情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

類似マルウェアとの差分の抽出による正確な挙動の解析手法についての検討

2012. 7. 6 (DICOMO 2012)

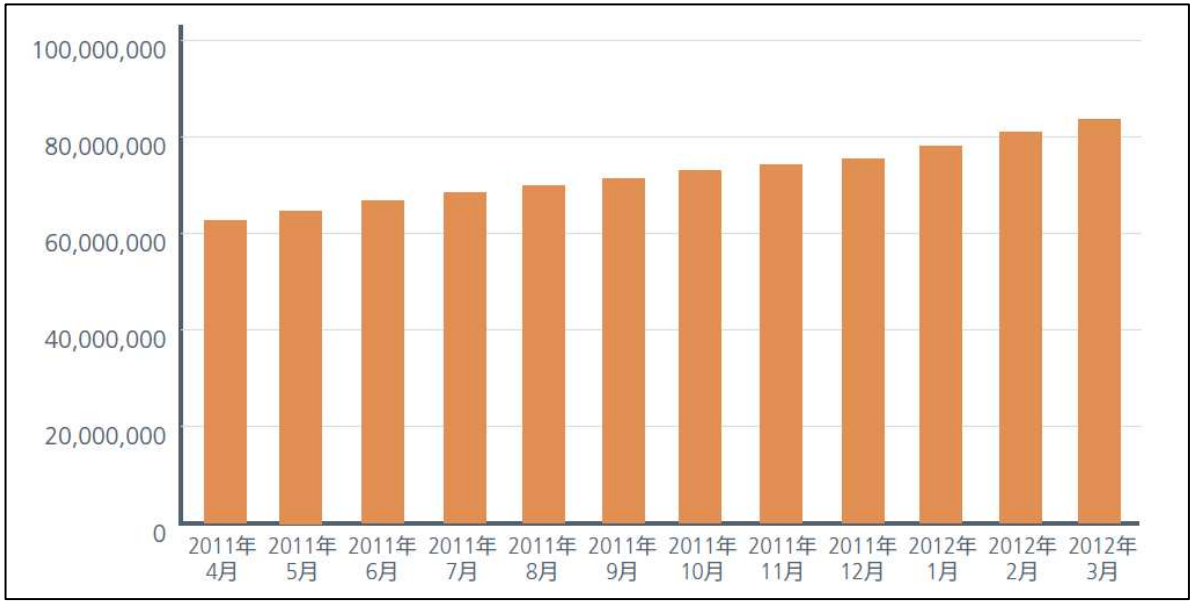
情報セキュリティ大学院大学
羽田 大樹, 後藤 厚宏

増加するマルウェア亜種

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

♣マルウェアの増加により対策はより困難に

- 発見したマルウェアは8000万種類超 (1.5秒に1個のペース)



出典: McAfee脅威レポート: 2012年 第1四半期

- 亜種生成技術の発達や頻繁なアップデート
  （完全にユニークなマルウェアの種類は多くはない）
- マルウェア亜種を網羅することは不可能

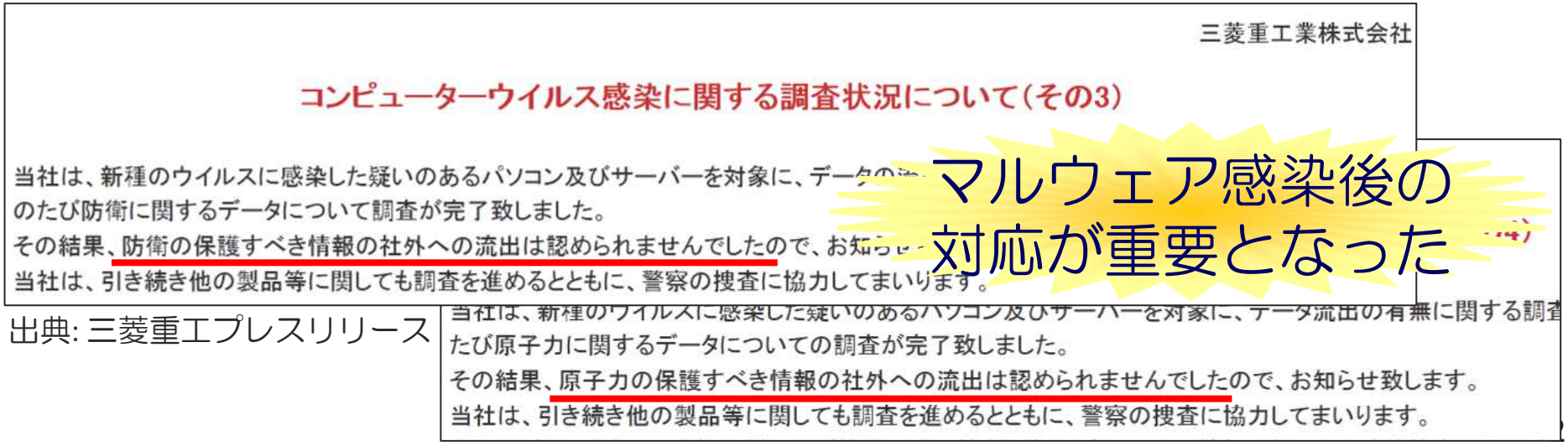
マルウェア被害の傾向と課題

♣手法やターゲットの変化

- 標的型攻撃など攻撃手法の高度化
- 政府や軍事機関などの重要インフラがターゲット

♣事例: 三菱重工株式会社

- 2011.9, 計83台のサーバーとPCにマルウェア感染を確認
- 専門家によるフォレンジック調査の結果, 影響範囲を特定
- 事件発生後の迅速かつ正確な解析が求められている



三菱重工業株式会社

コンピューターウイルス感染に関する調査状況について(その3)

当社は、新種のウイルスに感染した疑いのあるパソコン及びサーバーを対象に、データの
のたび防衛に関するデータについて調査が完了致しました。
その結果、防衛の保護すべき情報の社外への流出は認められませんでしたので、お知らせ
当社は、引き続き他の製品等に関しても調査を進めるとともに、警察の捜査に協力してまいります。

マルウェア感染後の
対応が重要となった

当社は、新種のウイルスに感染した疑いのあるパソコン及びサーバーを対象に、データ流出の有無に関する調査
たび原子力に関するデータについての調査が完了致しました。
その結果、原子力の保護すべき情報の社外への流出は認められませんでしたので、お知らせ致します。
当社は、引き続き他の製品等に関しても調査を進めるとともに、警察の捜査に協力してまいります。

出典: 三菱重工プレスリリース

マルウェア解析手法と研究テーマ

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

♣フォレンジックにおけるマルウェア解析手法

- 動的解析と静的解析に分類できる
- 動的解析で分からない挙動は静的解析で解析する

	動的解析	静的解析
解析方法	検体を実行して挙動を観測	実行コードを読み解く
解析精度	動作した部分だけ	原理的には完全に解析可能
作業時間	数分から数日	1週間以上
要するスキル	自動化できる場合もある	アセンブリ解析スキル 自動化は極めて困難

♣研究テーマ

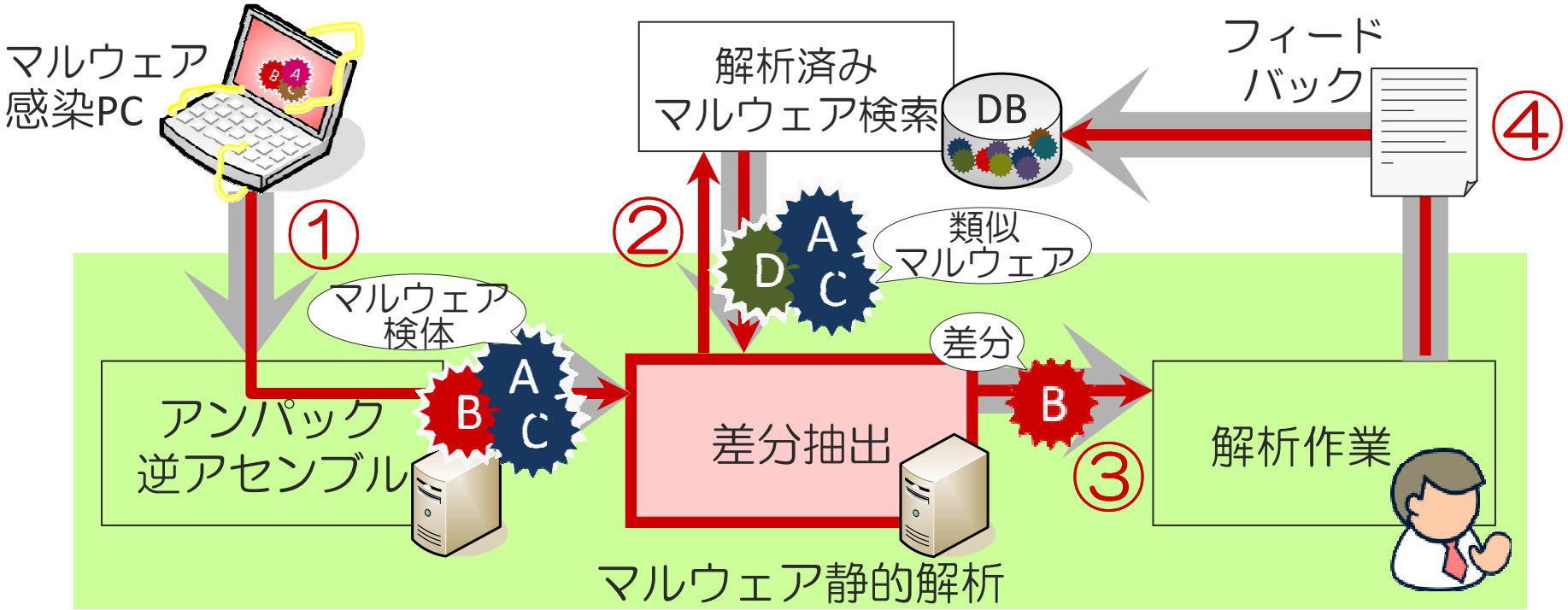
過去に解析した類似マルウェアの解析データを
利用して静的解析の作業時間を削減したい

提案システムアーキテクチャ (1)

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

♣ 全体像

① マルウェア検体のアセンブリコードを取得
② マルウェア検体と類似する解析済みマルウェアを検索
③ マルウェア検体と解析済みマルウェアから差分 (未解析部分) を出力
④ 手動解析結果をデータベースにフィードバック



提案システムアーキテクチャ (2)

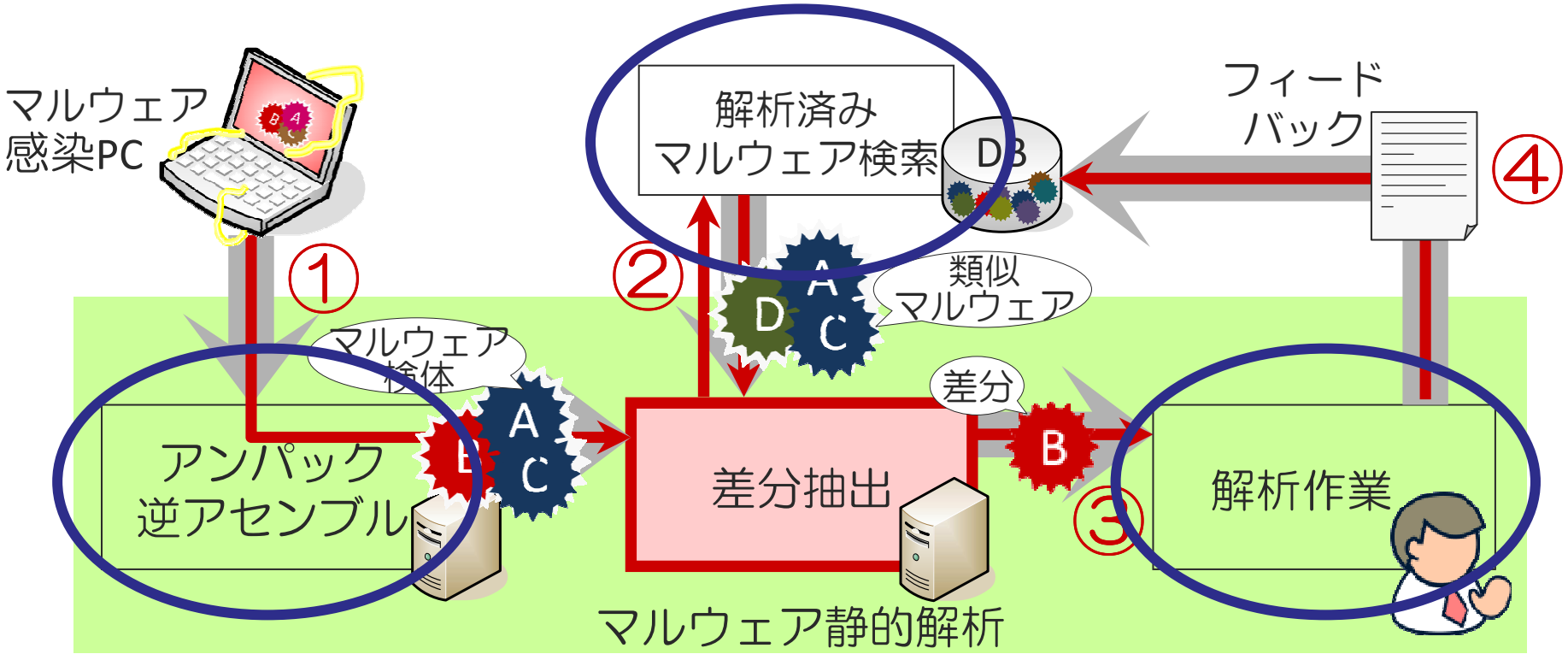
情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

♣ 仮定してよい条件

① 解析済みマルウェアデータベースが存在し, 類似する解析済みマルウェアを検索できること

② アンパックと逆アセンブルを理想的に行えること

③ 解析作業自体は手動で行うこと



関連研究: プログラムの差分抽出

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

♣ Structural Comparison of Executable Objects. H. Flake, 2004

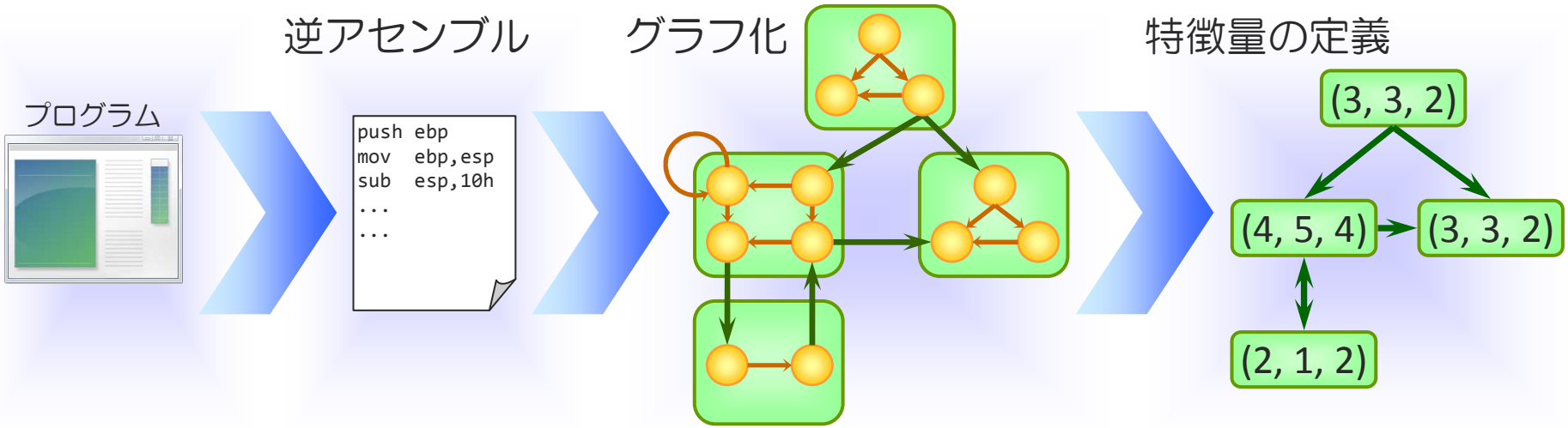
- リバースエンジニアリングにおけるパッチ修正箇所の特定手法
- アセンブリを入れ子の有向グラフで表現

  1. コールグラフ (緑色): 関数単位で分割
  2. 制御フローグラフ (黄色): if や for など制御命令単位で分割

- コールグラフのノードに特徴量を定義

  (ベーシックブロック数, ノード内の辺の数, ノード間の辺の数)

- 特徴量を用いて2つのグラフを比較して同一箇所を特定

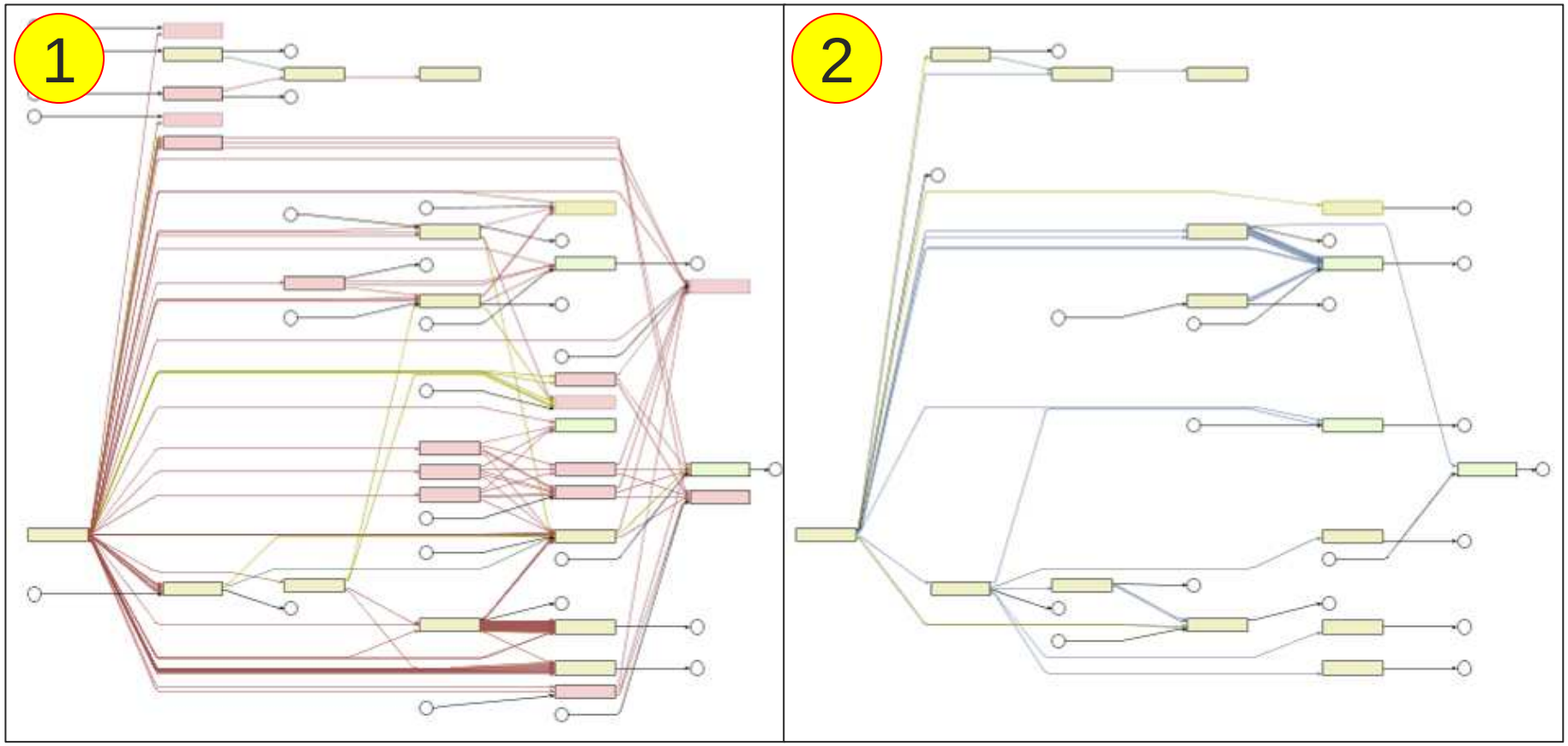


関連技術 – BinDiff (1)

♣ BinDiff

- H. Flakeの方式を実装したソフトウェア

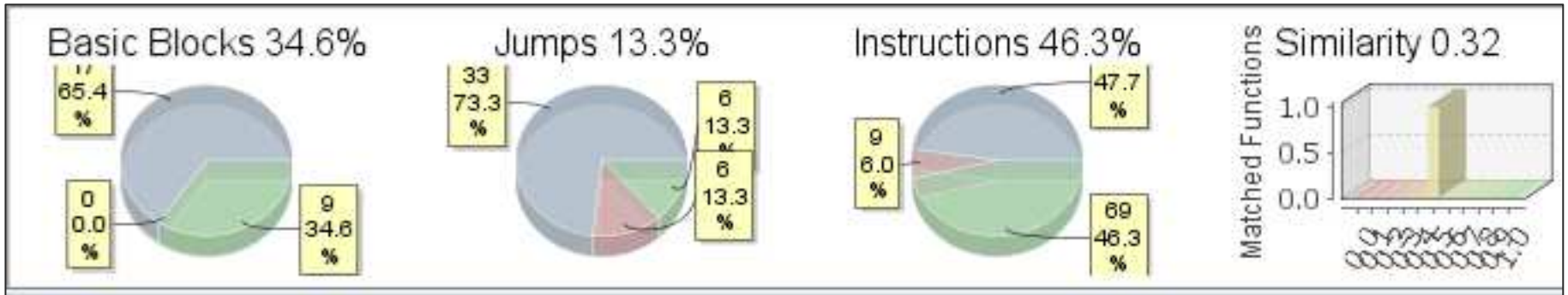
♣ 実行画面1: コールグラフの比較



関連技術 – BinDiff (2)

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

♣実行画面2: コールグラフの比較結果



134 / 134 Matched Functions

1

Show structural chang

2

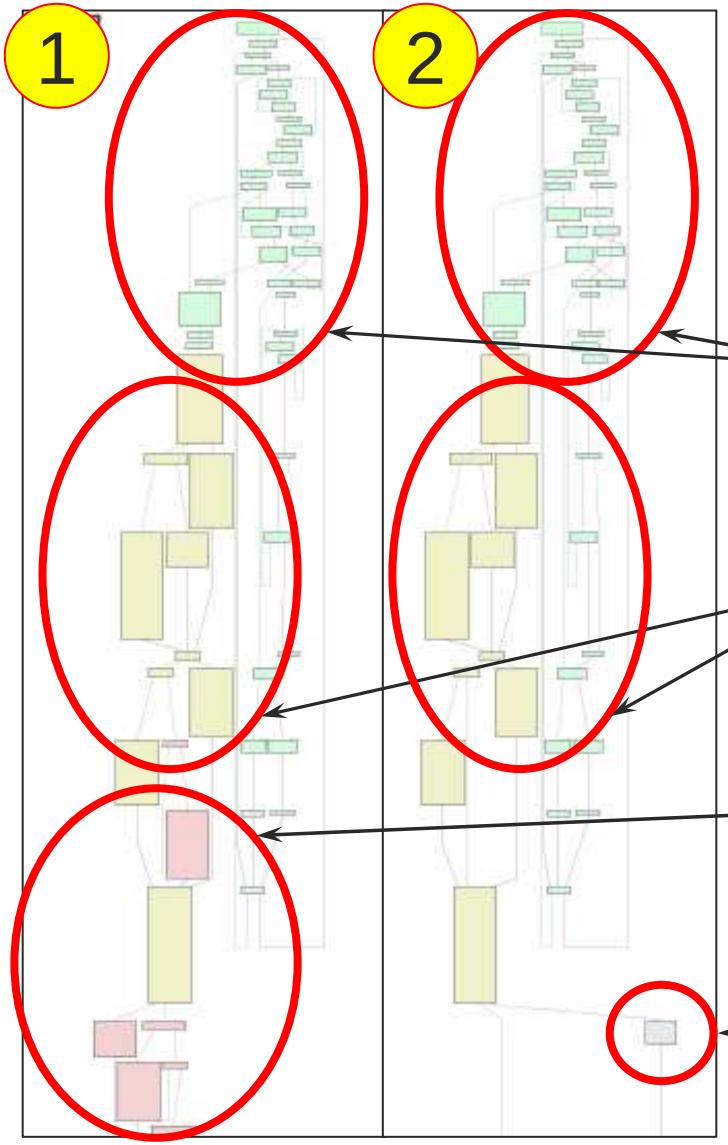
Show only instructions changed

Show identic

Sim...	Confid...	Address	Primary Name	Type	Address	Secondary Name	Type	Basic Blocks			Jumps		
0.99	0.99	00405F6D	sub_405F6D	No...	004010A7	sub_4010A7	No...	0	8	0	0	10	0
0.98	0.99	0040332D	sub_40332D	No...	00403E...	sub_403EAD	No...	0	4	0	0	5	0
0.98	0.99	00416C59	sub_416C59	No...	00405F88	sub_405F88	No...	0	71	2	1	107	4
0.98	0.99	0041649C	sub_41649C	No...	004057...	sub_4057CB	No...	1	39	0	2	60	1
0.96	0.97	0040A1E0	sub_40A1E0	No...	00402A31	sub_402A31	No...	0	17	0	0	25	0
0.96	0.96	0040A160	sub_40A160	No...	004029B5	sub_4029B5	No...	0	4	0	0	5	0
0.90	0.99	00428120	start	No...	00418F30	start	No...	7	49	8	14	70	14
0.90	**0.98**	**004160FB**	**sub_4160FB**	**No...**	**0040542A**	**sub_40542A**	**No...**	**1**	**13**	**0**	**5**	**13**	**4**
0.80	0.98	0040A6F3	sub_40A6F3	No...	00402C...	sub_402C7C	No...	0	15	6	3	19	12
0.79	0.95	00415FE3	sub_415FE3	No...	00405312	sub_405312	No...	0	5	2	0	6	4
0.76	0.78	00418CB7	sub_418CB7	No...	00403C...	sub_403C5B	No...	0	3	0	0	3	0
0.74	0.78	0040BF2F	sub_40BF2F	No...	00403F1C	sub_403F1C	No...	0	3	0	0	3	0
0.71	0.73	004113A4	sub_4113A4	No...	00401328	sub_401328	No...	0	4	0	0	4	0

関連技術 – BinDiff (3)

♣実行画面3: 制御フローグラフ



完全に一致するブロック
(緑色)

部分的に一致するブロック
(黄色)

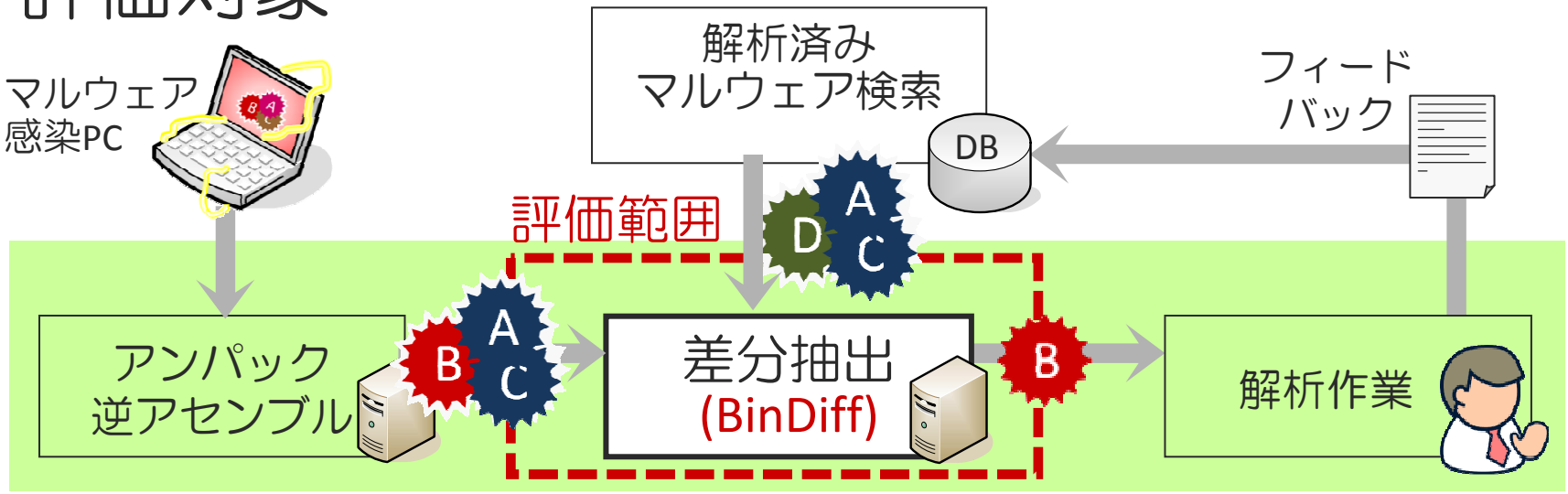
① のみに存在するブロック
(赤色)

② のみに存在するブロック
(青色)

マルウェア差分抽出の評価結果

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

♣ 評価対象



♣ 評価マルウェア検体 “SpyEye”

- 口座番号やクレジットカード情報を取得するマルウェア

検体	関数	ハッシュ値 (MD5)	想定する役割
検体1	523	9D2A48BE1A553984A4FDA1A88ED4F8EE	解析対象マルウェア検体
検体2	139	D64CA15261C53279A7288616B3CB1A92	解析済みマルウェア
検体3	609	DF04C2CD2B5F7E471CB0435FDB9B3014	解析済みマルウェア
検体4	218	42DACFBE2E5AF0C43D17356CA76F0271	解析済みマルウェア

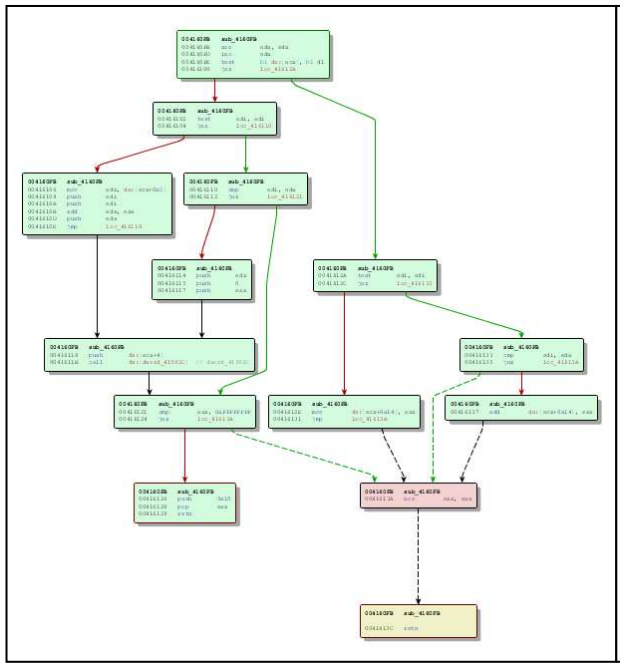
評価方法

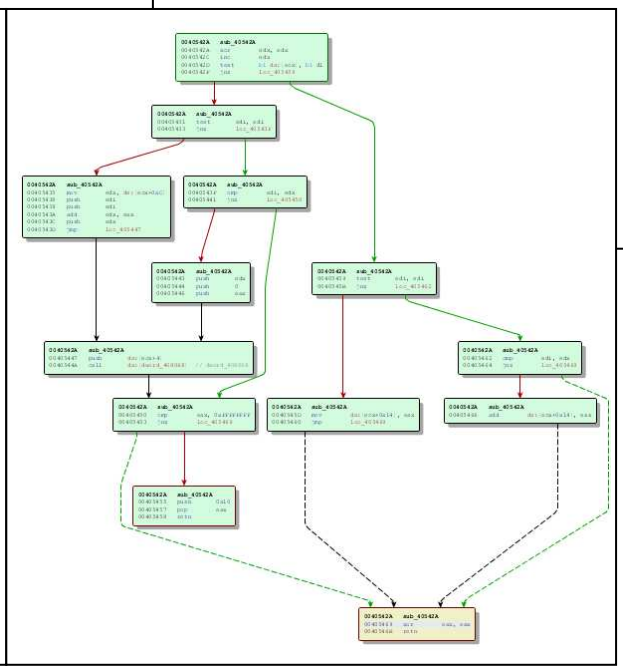
情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

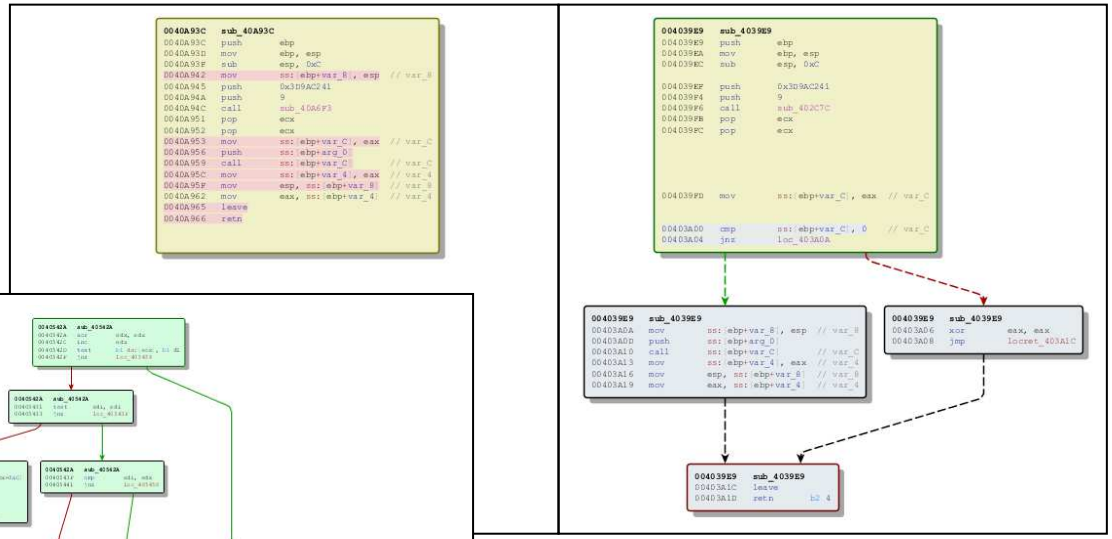
♣ 評価基準

- 解析済みマルウェアと一致する関数の数で評価
- 一致率が低いものは対象外 (今回は0.50を閾値)

一致率 0.90 の例







一致率 0.21 の例

評価1: 差分抽出の有効性

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

♣ 検体1, 2における比較結果

- BinDiff で比較した結果, 53 / 523 = 10.1% が一致

検体における関数の数

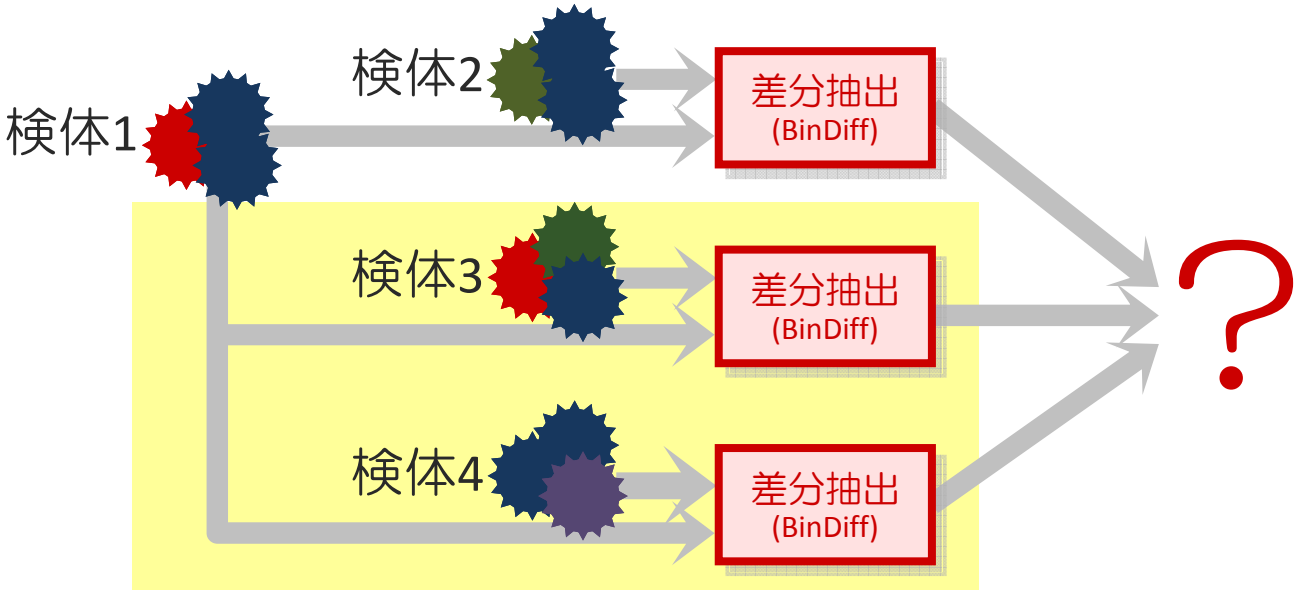
検体	関数
検体1	523
検体2	139

比較結果

関数	個数
検体1, 2に共通	53
検体1にのみ存在	470
検体2にのみ存在	58

評価2: 比較の組合せによる精度向上

♣ 検体1と検体2, 3, 4における比較結果

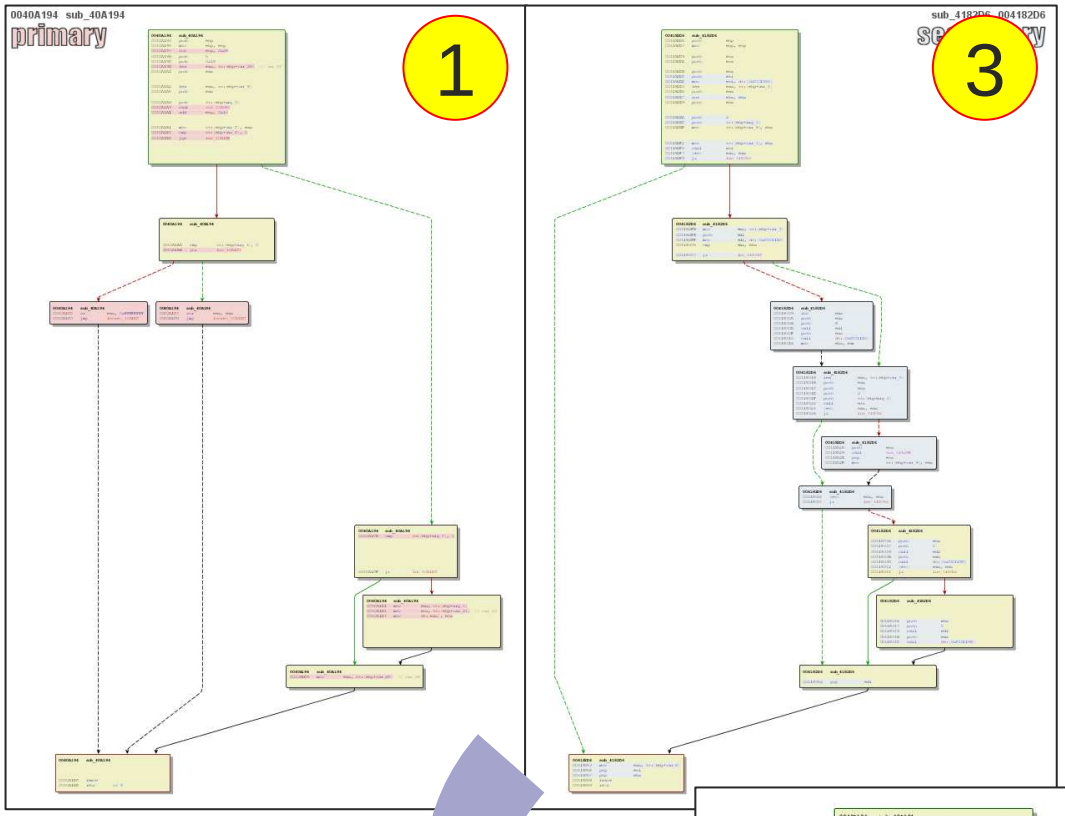


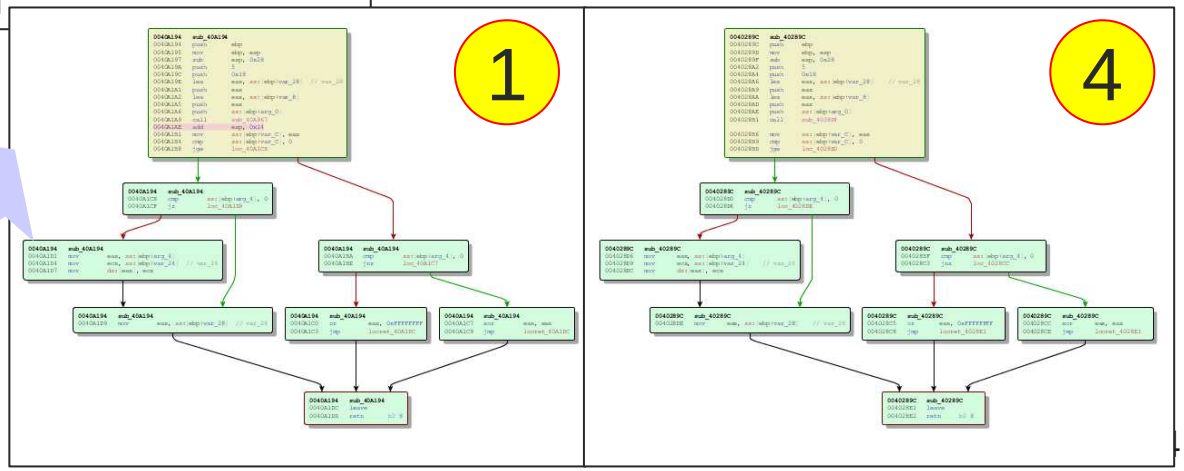
- 135 / 523 = 25.8% が一致
- 複数のマルウェアと比較することで, 解析対象をさらに限定可能

	検体2	検体3	検体4	検体2, 3, 4
検体1と共通する関数	53	78	85	135

評価2: 比較の組合せによる精度向上

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY





より一致率の高い
比較結果が利用可能

まとめと今後の課題

♣まとめ

- 類似マルウェアの解析結果を利用した静的解析支援アーキテクチャを提案
- グラフ構造の比較によるアプローチが有効である
- 複数のマルウェアとの比較により作業をさらに限定可能

♣今後の課題

- SpyEye以外のマルウェアについて評価
- 亜種ではなく全く異なるマルウェア同士の比較
- 逆アセンブリの差分ではなく，本質的に異なる機能を抽出