【MWS（ログ解析）】
2B1-1

企業内ネットワークの通信ログを用いたサイバー攻撃検知システム

○大谷 尚通　北野 美紗　重田 真義

（株）NTTデータ 品質保証部 情報セキュリティ推進室 NTTDATA-CERT

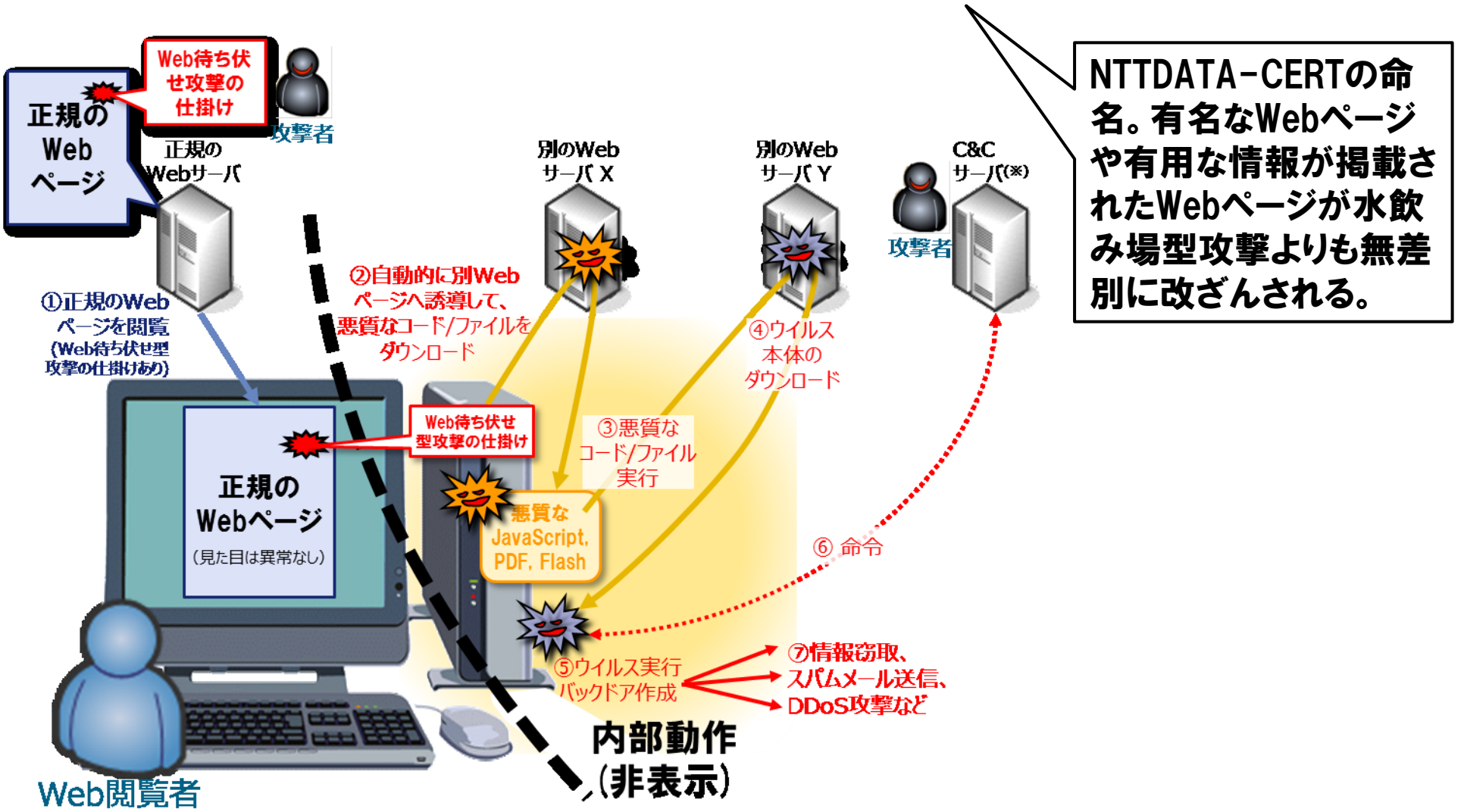
NTT DATA

1. サイバー攻撃の状況

1.1 最新のサイバー攻撃 ～Web待ち伏せ攻撃～

NTT DATA

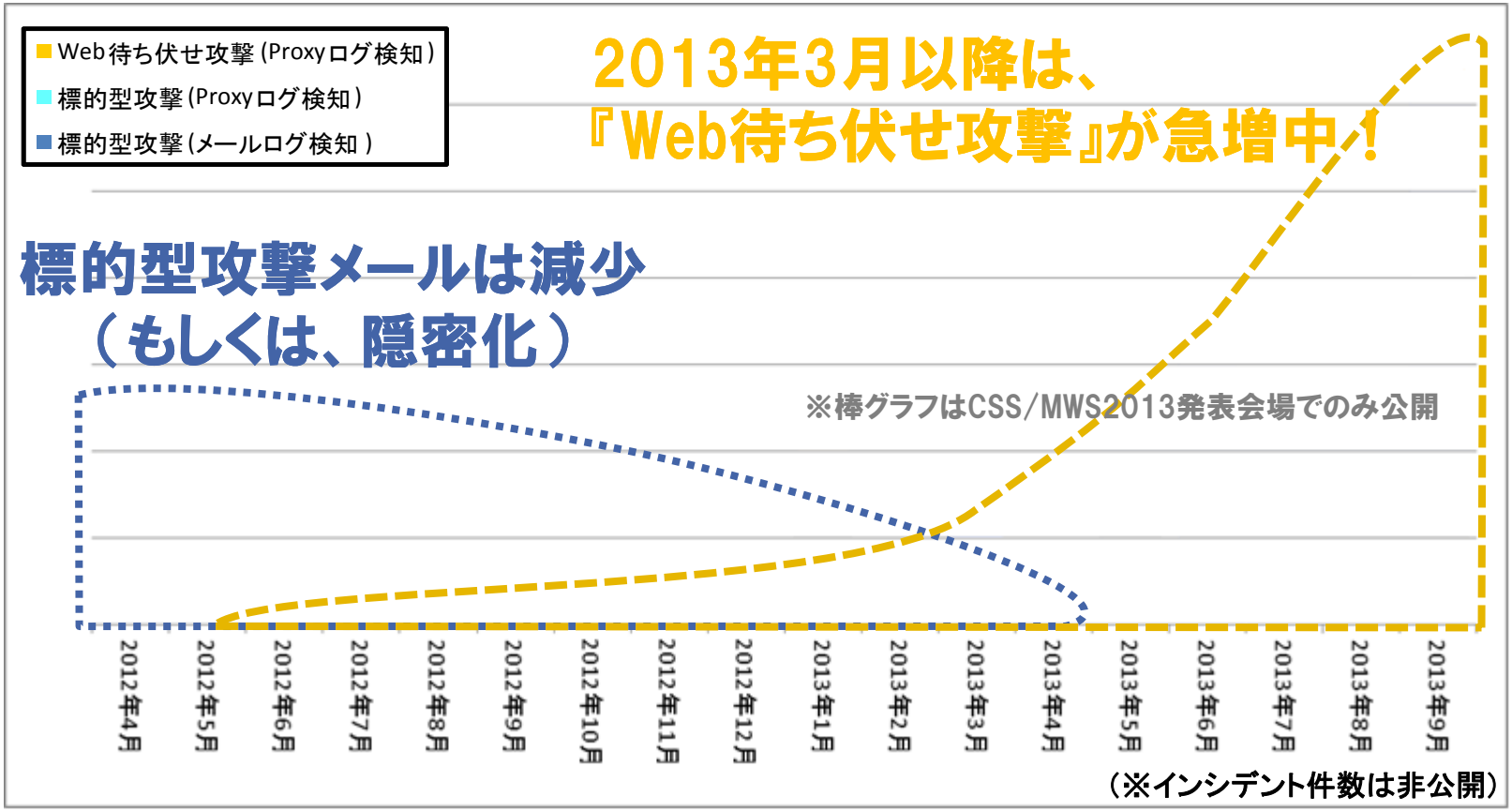
ユーザがアクセスする可能性の高いWebページへ
Drive-By-Download攻撃を仕掛ける<u>Web待ち伏せ攻撃</u>が大量発生。



NTTDATA-CERTの命名。有名なWebページや有用な情報が掲載されたWebページが水飲み場型攻撃よりも無差別に改ざんされる。

1.2 サイバー攻撃の検知状況

サイバー攻撃由来のインシデント数の推移（2012年4月～2013年9月）



Exploit Kitを使った攻撃が流行中 ⇒ 感染の早期検知、早期対応が急務！

1.3 サイバー攻撃対策の問題点

NTT DATA

1. ウイルス対策ソフトの限界
   - 定義ファイルの配布タイムラグ　　→ 1週間以上遅れる
   - 定義ファイル未対応による検知漏れ → 半分以上が検知不能

2. URLフィルタ遮断の限界
   - URLブラックリスト、URLレピュテーションリストが間に合わない

感染の未然防止および早期検知ができない
（感染、被害拡大後に検知）

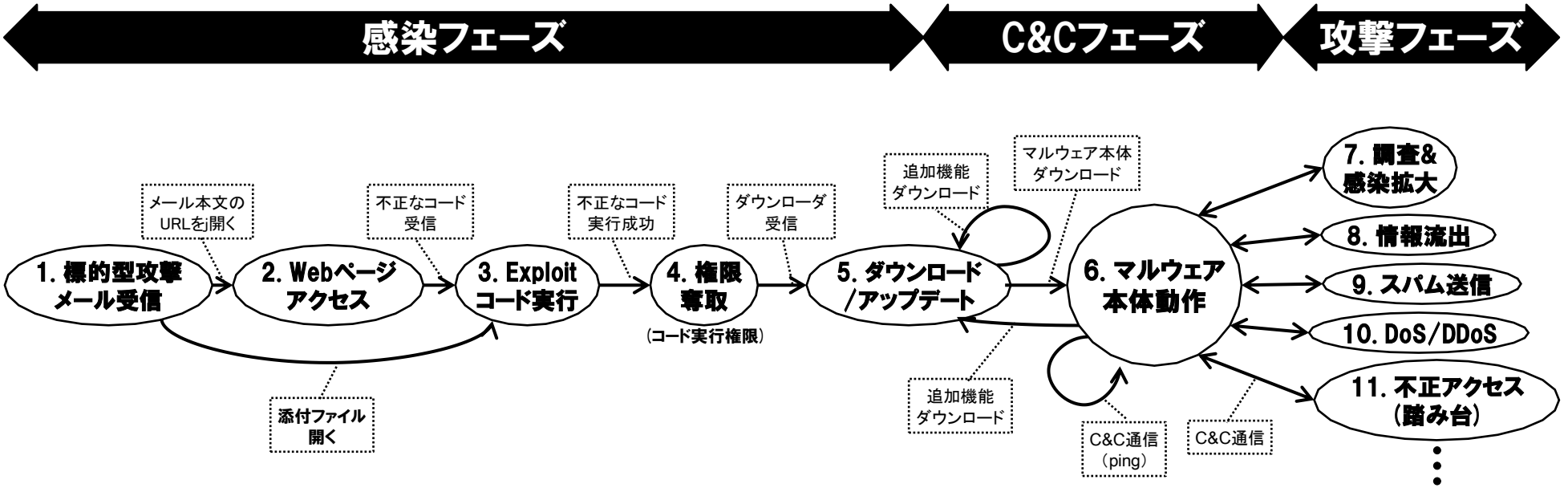
最新のサイバー攻撃によるマルウェア感染を
早期検知できるサイバー攻撃検知システムの開発！

2. サイバー攻撃の分析結果と検知方式の提案

2.1 サイバー攻撃の分析とモデル化（1）

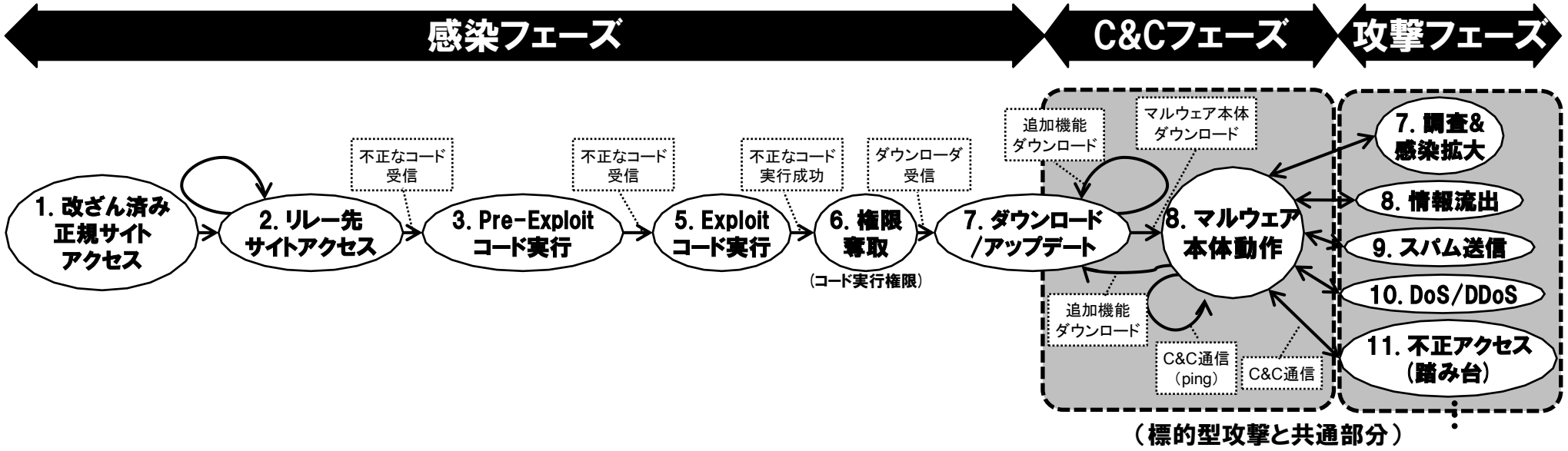
近年のサイバー攻撃は、攻撃動作が複雑化
→ 攻撃手法を解析し、検知方法を検討

標的型攻撃（電子メール）を分析して動作をモデル化



【標的型攻撃（電子メール）の状態遷移モデル】

Web待ち伏せ攻撃を分析して動作をモデル化



【Web待ち伏せ攻撃の状態遷移モデル】

近年の複雑で変化の早いサイバー攻撃の検知方法を提案

2.2 既存の攻撃手法に着目した検知

NTT DATA

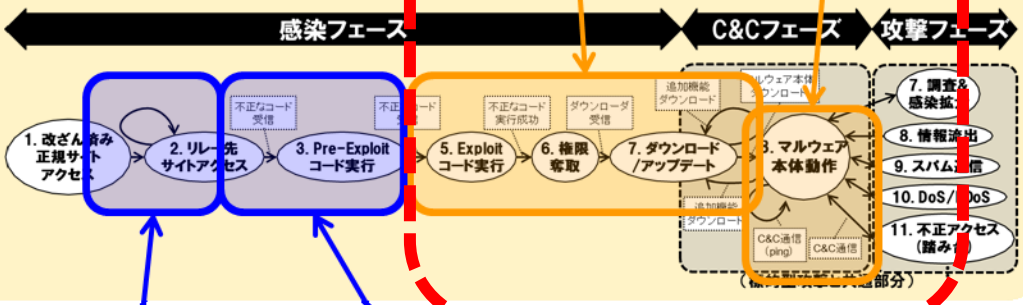
最新の高度化されたサイバー攻撃でも、既存の攻撃手法を持つ場合が多い

高度なサイバー攻撃は動作が複雑で開発コスト/期間が必要。攻撃手法の再利用により解決。

既存の攻撃手法

DBD動作　C&C通信

例）
Web待ち伏せ攻撃
状態遷移モデル



新しい攻撃手法

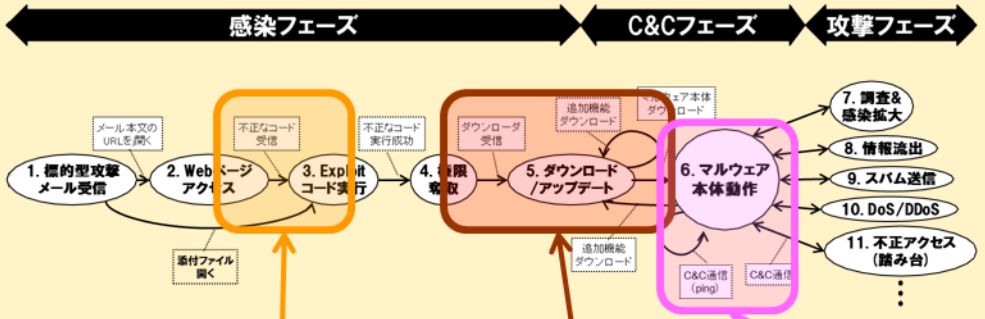
リダイレクト　Pre-Exploit
（脆弱性調査後に
攻撃手法を選択）

仮説）既存の攻撃手法の特徴を使えば、新しいサイバー攻撃も検知可能

2.3 共通する攻撃手法に着目した検知

サイバー攻撃は一部に共通した動作やしくみを持つ場合も多い

例）
標的型攻撃
状態遷移モデル

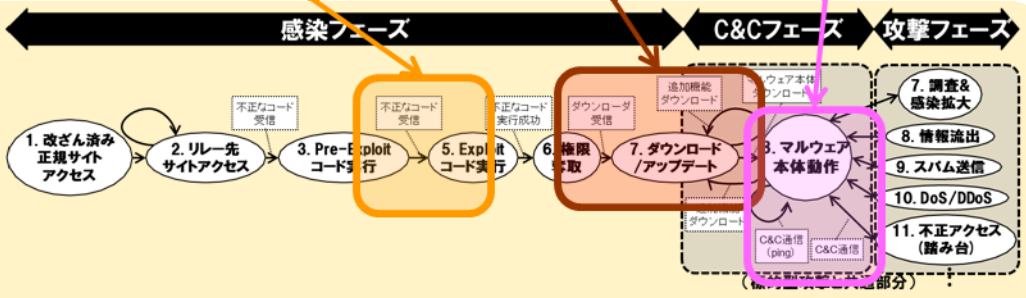


Exploitコード
Web受信

DBD動作

C&C通信

例）
Web待ち伏せ攻撃
状態遷移モデル



攻撃の検知が目的
異常検知でよい
(Anomaly Detection)

※攻撃種類の特定は必須でない

仮説）共通する特徴を使えば、別のサイバー攻撃も検知可能

2.4 検知方法 ～Step1：定性的な特徴～

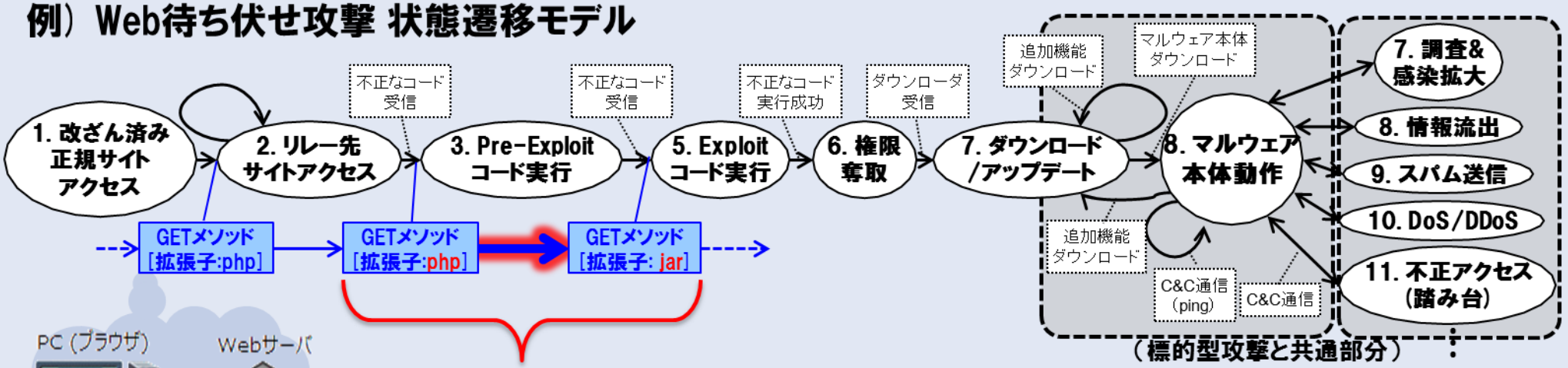
NTT DATA

検知に使用する特徴

- 変化しやすい特徴（例：URL文字列）
  - 変化の要因
    - 設定変更
    - バージョンアップ
    - 新しい攻撃ツール
- 変化にくい特徴 ⇒ “定性的な特徴” “定性的な特徴の遷移”

■ “定性的な特徴”を使った検知

例）Web待ち伏せ攻撃 状態遷移モデル



【定性的な特徴】
GETメソッド（拡張子、引数）の変化やUserAgentの変化など、攻撃動作の大きな変化＝状態変化

➡ 検知

※【定性的】対象の状態を不連続な性質の変化に着目してとらえること。（大辞林 第三版）

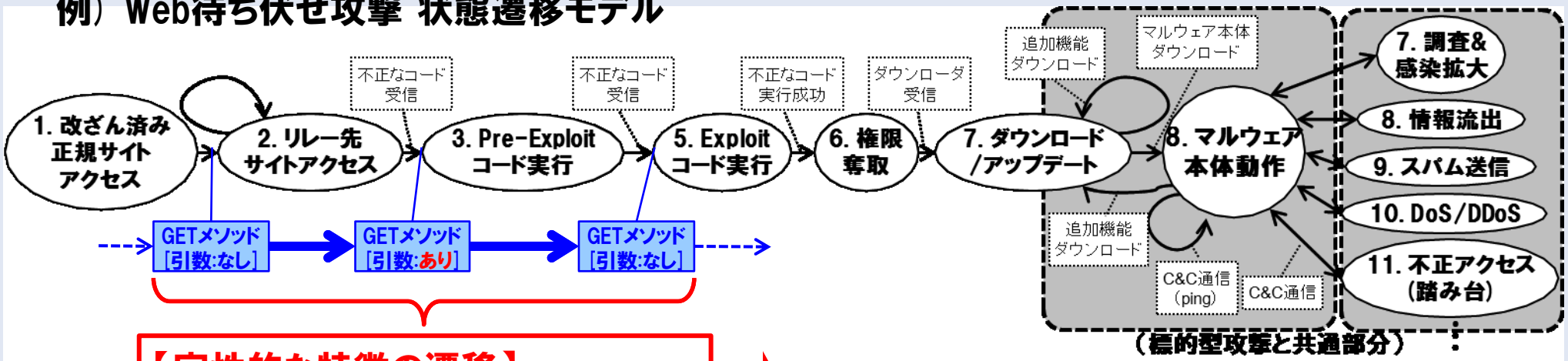
2.4 検知方法 ～Step2：定性的な特徴の遷移～

NTT DATA

“定性的な特徴”を使った場合
⇒ 正常な処理を誤検知(False Positive)する場合がある

■ “定性的な特徴の遷移”を使った検知

例）Web待ち伏せ攻撃 状態遷移モデル



【定性的な特徴の遷移】
定性的な特徴(状態変化)を複数組み合わせて特徴とする

➡ 検知

2.5 実装方式の提案

NTT DATA

当社社内へ導入を考慮して、実装方式を検討

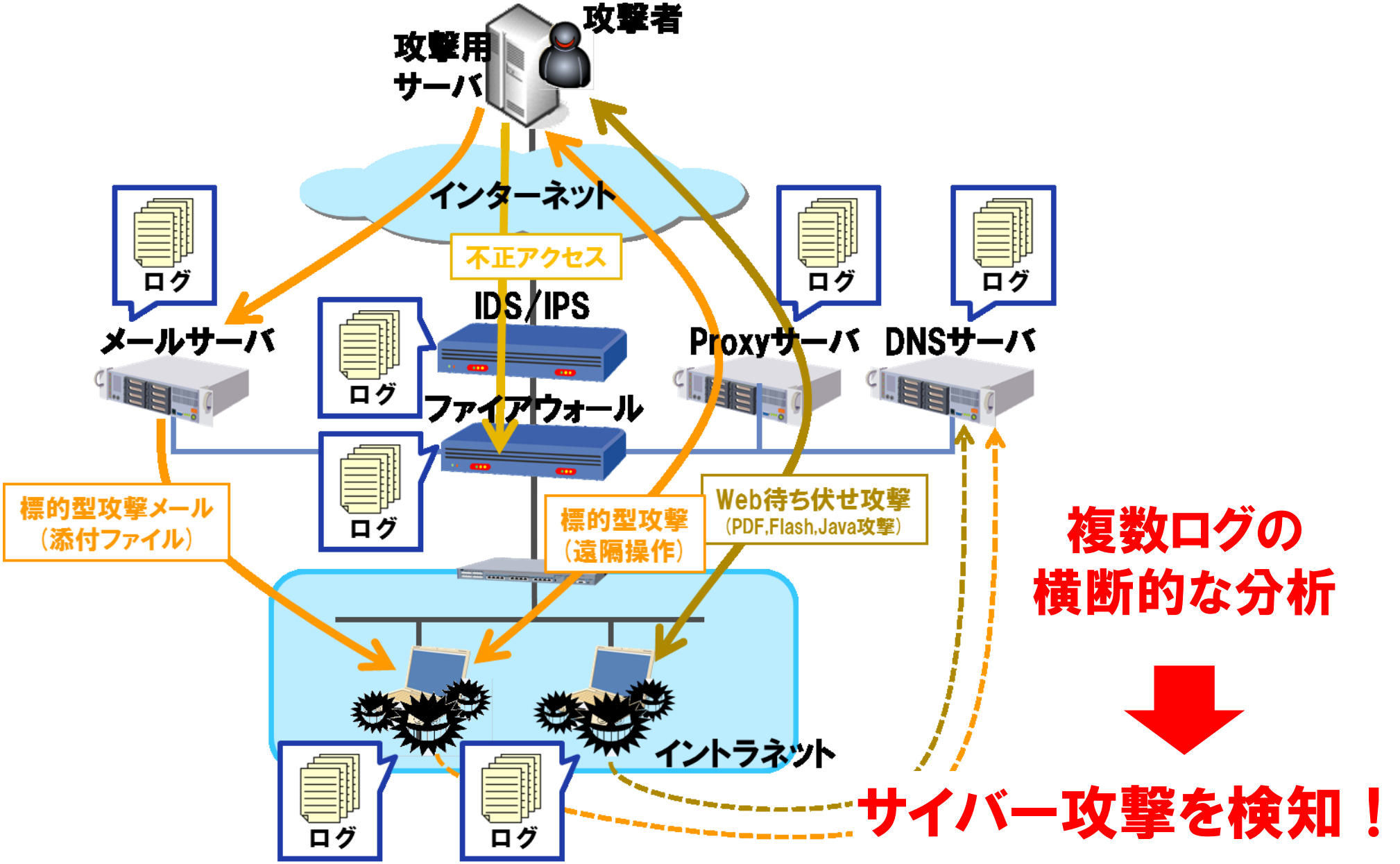
【制約条件】
A)「既存の社内システム(社員/協働者のOA端末含む)への影響が少ないこと」
B)「新規投資する対策コストを押さえること」

【方針】
□ 既存リソース(設置済みセキュリティ機器)の利活用・・・(A)(B)
□ 既存の検知/対策システムに加えて、本システムを追加導入
□ 独自のブラックリストや検知パターンを自社開発
□ 定常監視、および高度分析を合わせた継続的な運用

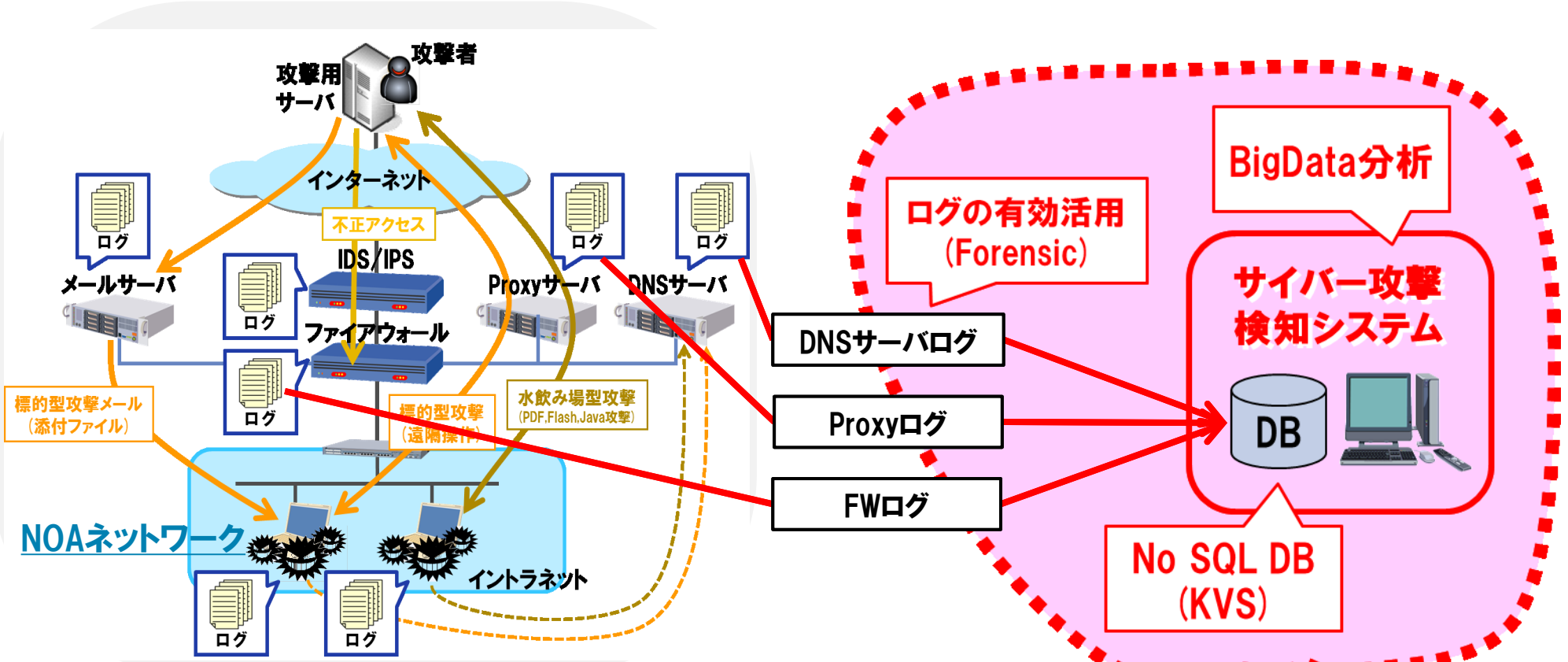
× 通信モニタ方式
× 端末ソフト方式

導入済みのネットワーク製品(DNS, Proxy)やセキュリティ製品(Firewall, IDS/IPS)のログを活用し、ログ上の定性的な特徴から検知する実装方式

2.6 ログを有効利用したサイバー攻撃検知システム



2.7 サイバー攻撃検知システムのアーキテクチャ



複数ログの統合分析の3つの基本処理

- □ 情報の集約（Aggregation）
- □ 正規化（Normalization）
- □ 相関分析（Correlation）　を考慮して設計

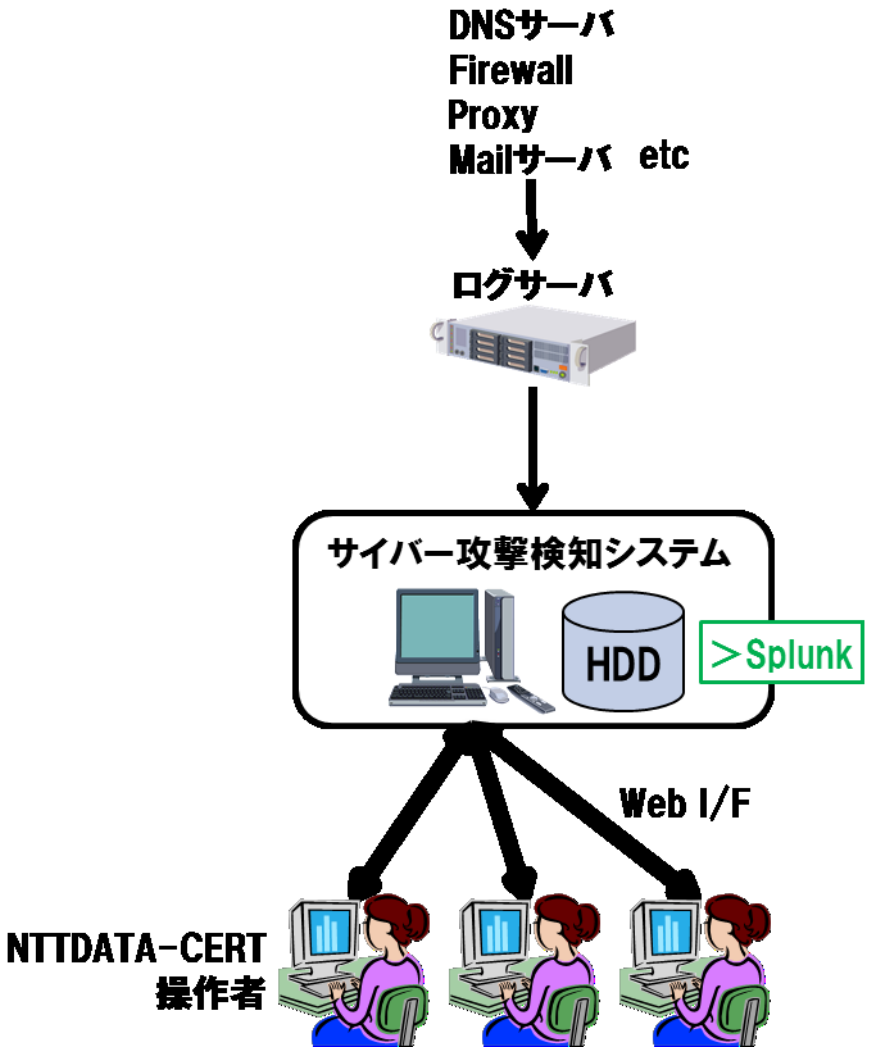
3. サイバー攻撃検知システムの実装と運用実績

3.1 サイバー攻撃検知システム仕様

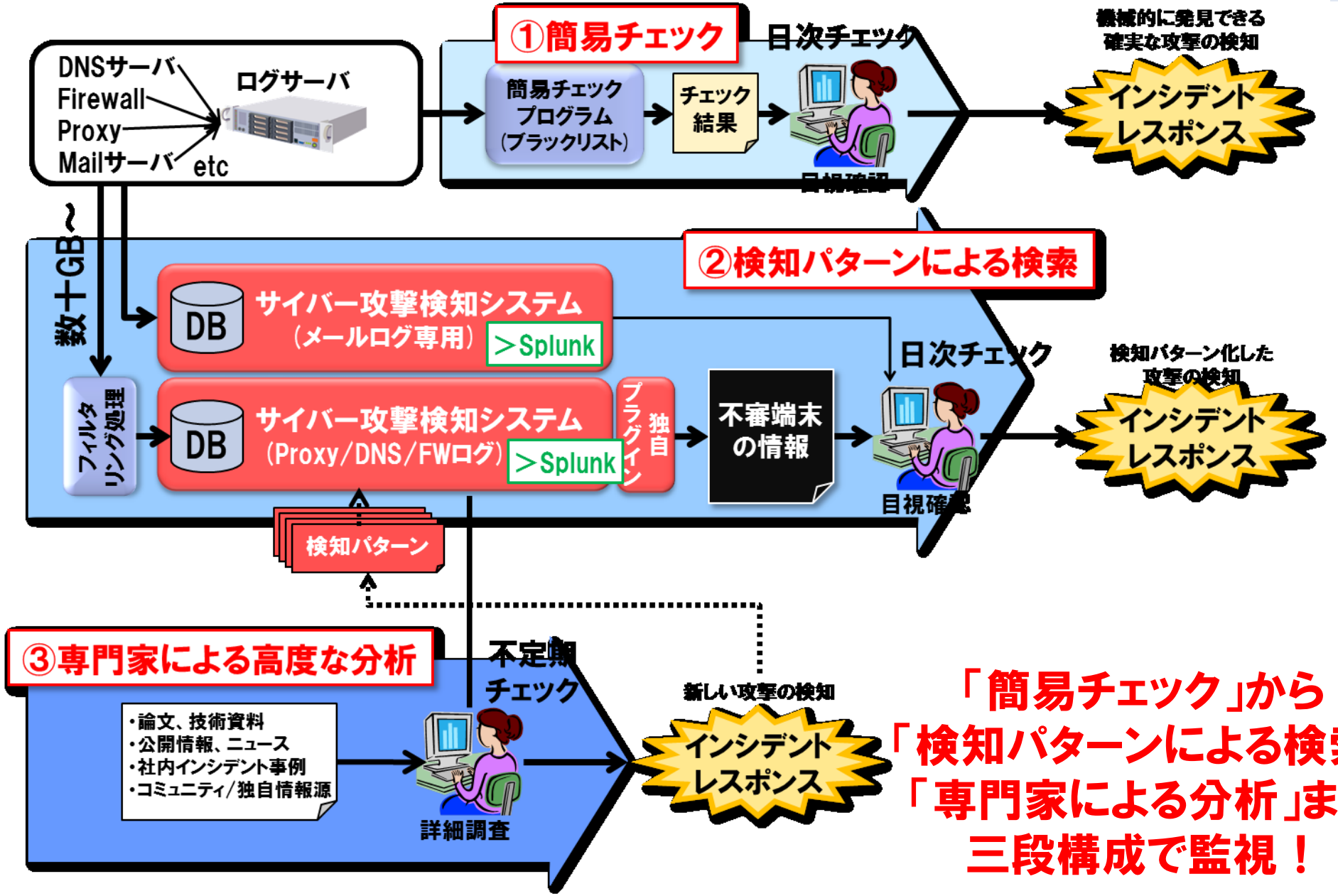
サイバー攻撃検知システム（試作機）のハードウェア/ソフトウェア仕様

■ ハードウェア
- □ CPU Intel Core i7
- □ メモリ 16GB
- □ 外付HDD RAID5 12TB

■ ソフトウェア
- □ OS: Ubuntu 11.10
- □ Splunk ver. 5
- □ ログ自動取得プログラム
- □ フィルタリング処理プログラム
- □ 簡易チェックプログラム



3.2 システム全体の処理フローと監視対応体制



「簡易チェック」から
「検知パターンによる検索」
「専門家による分析」まで
三段構成で監視！

3.3 検知ルール数

独自開発した運用中のSplunk用 (Proxy) の検知ルール

[3A1-1] /MWS (不正通信4)
「Drive-by-Download攻撃における通信の定性的特徴とその遷移を捉えた検知方式」
北野美紗, 大谷尚通, 宮本久仁男

Exploit Kitの調査

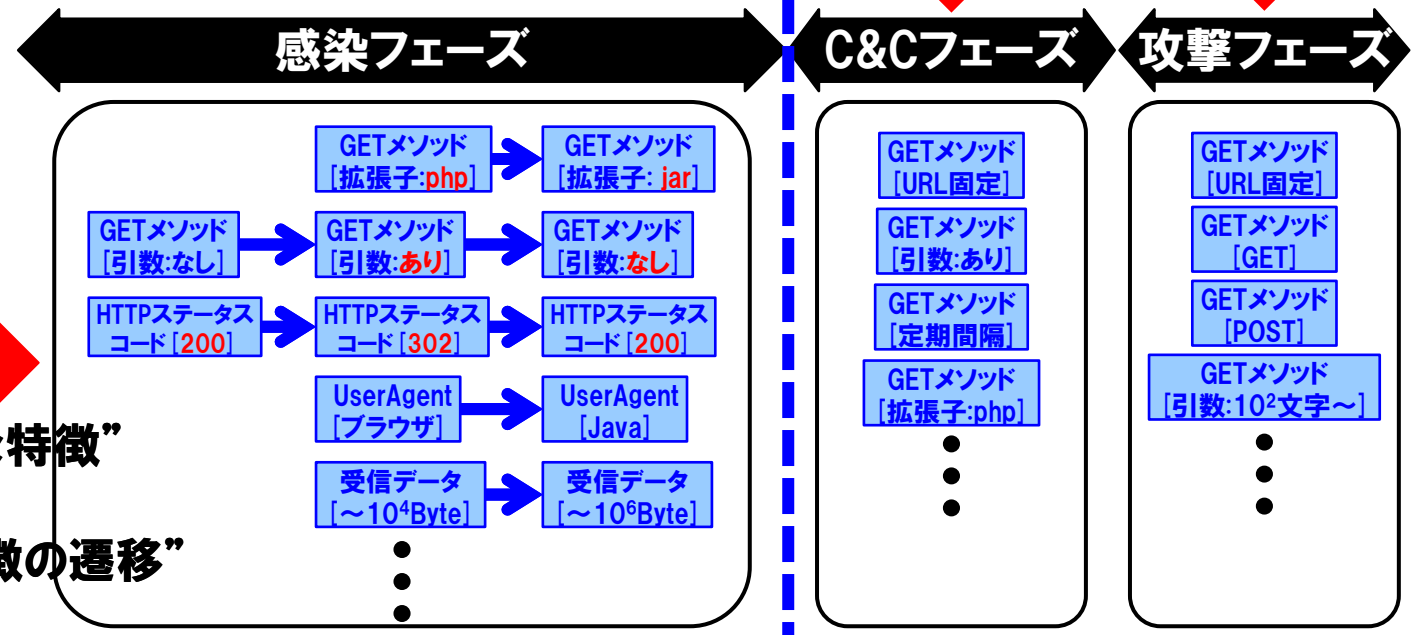
BlackHole ExploitKit, RedKit Exploit Kit
Neutrino ExploitKit, Glazunov ExploitKit
Sakura ExploitKit等 ・・・計8種類

RATの調査

PoisonIvy, Xtreme RAT, Cybergate RAT,
DarkComet RAT, uBOT, Zeus, Spyeye,
Mirkov4, BlackEnergy RAT等・・・14種類

“定性的な特徴”

“定性的な特徴”
＋
“定性的な特徴の遷移”



その他
文献
調査等

検知パターン 32個
(特徴を組み合わせたルール含む)

検知パターン
17個

検知パターン
7個

検知パターン
23個

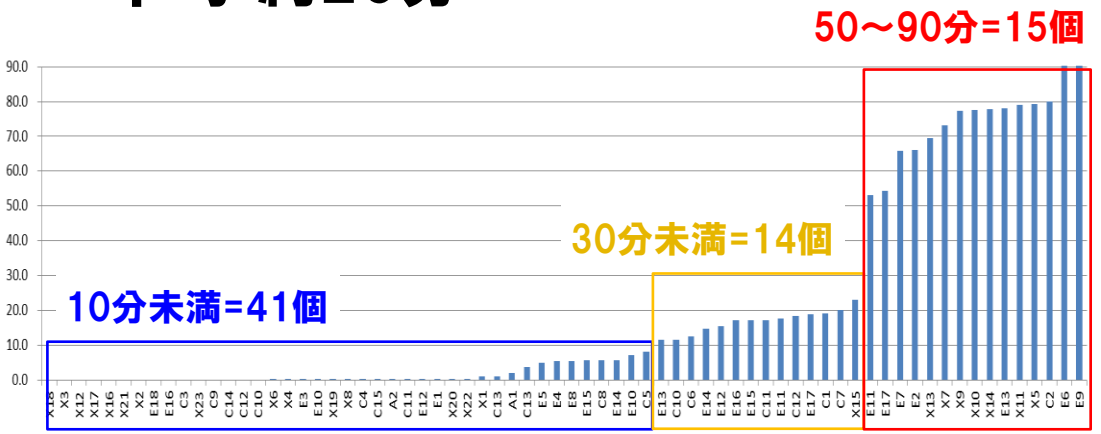
3.4 処理性能と検知実績

■ 処理性能　（平日のログを処理した場合）

□処理ログ行数 ＝ 約11.6×10^6 行/日

□1パターンあたりの処理時間 ＝ 平均 約20分

⇒ 実質5時間程度
※検知パターン70個は逐次並列処理



【検索パターン1個あたりの処理時間（分）】

■ 検知実績

□標的型攻撃メールに感染したオフィスPCの検知（C&C通信）
□Web待ち伏せ攻撃に感染したオフィスPCの検知（PreExploit,Exploit通信）

□ウイルス対策ソフトをすり抜けた標的型攻撃メール/添付ファイルの検知
（件名，差出人などの文字列）

4. まとめ

4.1 まとめ

NTT DATA

- 定性的な特徴とその遷移を用いた検知方式、および既存製品のログを活用し、検知する実装方式を提案した。
- サイバー攻撃検知システムを実装し、定常的に運用できることを確認した。

■ 新しいサイバー攻撃の早期検知・早期対応

- ➢ウイルス対策ソフトでは検知できない/遅れるインシデントを検知
- ➢ユーザが気づかないインシデントの検知（スパイ活動系のウイルス）
- ➢予防が困難で感染してしまうインシデントの検知（標的型攻撃メール、Web待ち伏せ攻撃など）

＋インシデント報告を受けて対応する受動的な体制から能動的な対応へ

■ 組織内に点在する様々なログを有効活用

＋既存システムに大きな影響を与えずに導入可能

4.2 課題と今後の予定

NTT DATA

□検知精度の向上

感染フェーズに特徴がない攻撃は検知できない。
→C&Cフェーズや攻撃フェーズで検知できる検知パターンを追加開発
→Firewall, DNS, IDS/IPS などの他のログとの相関分析検知パターンの開発

□スケールアウト構成

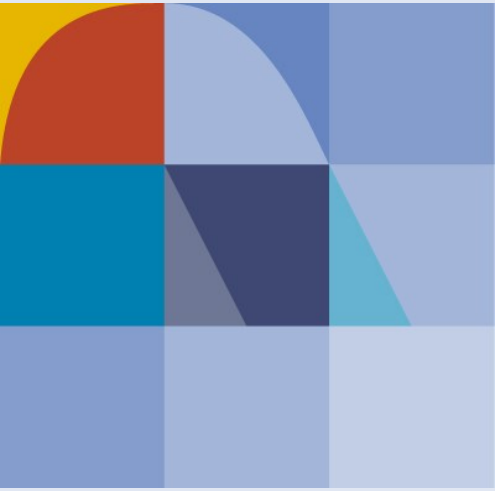
検知パターン数やログ量の増加に伴い検索処理時間が増加
日次の検索処理と検索結果のチェックが1日以内に完了できなくなる恐れ
→データベースのNAS配置、複数台のPCからの同時検索構成（スケールアウト構成）の導入

□スコア処理の高度化

スコア処理を高度化し、毎日の誤検知のチェック作業の工数を削減

□統計分析および機械学習

システムへ蓄積された大量のデータを有効利用し，統計分析や機械学習を応用した
検知方式を開発



Global IT Innovator

NTT DATA Group

NTT DATA