Abrindo a Caixa-Preta – Aplicando IA Explicável para Aprimorar a Detecção de Sequestros de Prefixo

Adriano B. de Carvalho, Brivaldo A. da Silva Jr,
Carlos Alberto da Silva e Ronaldo A. Ferreira

Universidade Federal de Mato Grosso do Sul

Agradecimentos

CAPES CNPq FAPESP

UFMS

18 de setembro de 2024

Motivação

ROUTING SECURITY | ROUTING SECURITY INCIDENTS

BGP Security in 2021

By Aftab Siddiqui • 21 Feb 2022

- October 25: AS212046 – MEZON – hijacked 3786 prefixes – Qrator Labs
- October 13: AS212046 – MEZON – hijacked 1029 prefixes – Qrator Labs
- September 21: AS62325 — HDHK – hijacked 89 prefixes – Qrator Labs
- May 18: AS48467 — PRANET – hijacked 454 prefixes – Qrator Labs
- April 16: AS55410 – Vodafone Idea Ltd – hijacked 30,000 prefixes – MANRS Blog
- February 5: AS136168 – Campana MYTHIC – hijacked Twitter prefixes – MANRS Blog

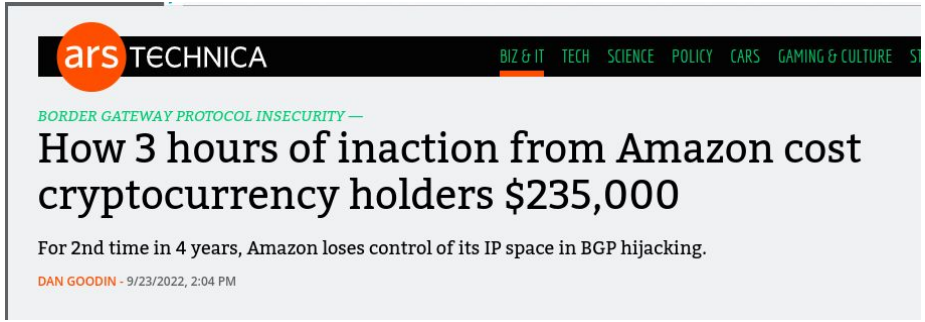
} Falha ou erro

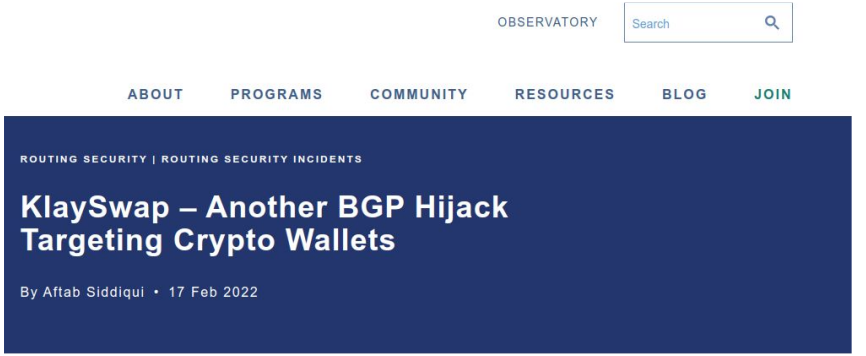
} Censura

https://www.manrs.org/2022/02/bgp-security-in-2021/

Motivação

Furto de criptomoedas





Aprox. 235 mil dólares

Aprox. 1,9 milhão dólares

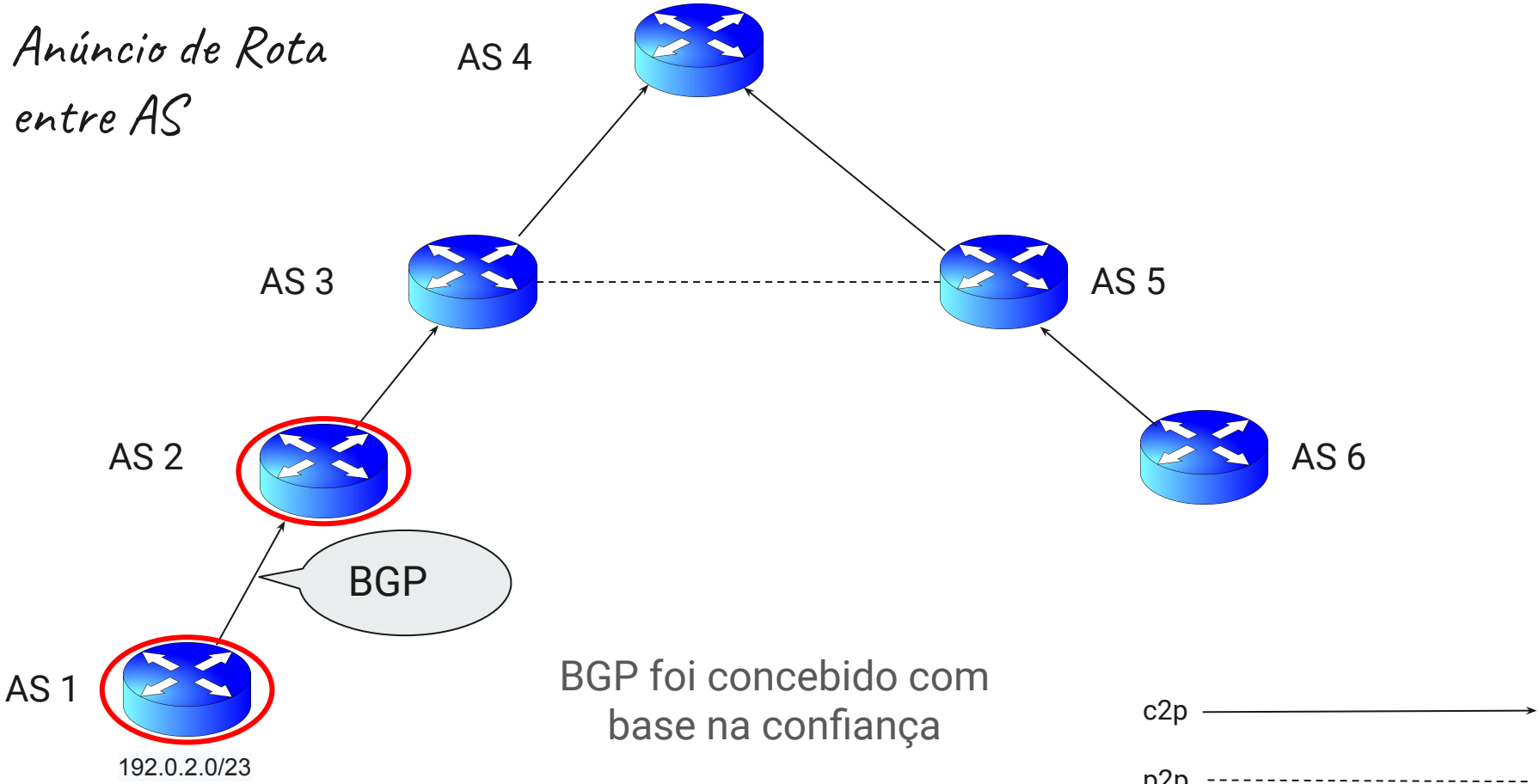
https://arstechnica.com/information-technology/2022/09/how-3-hours-of-inaction-from-amazon-cost-cryptocurrency-holders-235000/

https://www.manrs.org/2022/02/klayswap-another-bgp-hijack-targeting-crypto-wallets/

Como é possível ocorrer o sequestro de prefixo?

Primeiro é necessário saber como a Internet é formada

Anúncio de Rota entre AS

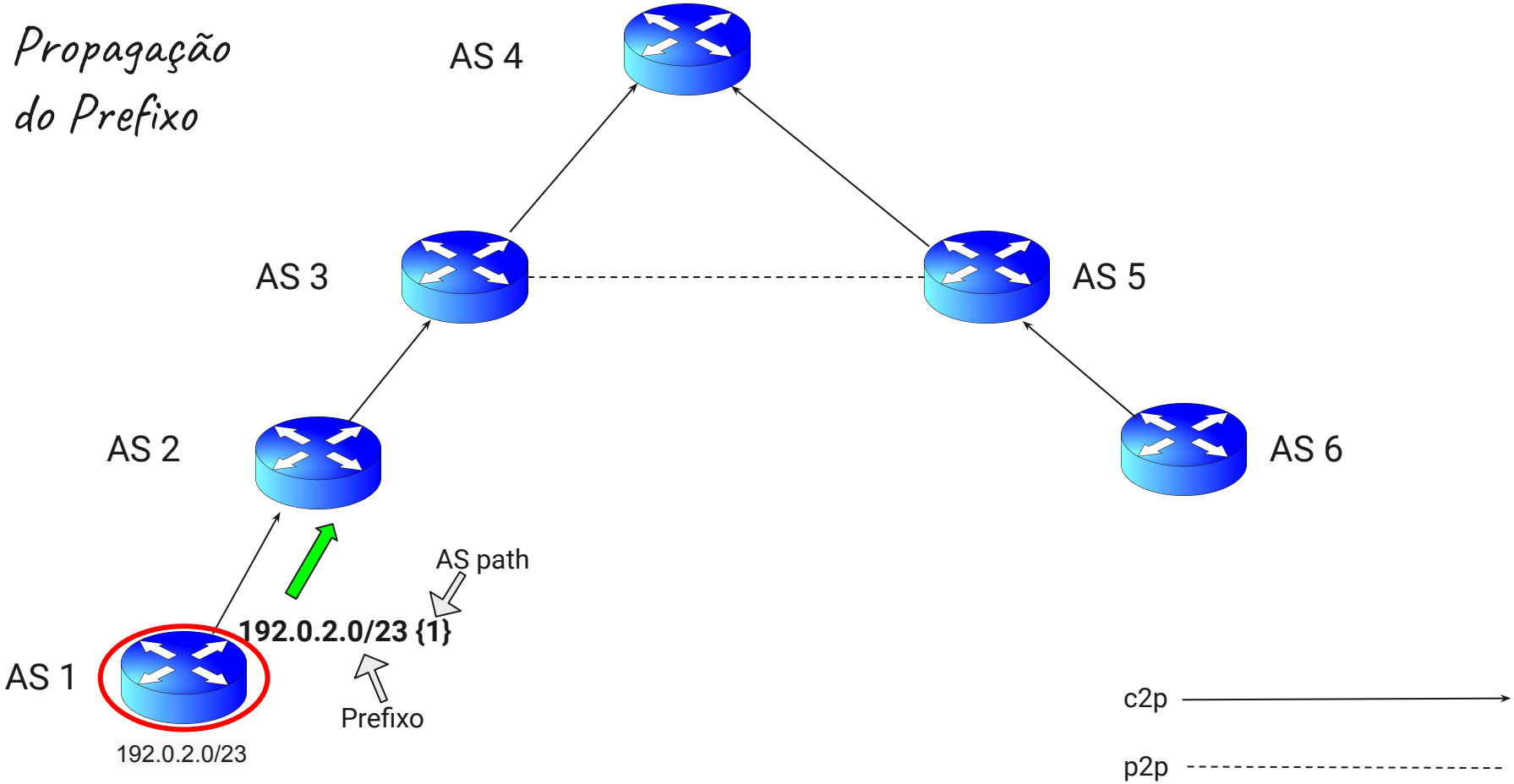


BGP foi concebido com base na confiança

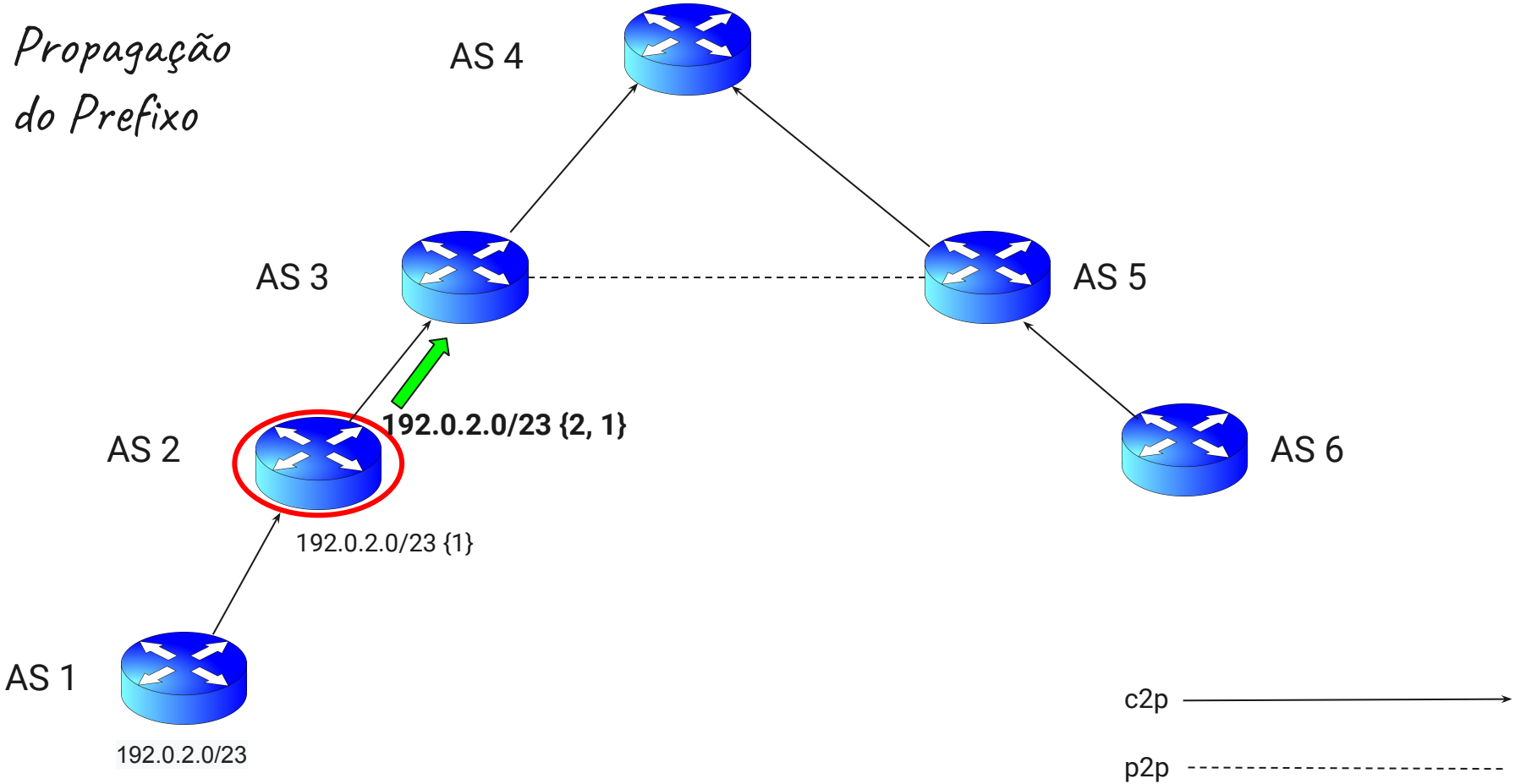
c2p ——————————→

p2p -

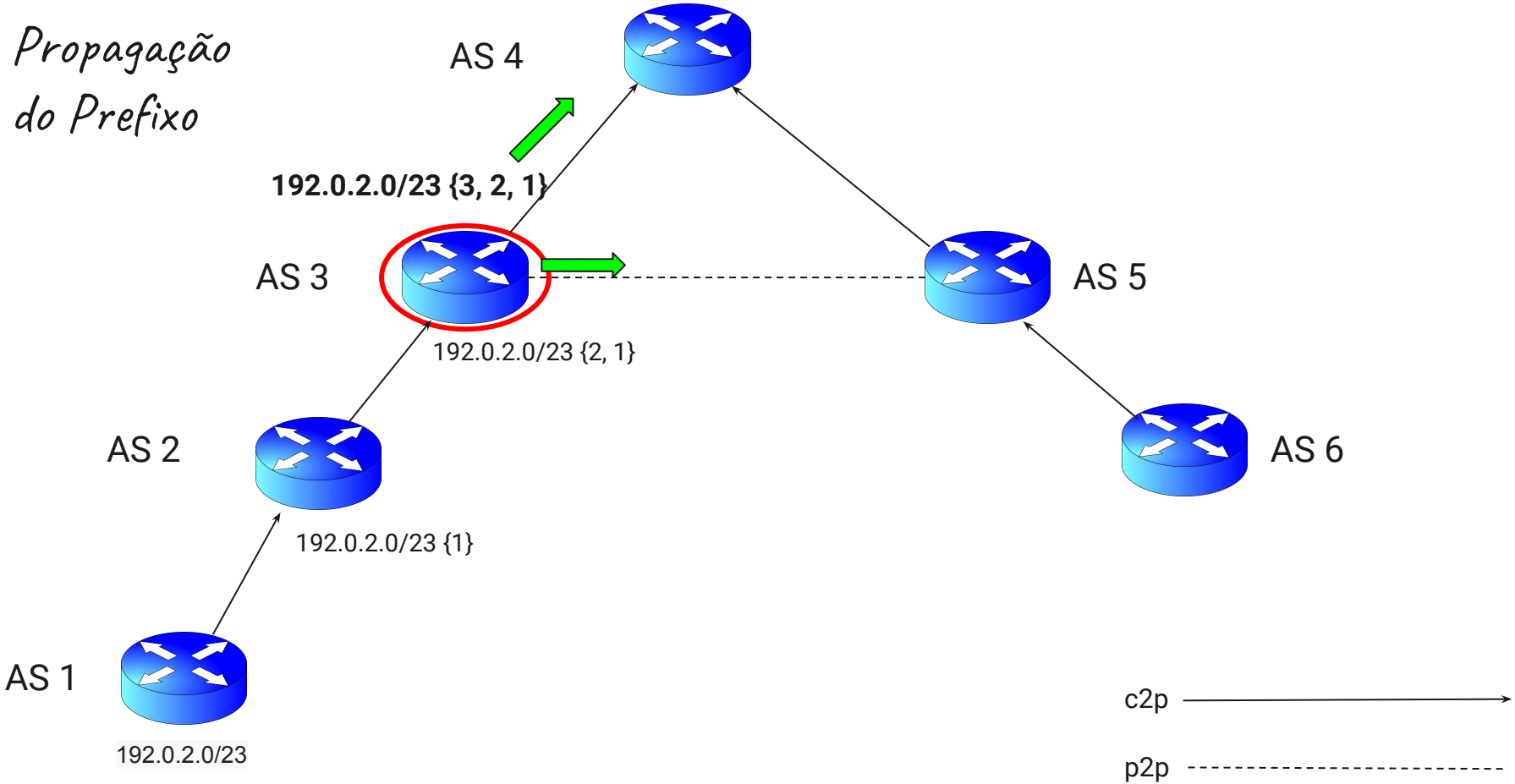
Propagação do Prefixo



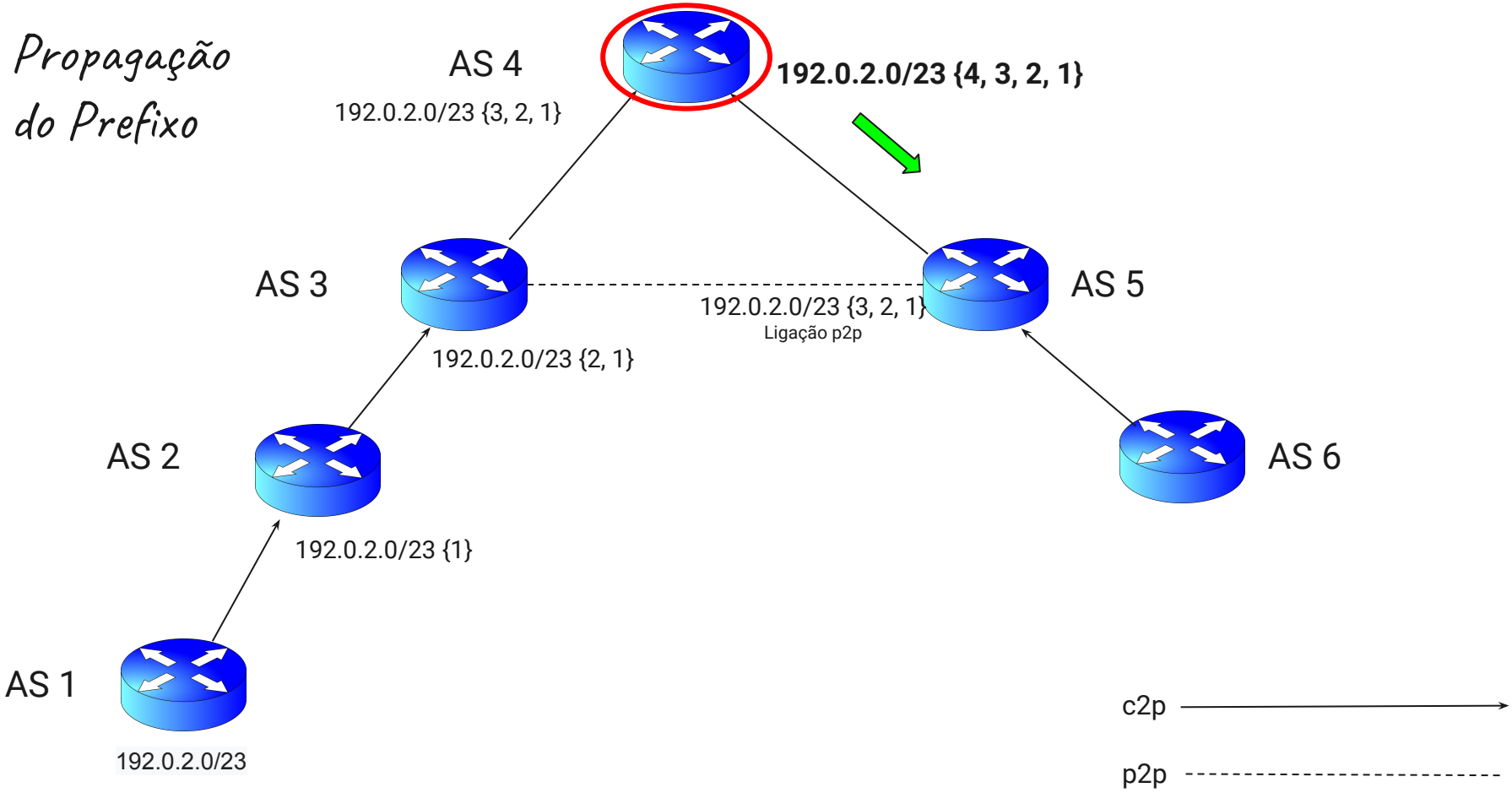
Propagação do Prefixo



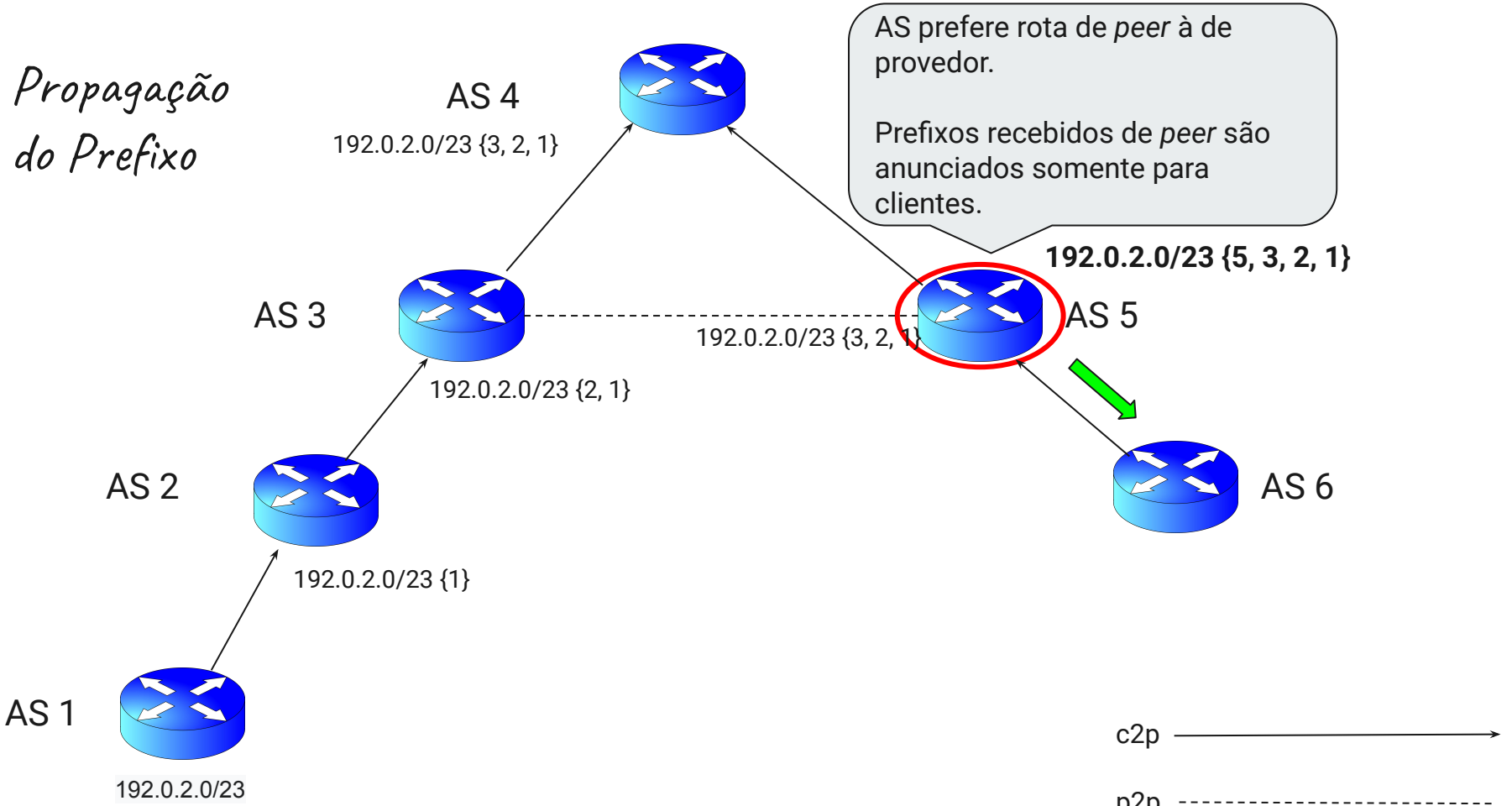
Propagação do Prefixo



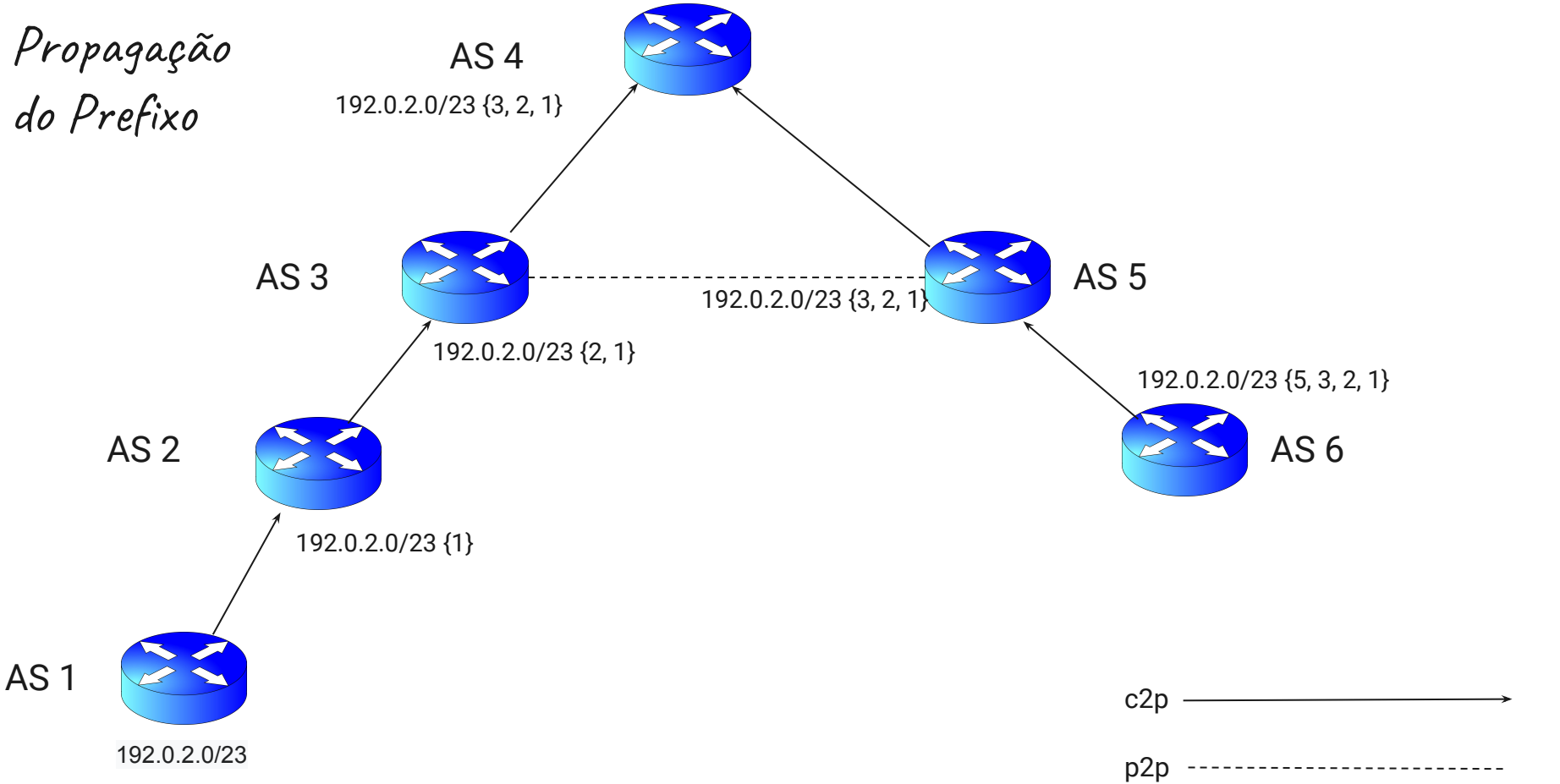
Propagação do Prefixo



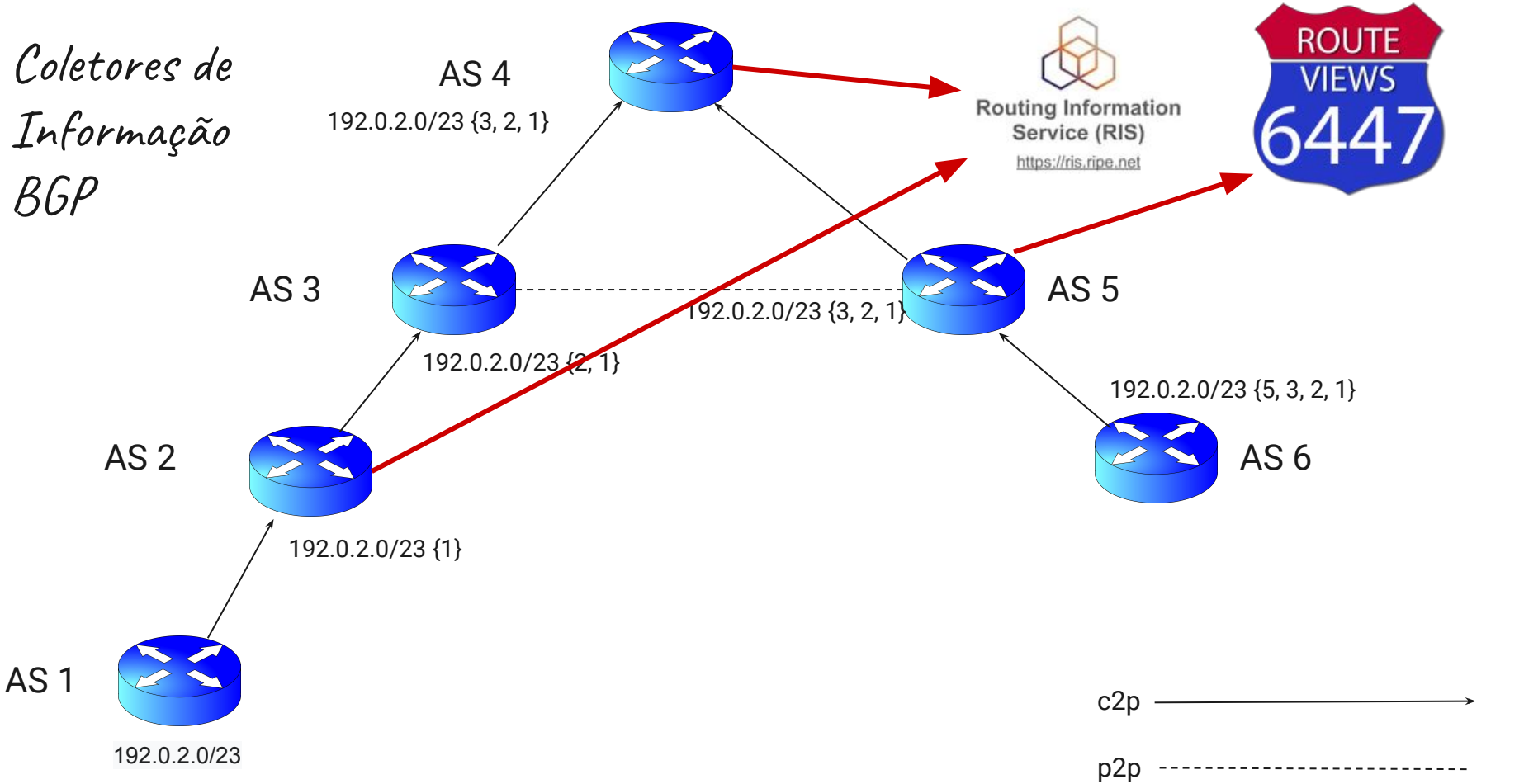
Propagação do Prefixo



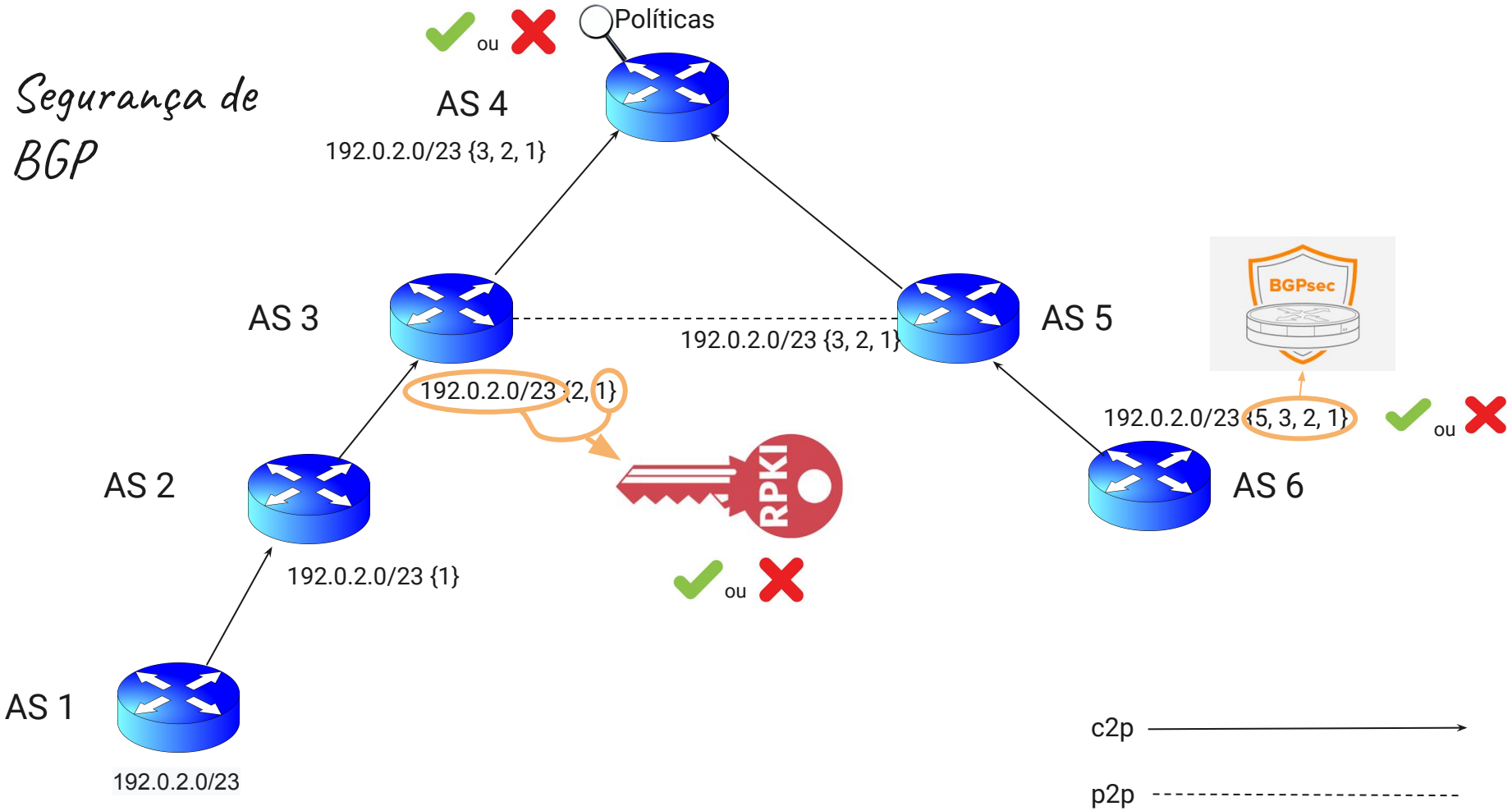
Propagação do Prefixo



Coletores de Informação BGP



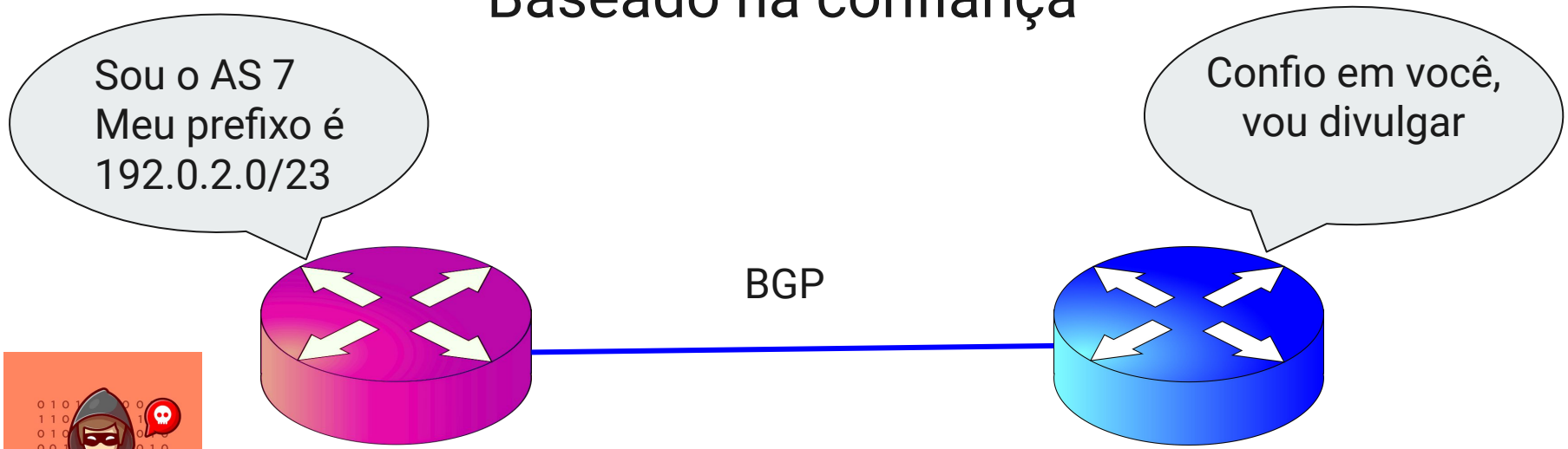
Segurança de BGP



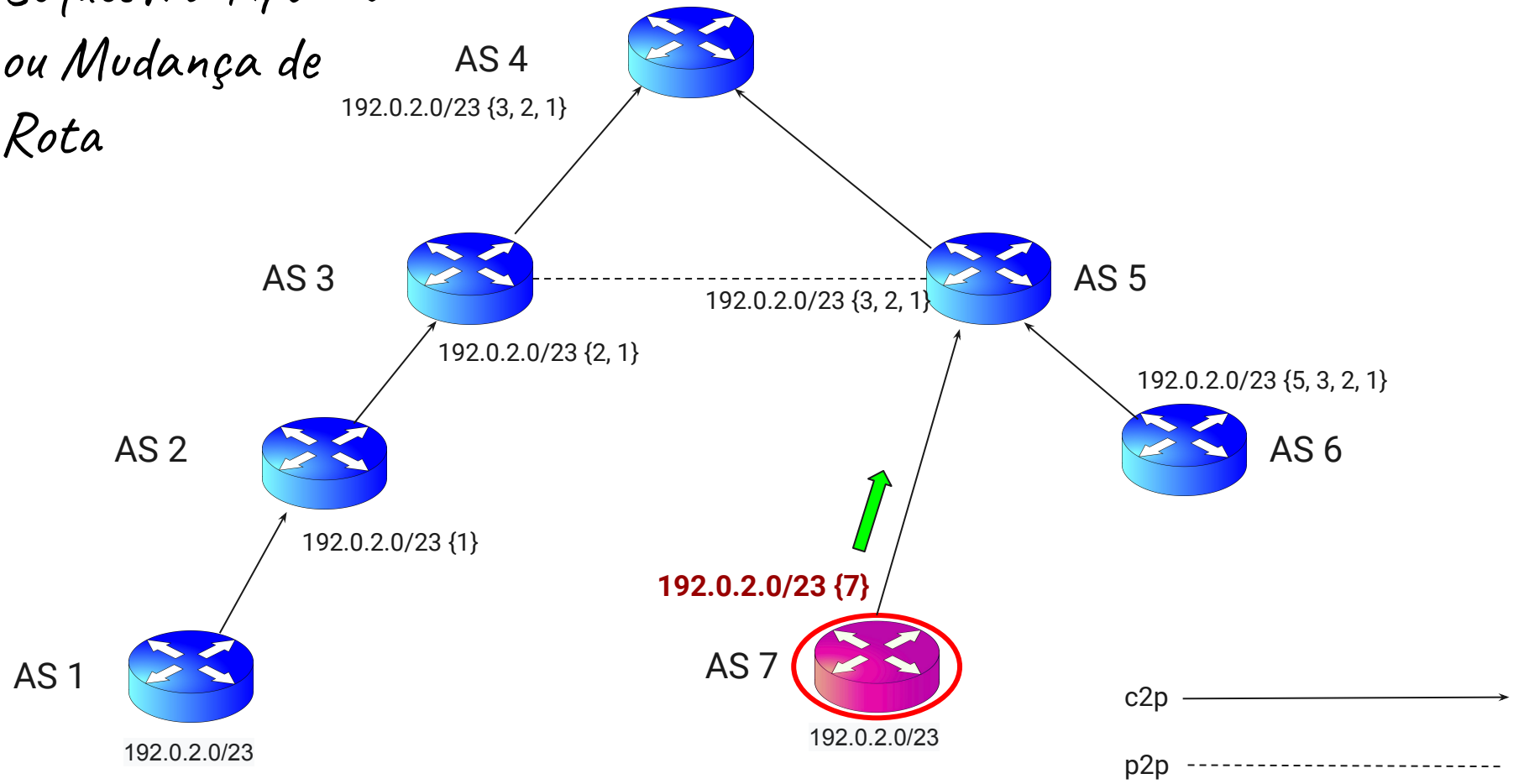
Voltamos a pergunta, como é possível ocorrer o sequestro de prefixo?

BGP carece de segurança, pois não possui mecanismos nativos de validação e autenticação

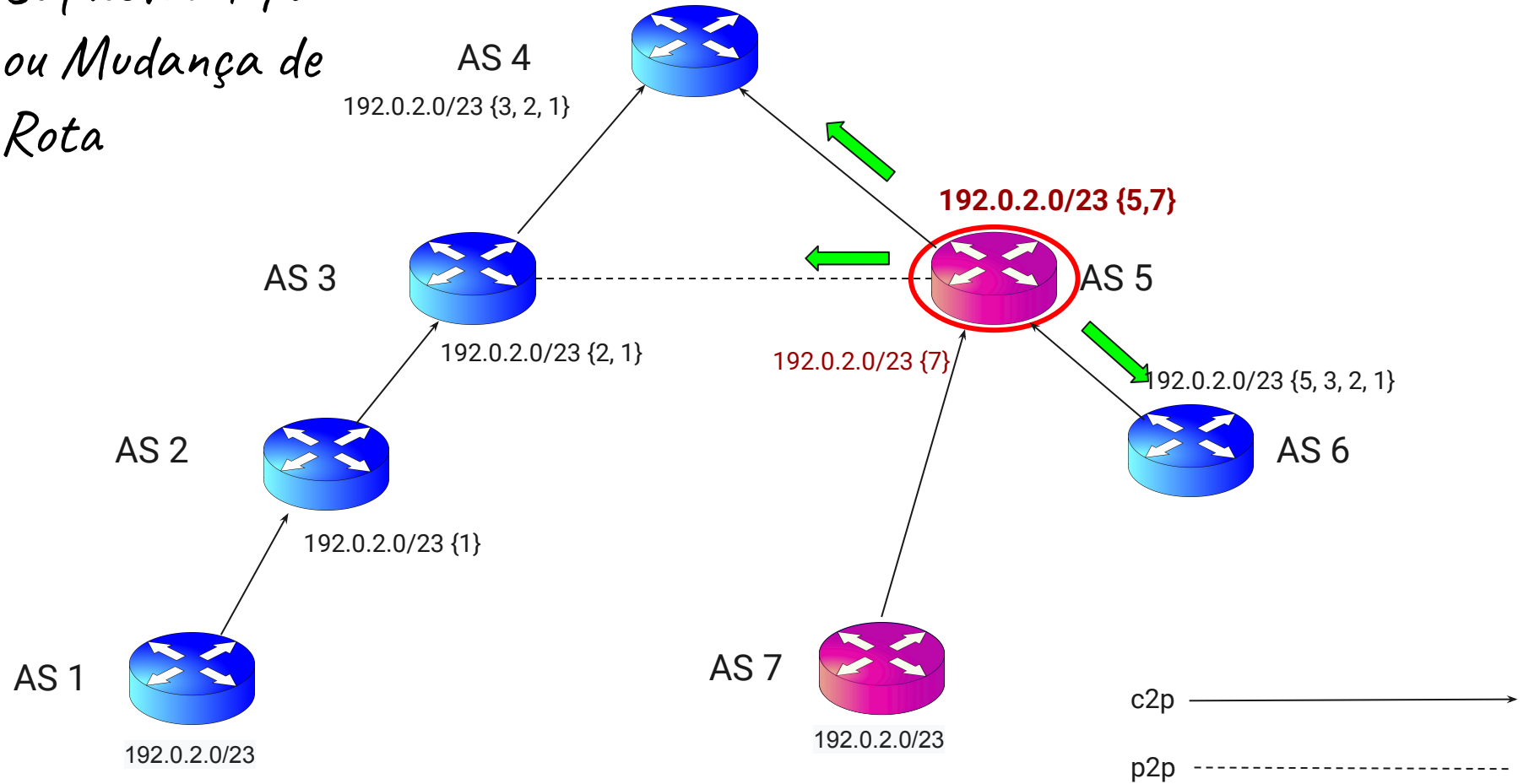
Baseado na confiança



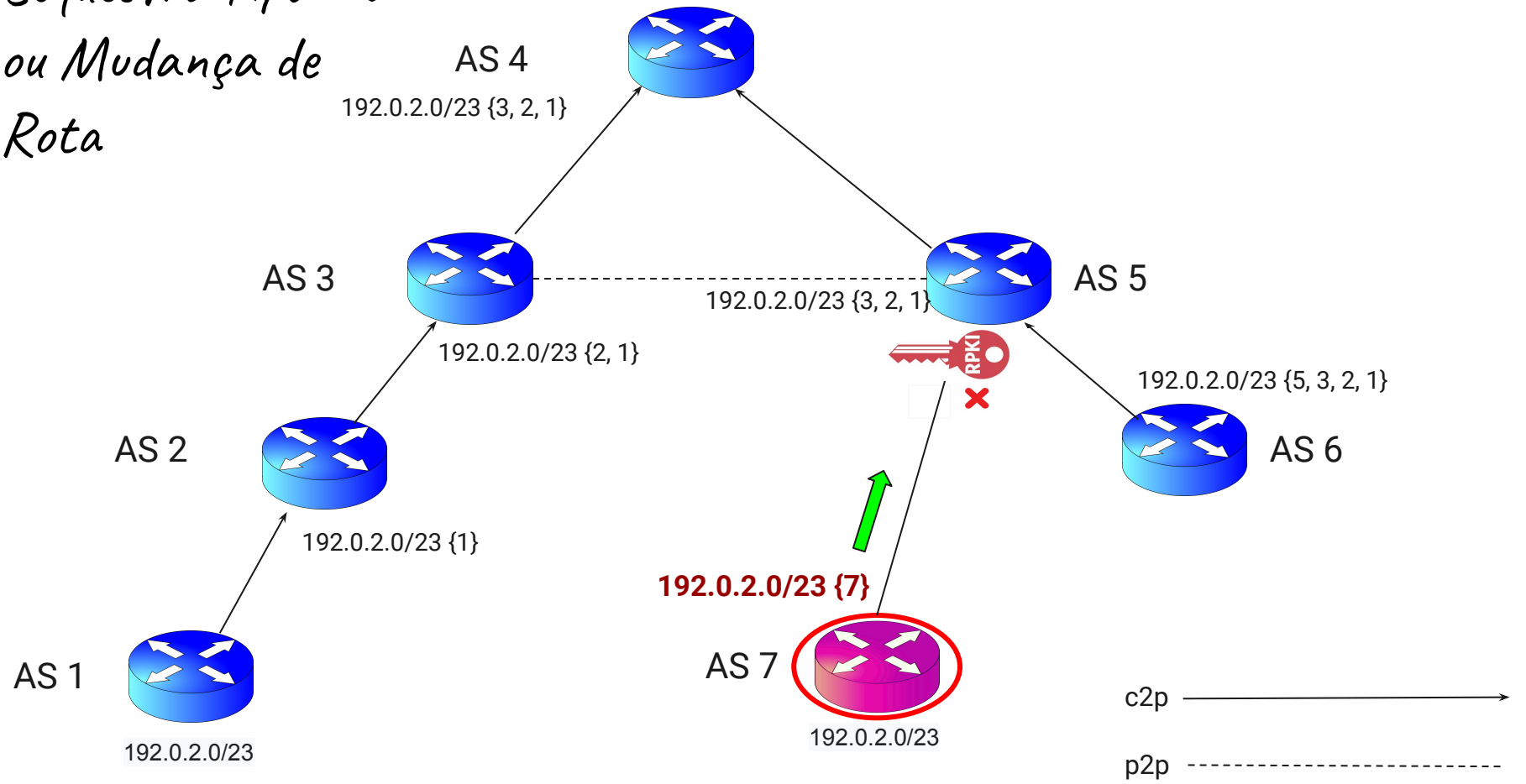
Sequestro Tipo - 0 ou Mudança de Rota



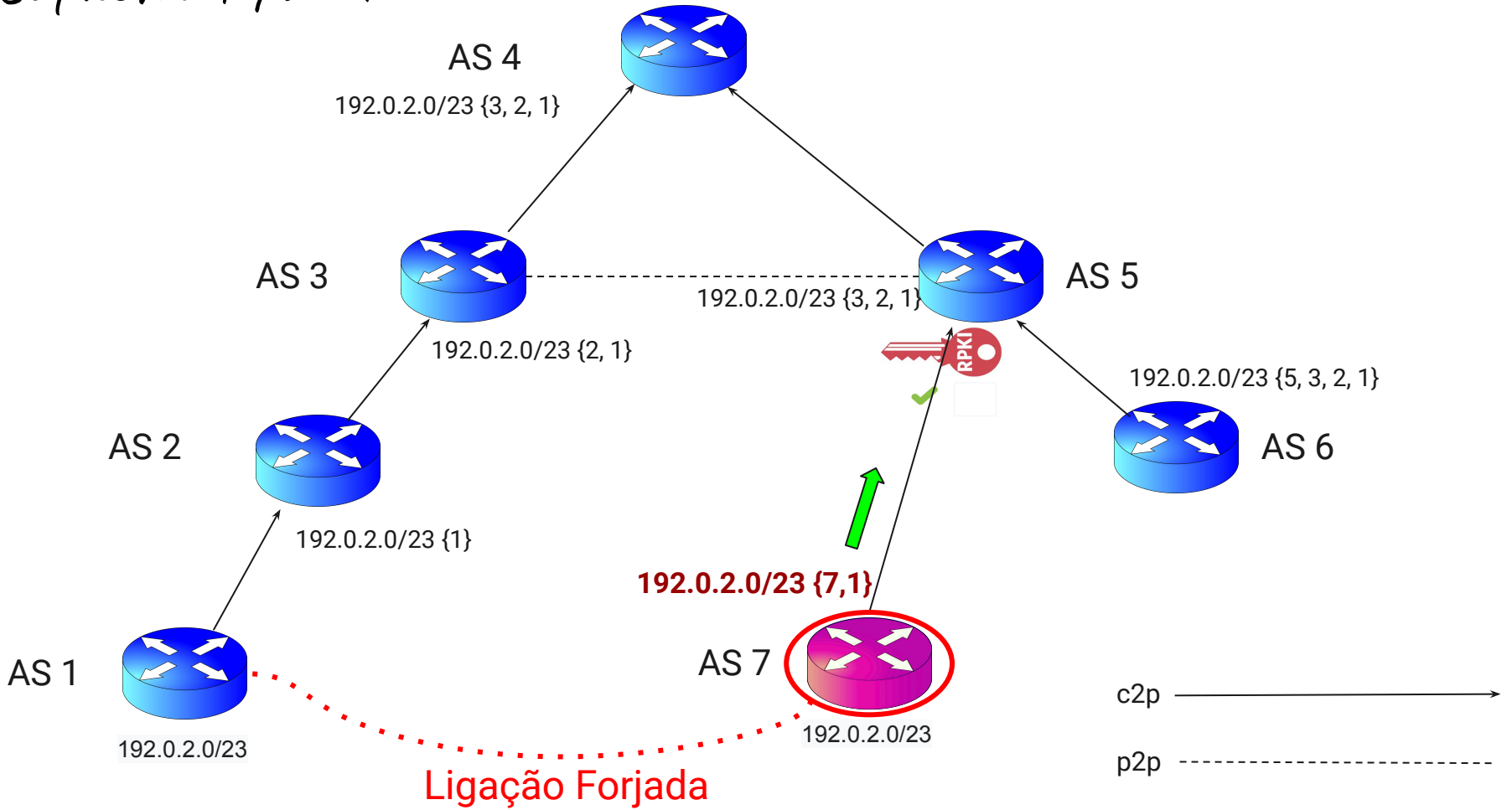
Sequestro Tipo - 0 ou Mudança de Rota



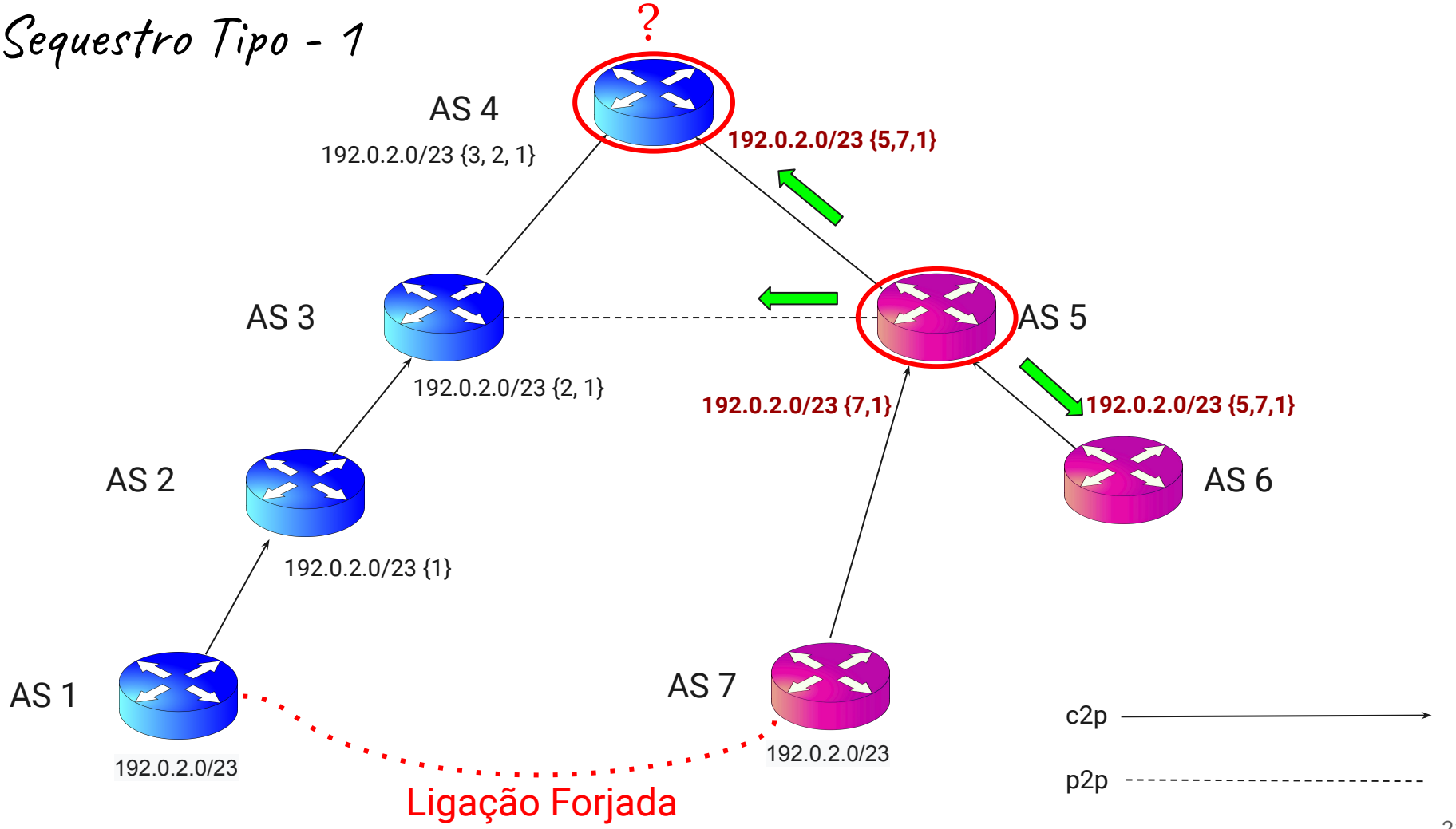
Sequestro Tipo - 0 ou Mudança de Rota



Sequestro Tipo - 1

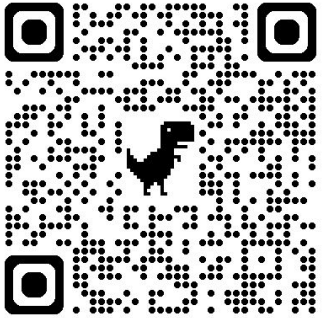


Sequestro Tipo - 1



A System to Detect Forged-Origin BGP Hijacks

- *Random Forest*
- 28 *Features*
  - 18 Topológicas
  - 5 Peering
  - 3 padrão do AS path (geradas por *Random Forest*)
  - 2 Bidirecionalidade
- Amostras de treinamento baseadas em agrupamento
- Uso de enlaces forjados de forma sintética para treinamento



https://dfoh.uclouvain.be/

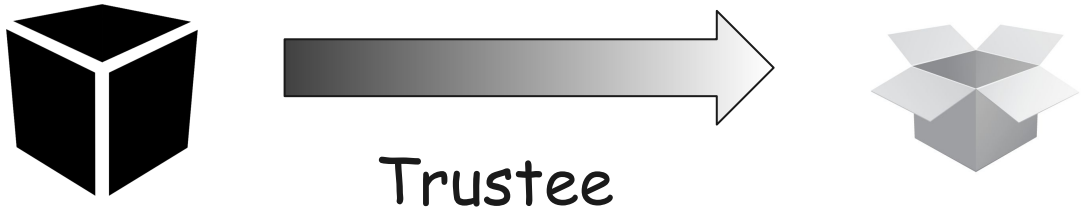
Thomas Holterbach, Thomas Alfroy, Amreesh Phokeer, Alberto Dainotti and Cristel Pelsser

NSDI 2024

A System to Detect Forged-Origin BGP Hijacks

- Conjunto de dados
  - 300 dias de observação (coletores (RIB + Updates) + CAIDA)
  - 60 dias de amostras para treinamento
  - 2.000 amostras por dia (mil de cada classe)
  - 200 VPs (Vantage Points)
- Foco na classificação dos novos enlaces

eXplainable Artificial Intelligence - XAI



- *Random Forest*
- *Extra-trees*

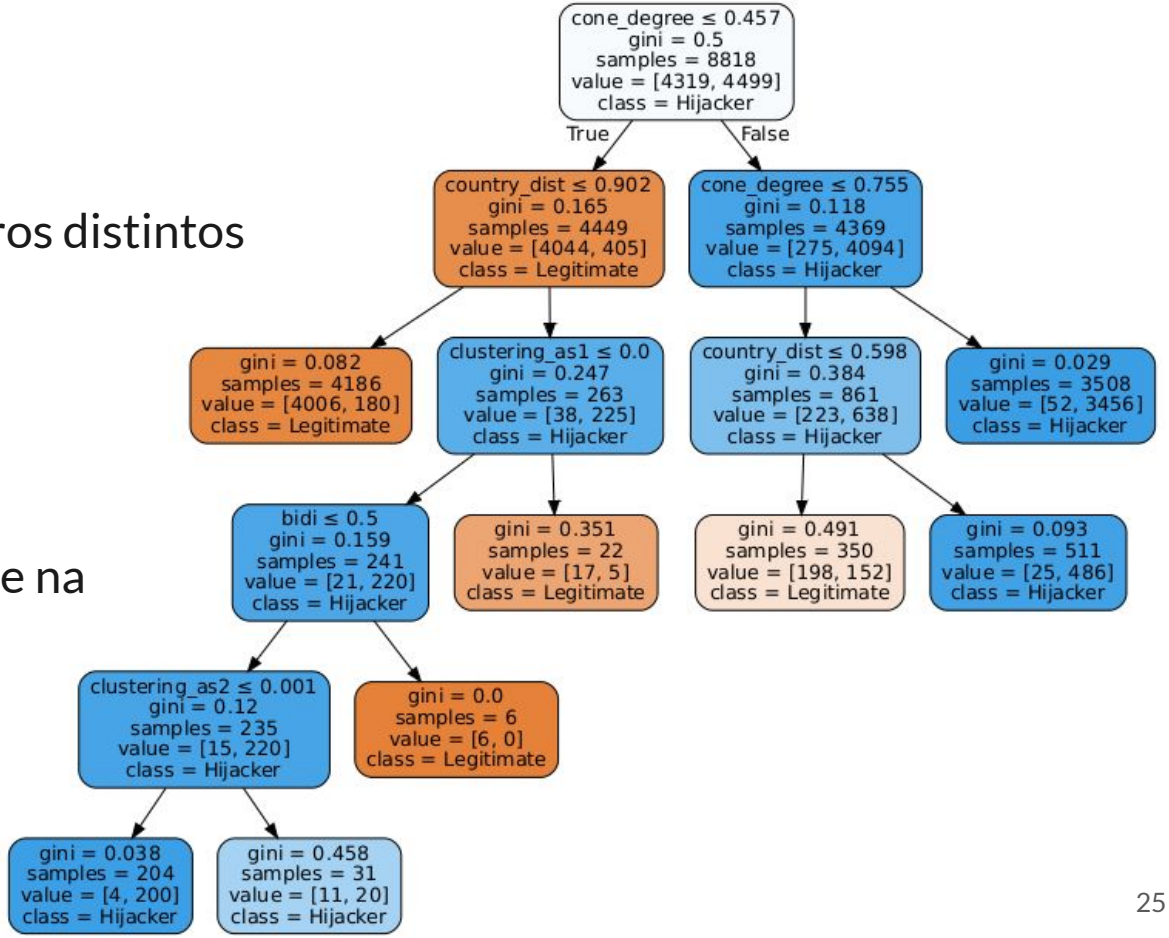
- Árvore de Decisão

- Aprendizado por imitação
- Amostras classificadas pelo modelo caixa-preta para treinamento da Árvore de Decisão

Abrindo o modelo Caixa-Preta do DFOH

Usando a ferramenta Trustee

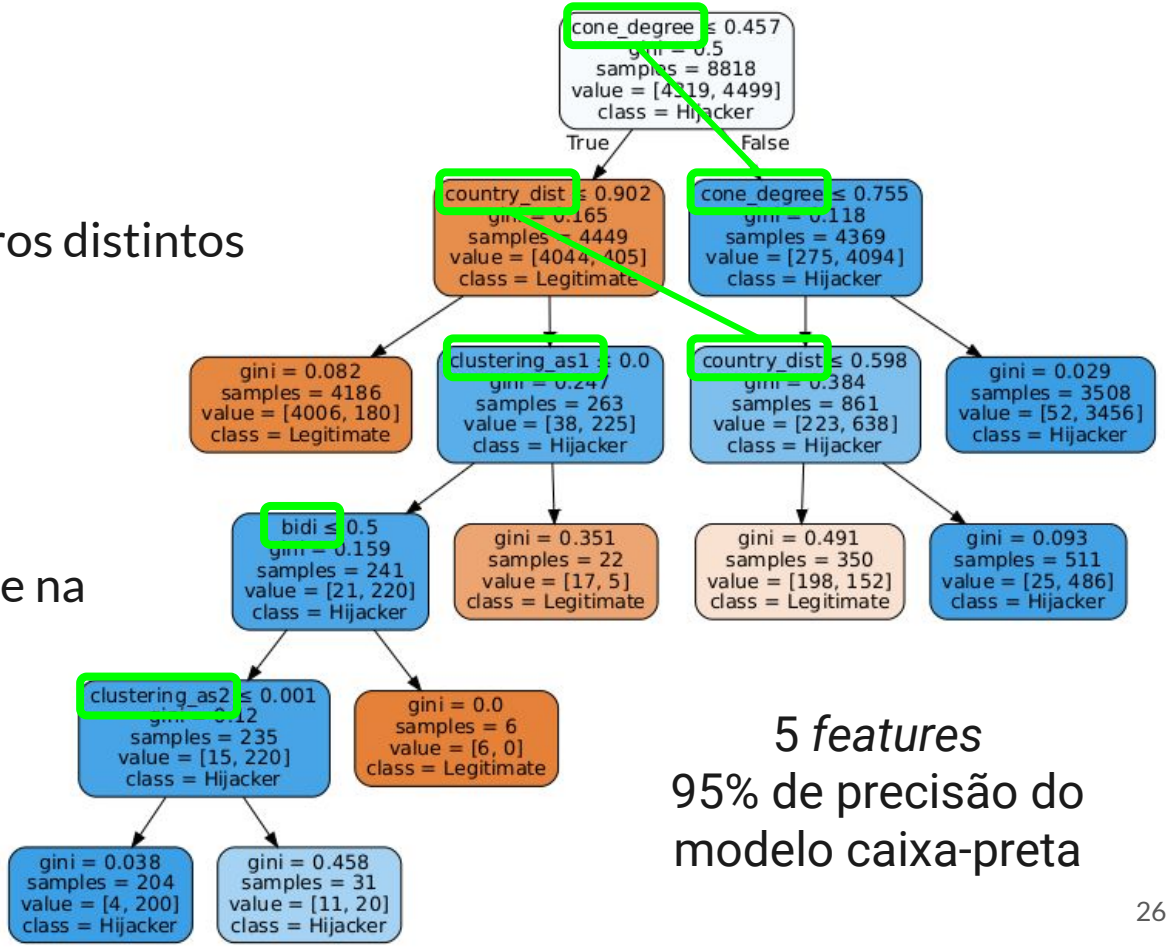
- 28 execuções com parâmetros distintos
- 56 Árvores de Decisão
  - 28 Árvores completas
  - 28 Árvores com podas
- *Features* ordenadas com base na presença nas Árvores



Abrindo o modelo Caixa-Preta do DFOH

Usando a ferramenta Trustee

- 28 execuções com parâmetros distintos
- 56 Árvores de Decisão
  - 28 Árvores completas
  - 28 Árvores com podas
- *Features* ordenadas com base na presença nas Árvores



5 *features*
95% de precisão do modelo caixa-preta

Avaliação dos modelos

- 4 modelos criados com redução de *features* (baseado na relevância);
- Todos os modelos (4 criados + original) obtiveram resultados dentro do Intervalo de Confiança;
- Selecionados os 2 menores para uma análise mais detalhada;

Categoria	M1	M4	Redução em M4	M5	Redução em M5
Topológica	18	5	72,22%	2	88,89%
Peering	5	2	60,00%	1	80,00%
Padrão AS path	3	2	33,33%	1	66,67%
Bidirecionalidade	2	2	0,00%	1	50,00%
Total de *features*	28	11	60,71%	5	82,14%

Avaliação dos modelos

- 40 dias de avaliação (2 periodos de 20 dias);
- 2.000 amostras por dia para avaliação dos modelos;
- Redução de tempo (média dos 40 dias, tempo em segundos):

	M1 ± IC	M4 ± IC	Redução em M4	M5 ± IC	Redução em M5
Cálculo das *features*	1875 ± 45	1291 ± 36	31,15%	1180 ± 32	37,04%

IC = Intervalo de Confiança (95%)

- Redução de espaço de armazenamento (160 dias = 2 x (60 dias para o treino + 20 dias de avaliação), valores em bytes):

	M1	M4	Redução em M4	M5	Redução em M5
Arquivos das *features*	144141619	57872896	59,85%	41575973	71,16%

Avaliação dos modelos

- Resultados obtidos com avaliação similar ao artigo original

	Legítimos					
Modelos	**Precisão ± IC**		**Recall ± IC**		**F1-Score ± IC**	
M1	0,9570	0,0018	**0,9590**	0,0022	0,9580	0,0014
M4	**0,9582**	0,0020	0,9580	0,0022	**0,9581**	0,0015
M5	0,9488	0,0020	0,9488	0,0025	0,9488	0,0015

	Suspeitos					
Modelos	**Precisão ± IC**		**Recall ± IC**		**F1-Score ± IC**	
M1	**0,9590**	0,0021	0,9569	0,0019	0,9579	0,0013
M4	0,9581	0,0021	**0,9581**	0,0021	**0,9581**	0,0015
M5	0,9488	0,0024	0,9488	0,0021	0,9488	0,0015

IC = Intervalo de Confiança (95%)

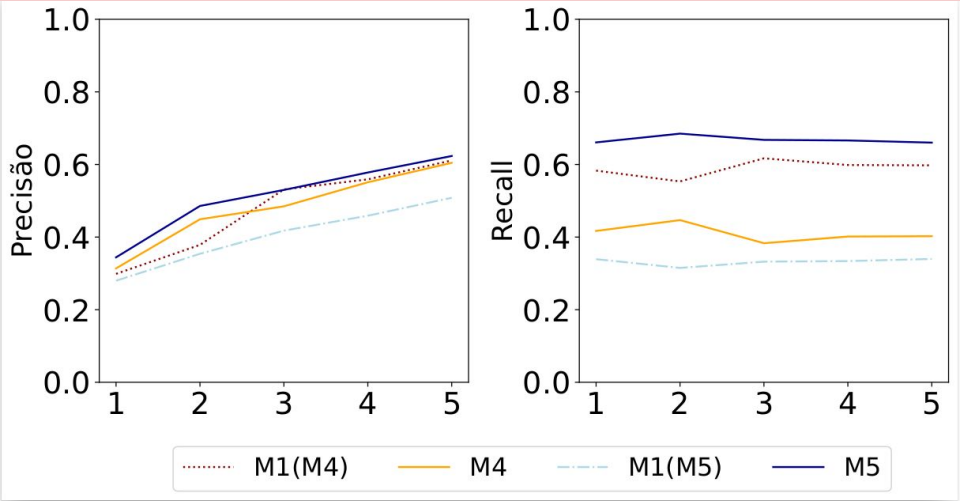
Avaliação dos modelos

- Novos enlaces observados nos 40 dias - 16.107;
- 968 enlaces inferidos diferente entre M1 e M4 (6,0%)
- 1.256 enlaces inferidos diferente entre M1 e M5 (7,8%)
- Como não há uma verdade sobre os novos enlaces observados, foi considerado o seguinte:
- Os sequestros costumam durar curtos períodos de tempo;
- Foi verificado se o novo enlace foi observado nos 5 meses seguintes;
- Foram analisadas somente as duas primeiras horas de RIB do mês.

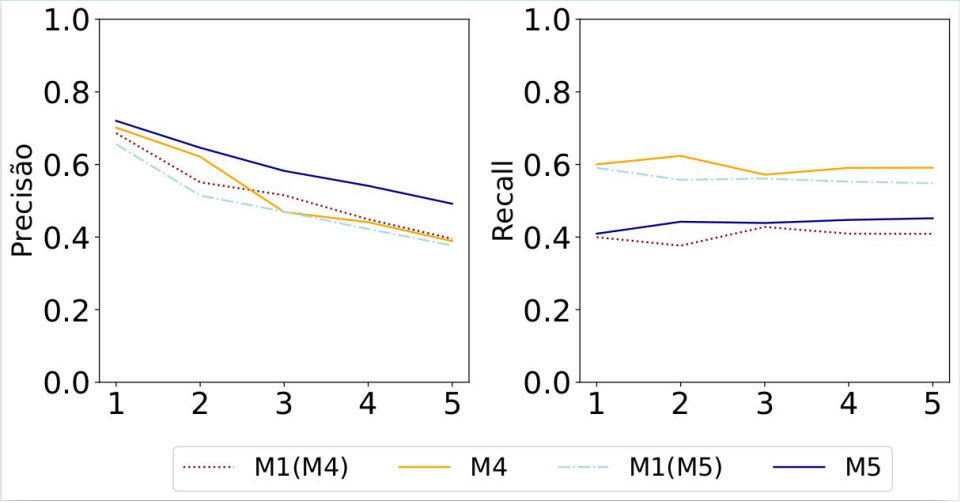
Avaliação dos modelos

- Foi considerada a necessidade do enlace ser observado de 1 a 5 meses posteriores para ele ser considerado legítimo (eixo X do gráfico).

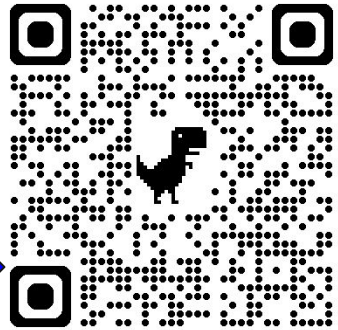
Suspeitos



Legítimos



Considerações finais



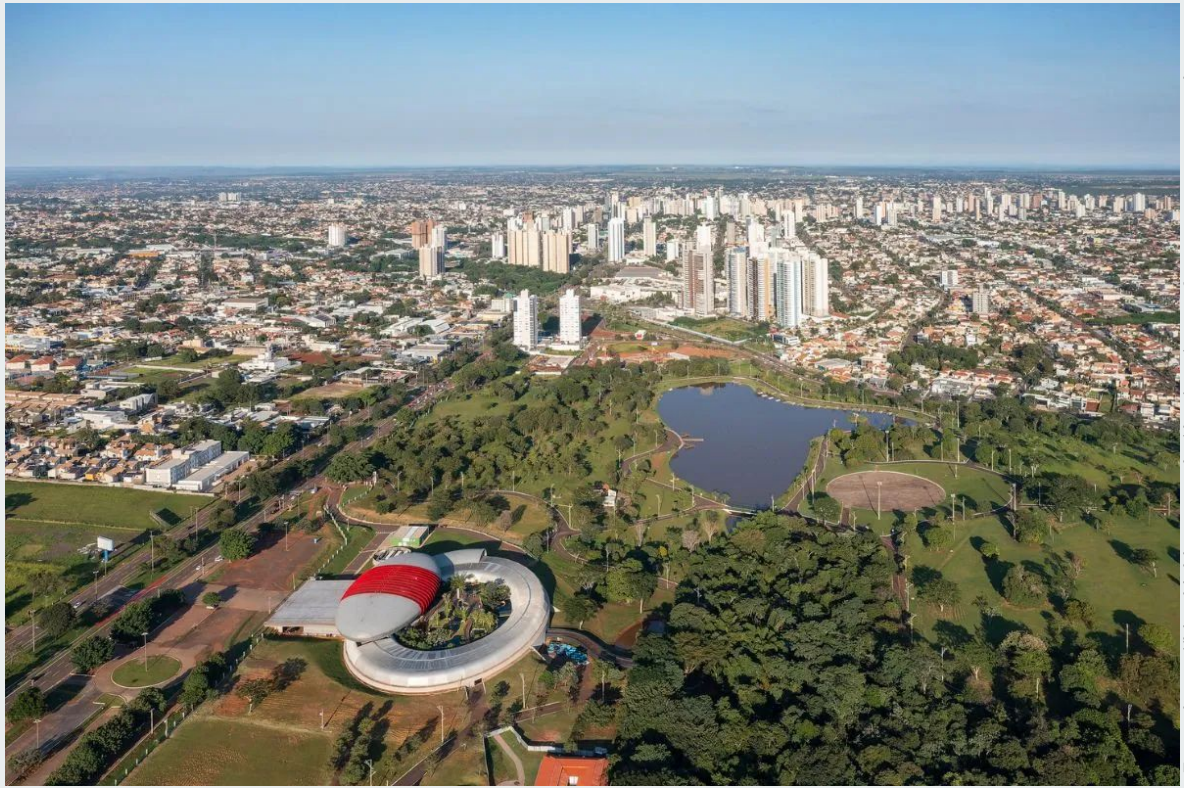
- Código utilizado disponível
- O sequestro de prefixo continua sendo um problema
- Os operadores de rede são relutantes no uso de modelos caixa-preta em cenários críticos
- XAI pode ajudar a entender o modelo caixa-preta (se há generalização)
- Usando XAI, pode-se observar que diversas *features* do atual estado-da-arte na detecção de sequestro com origem forjada são irrelevantes.

Trabalhos futuros

- Comparar os resultados obtidos com outras técnicas de seleção de *features*
- Buscar novas fontes de dados para incrementar o modelo
- Verificar a possibilidade de se obter uma outra forma de treinamento do modelo

Obrigado!

- Adriano B. de Carvalho
(adriano.bastos@ufms.br)
- Brivaldo A. da Silva Jr
(brivaldo.junior@ufms.br)
- Carlos Alberto da Silva
(carlos.silva@ufms.br)
- Ronaldo A. Ferreira
(ronaldo.ferreira@ufms.br)



Fonte: https://blog.blablacar.com.br/destinos/o-que-fazer-em-campo-grande

Campo Grande - MS