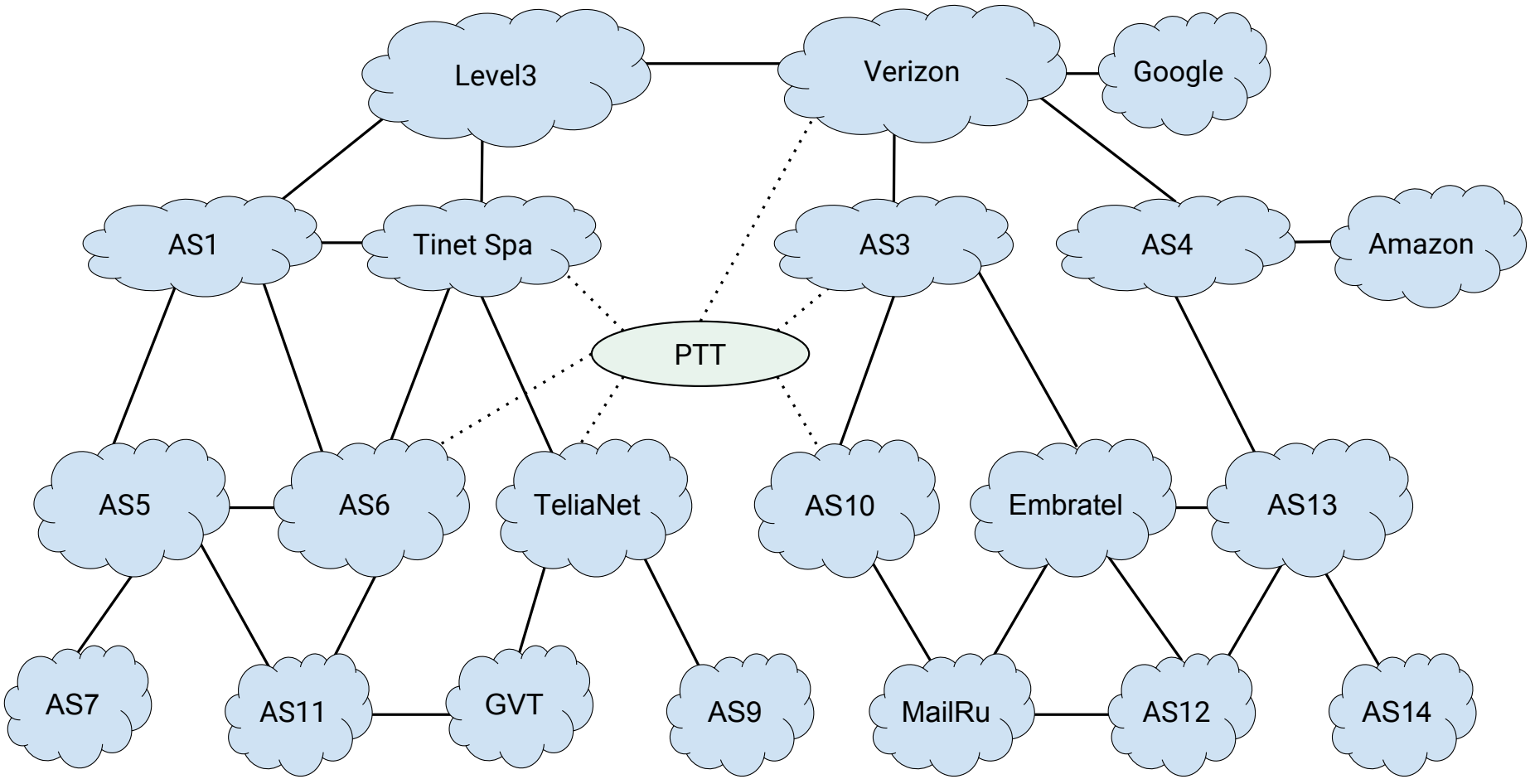
# MEDIÇÃO, CARACTERIZAÇÃO E REDUÇÃO DOS CUSTOS ASSOCIADOS AO TRÁFEGO DE SPAM

Osvaldo Luís H. M. Fonseca

DCC-UFMG
NIC.br

13 de maio de 2016

# Internet e sistemas autônomos



# Mensagens de spam e o tráfego

- 70% das mensagens de e-mail
- Frequentemente associadas a:
  - propagação de malware
  - divulgação de produtos ilegais
  - disseminação de mensagens de phishing

# Combate ao spam e custos associados

Batalha contra os spammers:

- combate ao envio de spam na origem
- evitar custo gerado pelo tráfego de spam

Custo do spam:

- spam gera um alto custo para as entidades da Internet
- apenas custo agregado

# Objetivos

Analisar o custo gerado pelo tráfego de spam para cada sistema autônomo

- incentivar a filtragem do spam

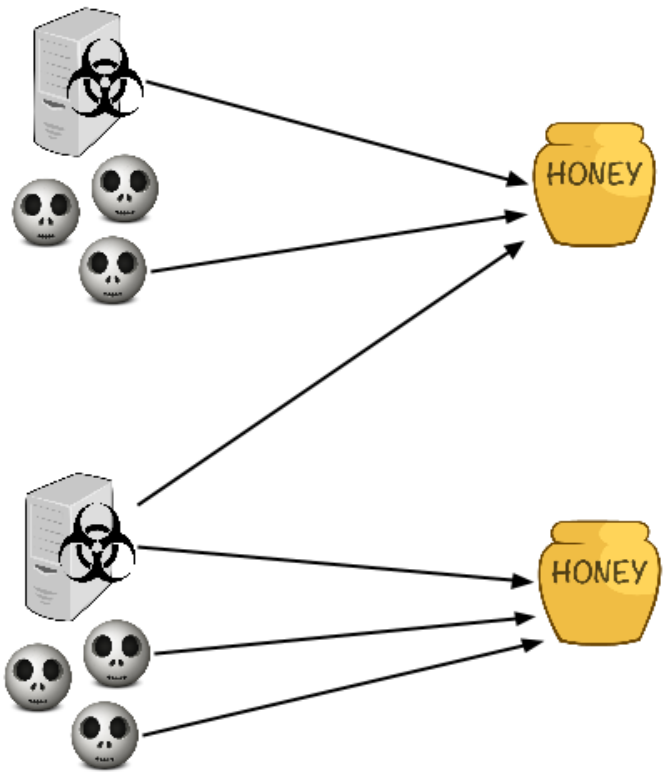
Mitigar o custo do tráfego de spam

# Sumário

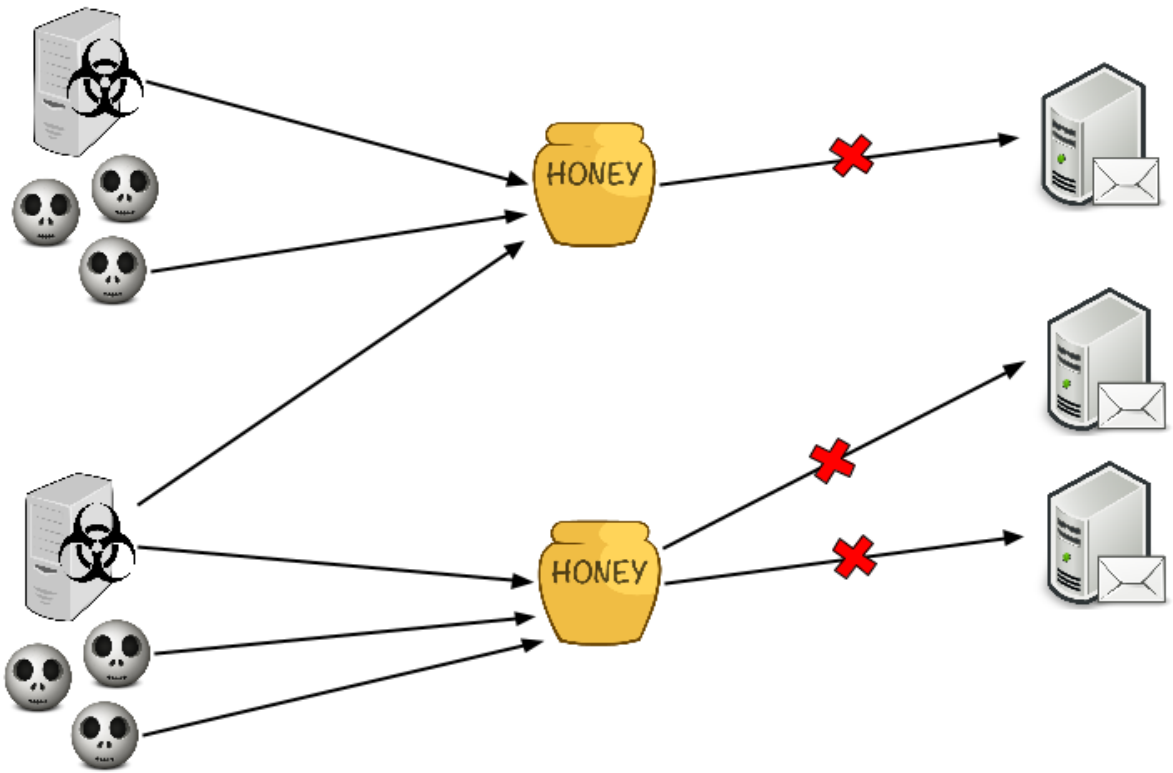
# Sumário

# Infraestrutura de coleta


# Infraestrutura de coleta



# Infraestrutura de coleta



# Infraestrutura de coleta

Foram utilizados cinco honeypots:

- AT, BR, US, NL e UY
- redes com características diferentes

As mensagens foram coletadas entre 8 de setembro e 8 de outubro de 2015.

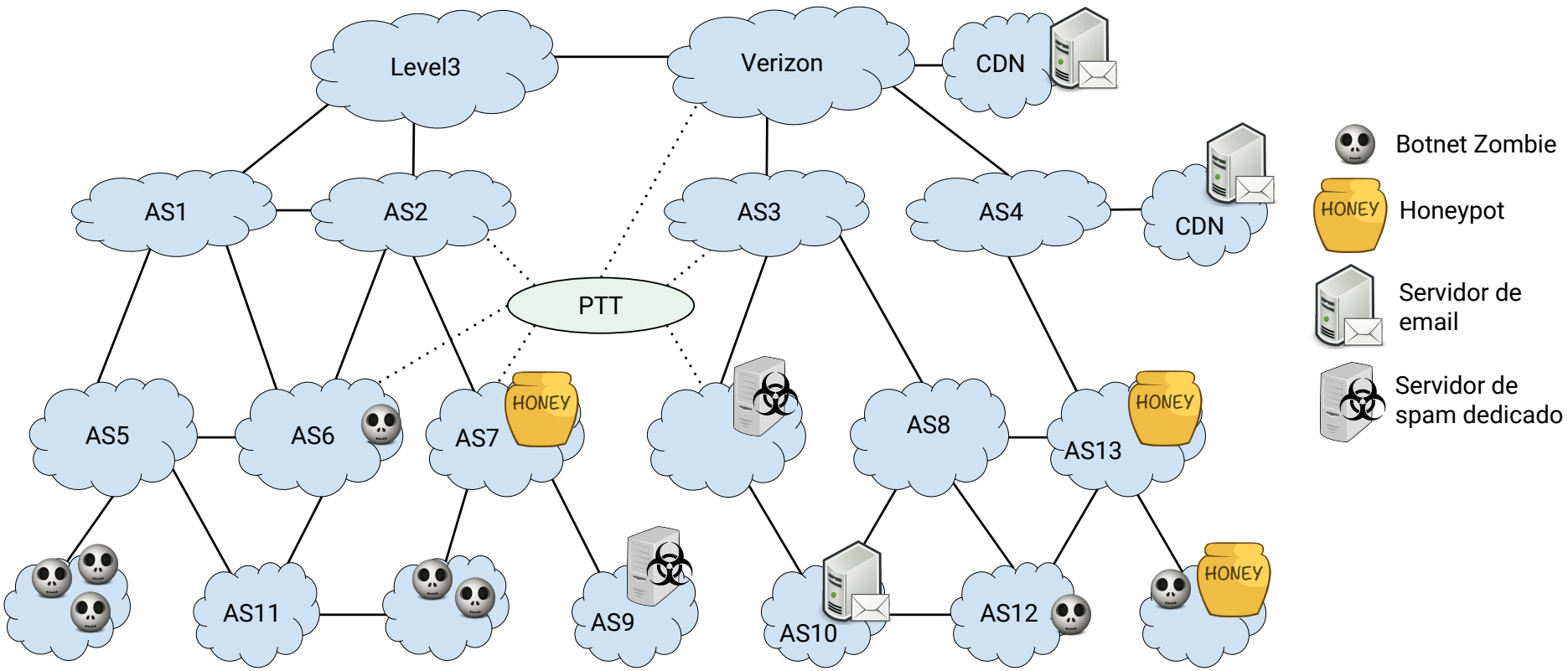
# Infraestrutura de coleta

Caracterização das 133 milhões de mensagens de spam capturadas pelos honeypots:

- mais 56.000 endereços IP distintos
- 879 ASes
- 115 países

| | Endereços IP (%) | Mensagens (%) |
|---|---|---|
| Open relay | 82 | 27 |
| Open proxy | 18 | 73 |

# Infraestrutura de coleta



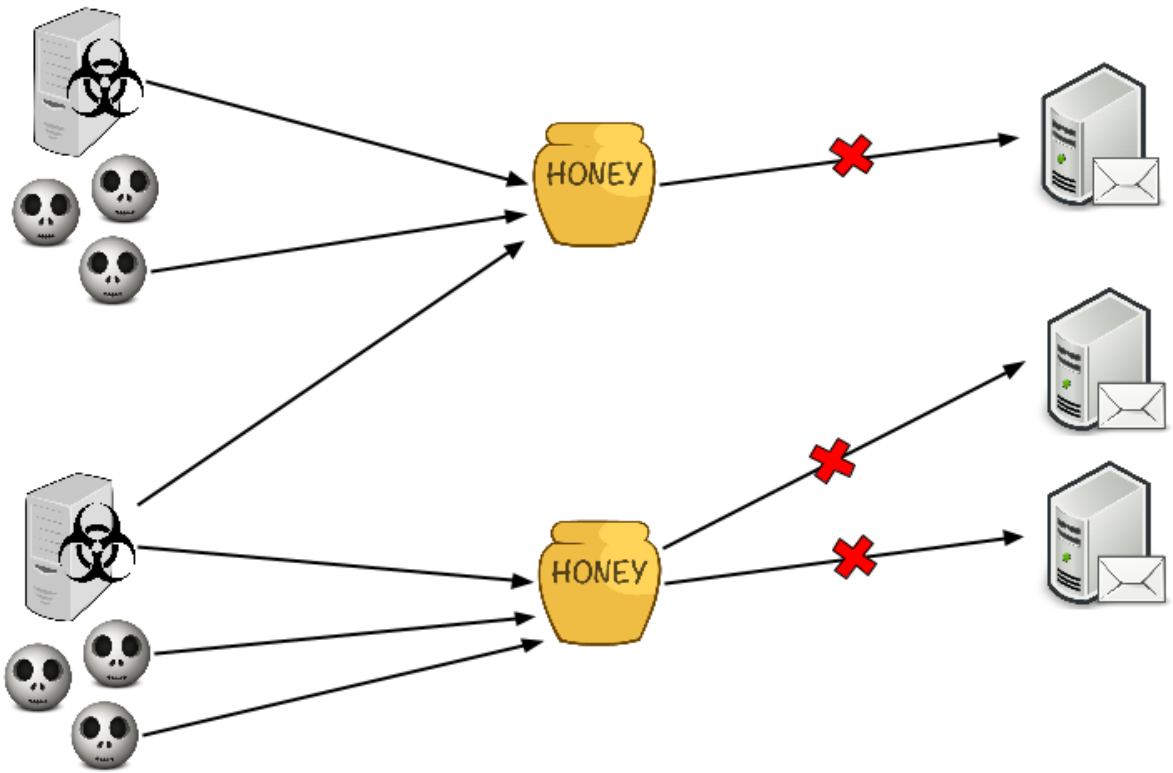
# Sumário

# Medição das rotas percorridas pelas mensagens

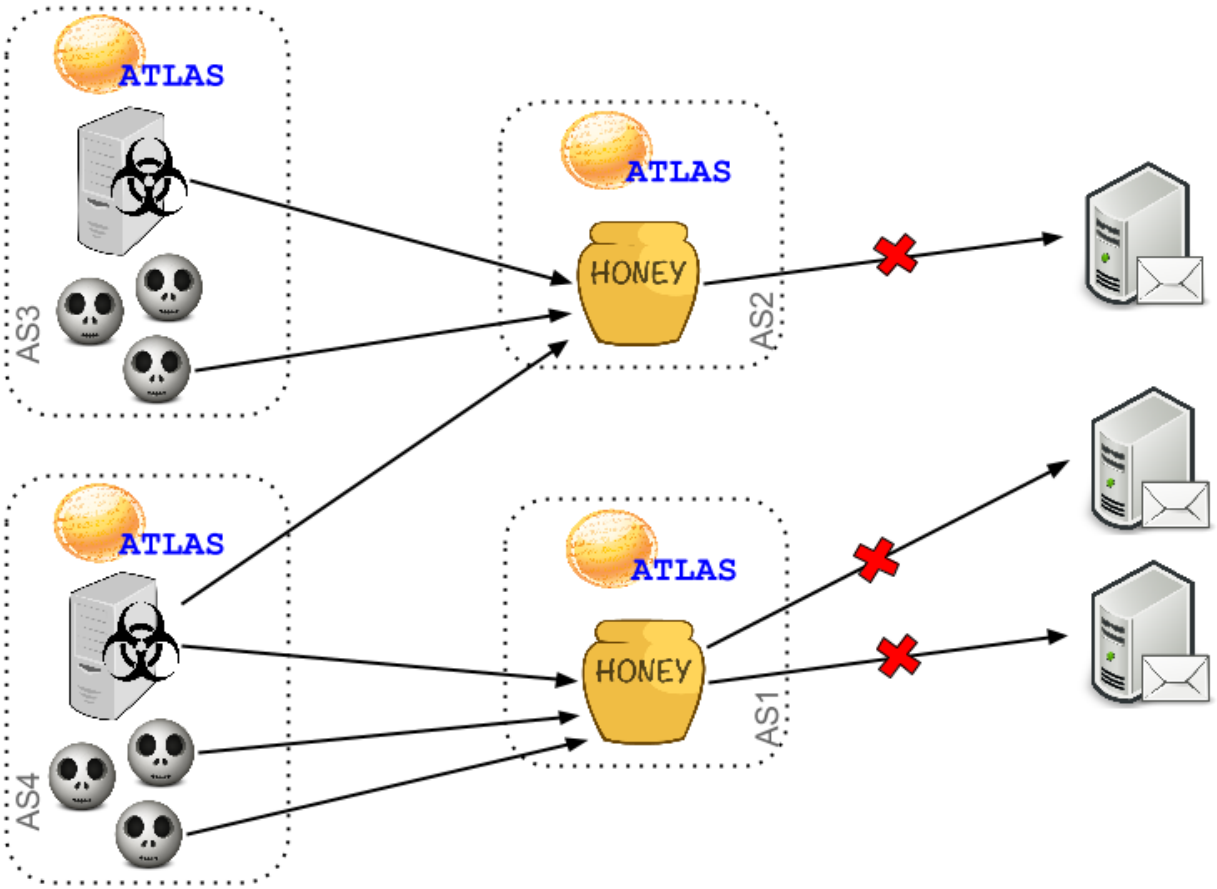
Ripe Atlas é uma plataforma de medição distribuída:

- mais de 9.000 máquinas em diferentes países

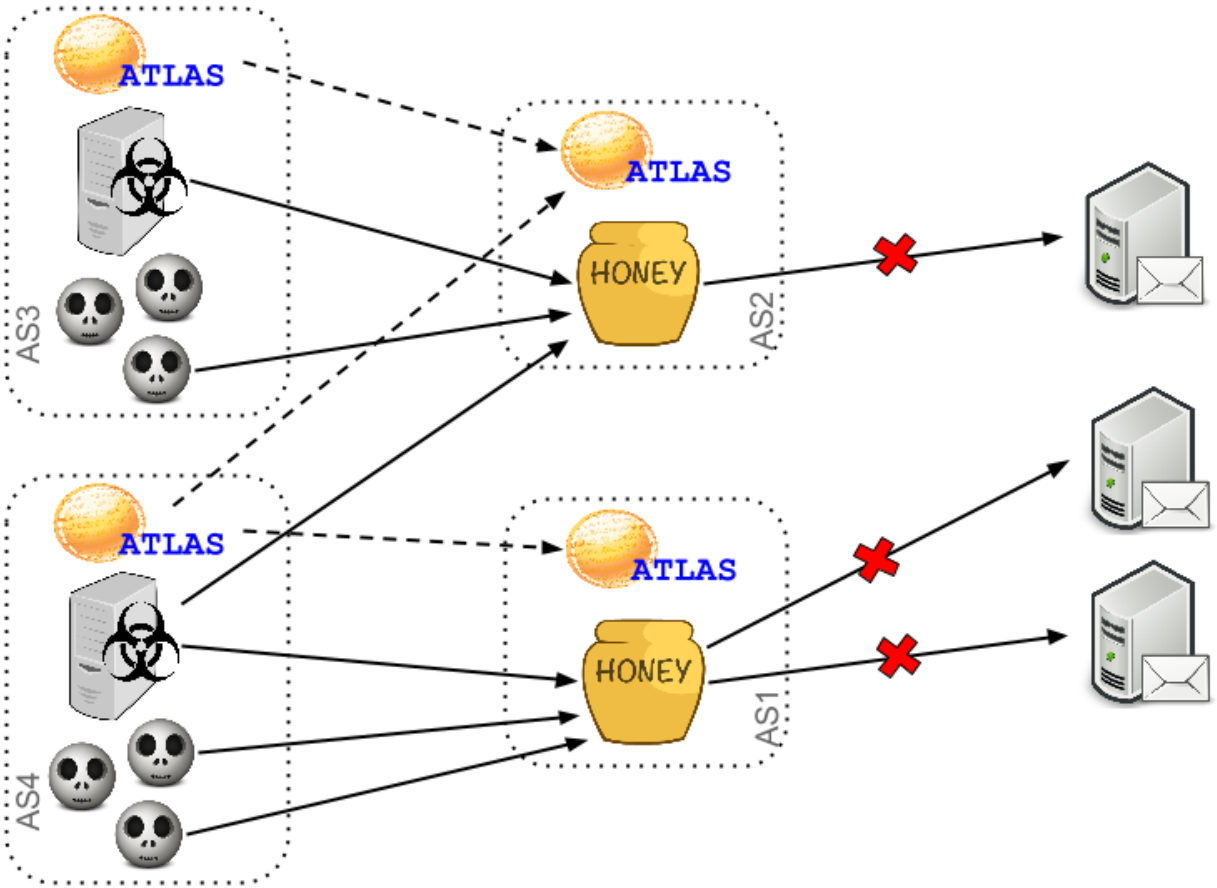

# Medição das rotas percorridas pelas mensagens



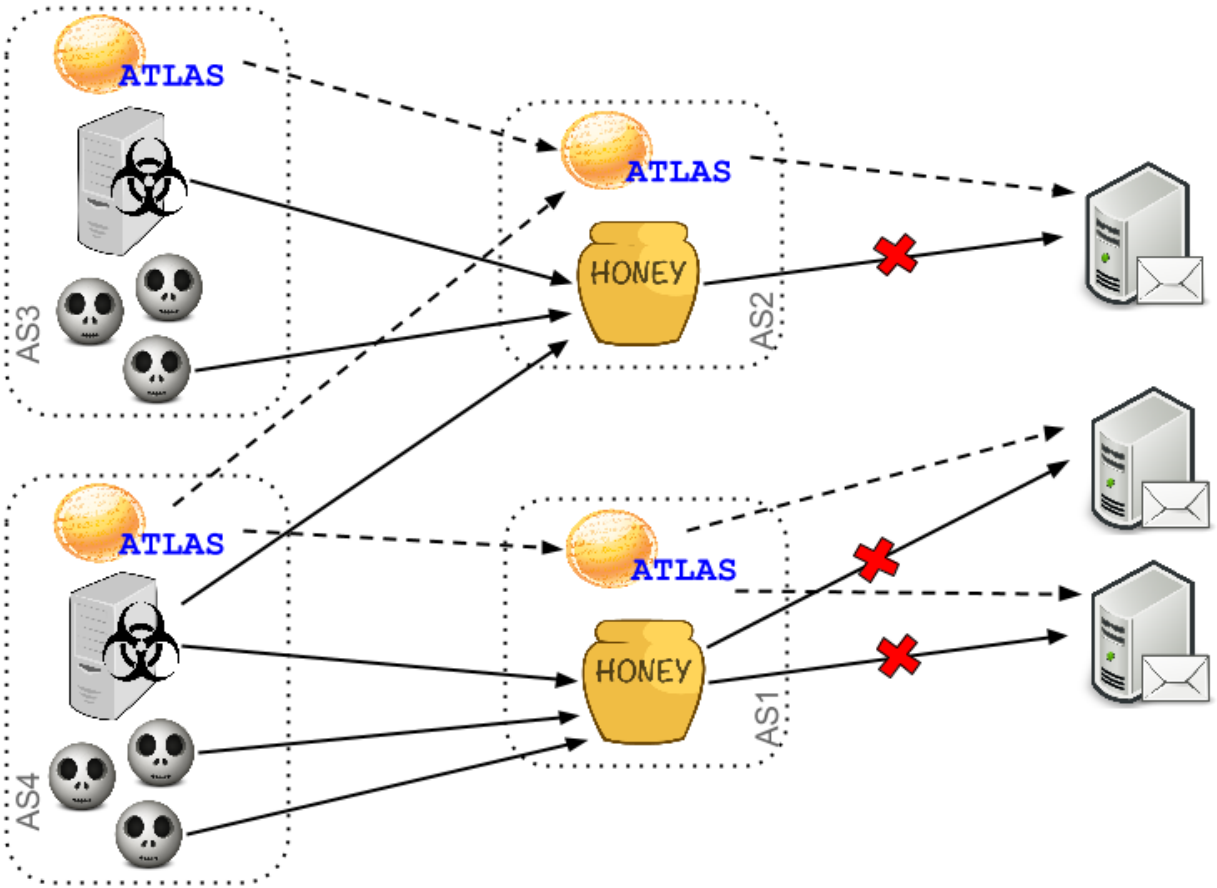
# Medição das rotas percorridas pelas mensagens



# Medição das rotas percorridas pelas mensagens



# Medição das rotas percorridas pelas mensagens



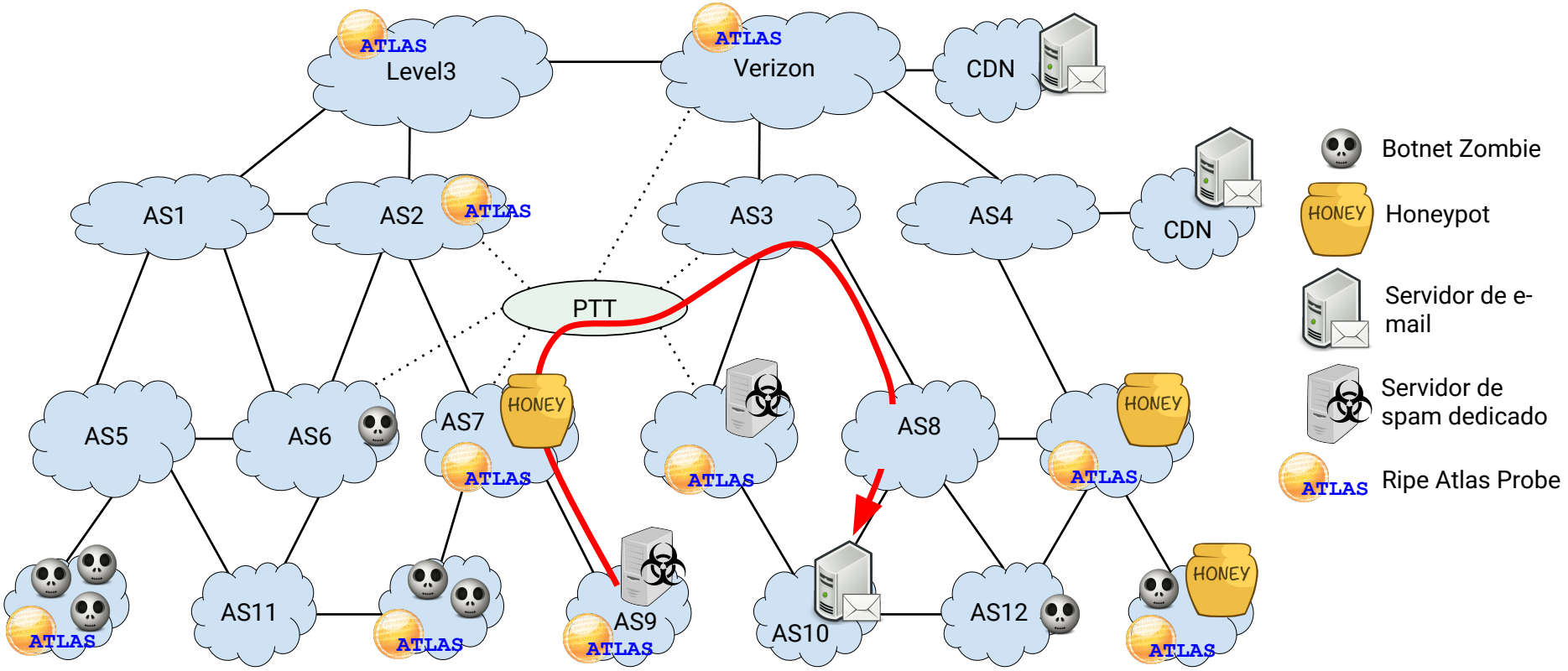
# Medição das rotas percorridas pelas mensagens

- Cota de 500 medições por honeypot:
  - 50 medições para as rotas entre spammers e o honeypot
  - 450 medições para as rotas entre o honeypot e os servidores de e-mail
- 57.419 medições de traceroute

# Medição das rotas percorridas pelas mensagens

- 40% dos 879 ASes são cobertos por máquinas do RIPE Atlas
- o envio de spam está concentrado em poucos ASes
- 92% das mensagens são cobertas (spammers e honeypots)
- 70% das mensagens são cobertas (honeypots e servidores de e-mail)

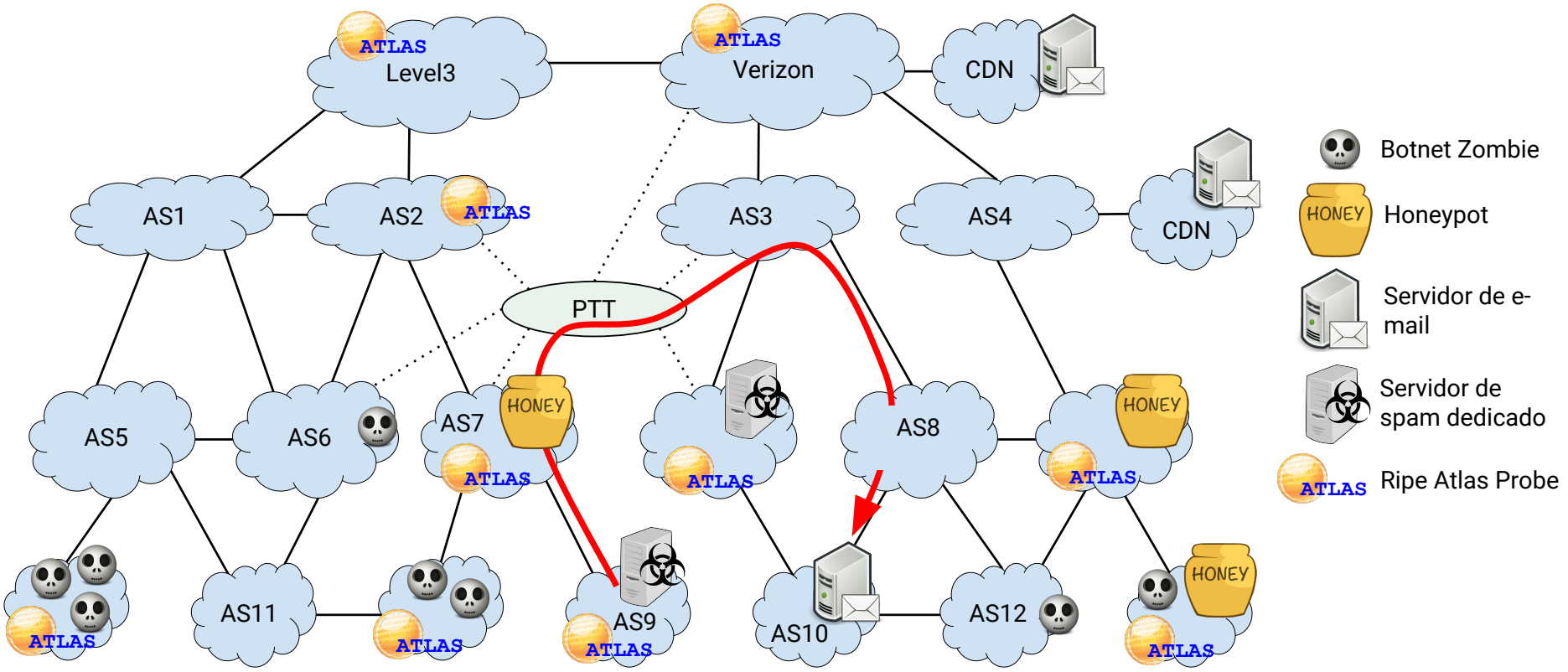
# Medição das rotas percorridas pelas mensagens




- Proteger a identidade dos honeypots

# Sumário

# Mapeamento do traceroute para o caminho no nível de AS




- Caminho no nível de AS

# Mapeamento do traceroute para o caminho no nível de AS

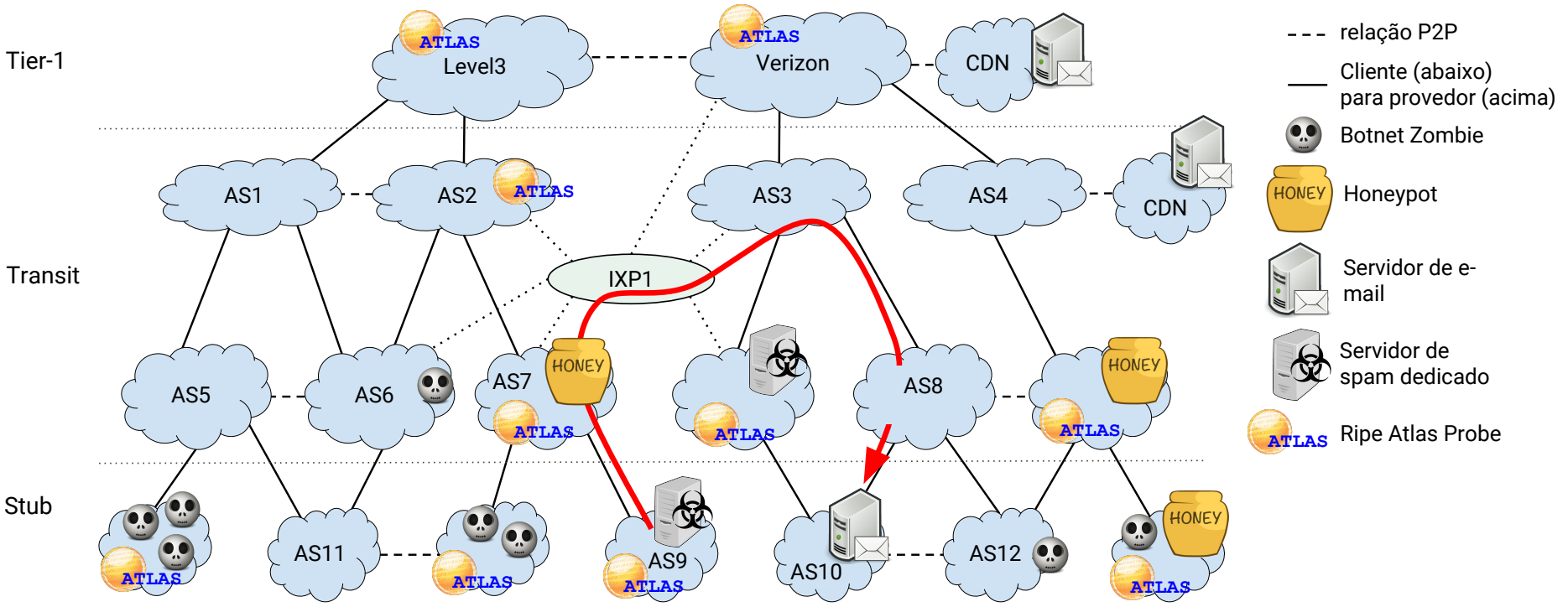
- Base de dados iPlane
  - mapeamento de endereço IP para AS
  - mapeia um prefixo IP para o AS que anunciou aquele prefixo

# Mapeamento do traceroute para o caminho no nível de AS

- Avaliação para um dos honeypots:
  - caminhos no nível de AS obtidos pelo mapeamento das rotas
  - caminhos no nível de AS retiradas da tabela BGP
- 89,16% das rotas no nível de AS são idênticas
- 99,39% das rotas no nível de AS diferem em no máximo um AS

# Sumário

# Cálculo do custo do tráfego de spam



# Cálculo do custo do tráfego de spam

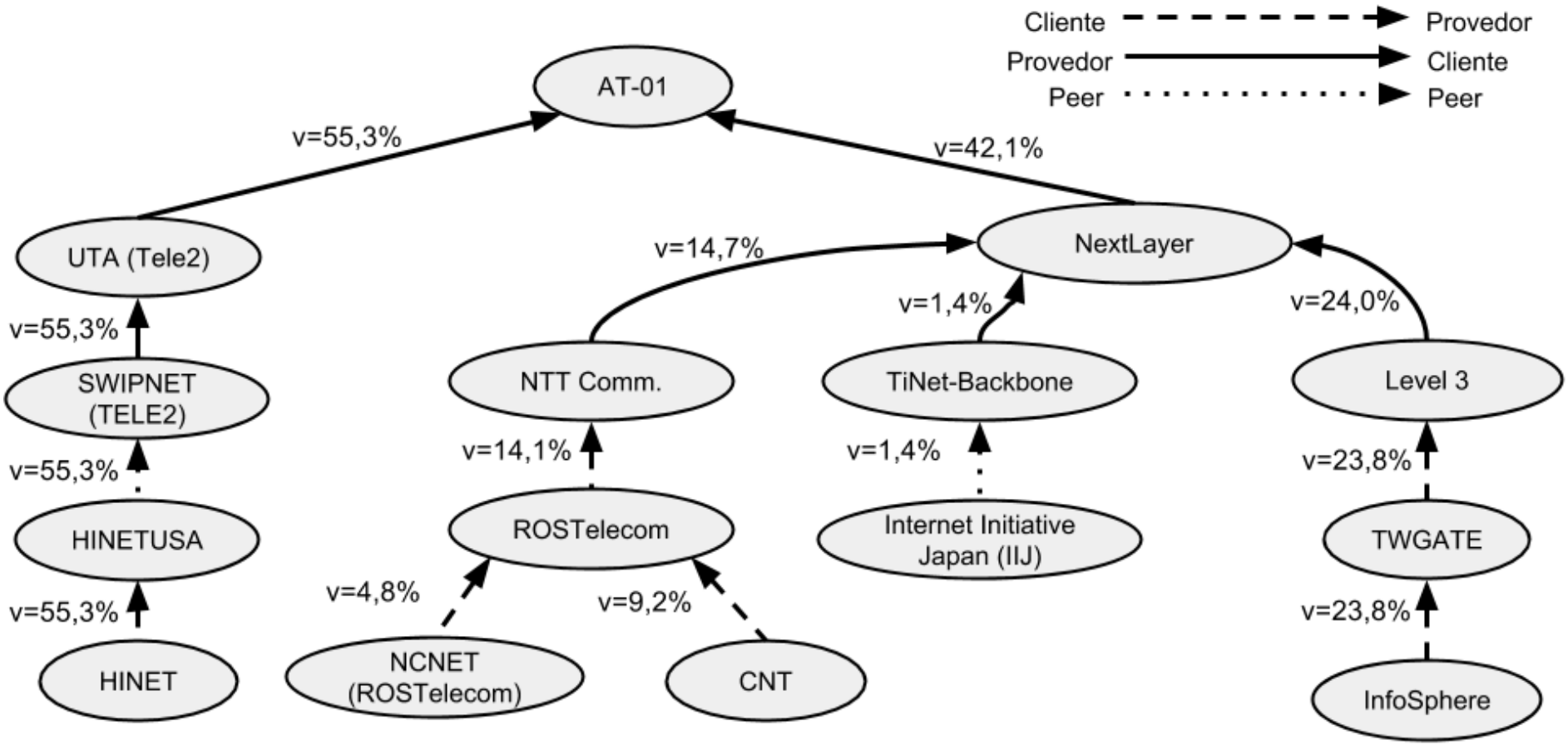
- Base de dados de relações entre ASes da CAIDA
  - 95% de precisão
  - cliente-provedor
  - ASes parceiros
- Assumimos que:
  - clientes pagam a provedores
  - parceiros trocam tráfego livres de cobrança

# Cálculo do custo do tráfego de spam

- Contratos das relações comerciais entre ASes são privados:
  - não conseguimos estimar o valor de cada byte
  - estimamos o custo como o volume de spam trafegado entre provedores e clientes

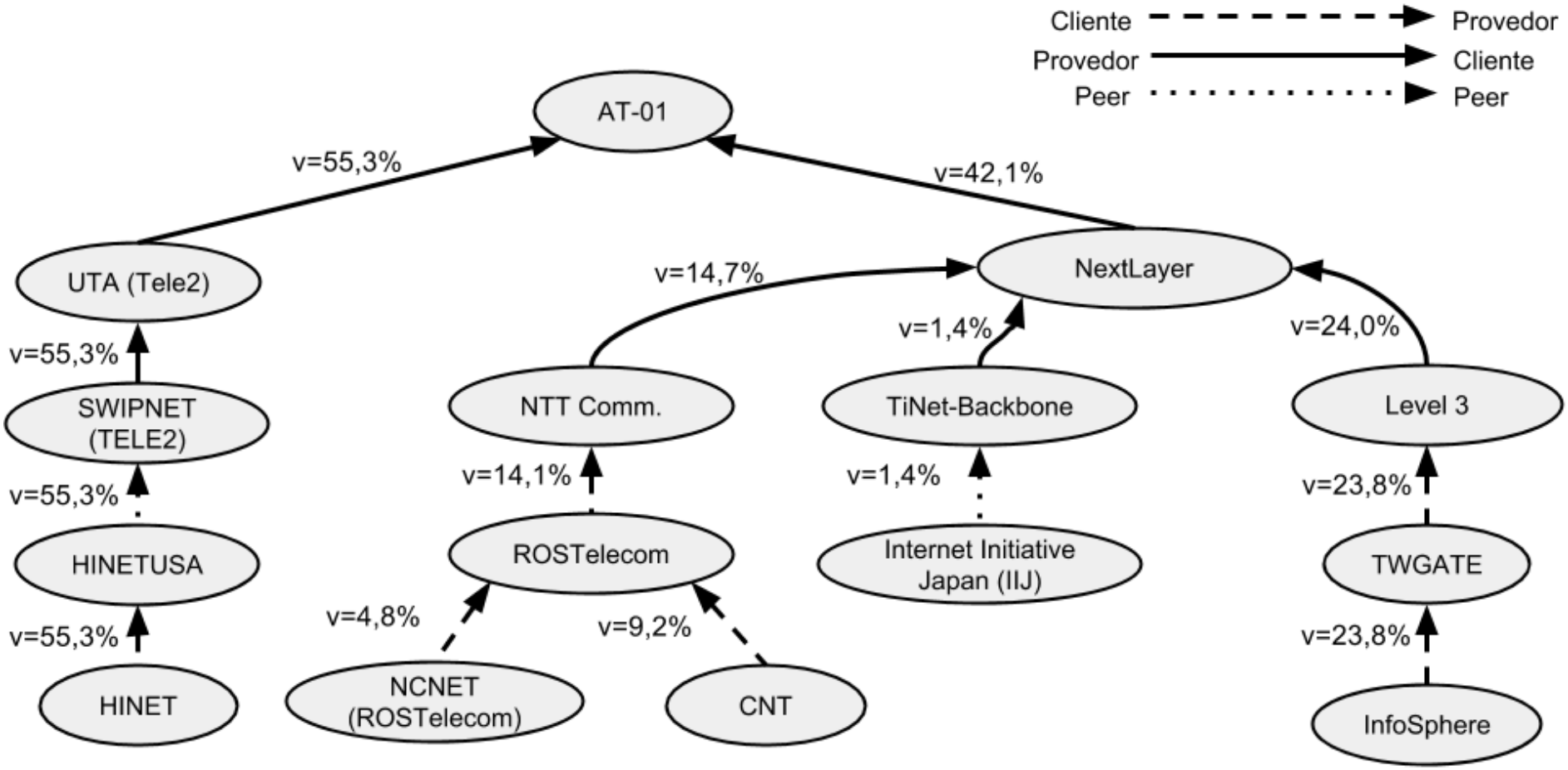
# Sumário

# Estudo de caso - Spammers ao honeypot

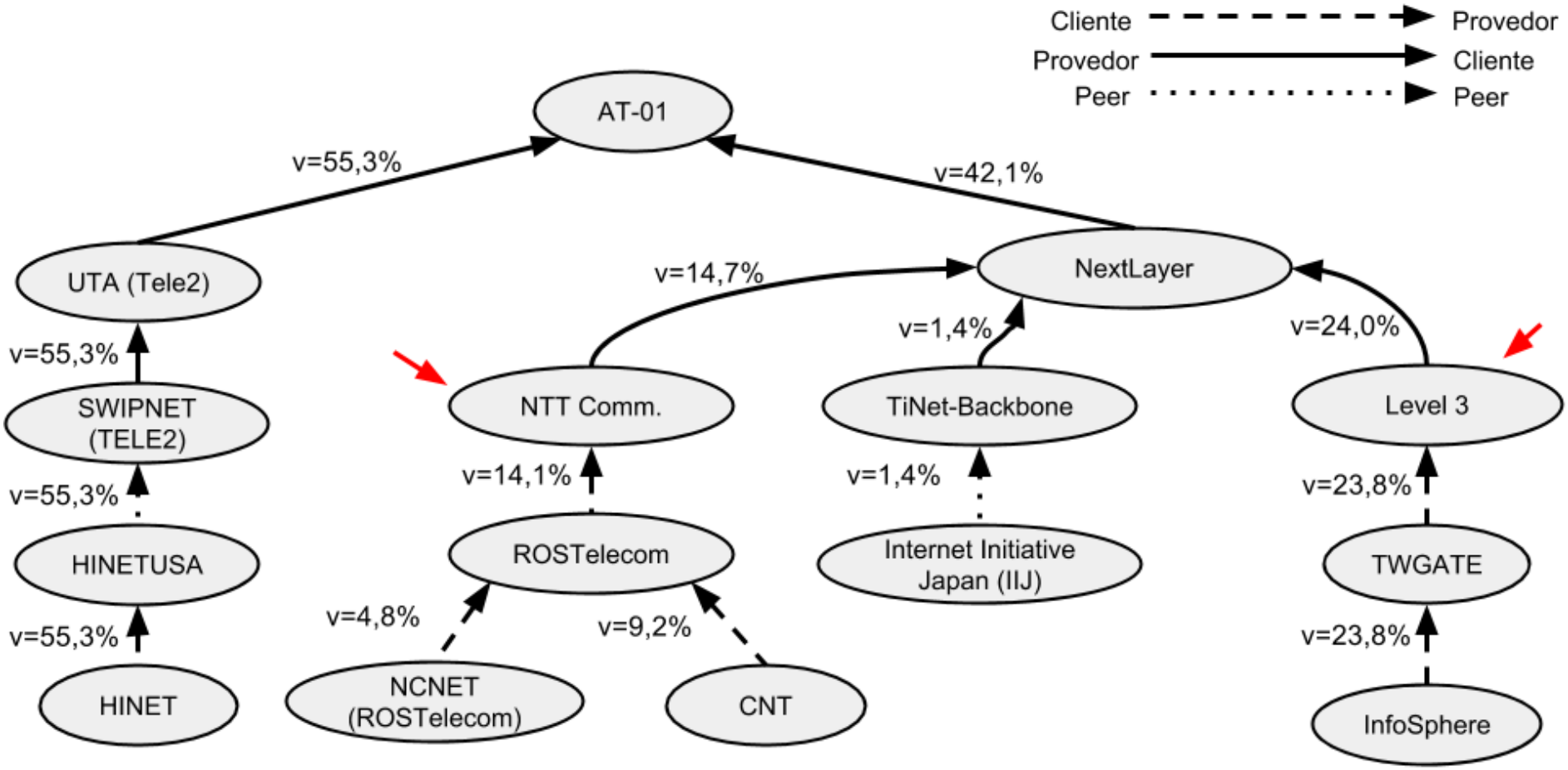



- O sentido das arestas indicam a direção das medições
- O tipo da linha indica o tipo de relação entre os ASes

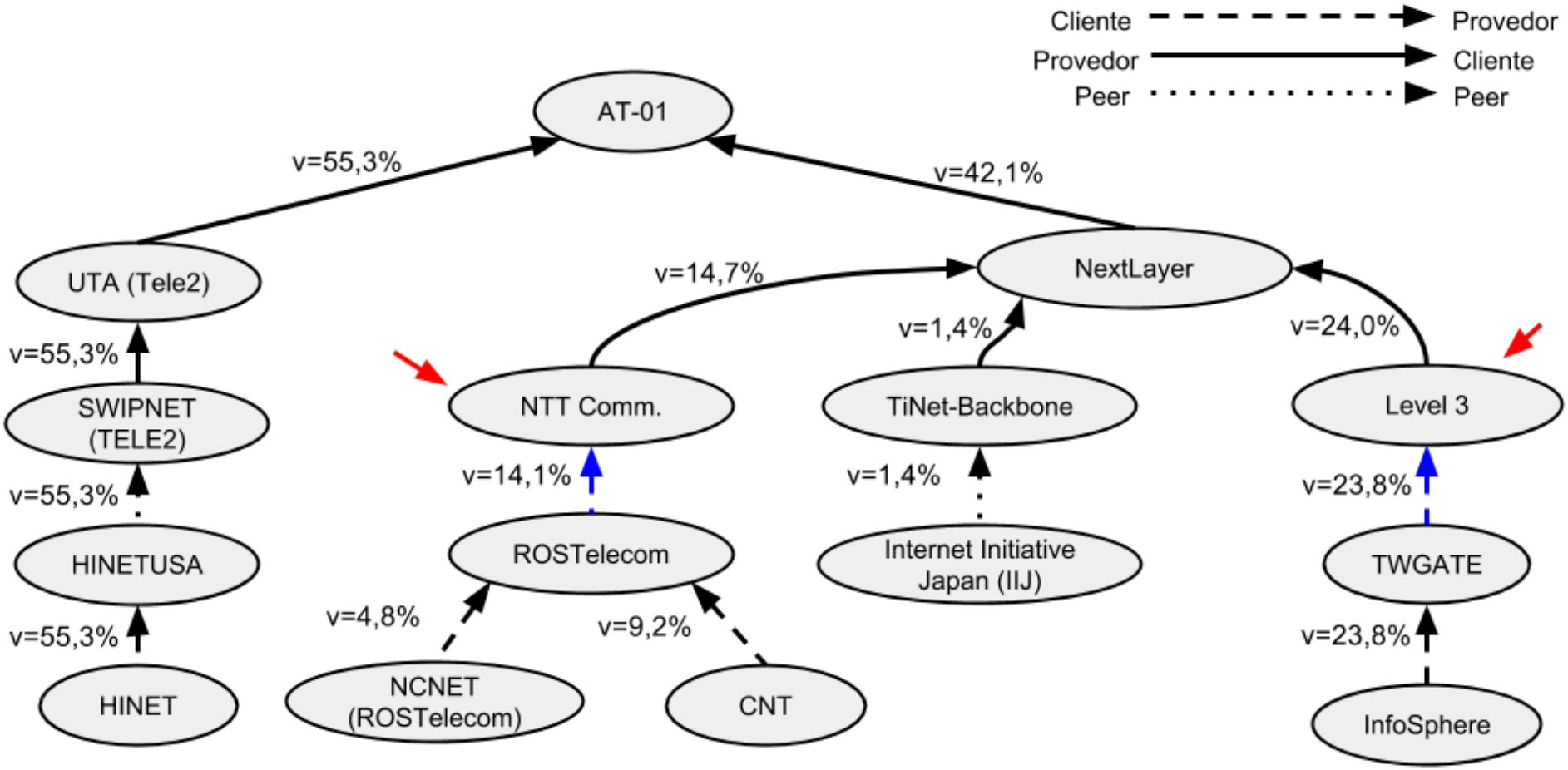
# Estudo de caso - Pouco incentivo



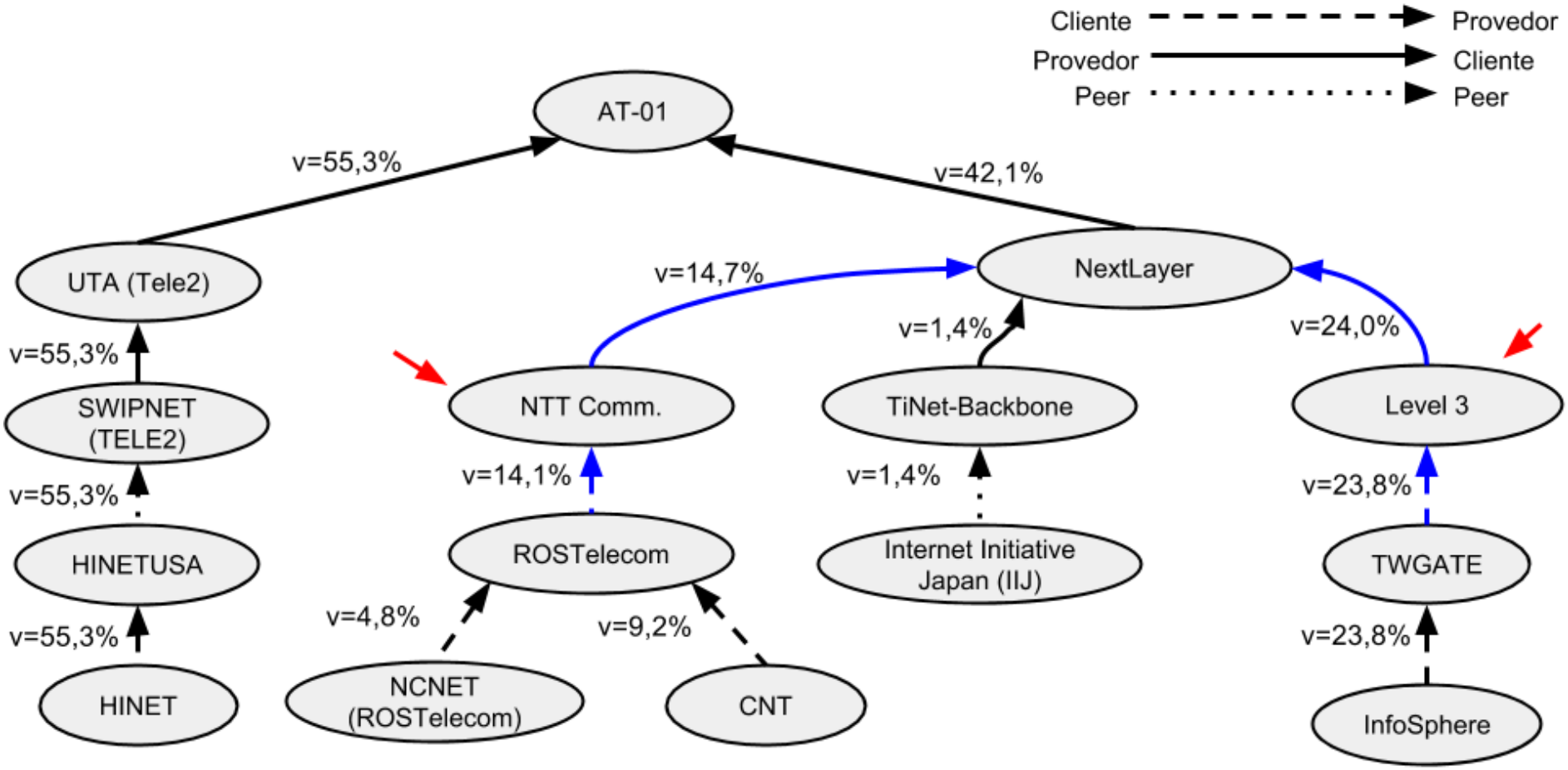
# Estudo de caso - Pouco incentivo



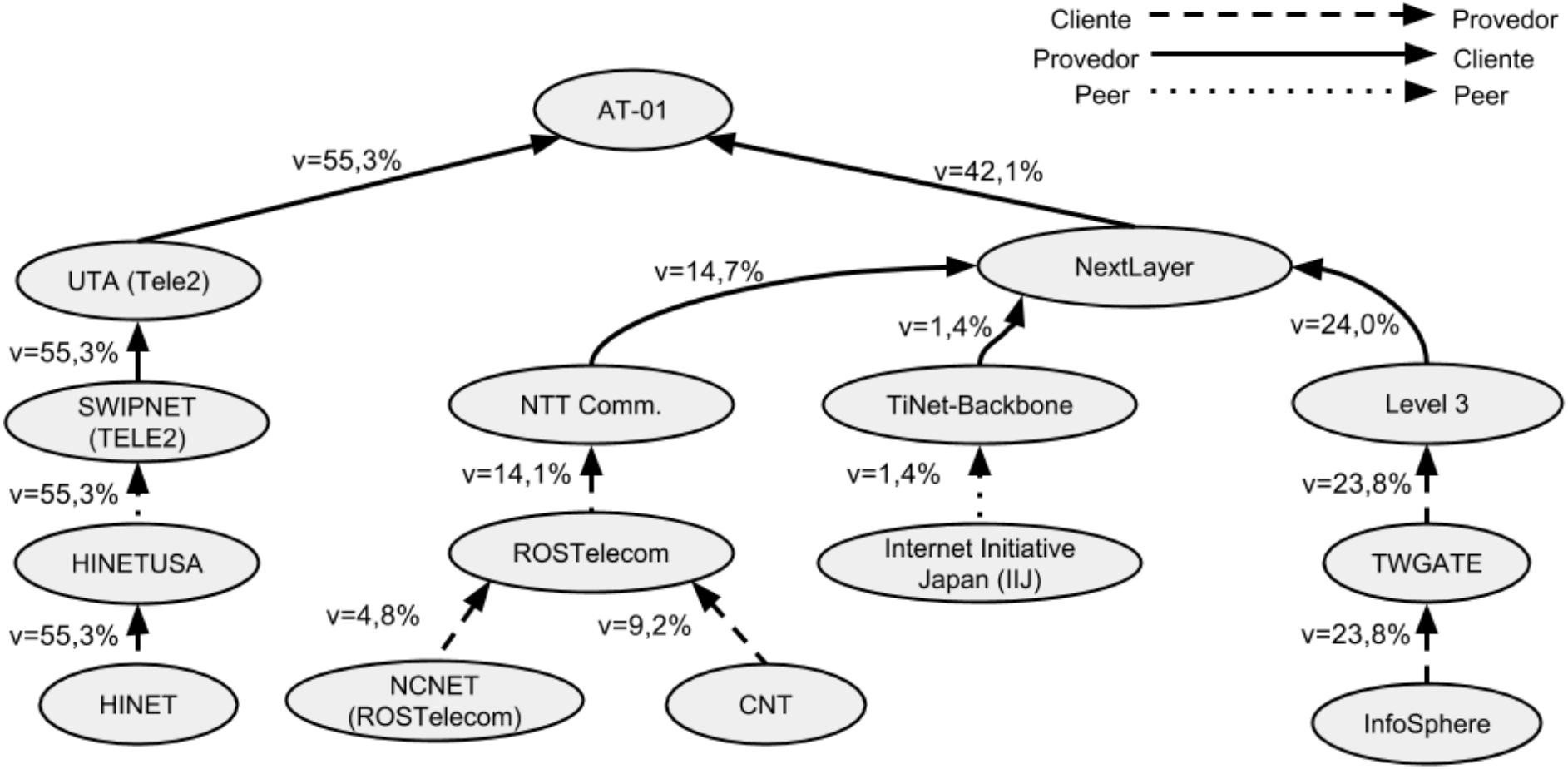
# Estudo de caso - Pouco incentivo



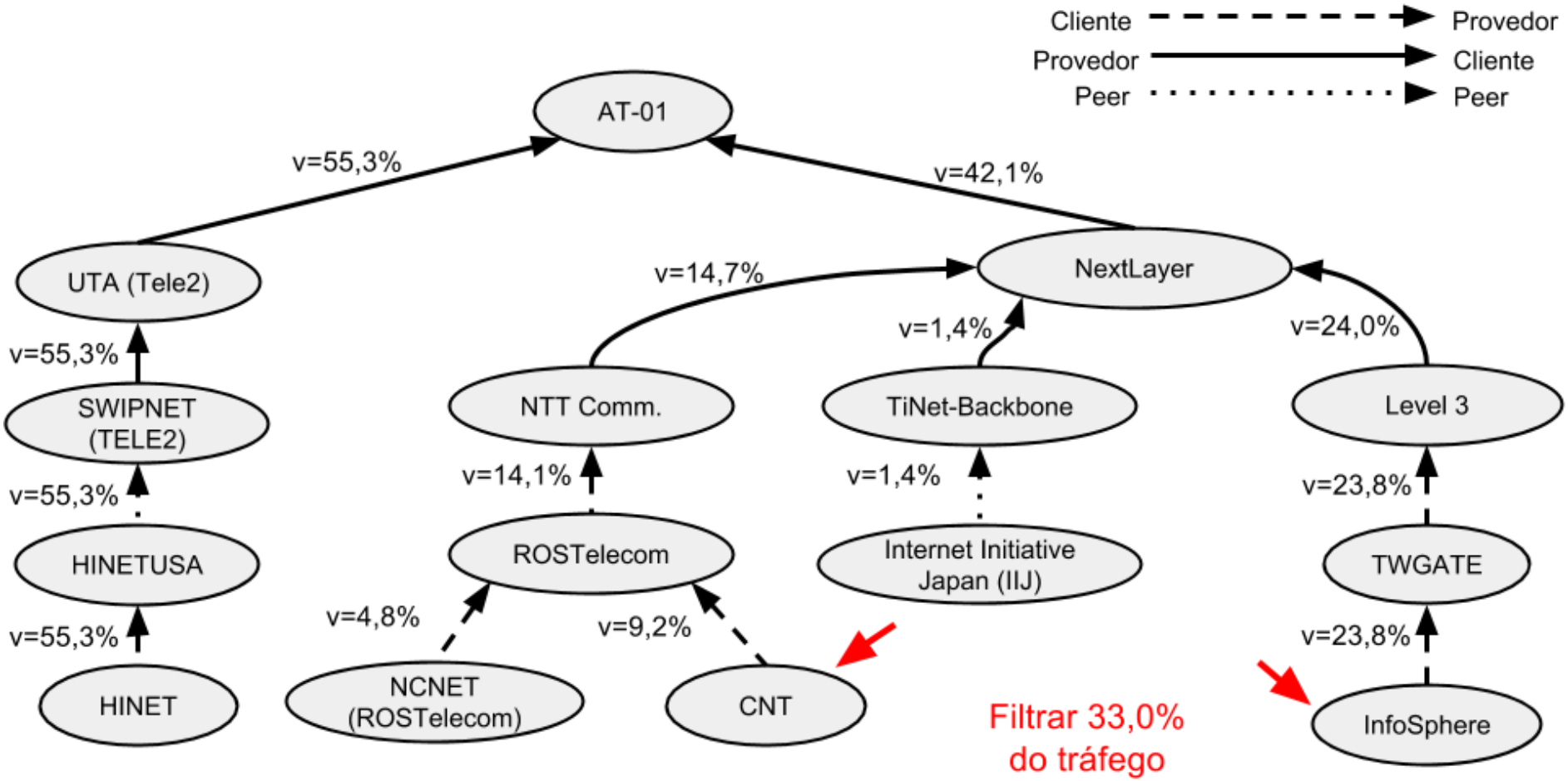
# Estudo de caso - Pouco incentivo



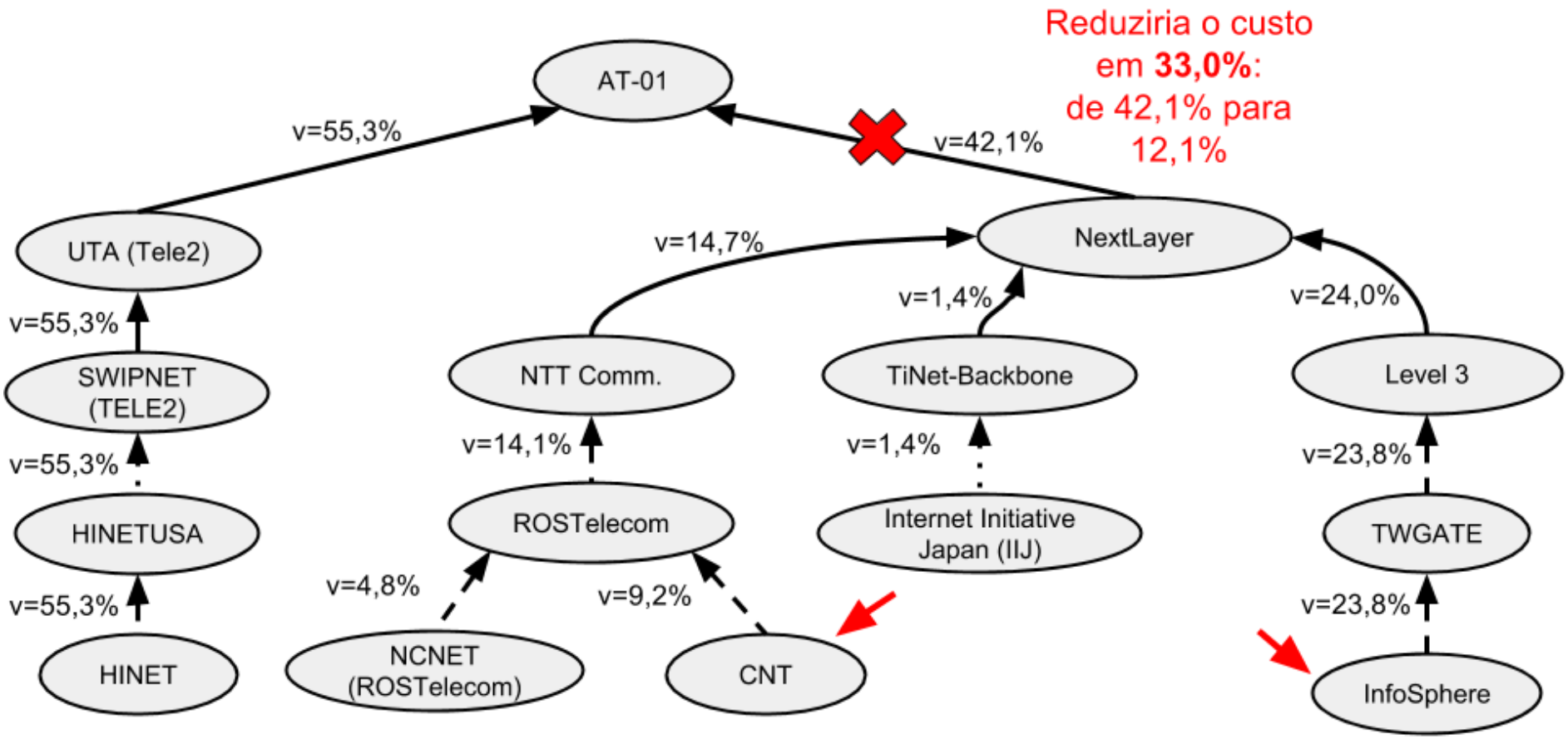
# Estudo de caso - Filtro de spam na origem



# Estudo de caso - Filtro de spam na origem



# Estudo de caso - Filtro de spam na origem



# Sumário

# Análise geral dos honeypots

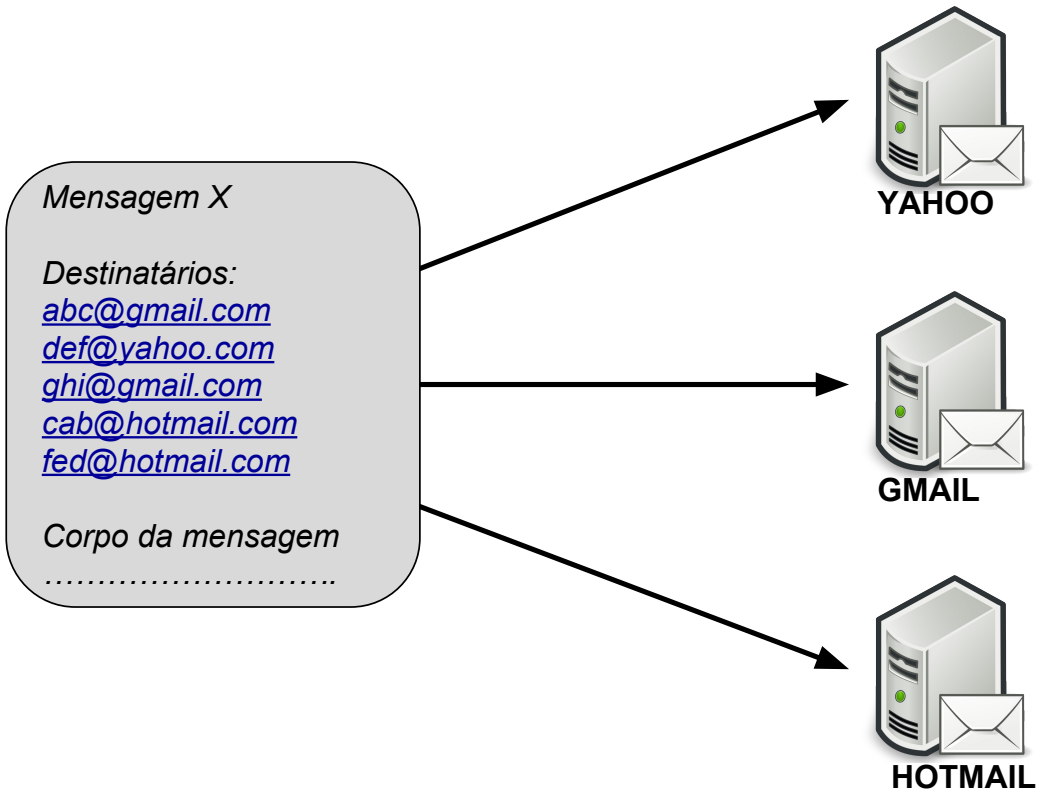
- Os ASes que hospedam os honeypots são os mais onerados
  - estão hospedados em redes pequenas
  - pagam para receber as mensagens de spam
  - pagam para encaminhar as mensagens de spam
- Redes menores que hospedam servidores vulneráveis
  - proxies e relays abertos
  - em geral, são as mais oneradas

# Análise geral dos honeypots

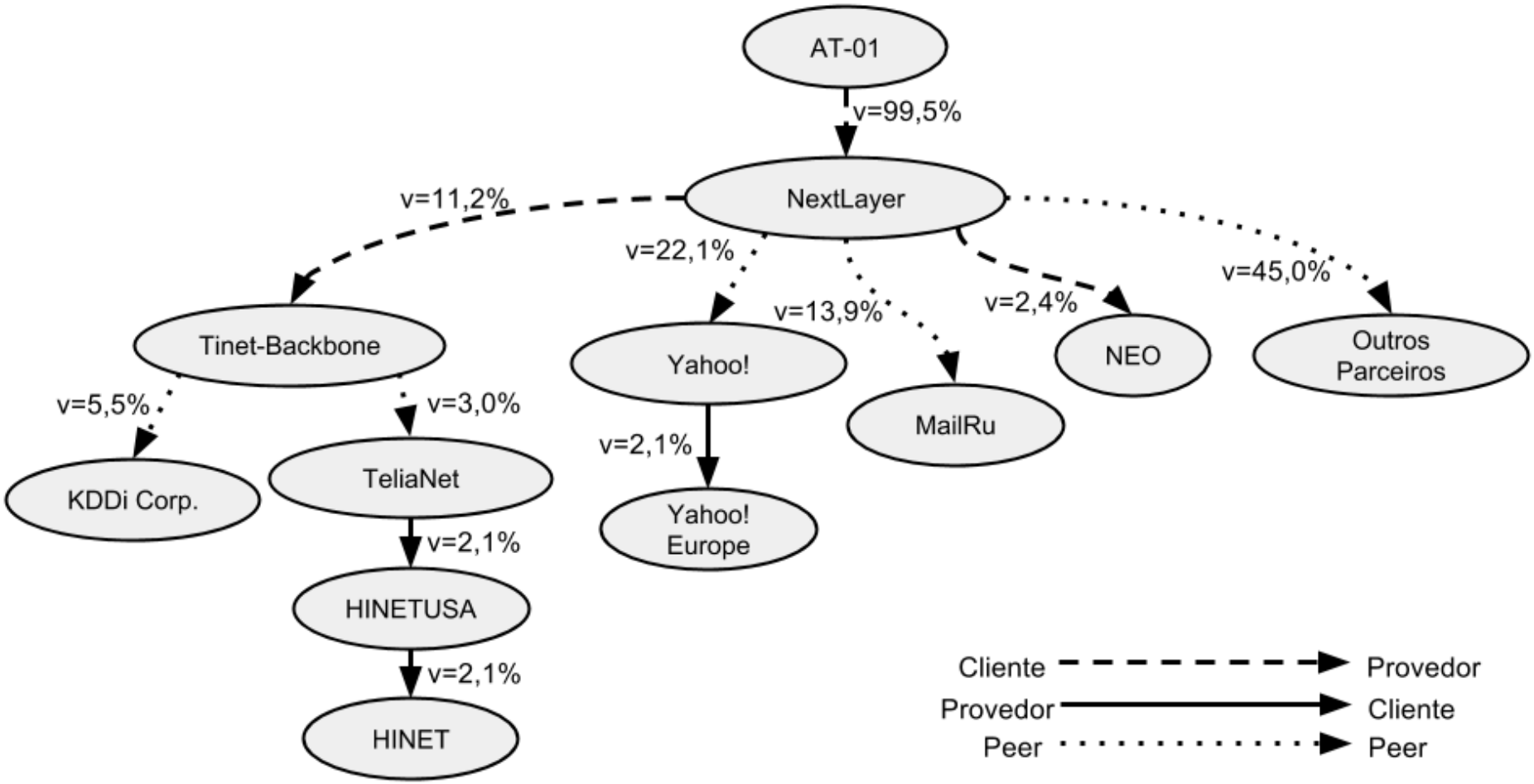
- O honeypot US-03 está hospedado em uma rede maior
- O AS do honeypot US-03 consegue evitar custos
  - 56% do tráfego chega ao honeypot sem custo
  - 84% do tráfego é encaminhado a ASes parceiros

# Amplificação do tráfego de spam

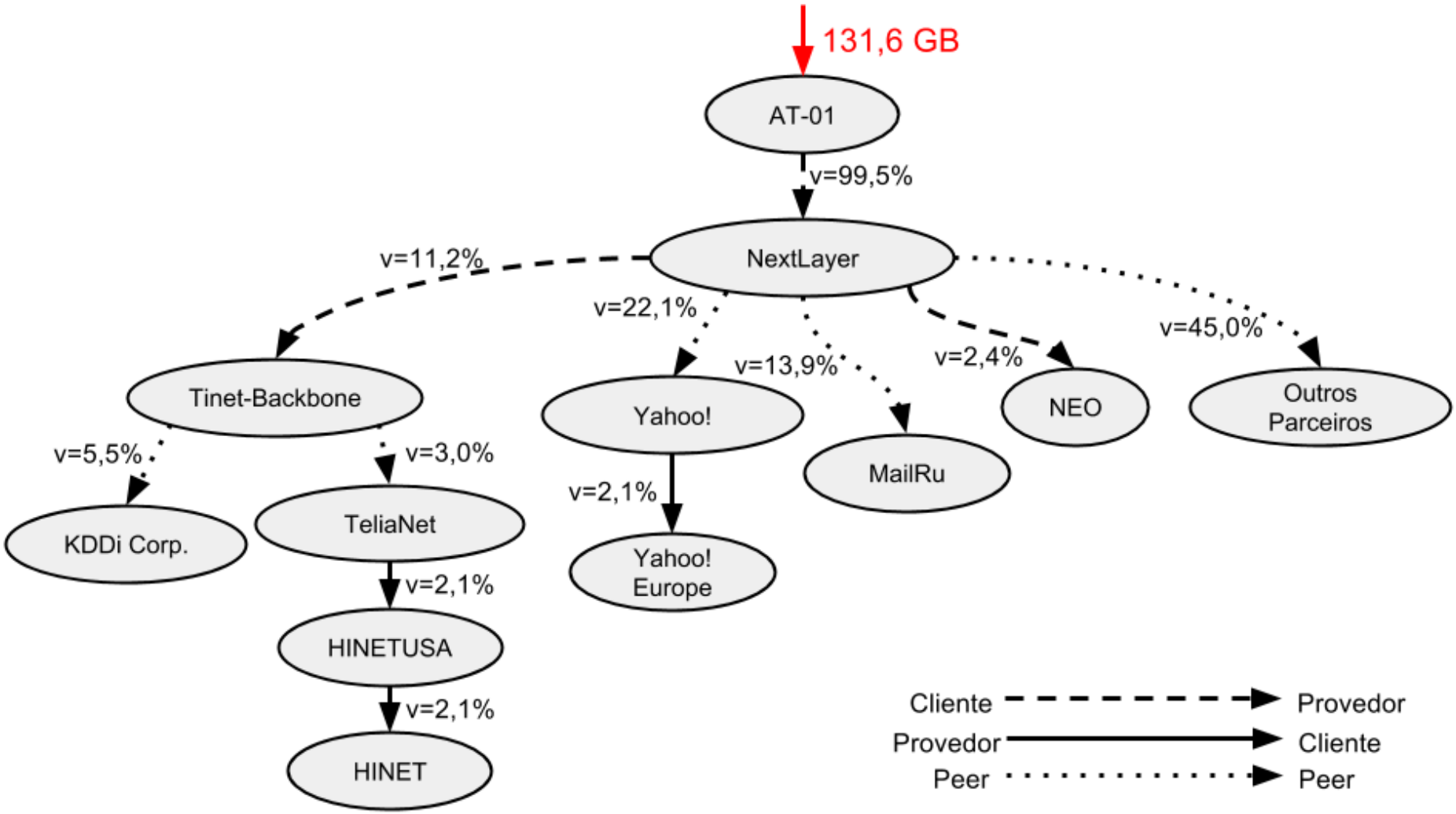
- Amplificação do tráfego ao alcançar servidores SMTP



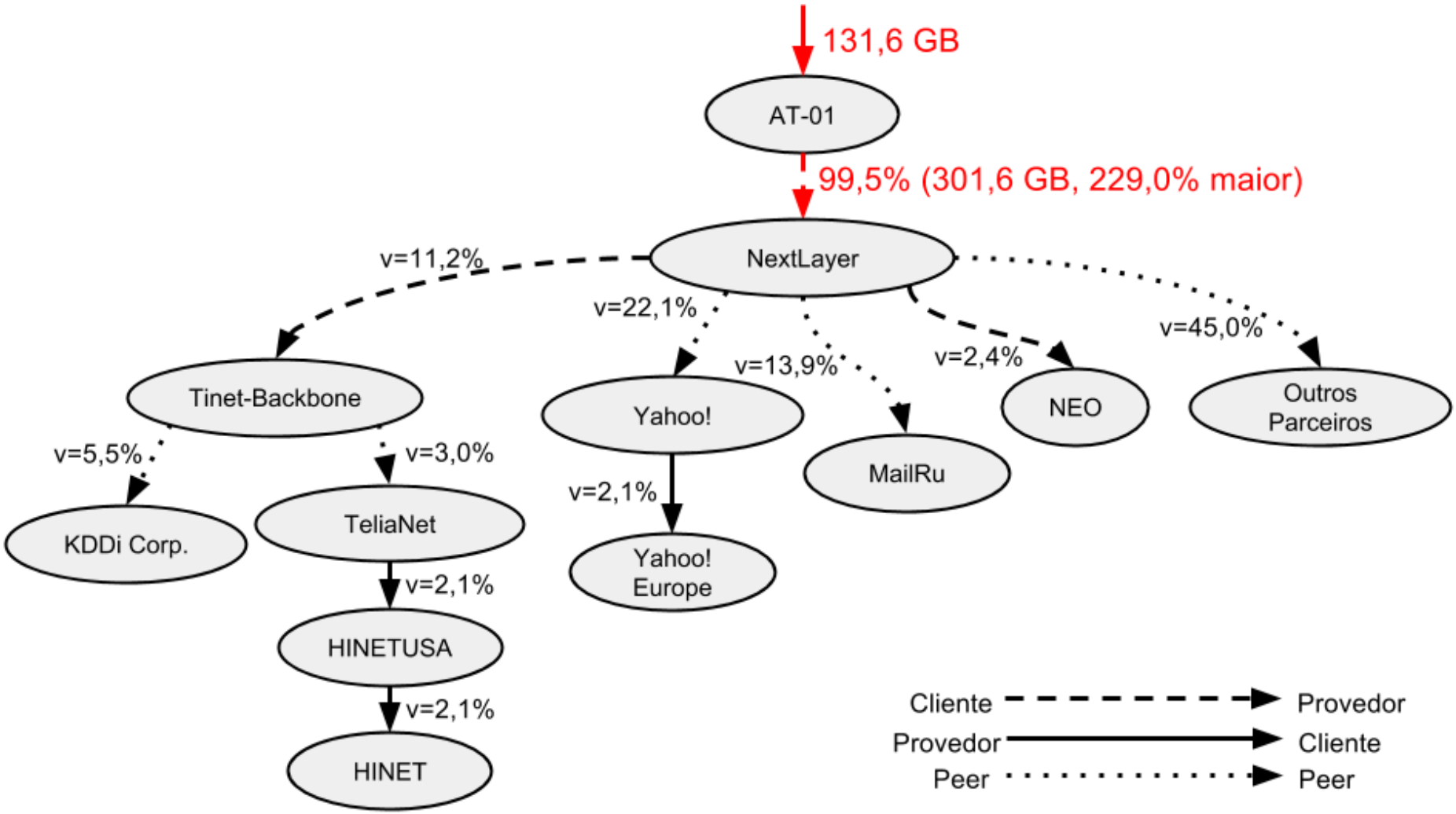
# Estudo de caso - Amplificação do tráfego



# Estudo de caso - Amplificação do tráfego



# Estudo de caso - Amplificação do tráfego



# Amplificação do tráfego de spam

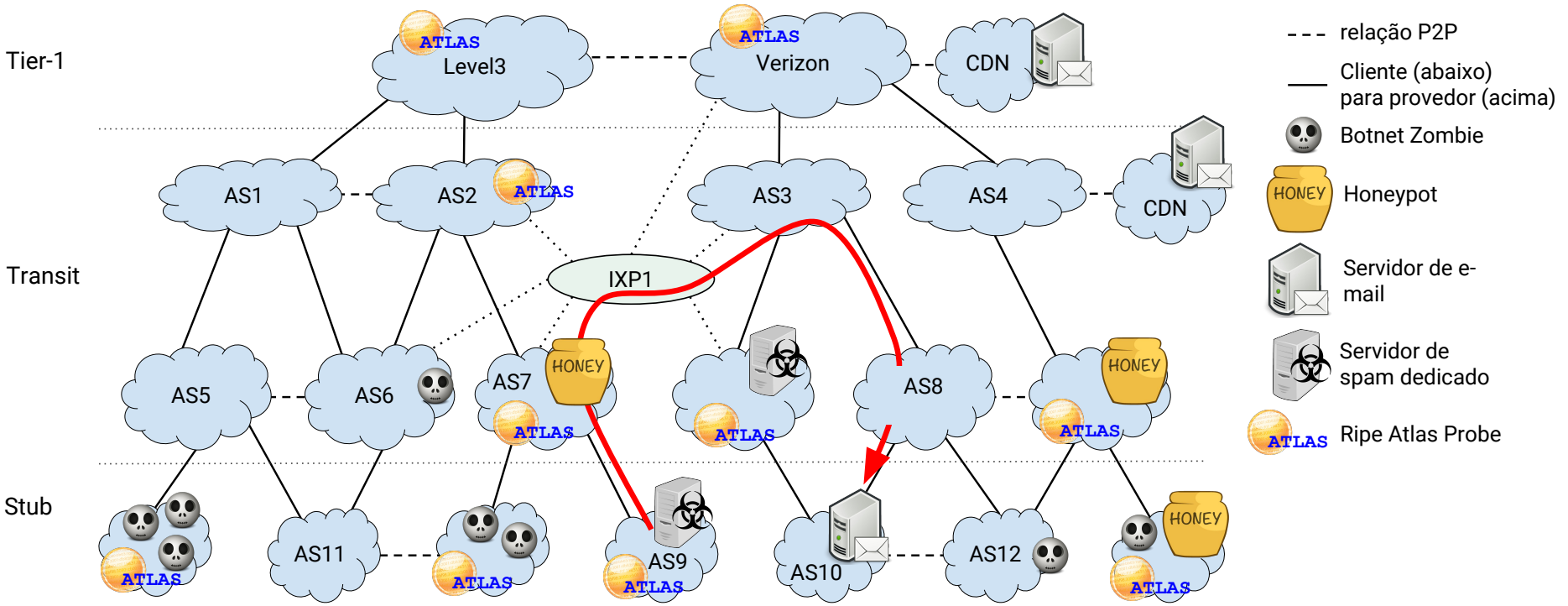
- AS do honeypot US-03:
  - maior amplificação
  - tráfego de entrada igual a 94,47 GB
  - tráfego de saída é igual a 511,75 GB (5,4 vezes maior)
  - deixa de pagar por 432 GB (84,48%)

# Sumário

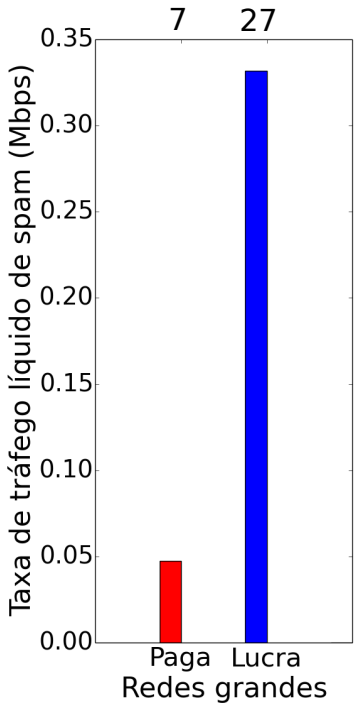
# Análise do tráfego líquido de spam

- Tráfego líquido de spam
- Um AS lucra quando o tráfego líquido de spam é positivo
- Um AS paga quando o tráfego líquido de spam é negativo
- Classificamos os ASes em três categorias
  - ASes grandes
  - ASes médios
  - ASes pequenos

# Análise do tráfego líquido de spam

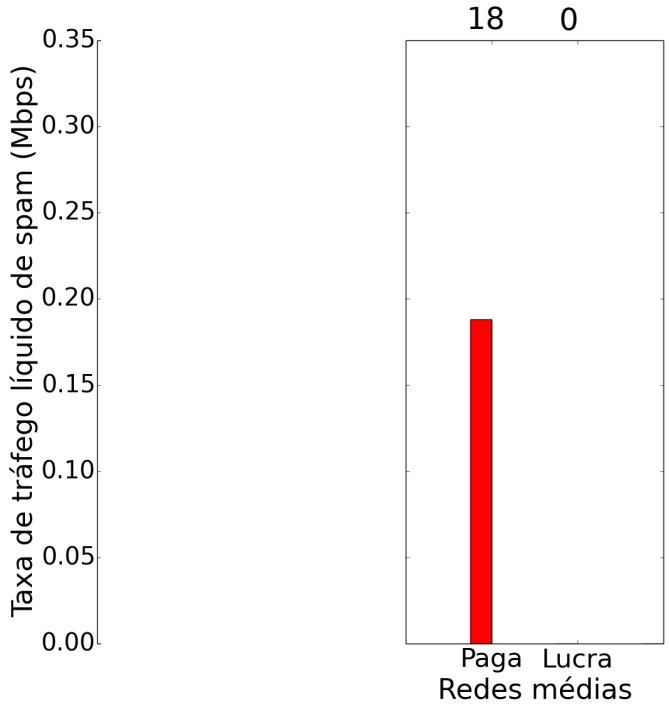


# Análise do tráfego líquido de spam



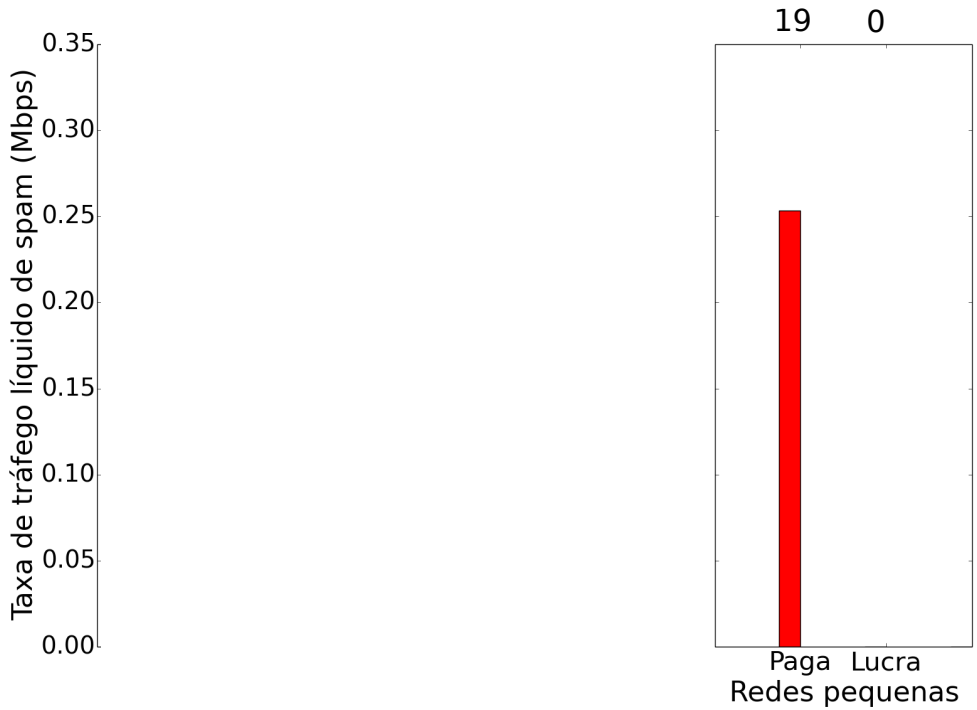

- raramente pagam pelo tráfego de spam
- lucram com a troca de tráfego com seus clientes
- podem lucrar duas vezes com o tráfego de algumas mensagens

# Análise do tráfego líquido de spam



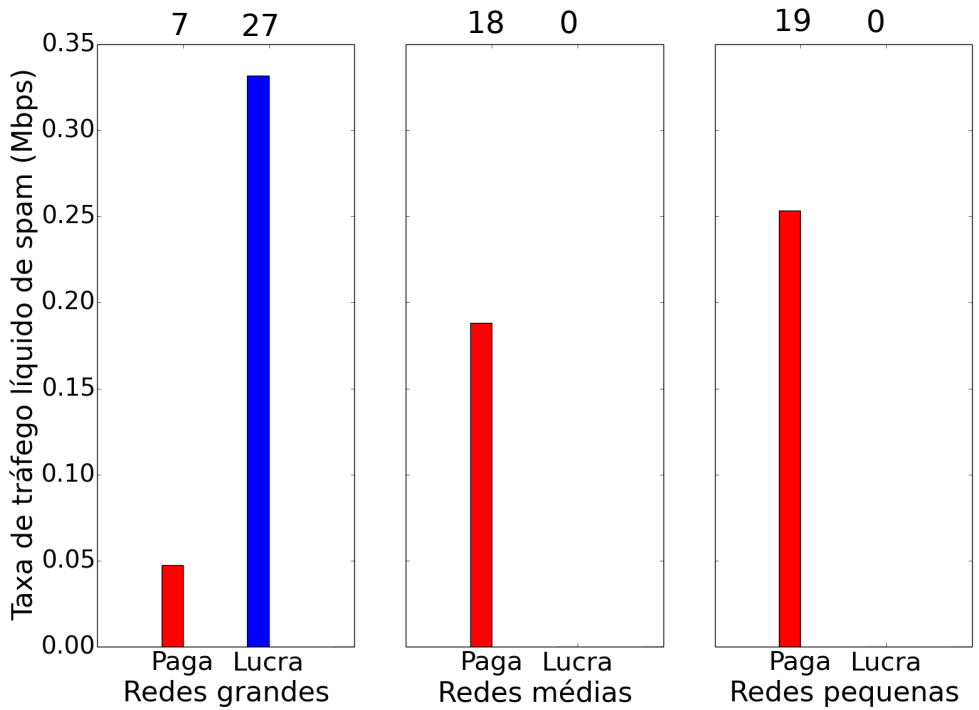

- lucram com a troca de tráfego com seus clientes
- pagam com a troca de tráfego com seus provedores
- tem tráfego líquido de spam negativo

# Análise do tráfego líquido de spam




- pagam pela maior parte do tráfego
- recebem e originam tráfego de spam
  - pagam pela troca de tráfego com seus provedores

# Análise do tráfego líquido de spam

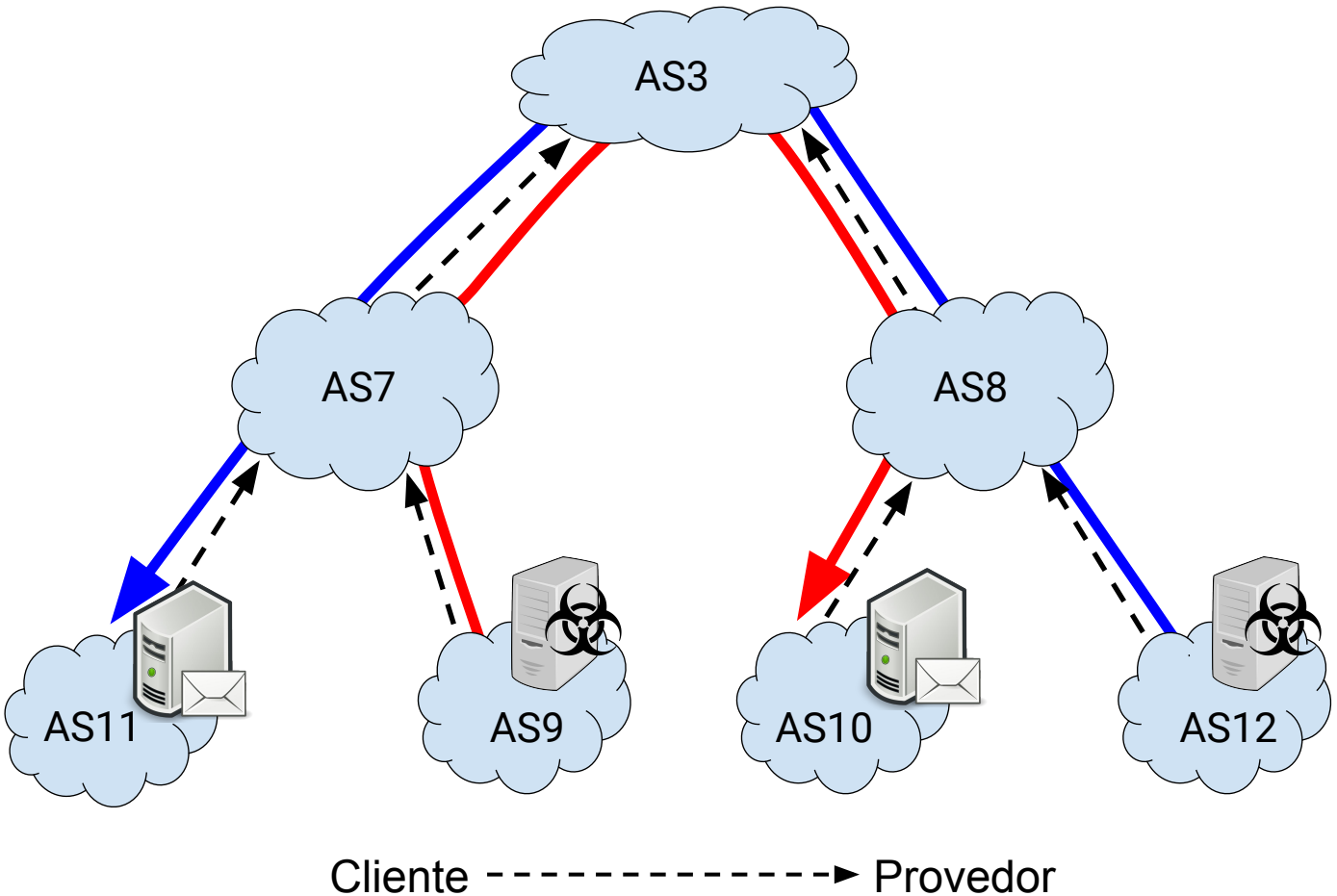



- redes grandes não se beneficiam da filtragem de spam
- esforços na filtragem de spam devem se concentrar em:
  - redes médias e redes pequenas

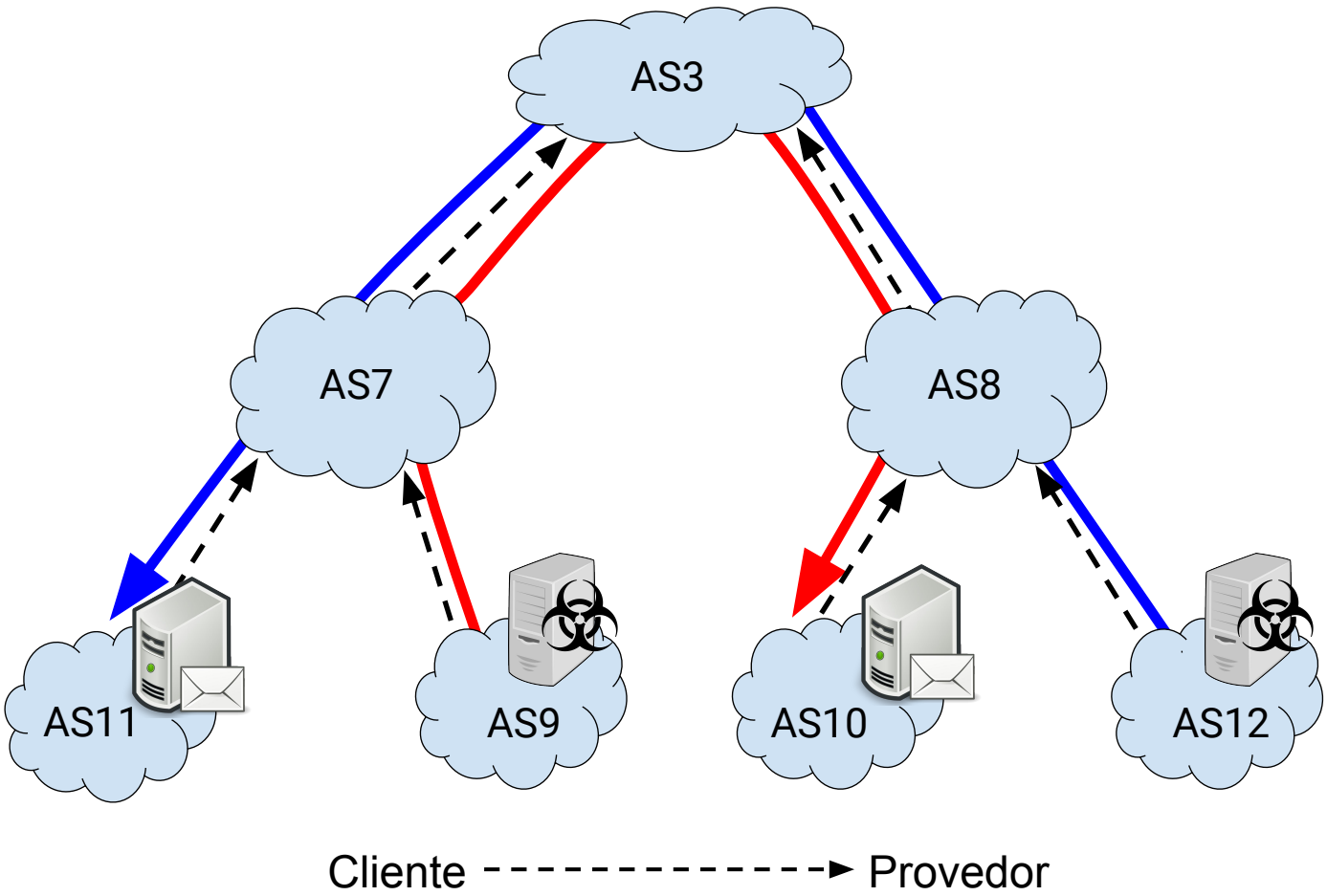
# Sumário

# Filtrando o tráfego de spam

- Alguns ASes lucram com o tráfego de spam
- Outros pagam pelo tráfego de spam, mas podem não estar interessados em filtrar as mensagens de spam na descida do caminho

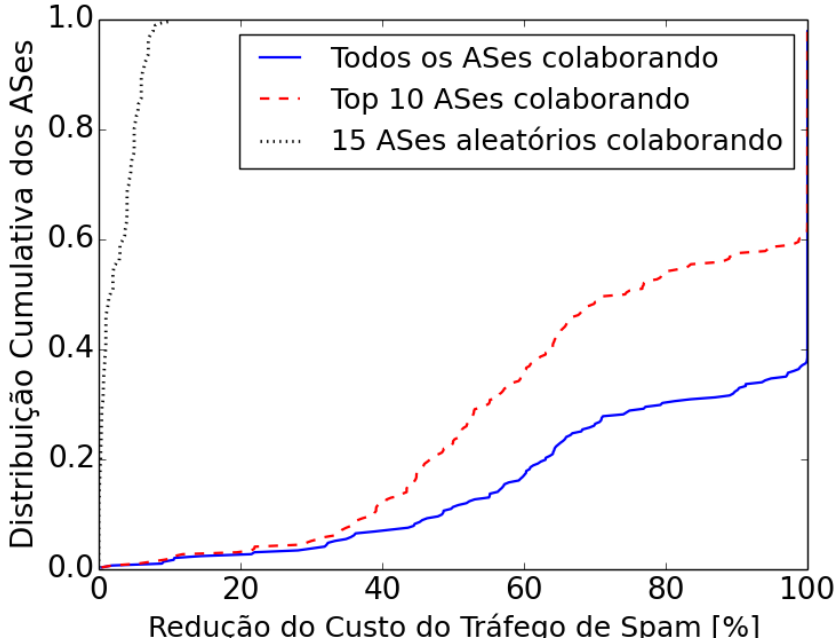


# Algoritmo proposto



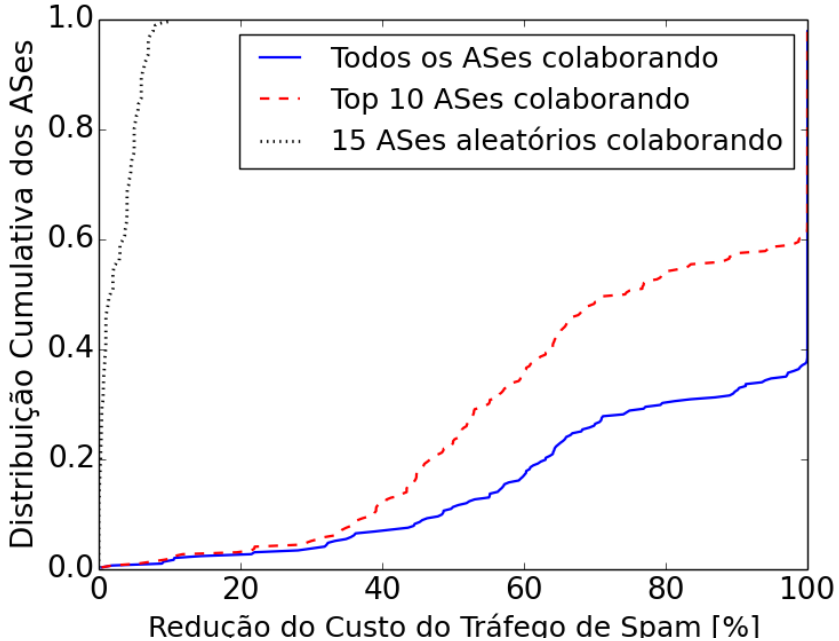

- ASes que pagam para receber as mensagens de spam
- ASes que pagam para encaminhar as mensagens de spam

# Filtrando o tráfego de spam



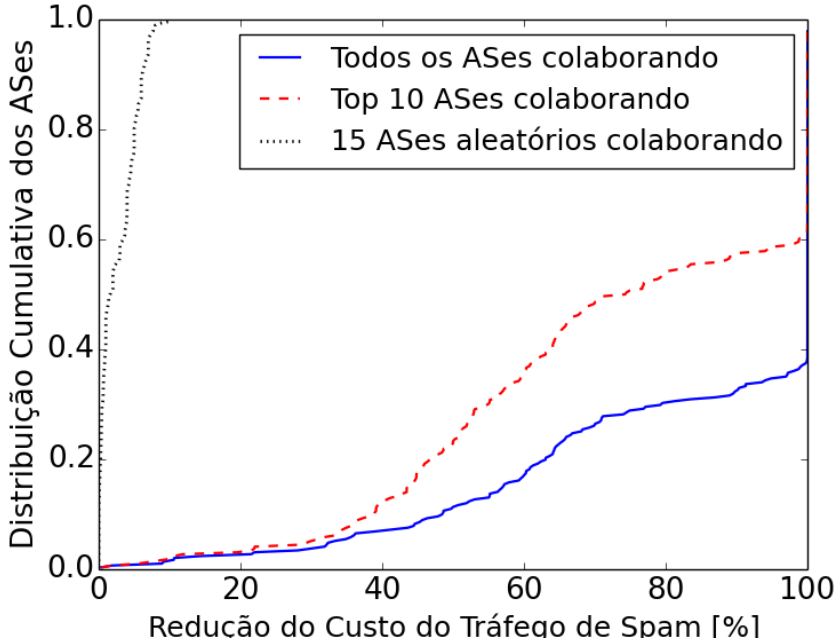

- A curva pontilhada em preto mostra o cenário em que cada AS faz acordo com quinze ASes aleatórios, sem utilizar nosso algoritmo
- Os ASes escolhidos podem não se beneficiar com a filtragem
- Não reduz significativamento o custo do tráfego de spam

# Filtrando o tráfego de spam




- A curva tracejada em vermelho mostra o cenário em que cada AS faz acordo com os dez ASes que levarão a uma maior redução dos custos
- Reduz o tráfego de spam em 100% para 37% dos ASes
- Reduz o tráfego de spam pelo menos 50% para 77,06% do ASes

# Filtrando o tráfego de spam




- A curva em azul mostra o cenário em que todos os pares de ASes apontados pelo algoritmo aceitam colaborar com a filtragem de spam
- Reduz o tráfego de spam em 100% para 60,14% dos ASes
- Reduz o tráfego em pelo menos 50% para 88,66% dos ASes

# Conclusões

- Redes grandes têm pouco incentivo econômico para filtrar mensagens de spam
- Mostramos que a instalação de filtros que impeçam as mensagens de sair dos servidores SMTP evita a replicação das mensagens e reduz drasticamente o tráfego de spam
  - gerência de porta 25
- Os resultados mostram que a escolha de apenas dez parcerias utilizando o algoritmo proposto é capaz de reduzir grande parte dos custos associados ao tráfego de spam

# Trabalhos Futuros

- Estender essas análises para outras fontes de tráfego indesejado
- Criar um serviço de monitoramento constante:
  - Alertar os sistemas autônomos quando os custos associados ao tráfego de spam estiverem aumentando para suas redes
  - Incentivar mais operadores de rede a adotarem medidas para mitigar o tráfego de spam

# MEDIÇÃO, CARACTERIZAÇÃO E REDUÇÃO DOS CUSTOS ASSOCIADOS AO TRÁFEGO DE SPAM

Osvaldo Luís H. M. Fonseca

DCC-UFMG
NIC.br

13 de maio de 2016

# Publicações

*Uma Análise do Custo do Tráfego de Spam para Operadores de Rede*
Apresentado no Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos - SBRC 2015

*Measuring, Characterizing, and Avoiding Spam Traffic Costs*
Aceito para publicação no Internet Computing.

# Filtrando o tráfego de spam

- Cada *AS x* que paga para receber mensagens de spam:
  - coletar as mensagens de spam recebidas de provedores
  - identificar os *ASes* que pagam para encaminhar essas mensagens
  - oferecer acordos para filtragem de spam
- Cada *AS x* que paga para encaminhar mensagens de spam:
  - coletar as mensagens de spam encaminhadas por provedores
  - identificar os *ASes* que pagam para recebê-las
  - oferecer um acordo para filtrar essas mensagens