マルウェアの通信履歴と定点観測の相関について

〇小堀　智弘、 菊池　浩明(東海大)
寺田　真敏(日立製作所)

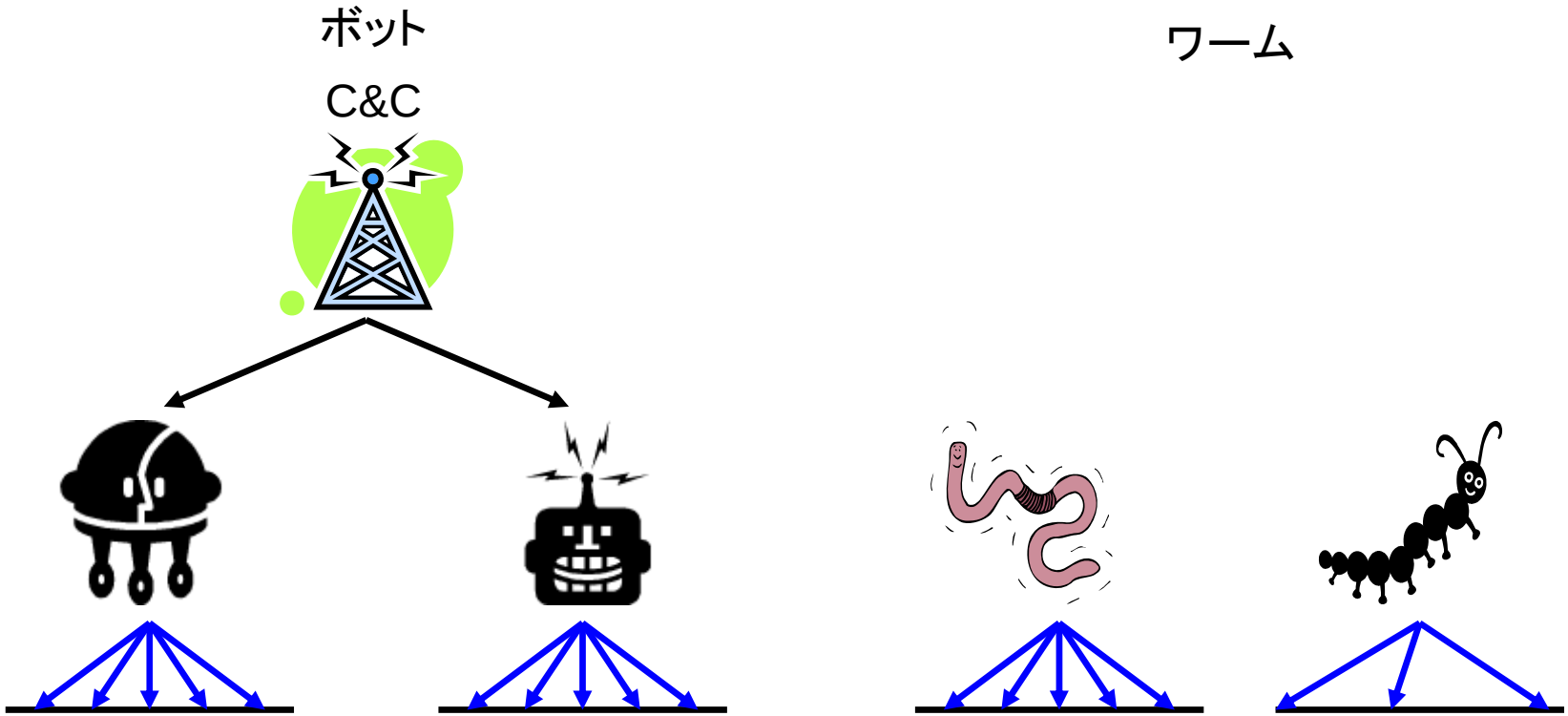
背景

- ネットワーク上に蔓延するコンピュータウィルスの80%はボットである

  http://www.blackhat.com/presentations/bh-jp-05/bh-jp-05-koyama.pdf

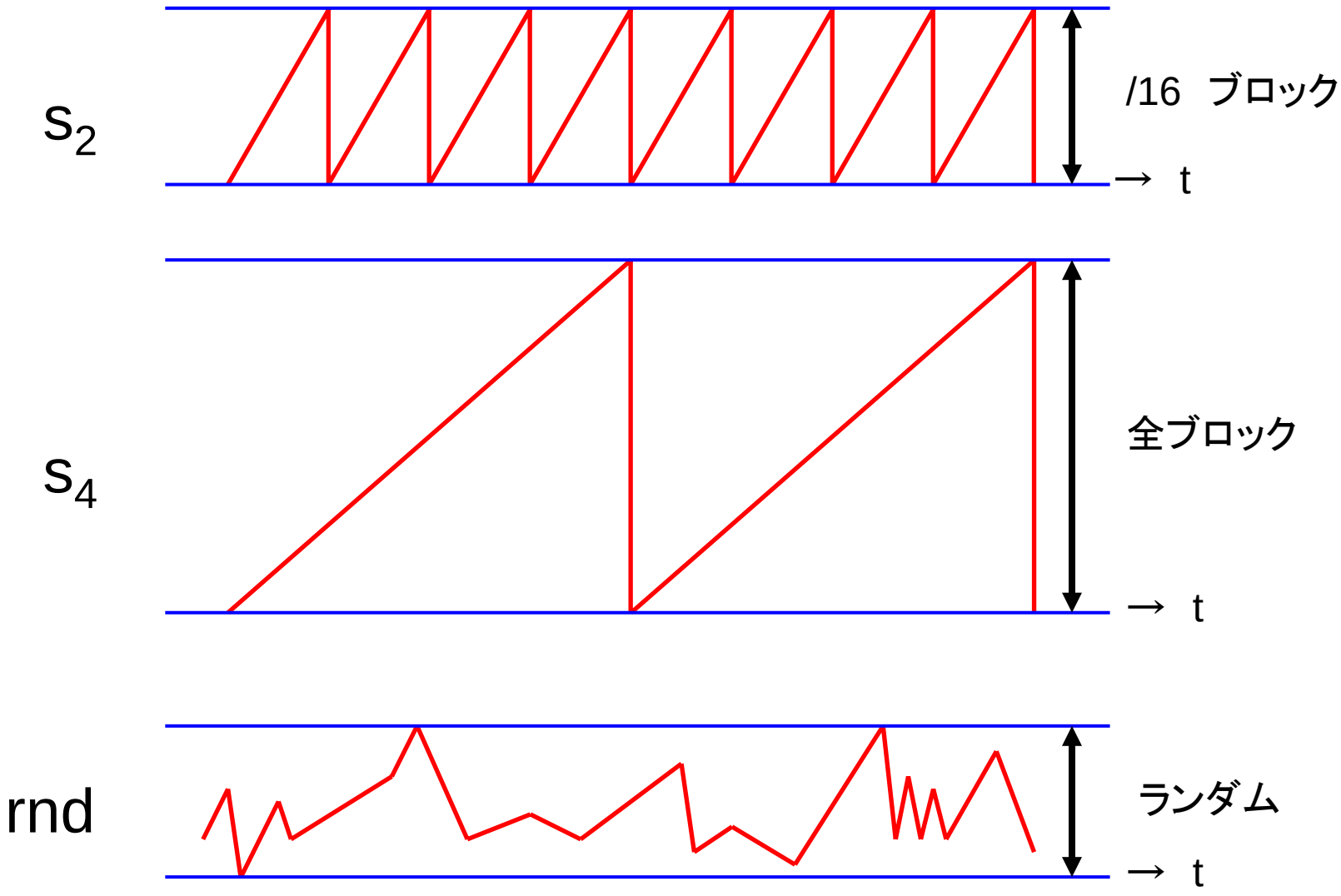
仮説1. ポートスキャンはすべてボットが行っている

ボットとワームの違い



仮説2. スキャンパターン　≠　MW（依存しない）

スキャンパターン



研究目的

- 仮説1.　ポートスキャンはすべてボット
- 仮説2.　スキャンパターン≠MW

  は本当だろうか？

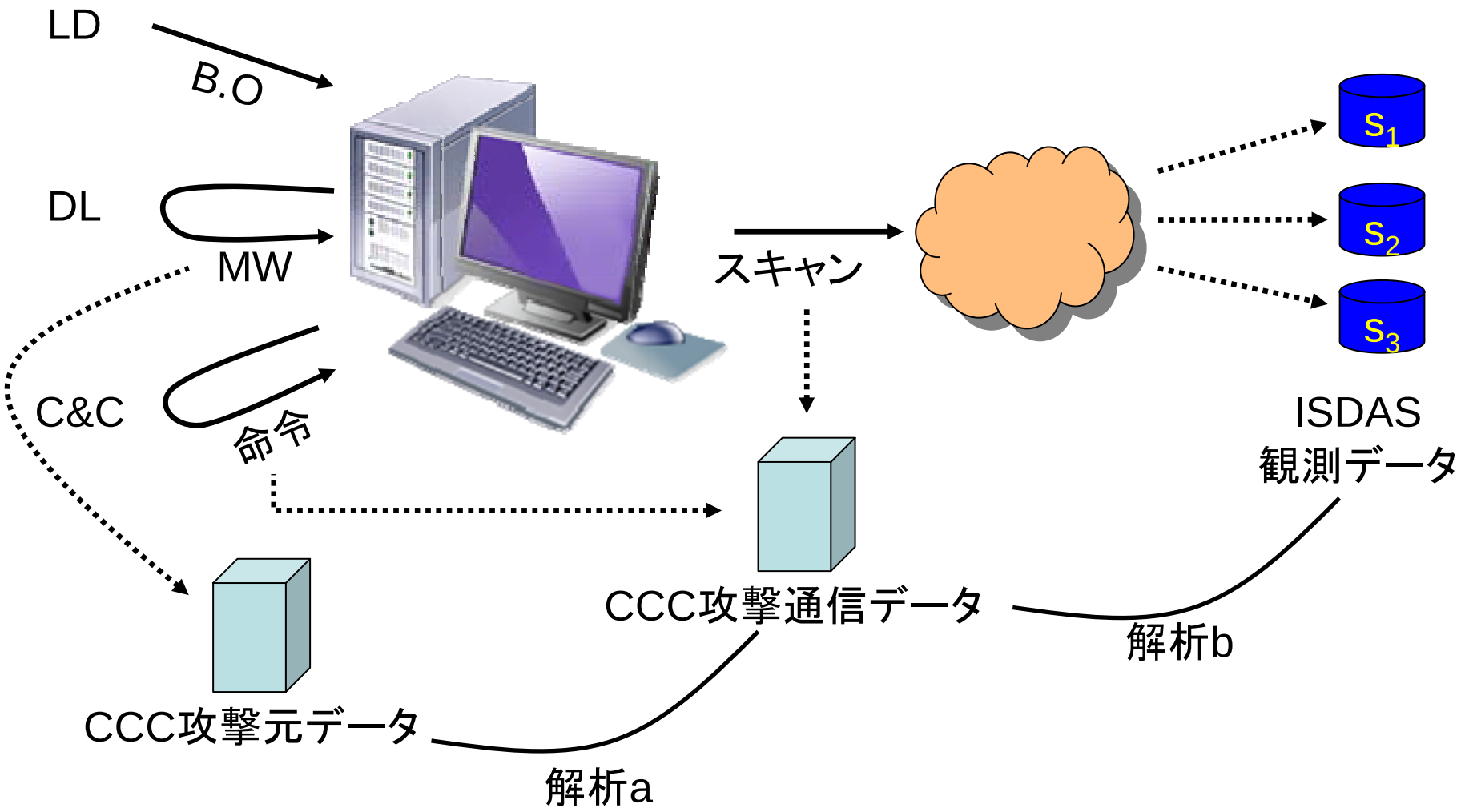
  実験的に検証する

研究方法

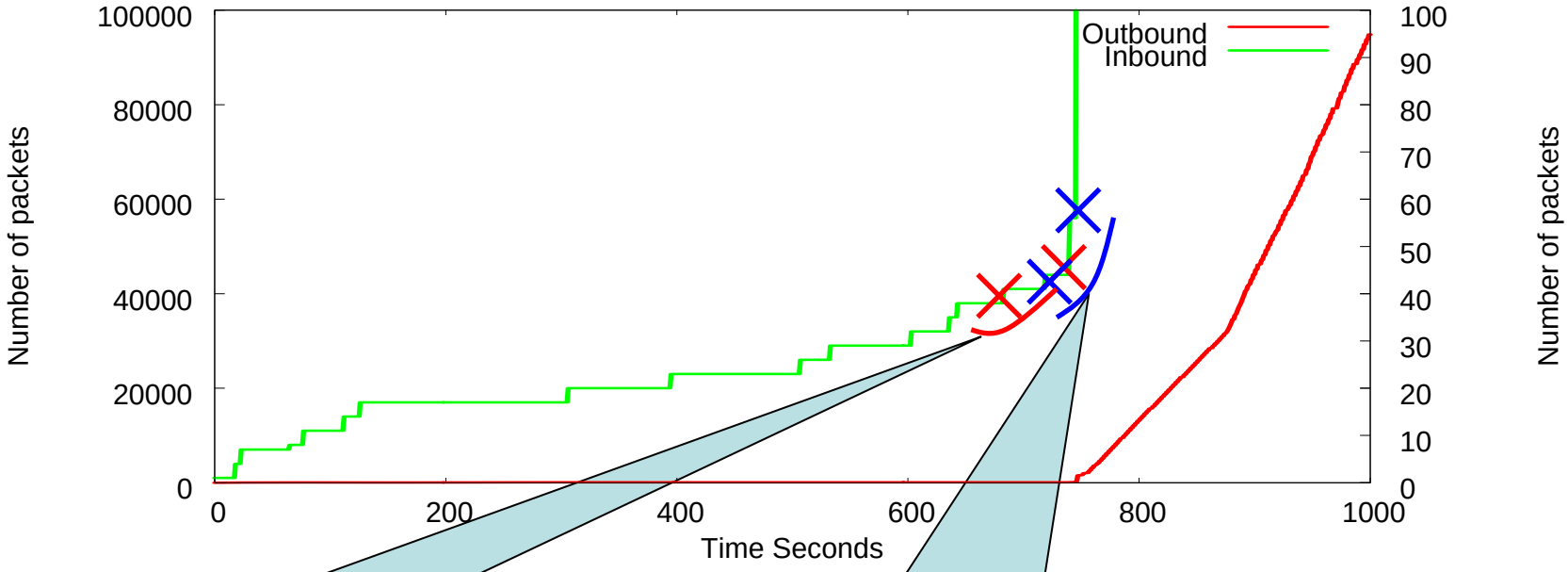
解析 a.　CCC攻撃元通信データ(3)と攻撃通信データ(2)の比較

解析 b.　攻撃通信データとISDASの比較

ボットネットのしくみ


LD
B.O
DL
MW
C&C
命令
スキャン
S1
S2
S3
ISDAS
観測データ
CCC攻撃通信データ
解析b
CCC攻撃元データ
解析a

実際の通信



07:18:40 PE_BOBAX.AK
07:18:47 TROJ_PACK.DT

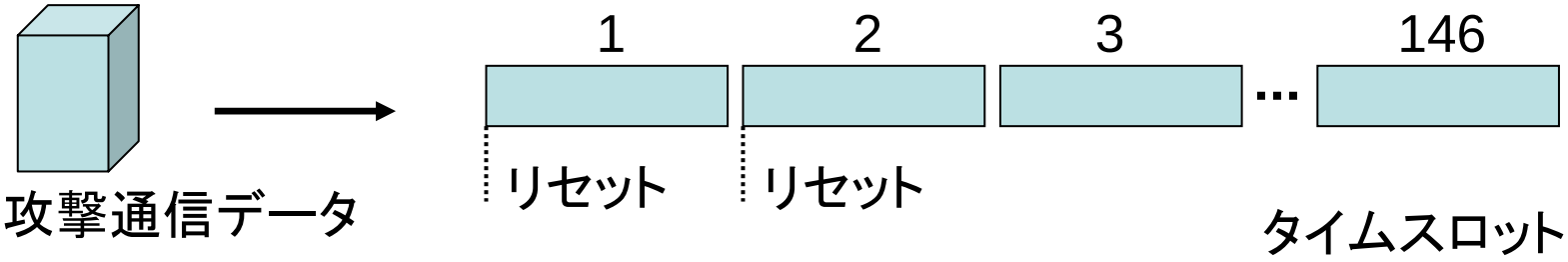
07:18:46 ipscan s.s.s.s dcom2 -s
07:20:00 ipscan s.s.s.s dcom2 -s

問題1. MWの多重感染(DL)

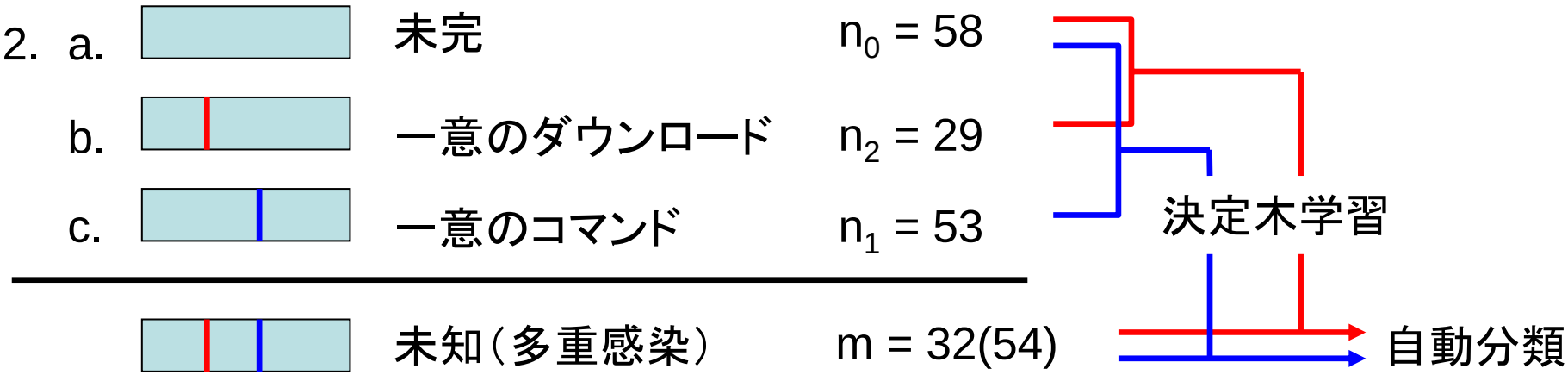
問題2. 多重コマンドとスキャンの一意性

提案方法：仮説2の検証

1.



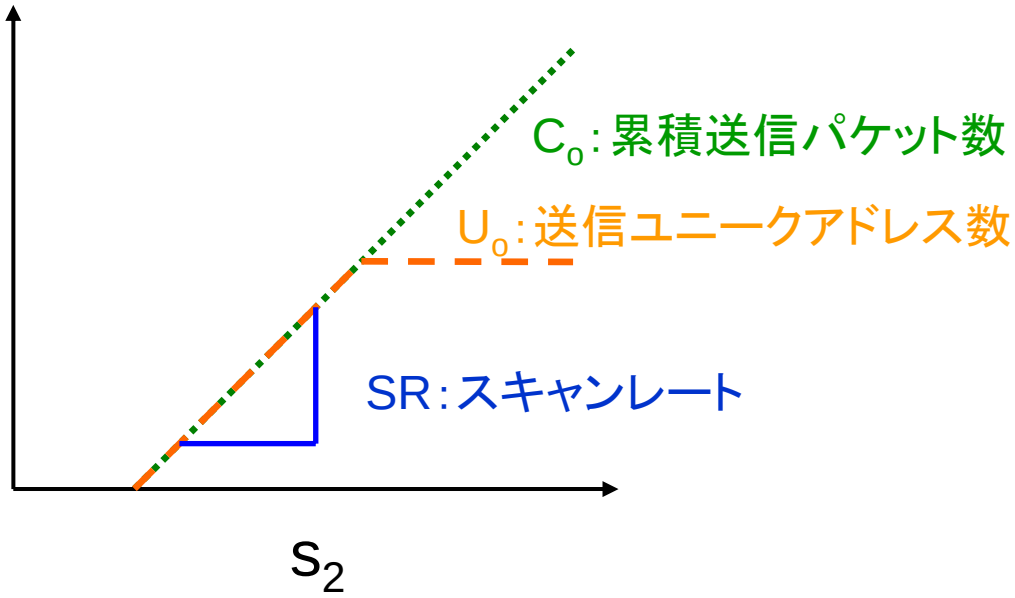
2.

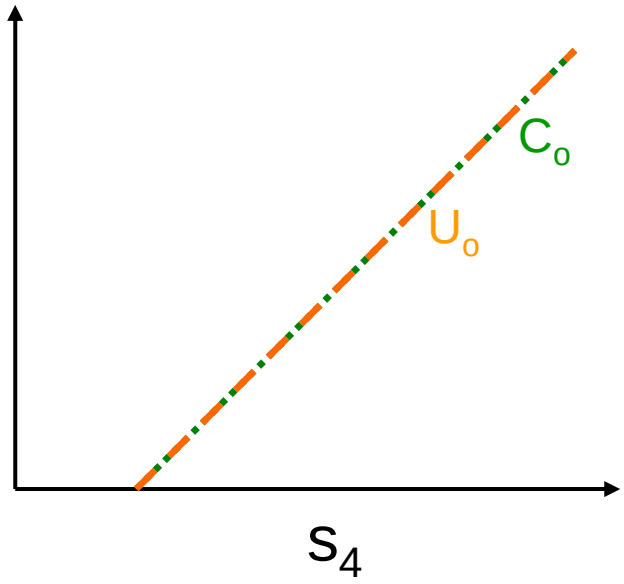


決定木

- 機械学習アルゴリズムC4.5
  - 情報量利得に基づいた識別
  - 決定木を作成
  - 連続値に対応

スキャンパターンの特徴量　1/2





スキャンパターンの特徴量　2/2

IN	C_I：総入力パケット数[pkt] U_I：ユニーク発信元アドレス数 V_{PS}：d_{SP} の標準偏差 MW：マルウェア名 DL：ダウンロードの有無(0,1)	MWを決定
OUT	C_O：総出力パケット数[pkt] U_O：ユニーク宛先 アドレス数 DP：宛先ポート(135, 445, ICMP) SR：スキャンレート[pkt/s]	スキャンパターンを決定

提案方法：仮説1の検証
ISDASとの比較

- ハニーポットのスキャンがISDASで観測しているかを調査

確認できず

1. ポートの比率
2. スキャンレートの比率

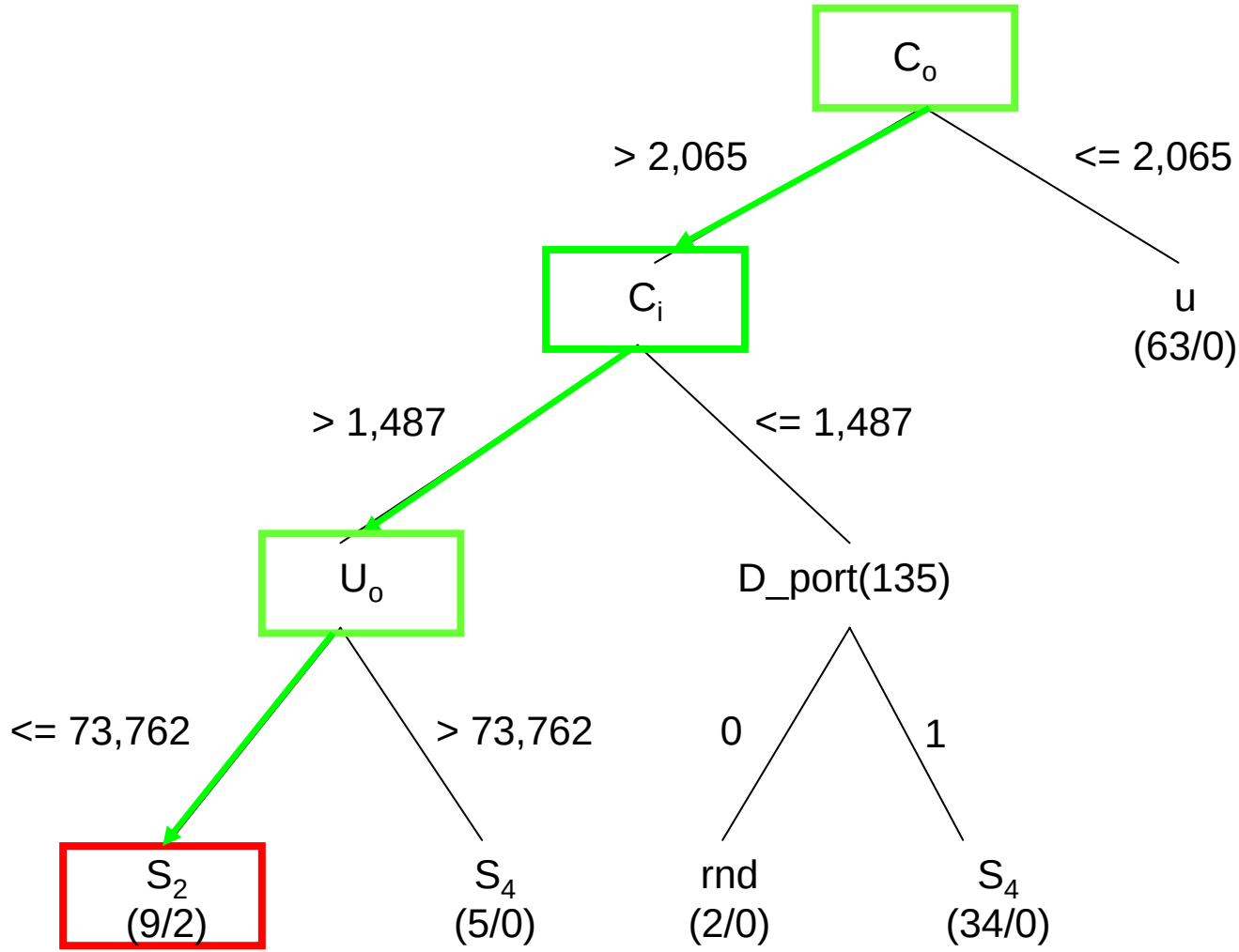
実験方法

実験 1. スキャンパターンの識別

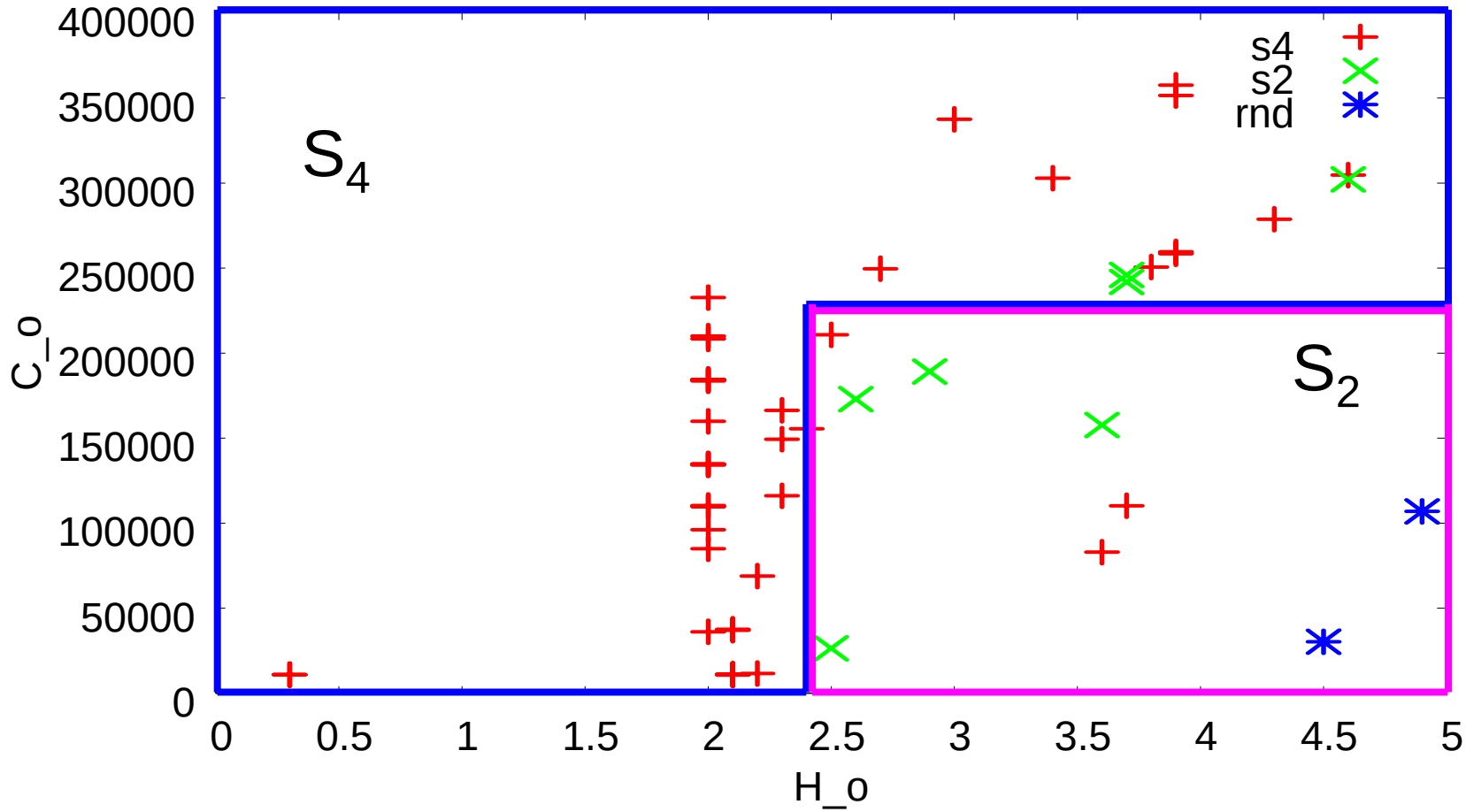
実験 2. MWの識別

実験 3. ISDASとの比較

実験1：C4.5による スキャンパターンの決定木



実験1:スキャンパターンの散布図



実験1:スキャンパターンの決定木の精度

ST/評価値	rnd	s_2	s_4	u	total
未分類 m	2	5	24	1	32
rnd	2				2
s_2		7		4	11
s_4		2	38	1	41
u	1			57	58
total	5	14	62	63	144
適合率 P_{ST}	0.67	0.78	1.0	0.92	0.93

- 再現率　0.93、平均適合率　0.94

実験1:マルウェアとスキャンタイプの相関

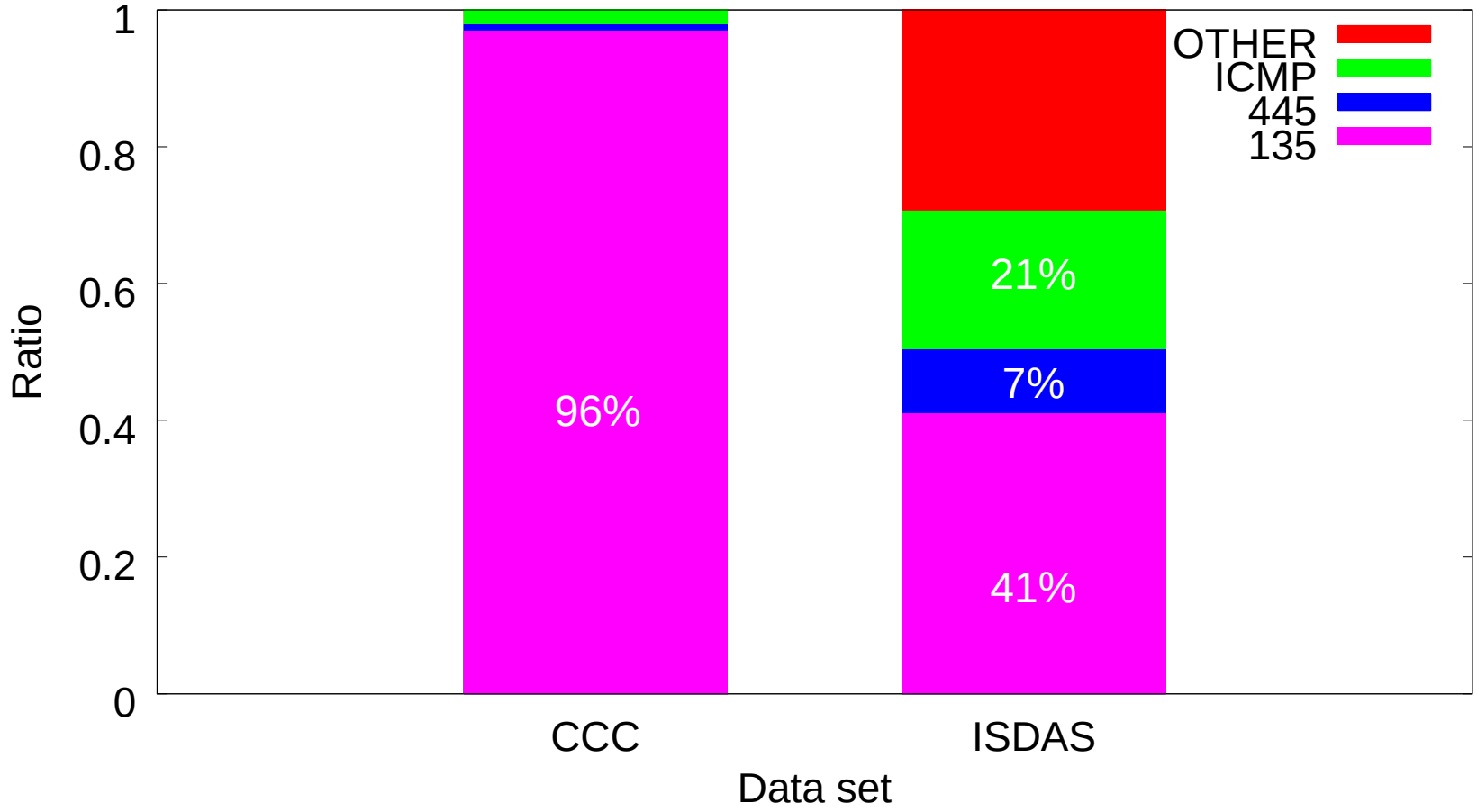
- マルウェアとスキャンタイプ

MW/ST	s_2	s_3	s_4	rnd	total
BOBAX	17	2	28	3	35
KOLABC	7	2	14	5	18
VANBOT	30	4	42	5	58
VIRUT	7	0	20	1	25
OTHER	8	2	20	2	32

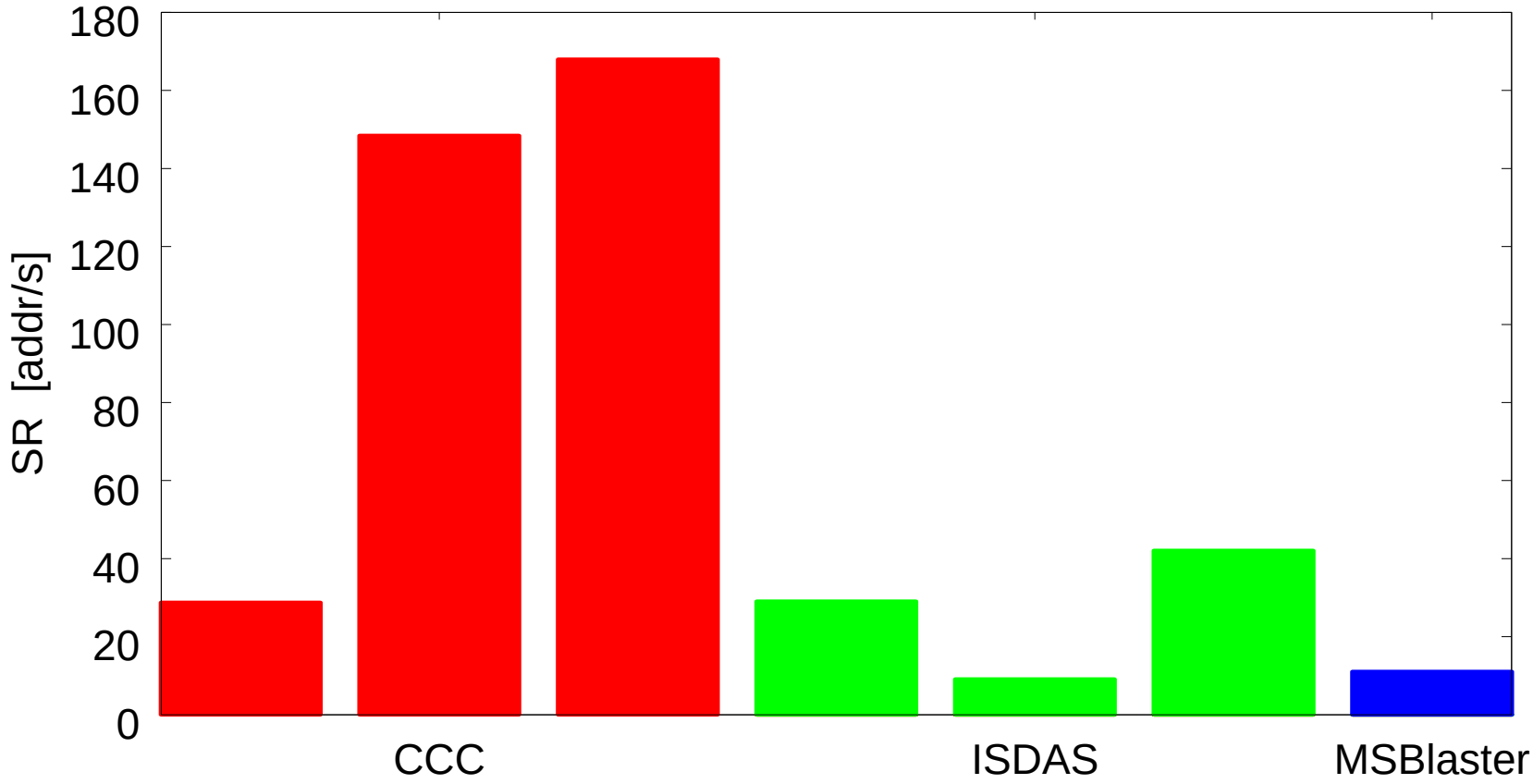
結論:仮説2成立

実験3:定点観測との比較

- 宛先ポートの分布の比較



3つの観測データによる スキャンレート　SR



結論：仮説1不成立

まとめ

- 仮説1
「ポートスキャンは全てボットによって行われている」は成立しない
- 仮説2
「ポートスキャンのタイプはマルウェアに依存しない」は成立する
- 決定木学習により、ペイロードを見ることなくスキャンパターンを同定することは可能である

ご清聴ありがとうございました