1A3-3: MWS ドライブ・バイ・ダウンロード

企業での実環境を考慮したサイバー攻撃検知システムの有効性評価

2015年10月21日

株式会社ＮＴＴデータ
○ 重田　真義　　大嶋 真一　　大谷 尚通

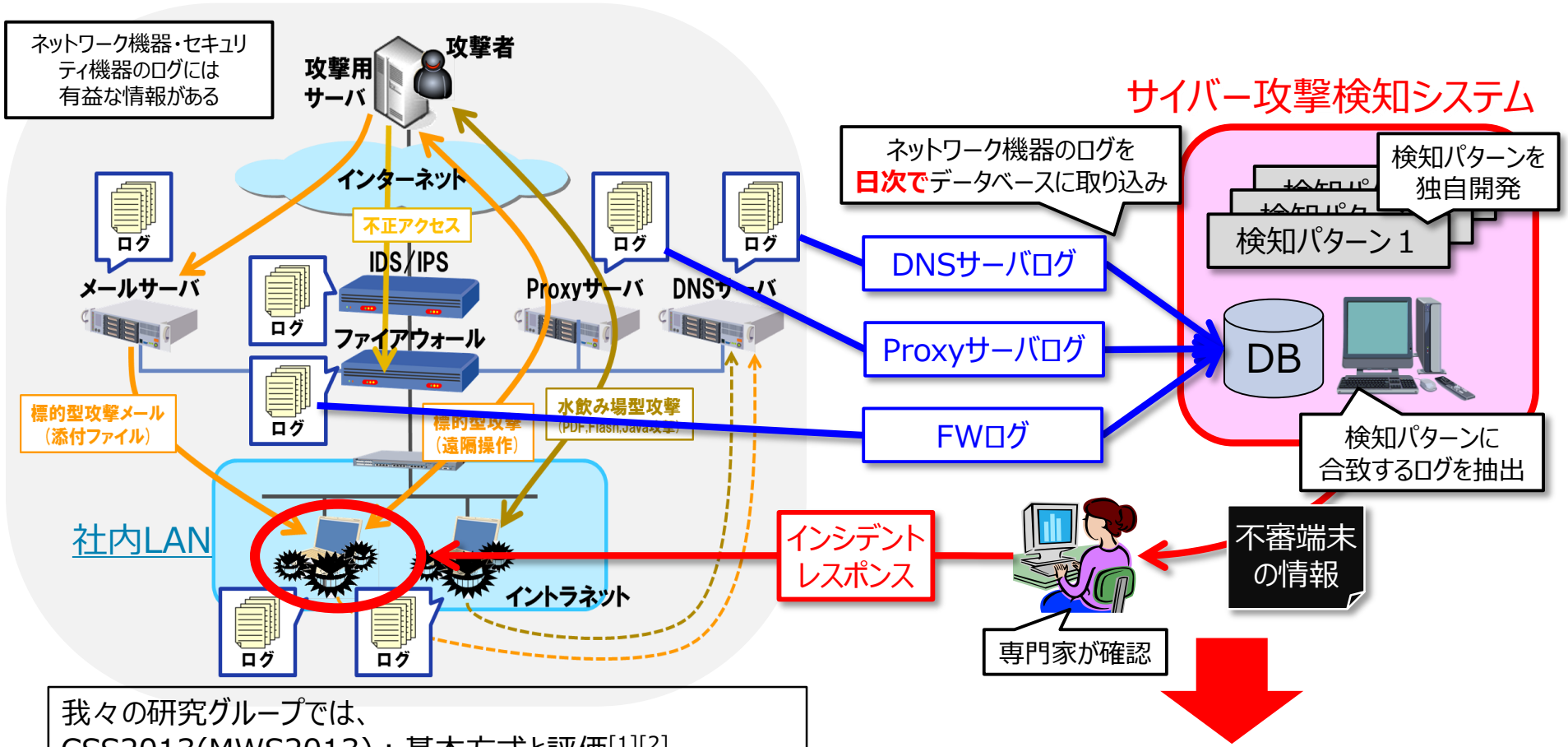
NTT DATA

目次

1. サイバー攻撃検知システムとは

NTT DATA

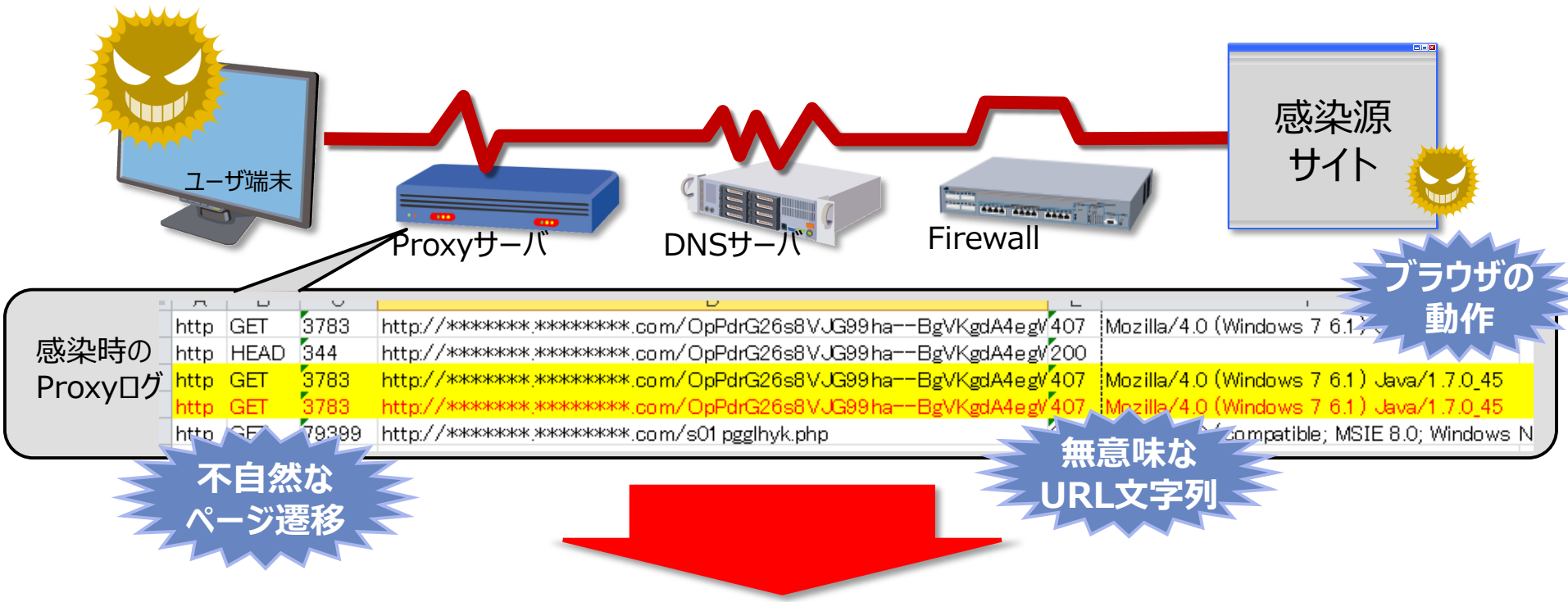
設置済みのネットワーク機器の通信ログを有効利用してマルウェア感染端末を検知



我々の研究グループでは、
CSS2013(MWS2013)：基本方式と評価[1][2]
CSS2014(MWS2014)：検知範囲拡大の取組と評価[3]
を発表してきた。

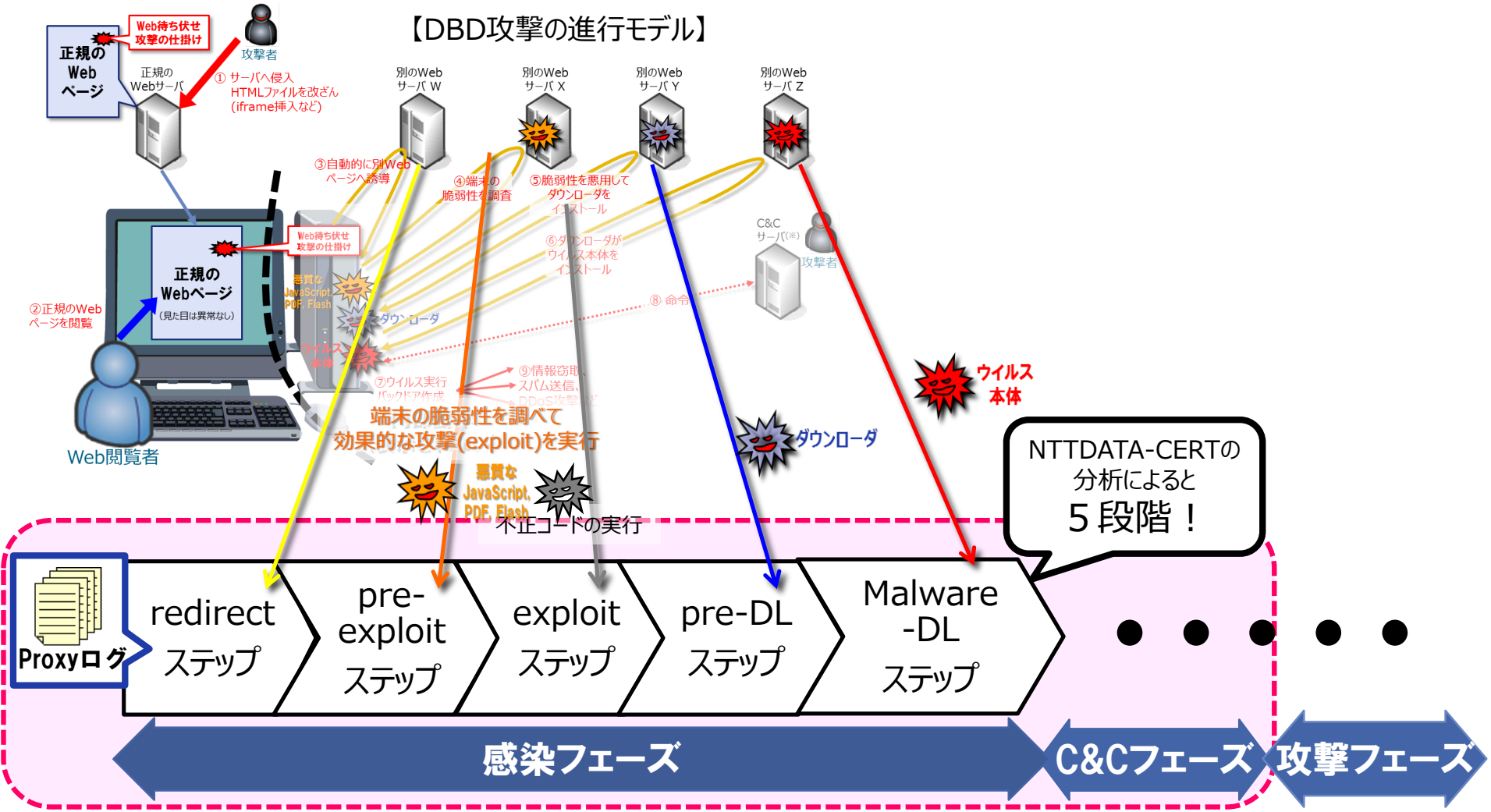
[1] 北野 美紗, 大谷 尚通, 宮本 久仁男, Drive-by-Download攻撃における通信の定性的特徴とその遷移を捉えた検知方式, MWS2013.
[2] 大谷 尚通, 北野 美紗, 重田　真義, 企業内ネットワークの通信ログを用いたサイバー攻撃検知システム, MWS2013.
[3] 大谷 尚通, 益子 博貴, 重田　真義, 実環境におけるサイバー攻撃検知システムの有効性評価および検知範囲の拡大に向けた検討, MWS2014.

マルウェア感染時/感染後は、正常な通信とは異なる特徴的な通信が発生



通信ログに残った特徴的な通信の痕跡をもとに
マルウェアに感染している端末を検知する

Drive-by-Download攻撃(DbD攻撃)の**感染フェーズに現れる定性的な特徴の遷移**や、感染後の**C&Cフェーズの特徴**を用いて検知



2. 我々が提案したシステムにおける運用上の問題

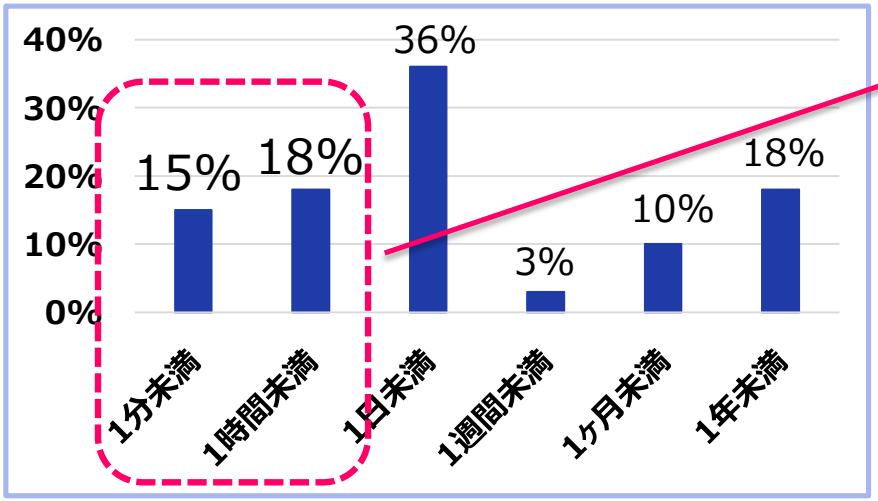
我々が提案したシステムを運用する上で、下記の2点の問題が存在する。

問題点①：検知率の維持・向上

→ 1A3-4「Exploit Kitの変化への適応を目的としたサイバー攻撃検知システムの改良」にて、問題点①に関わる取り組みとその成果を説明　→　**次の時間枠で発表する**

問題点②：検知タイムラグ

→ 提案したシステムでは、実環境の通信ログを1日1回まとめて収集・分析する方式であったため、1日分の通信ログをその翌日に分析する運用では、マルウェアに感染してから検知するまでに最大24時間経過してしまう恐れがある。　→　**本発表で扱う**



インシデントの**33%**が、最初の侵害から1時間以内に、機密情報が取り出されていたとの報告がある。

リアルタイムにDbD攻撃を検知しなければならない

最初の侵害から機密情報が取り出されるまでにかかる時間[4]

[4] ベライゾンジャパン,2013年度データ漏洩/侵害調査報告書, https://www.verizonenterprise.com/jp/DBIR/2013/, accessed Aug 24,2015.

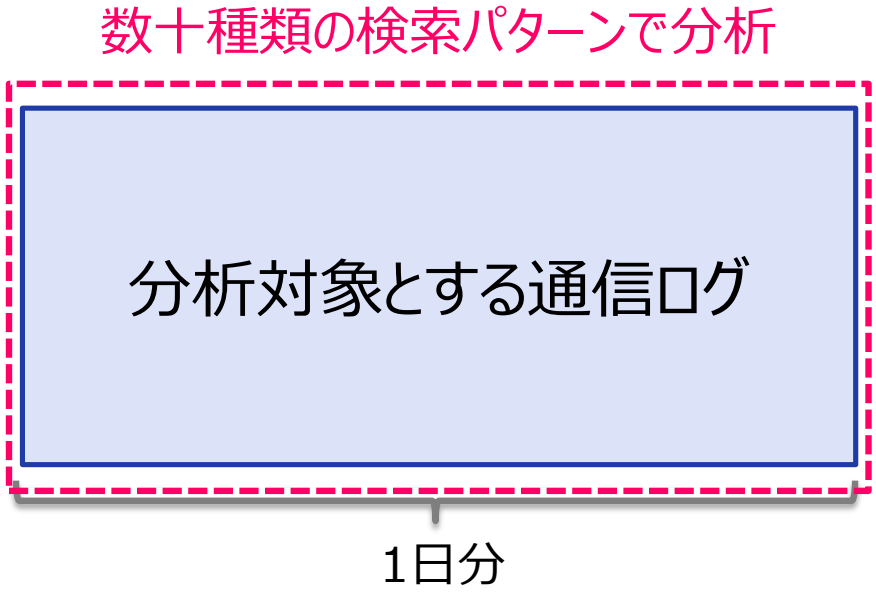
本研究では、検知タイムラグの問題の解決に取り組み、
目標として、下記を設定した。

目標	ネットワーク機器の通信ログをリアルタイムに分析することで、DbD攻撃によるマルウェアに感染した端末を**できる限り早期に検知**する

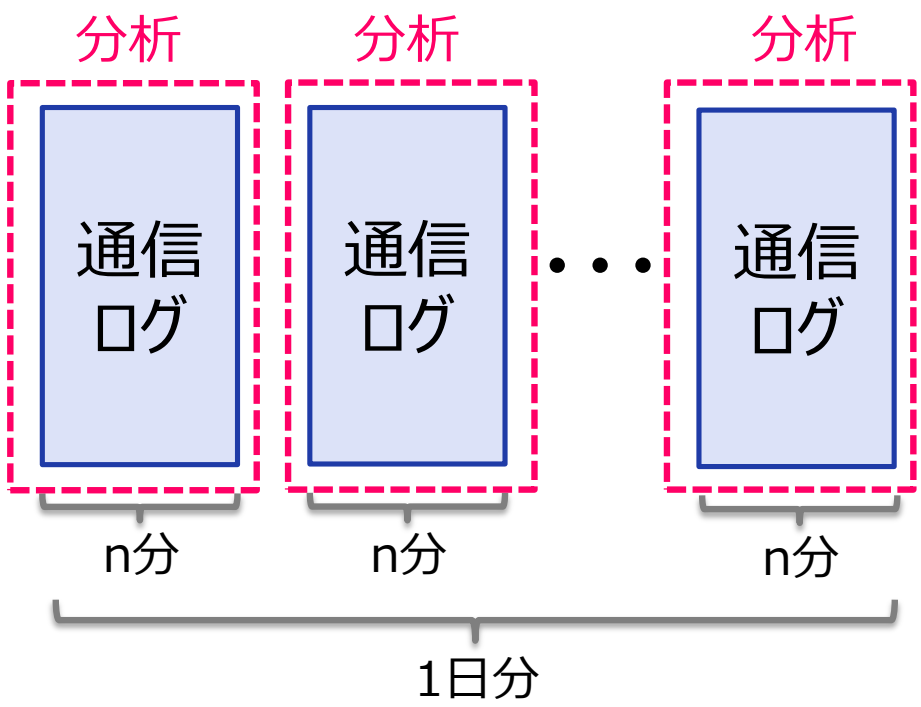
3. 実現方式の検討

リアルタイムログ分析方式を実現するに当たって、
ログを単位時間あたりで分割し、分析する方式を考える。

■ 以前の方式(日次版ログ分析方式)



■ 今回の方式(リアルタイムログ分析方式)

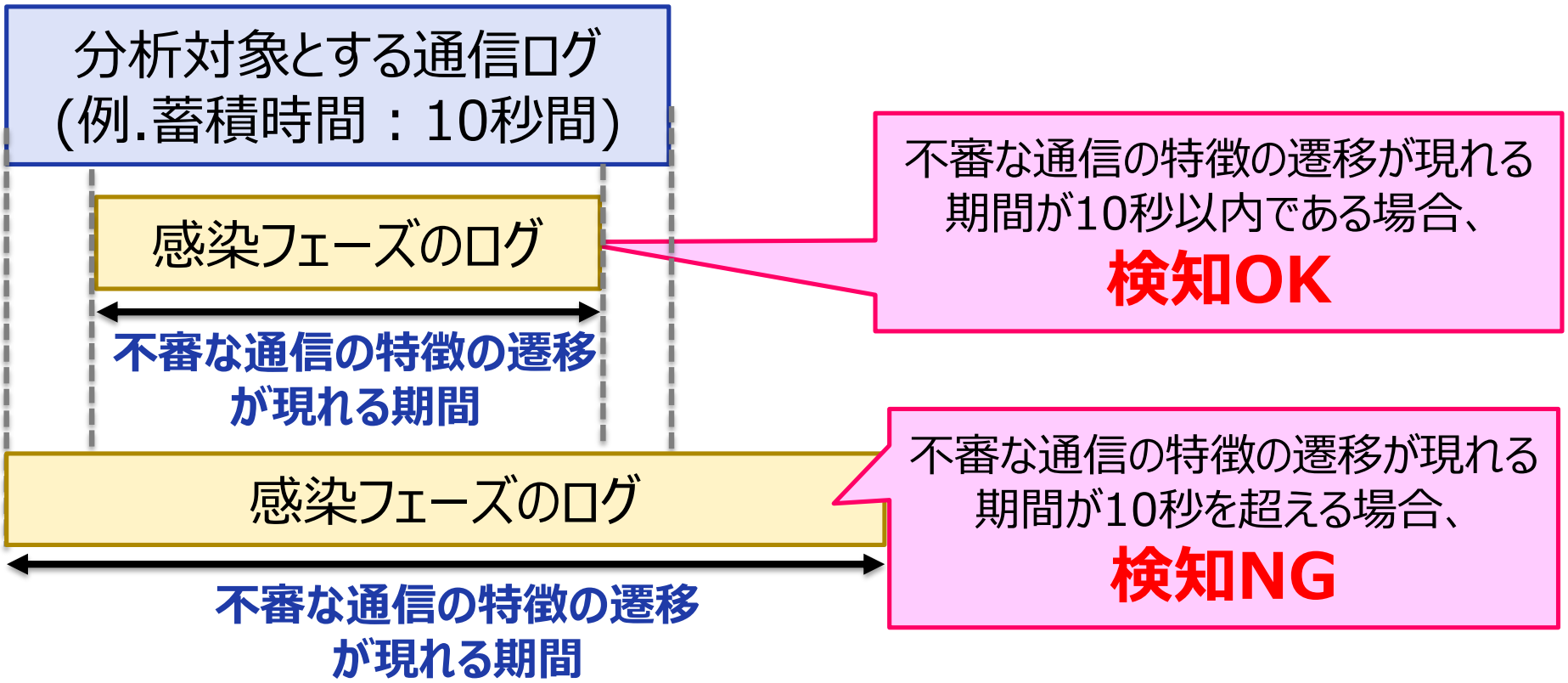


0 < n ≦ 720

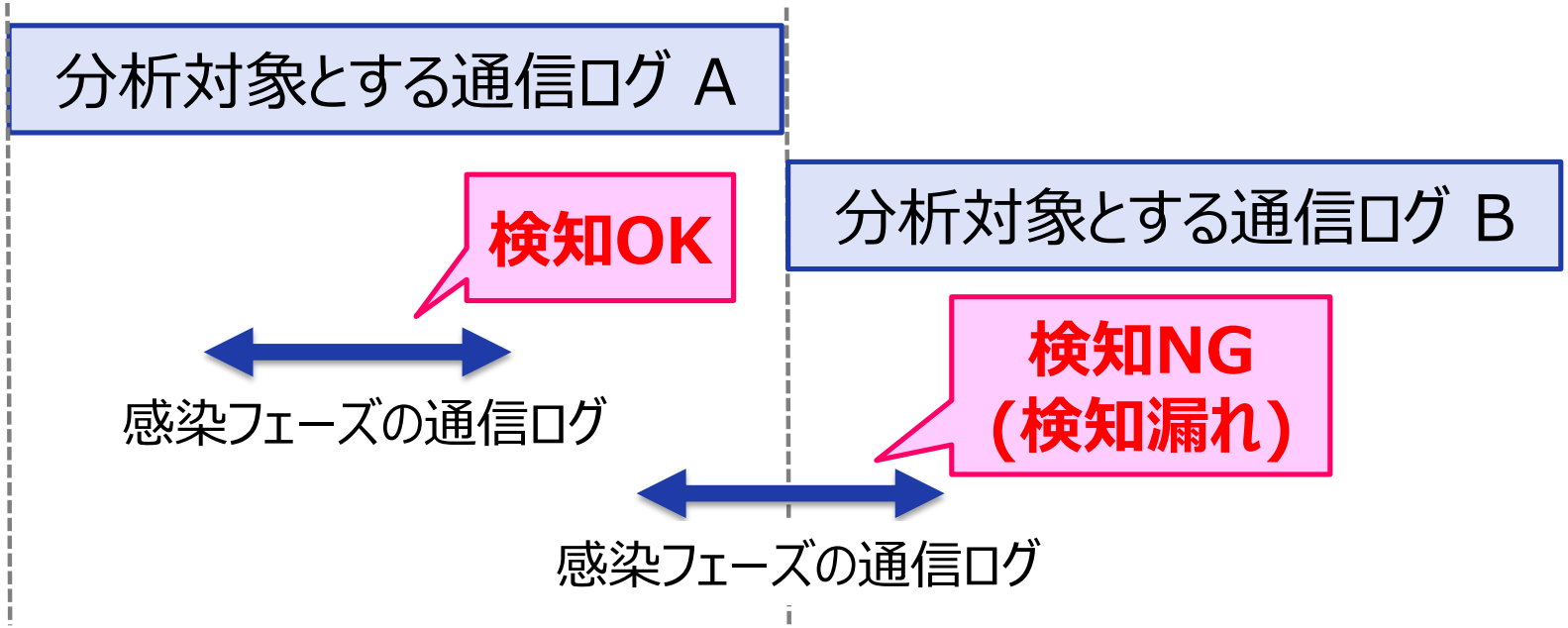
リアルタイム検知処理の課題は、3点存在する。

✓ 課題①：通信ログの蓄積時間と検知漏れ

分析対象とする通信ログの蓄積時間は、リアルタイム性を考慮すると、**できる限り短い方が望ましい**が、検知システムでは、DbD攻撃の感染フェーズに現れる定性的な**特徴の遷移**をもとに検知しているため、ある程度の時間はログを蓄積する必要がある。



✓ 課題②：通信ログの検索方式による検知漏れ



✓ 課題③：検索処理の遅延

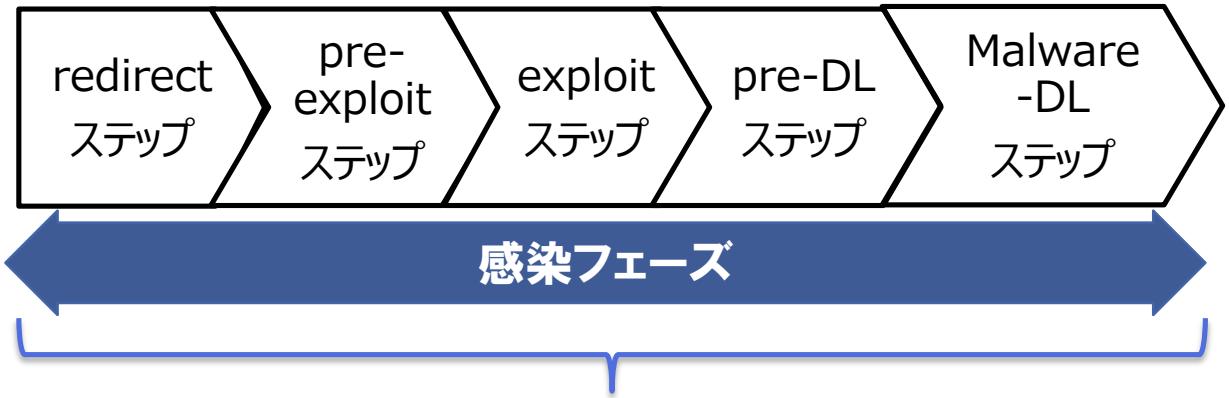
・リアルタイム化した場合においては、特に、**トラフィックのピーク時間帯**でも、最低限、**通信ログの蓄積時間以内に検索処理を完了**させる必要がある。

4. 各課題に対する対応方針

検知パターンによる検知漏れがない、かつリアルタイム性が高い
適切な通信ログの蓄積時間を決定する

約2年間※におよぶDbD攻撃 **約200件の通信ログを分析**した
※対象ログの取得期間：2013年03月 ～ 2015年03月

■ 分析方法



redirectステップからmalware-DLステップが終了するまでにかかる時間を計測

Exploit Kitごとに、最大遷移時間、最小遷移時間を測定した。

■ 分析例 (Nuclear Exploit Kitと思われる事例)

[2015/10/21 13:55:00] 10.10.10.10 GET 29723 http://example.com 200
※改ざんされているWebページへのアクセス

[2015/10/21 13:55:04] 10.10.10.10 GET 640 http://xxxxxxxxx.xx.lt/blog.php?id=SHIGsasd368das8Uaus 200
※不審なWebページへの誘導

[2015/10/21 13:55:05] 10.10.10.10 GET 54893 http://aa.black**list.com/S2sssaIBHgUsGA8UIGEU.html 200
※攻撃者が用意したWebサイトへの誘導

[2015/10/21 13:55:08] 10.10.10.10 GET 40104 http://aa.black**list.com/SAUI12sssaIBHUIOAa8UIGEU 200
※Exploit:クライアント端末の脆弱性を攻撃

[2015/10/21 13:55:16] 10.10.10.10 GET 320949 http://aa.black**list.com/7AsfWss989DshHHIH7ug9g 200
※マルウェアをダウンロード

redirectステップから、malware-DLステップまでの遷移にかかる時間

12秒

NTT DATA

分析結果

Exploit Kit(EK)名	最小遷移時間	最大遷移時間
Blackhole EK	4秒	63秒
Angler EK	18秒	50秒
Cool EK	30秒	30秒
Fiesta EK	18秒	28秒
Gongda EK	8秒	34秒
Goon EK	18秒	41秒
Neutrino EK	4秒	64秒
Nuclear EK	26秒	47秒
Redkit EK	13秒	13秒
RIG EK	21秒	52秒
SweetOrange EK	13秒	36秒
EK不明	4秒	**69秒**

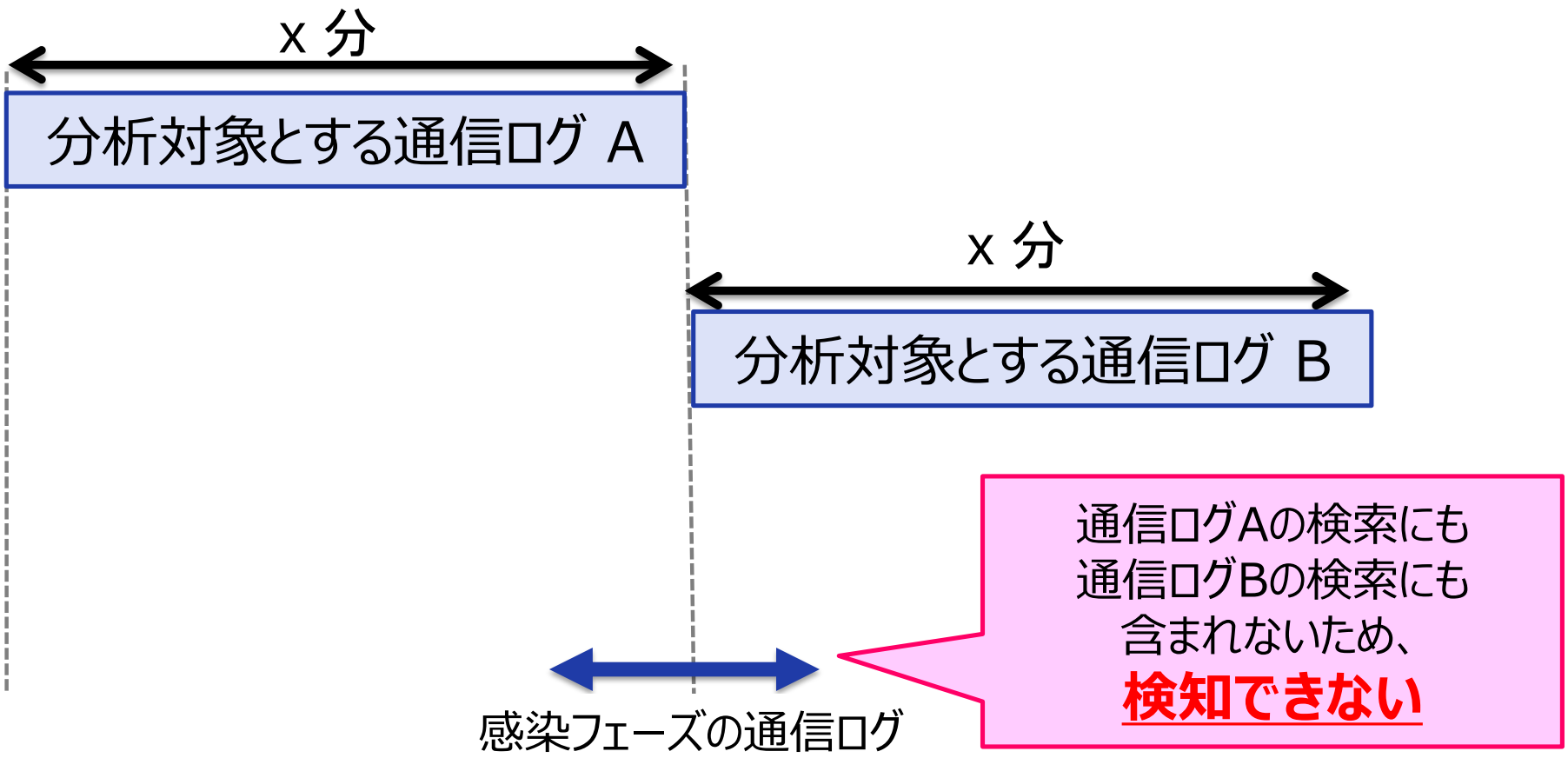
少なくとも分析対象とする通信ログは、**69秒以上**に設定する必要があることが判明

通信ログの蓄積時間を**2分以上(120秒以上)**に設定

※Exploit Kit名は、通信ログをもとに推測したもの
分析した通信ログの中には、Exploit Kitによる攻撃が途中で失敗したものも存在するが、その場合は、Exploitステップやpre-DLステップまでの時間を記載している。

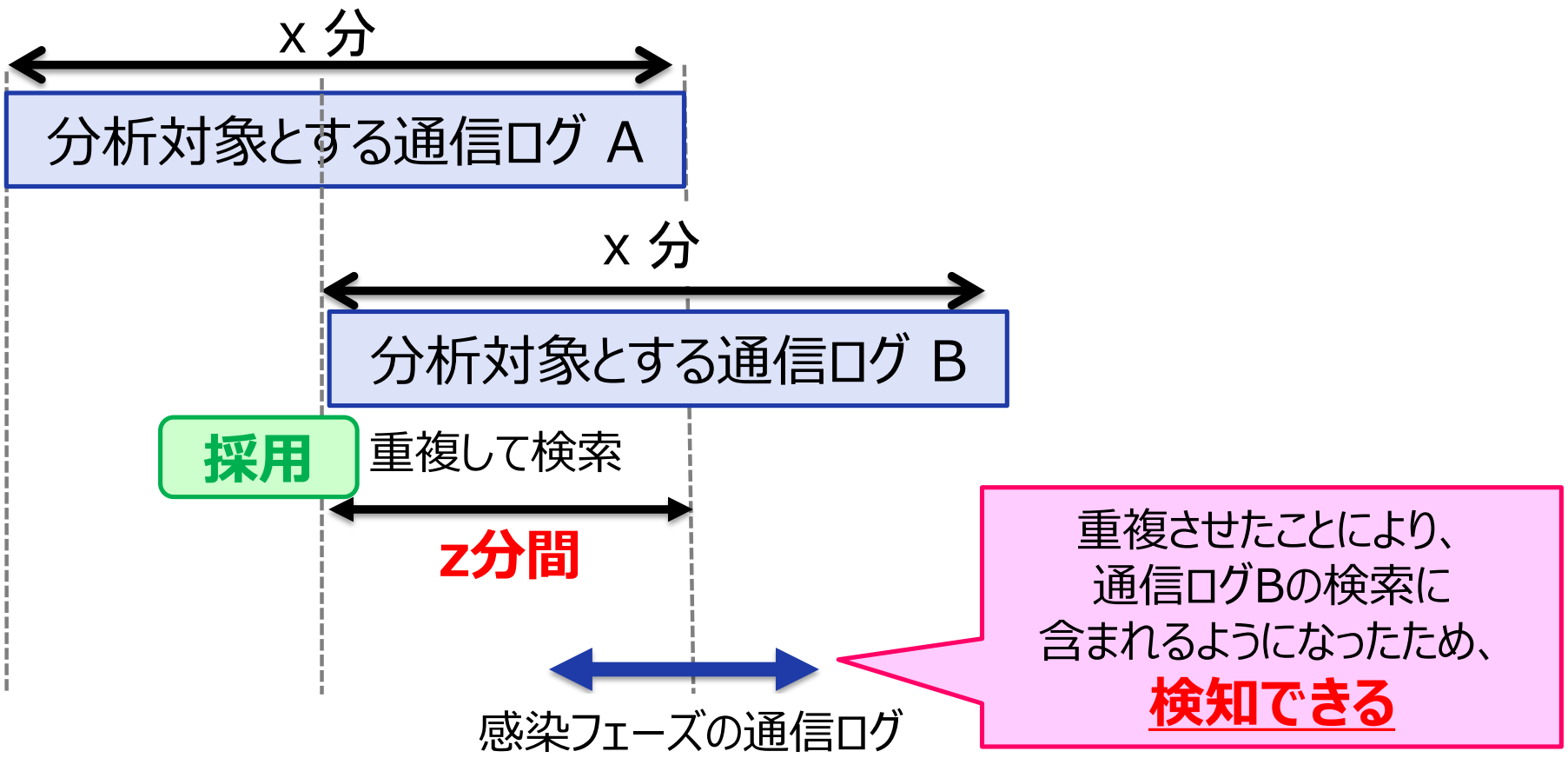
4.2 課題②への対応：通信ログの検索方式の工夫(1/4)

感染フェーズの通信ログが、検索対象ログにまたがって検知できない問題を
解決するために、検索対象ログを一定時間(z分間)重ねて検索する方式を採用した



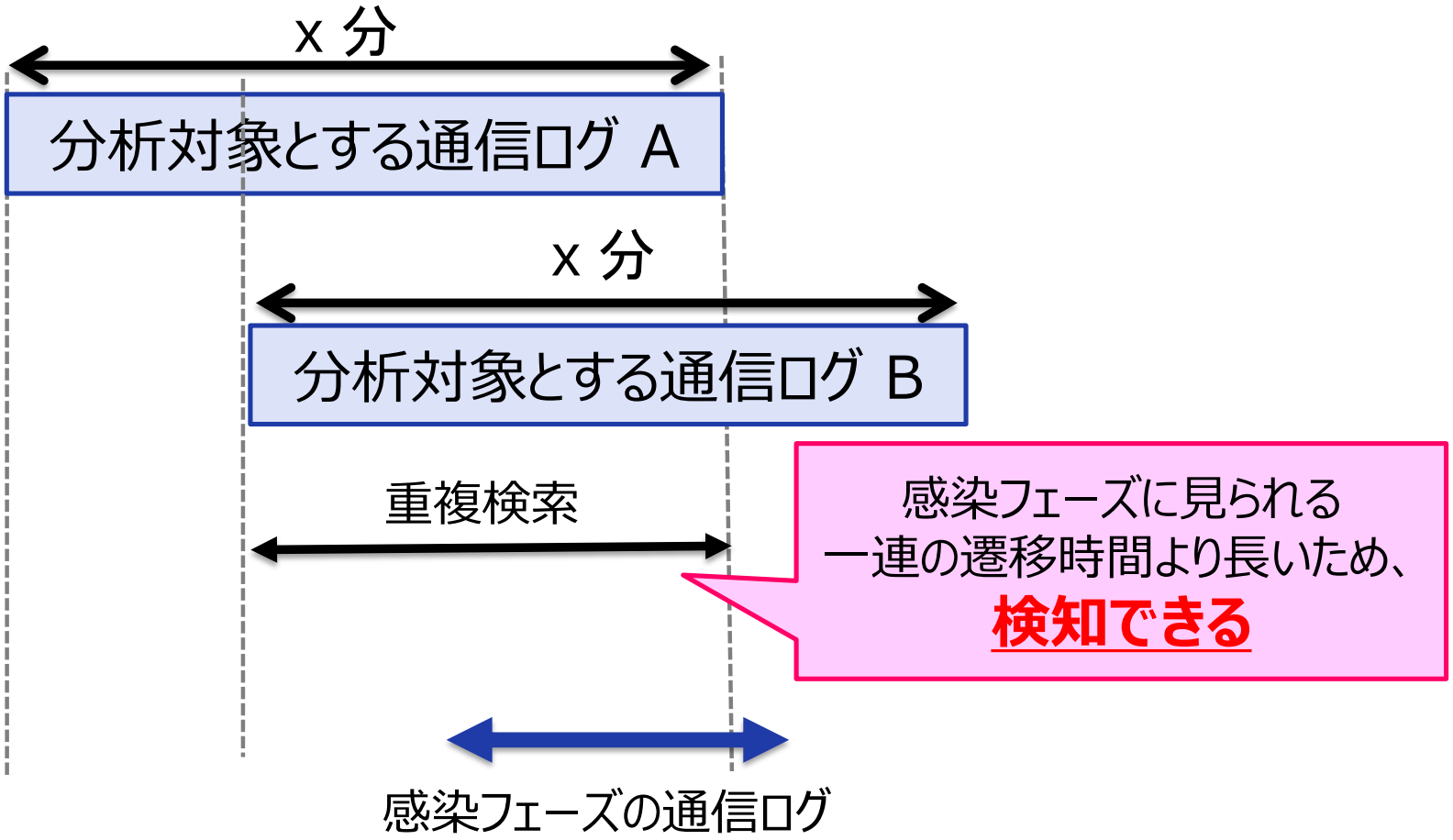
4.2 課題②への対応：通信ログの検索方式の工夫(2/4)

感染フェーズの通信ログが、検索対象ログにまたがって検知できない問題を
解決するために、検索対象ログを一定時間(z分間)重ねて検索する方式を採用した

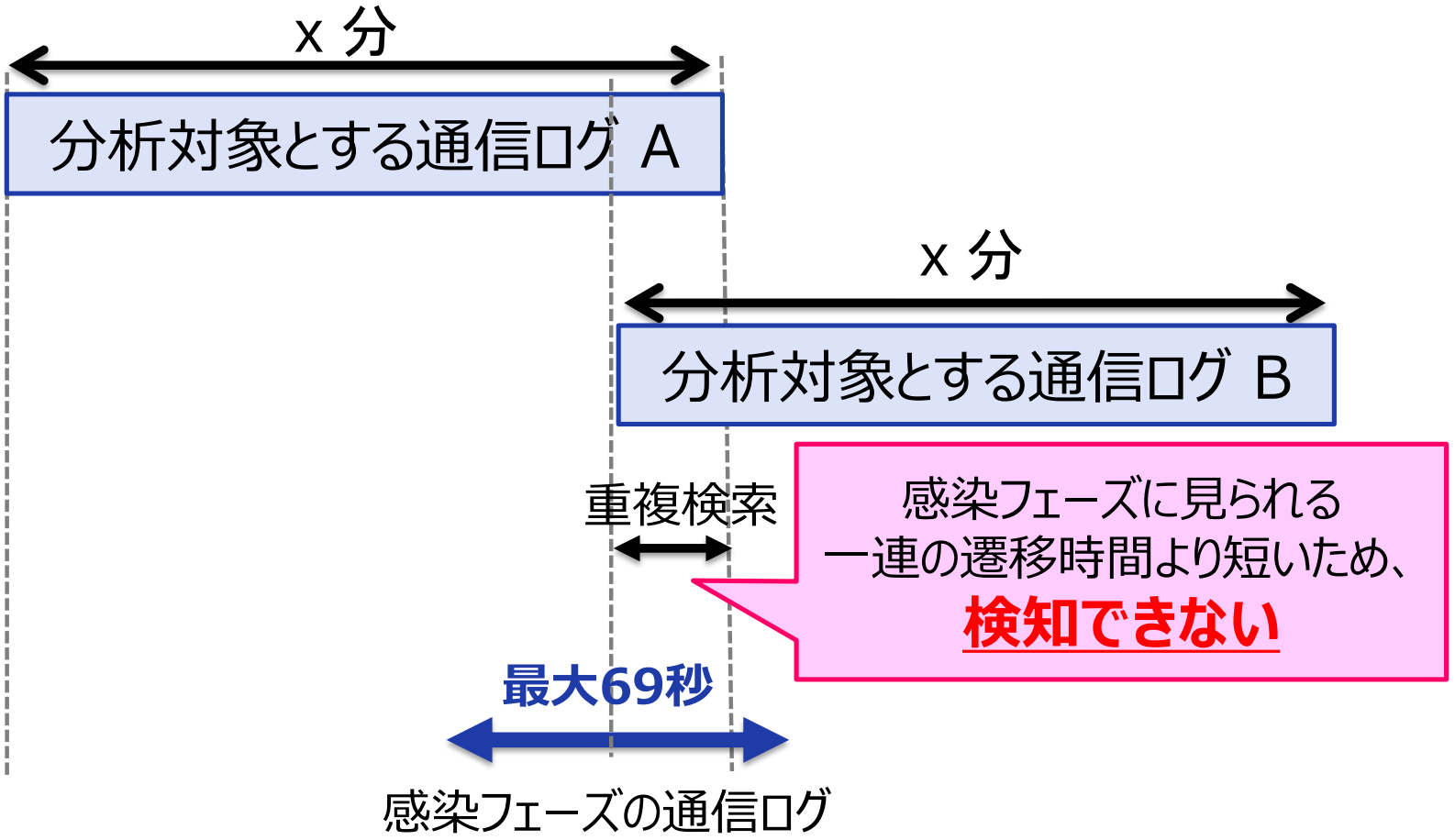


NTT DATA

重複させる期間の検討



重複させる期間の検討



この検知漏れを考慮し、**重複期間を2分間**、分析対象とする
通信ログの蓄積期間を5分間、検知パターンの**実行間隔を3分間**と設定した。

4.3 課題③への対応：検索処理の遅延への対応

前述した設計値で、検知パターンを実行させたところ、**トラフィックのピーク時間帯において**、
特定の検知パターン2個の実行時間が検知パターンの**実行間隔(3分)を上回ることがわかった**。

検知パターン	計測結果
検知パターンA	6分28秒
検知パターンB	7分8秒

※ログ行数：384,000件、データサイズ：0.4GB

採用

	①ハードウェア性能の増強	②検索処理のアルゴリズムの変更
ポイント	・高スペックな分析マシンを調達できれば、容易に実現できる。 ・高スペックな分析マシンを利用する場合、調達コストが高く、限界も存在する	・ボトルネックを見極めて、そのボトルネックを解消する必要があるため、調査・検討に時間が掛かる

ログをグルーピングする処理に大きく時間が掛かっていることがわかったため、事前に単純なソート処理を追加することとした。

5. リアルタイム版サイバー攻撃検知システムの実装と評価

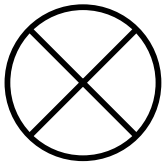
5.1 リアルタイム版サイバー攻撃検知システムの実装

NTT DATA

1. Proxyサーバ-クライアント端末間の通信をキャプチャ

2. キャプチャしたデータをProxyログと同等の形に整形する処理を実施

3. 数十種類の検知パターンで、前処理用サーバから送られたログを検索



ネットワーク装置

パケットキャプチャ装置

前処理用サーバ

ログ分析エンジン(Splunk)

Master

Peer1

・・・・・・

Peer4

クラスタ構成

冗長化を実現

アラートメール

4. 検知パターンに一致するログが存在した場合、メールが送付される。

！ ！

運用担当者

5. アラートメールをトリガとして、運用担当者はログを確認し、検知/誤検知を判断。

5.2 C&Cフェーズの検知パターンの検索方式の実装

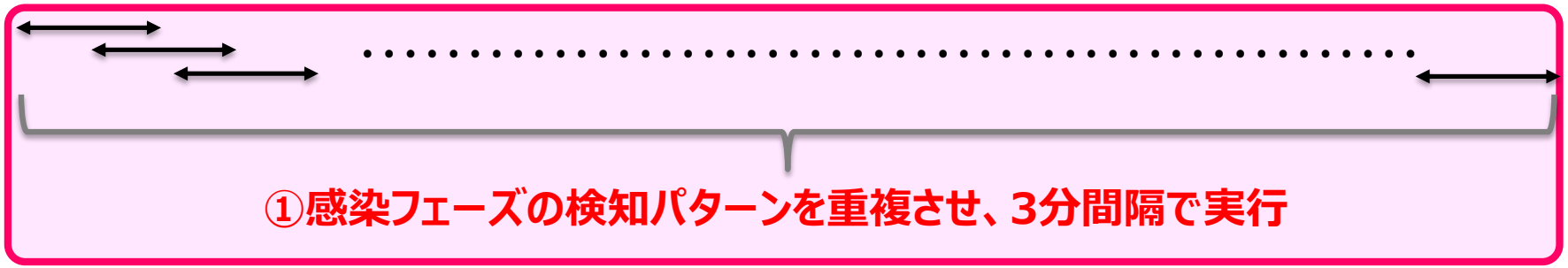
C&Cフェーズの検知パターンには、2種類のパターンが存在するため、
検知パターンの種類に応じて、ログ蓄積期間を設定

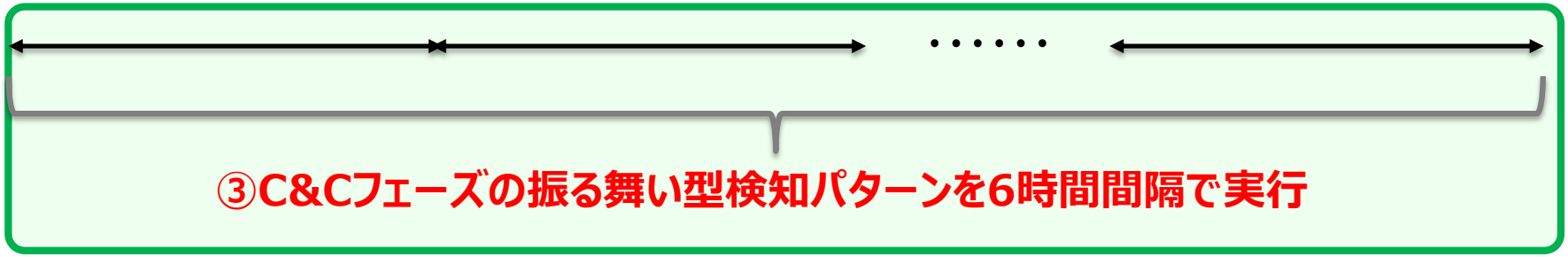
1. URLのパス等をもとに検知する「ブラックリスト型検知パターン」
   ⇒ ログ1行のみで判断できるため、長期間のログを蓄積する必要がない。
   ⇒ 検索対象ログの蓄積時間：**5分間** と設定

2. 複数行の通信ログの振る舞いを定義した「振る舞い型検知パターン」
   ⇒ 長期間のログを蓄積し、分析する必要がある。
   ⇒ 検索対象ログの蓄積時間：**6時間** と設定

各検知パターンを下記のように実行させた。

分析対象とする通信ログ





検知漏れと誤検知の評価

評価方法

・ある実環境のネットワークへリアルタイム版検知システムを導入し、
2015年4月～8月の5ヶ月間、DbD攻撃を検知するか確認した。

・あわせて日次版検知システムを上記のリアルタイム版検知システムと同じネットワークで同一期間運用し、検知数・検知結果・誤検知数が一致するか確認した。

結果

Exploit Kit	4月	5月	6月	7月	8月	合計
Nuclear EK	7件	1件	2件	4件	2件	16件
Angler EK	4件	4件	6件	4件	1件	19件
不明	0件	0件	4件	14件	0件	18件
合計	11件	5件	12件	22件	3件	53件

日次版検知システムと検知数・検知結果が一致した
また、リアルタイム化に伴う誤検知が存在しないことも確認した

検知パターンの処理時間による評価

評価方法

・検索処理遅延への対応を行う前に検索処理時間が、「目標時間：3分」を超えていた感染フェーズの検知パターン2個について、前述の高速化手法適用後、3分以内に検索処理が完了するかを確認

結果

検知パターン	高速化手法適用前	高速化手法適用後
検知パターンA	6分28秒	2分16秒
検知パターンB	7分8秒	2分19秒

※ログ行数：384,000件、データサイズ：0.4GB

設定した目標時間以内に検索処理がすべて完了することを確認

6. まとめと今後の課題

まとめ

・DbD攻撃によりマルウェアに感染した端末を**迅速に検知**するために、
日次版サイバー攻撃検知システムをリアルタイム化するための手法を検討し、実装した。

・通信ログの蓄積時間や重複検索など検索方式を工夫することで、
既存研究(日次版検知システム)と同等の精度で、検知できることがわかった。

副次効果

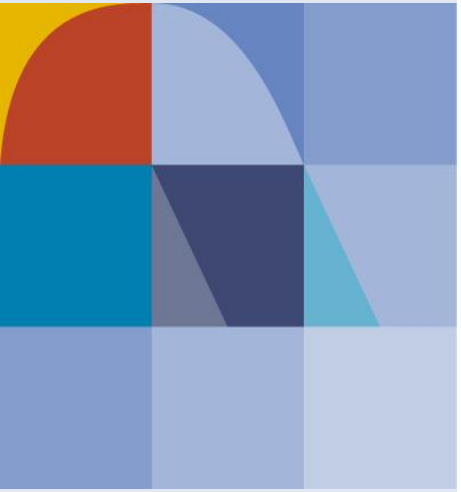
・日次版検知システムでは、検知した時点では、既に改ざんされたWebサイトや
Exploitコード、マルウェアを配布しているWebサイト等が存在しない場合が多かった。

⇒ リアルタイム版検知システムを導入することで、**改ざんされたWebサイトやExploitコード、マルウェア配布サイト等が存在している状態で検知できる**ようになった。

⇒ 攻撃手法に関する情報やインシデント対応に有益な情報を収集できるようになり、**次の研究の足掛かりにすることができるようになった**。

今後の課題

・本発表では、マルウェアに感染した端末を迅速に検知する仕組みを実装・評価したが、
今後は、マルウェアに感染した端末を検知した後の対応にかかる検討を進めていきたい



NTT DATA

Global IT Innovator