telesp.net.br)
7	0.9	terra.com (quase tudo mta.terra.com)
8	0.8	aknamail000.com.br
9	0.8	mailsender.com.br
10	0.7	virtua.com.br

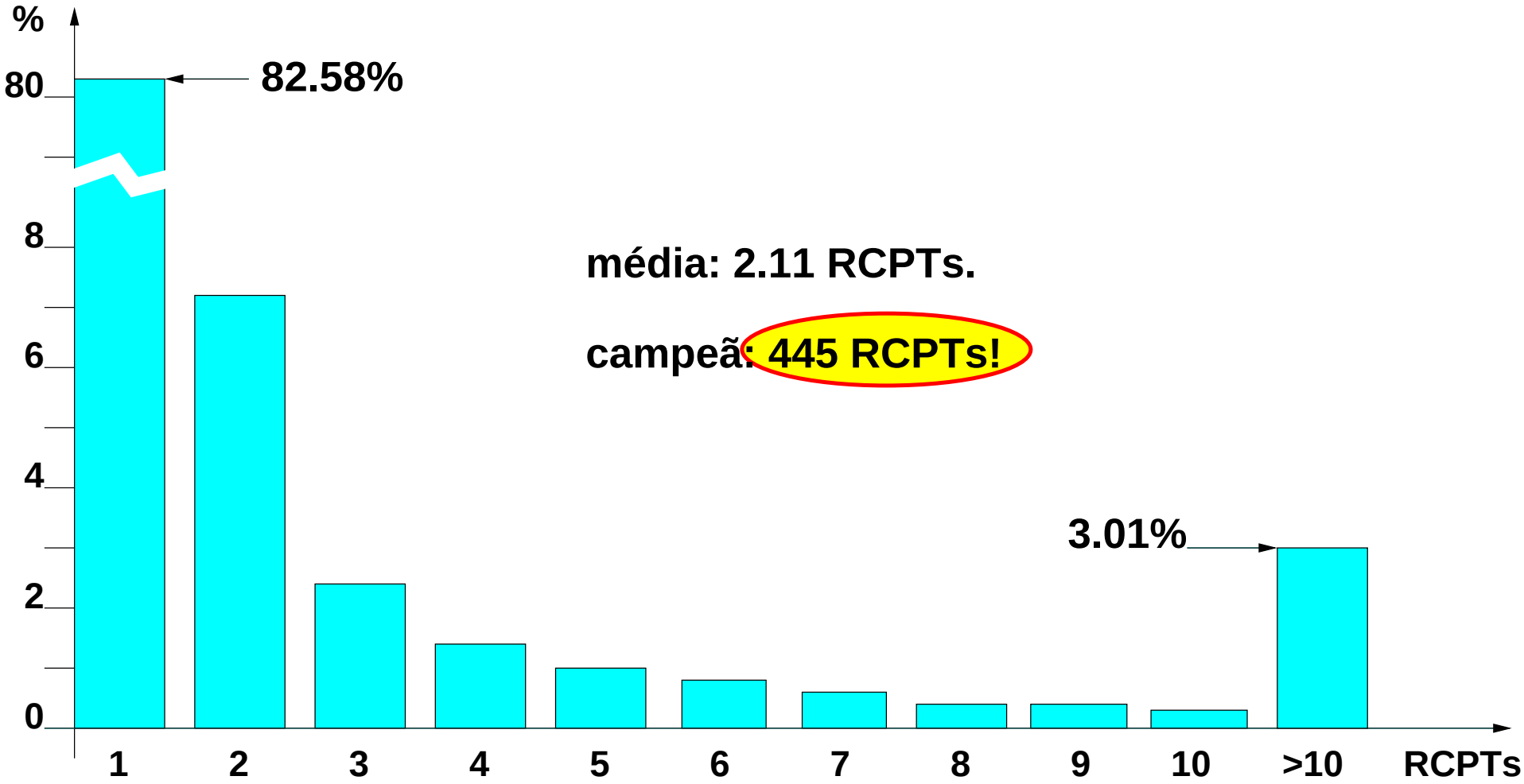
Os dois primeiros lugares são da LocaWeb.

mailsender.com.br é um nome bem sugestivo de qual deva ser sua finalidade.

Divisão por Sistema Autônomo



Curiosidade: RCPT/mensagem



média: 2.11 RCPTs.

campeã: 445 RCPTs!

Limitações deste estudo

1. **Consideramos somente as origens dos spams, e não o volume originado de cada origem;**

2. **A pré filtragem (listas de bloqueio e SPF) eliminam várias amostras que poderiam mudar os resultados;**

3. **Não guardamos as mensagens, nem mesmo amostras, que poderiam ser usadas para treinar algum filtro Bayesiano;**

4. **Não fizemos qualquer análise dos envelopes (tirando SPF), o que poderia trazer dados úteis.**

Conclusões e desdobramentos

1. **SPF saiu "chamuscado" do experimento. Muito pouco spam foi rejeitado com SPF fail.**

2. **Apenas 7 sistemas autônomos respondem por 25% dos endereços IP enviadores de SPAM.**

3. **Prevenir spam com base na tradução reversa do IP é ineficaz, pois mais de 70% dos endereços enviadores de spam tinham tradução reversa correta.**

4. **Cerca de 8 mil endereços enviam 750 mil mensagens por mês, o que mostra uma certa concentração apesar da tendência ao uso de botnets.**

5. **Vale a pena repetir este estudo com base no volume de mensagens recebidas.**

Referência

http://www.antispam.br/

sítio com recomendações e práticas contra a proliferação dessa praga;