数量化理論とCCCDATAset2009を利用したボットネットのC&Cサーバの特定手法の提案と評価

東京電機大学大学院
情報セキュリティ研究室
三原　元　　佐々木　良一

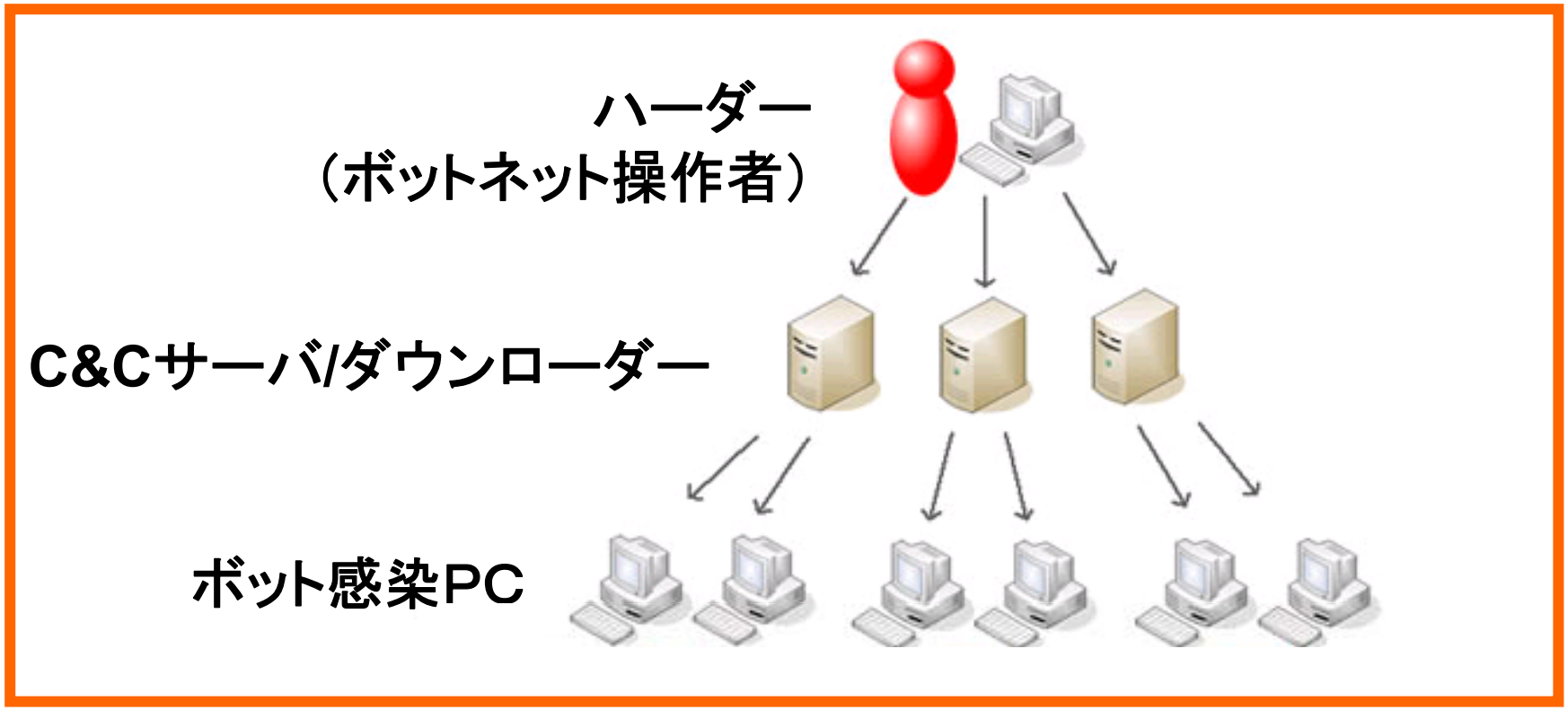
目次

1. はじめに
2. CCCDATAset2009の解析結果
3. 実験
4. 提案システム概要
5. まとめと今後

目次

1.1 ボットネットに対する既存対策手法の問題点

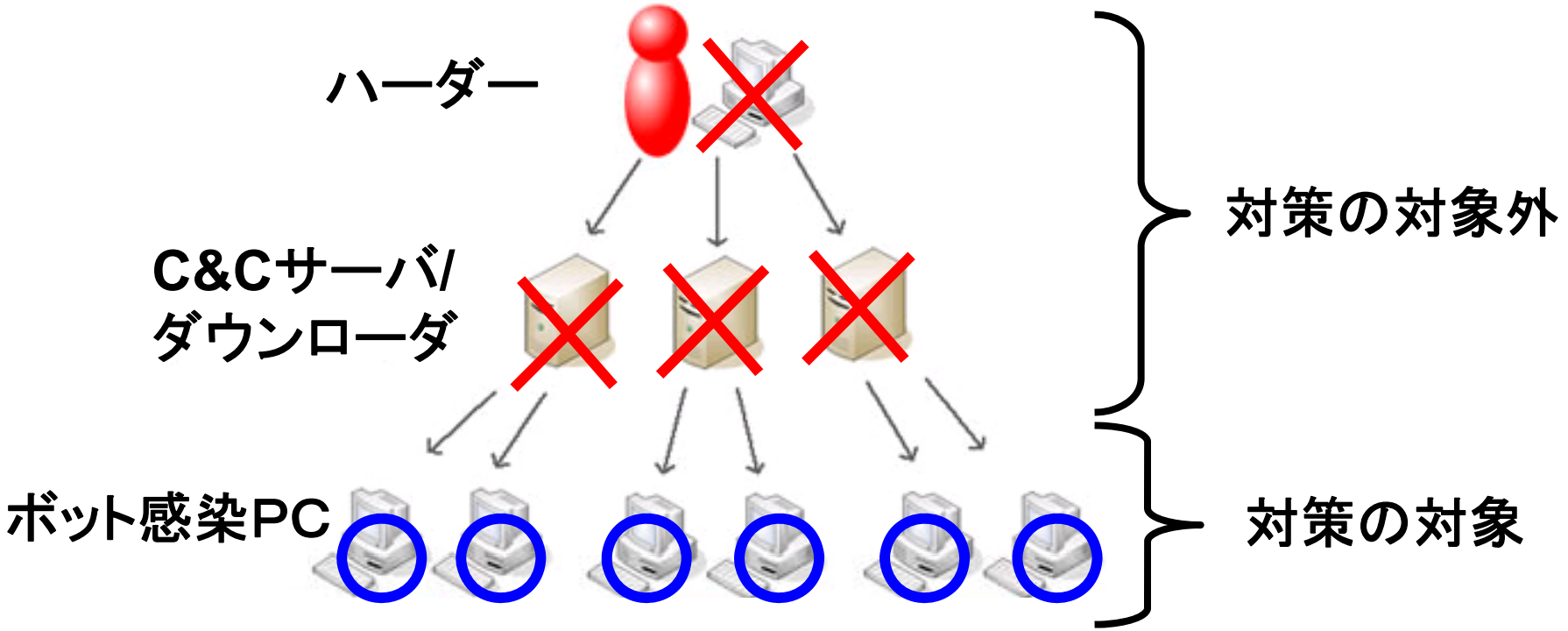
現在問題となっている「ボットネット」



1.1 ボットネットに対する既存対策手法の問題点

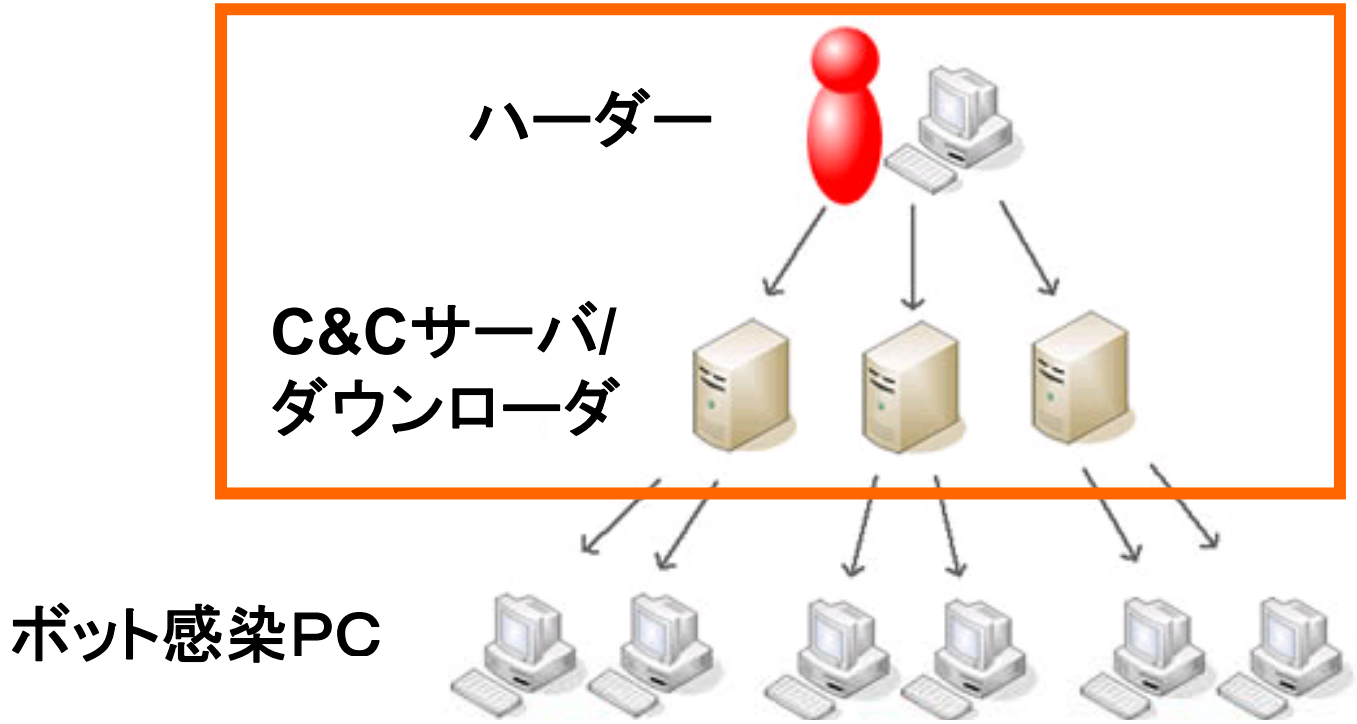
代表的な既存対策手法：アンチウイルスソフト

➡ 感染PCの駆除が目的



1.1 背景

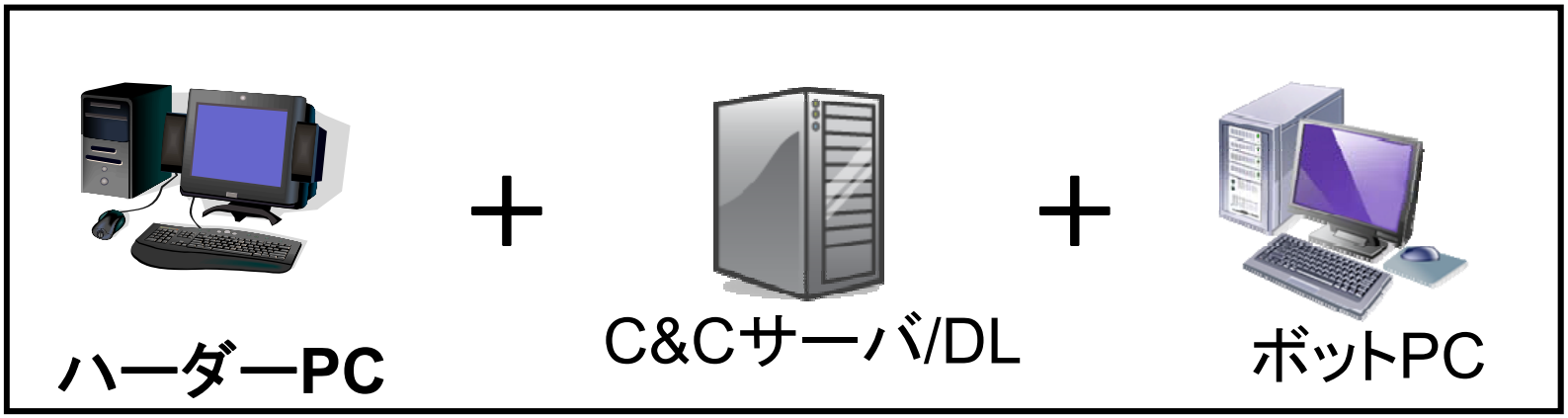
- ボットネットの問題点
  - ボットPCのみの対策では解決しない
  - ハーダー，C&Cサーバ等を対処しなければ無効が困難



➡ ボットネットに対する根本的な解決手法が必要

1.1 研究目的

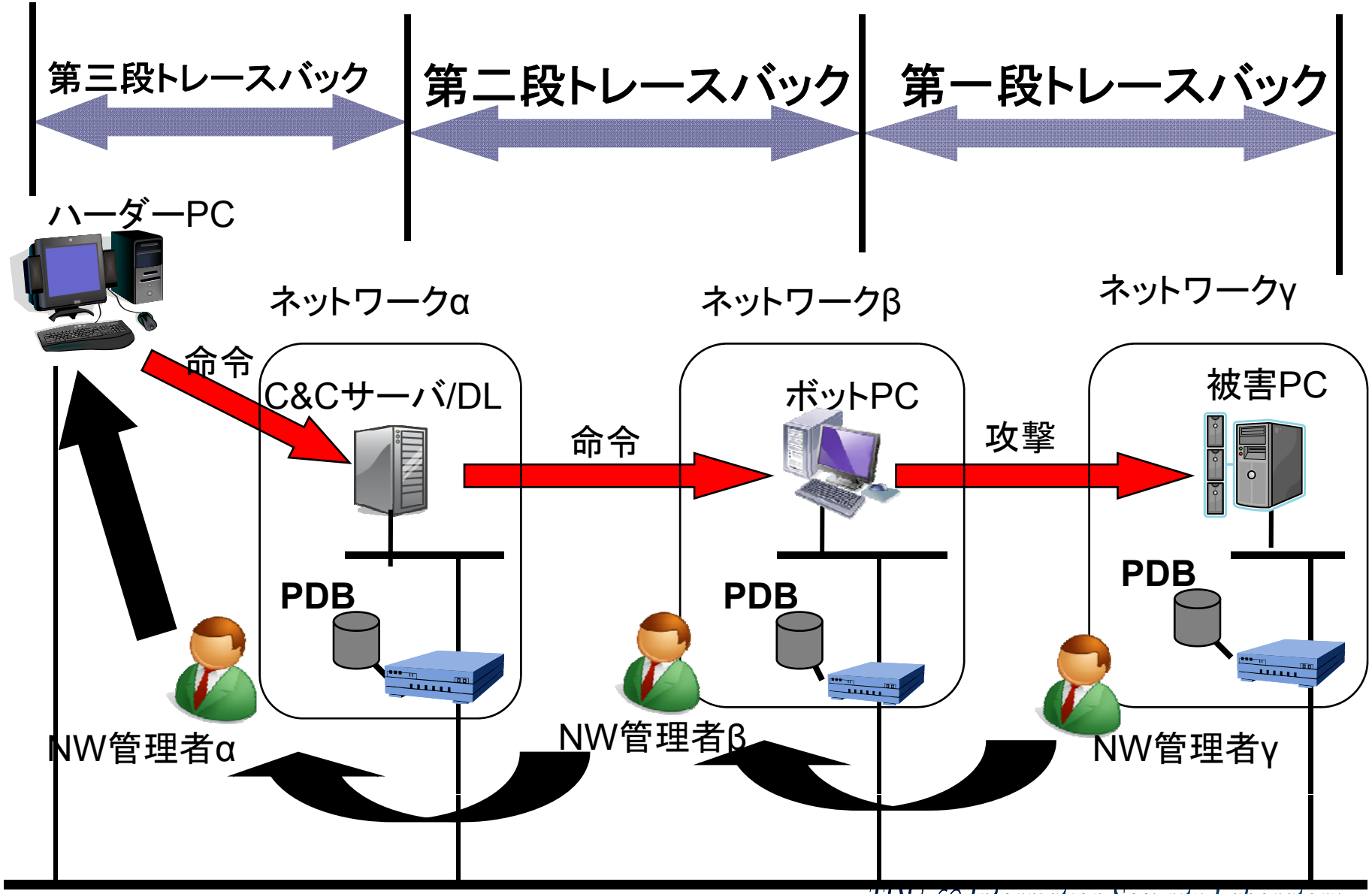
ネットワーク管理者同士が情報共有を行い，
3つの特定を行う



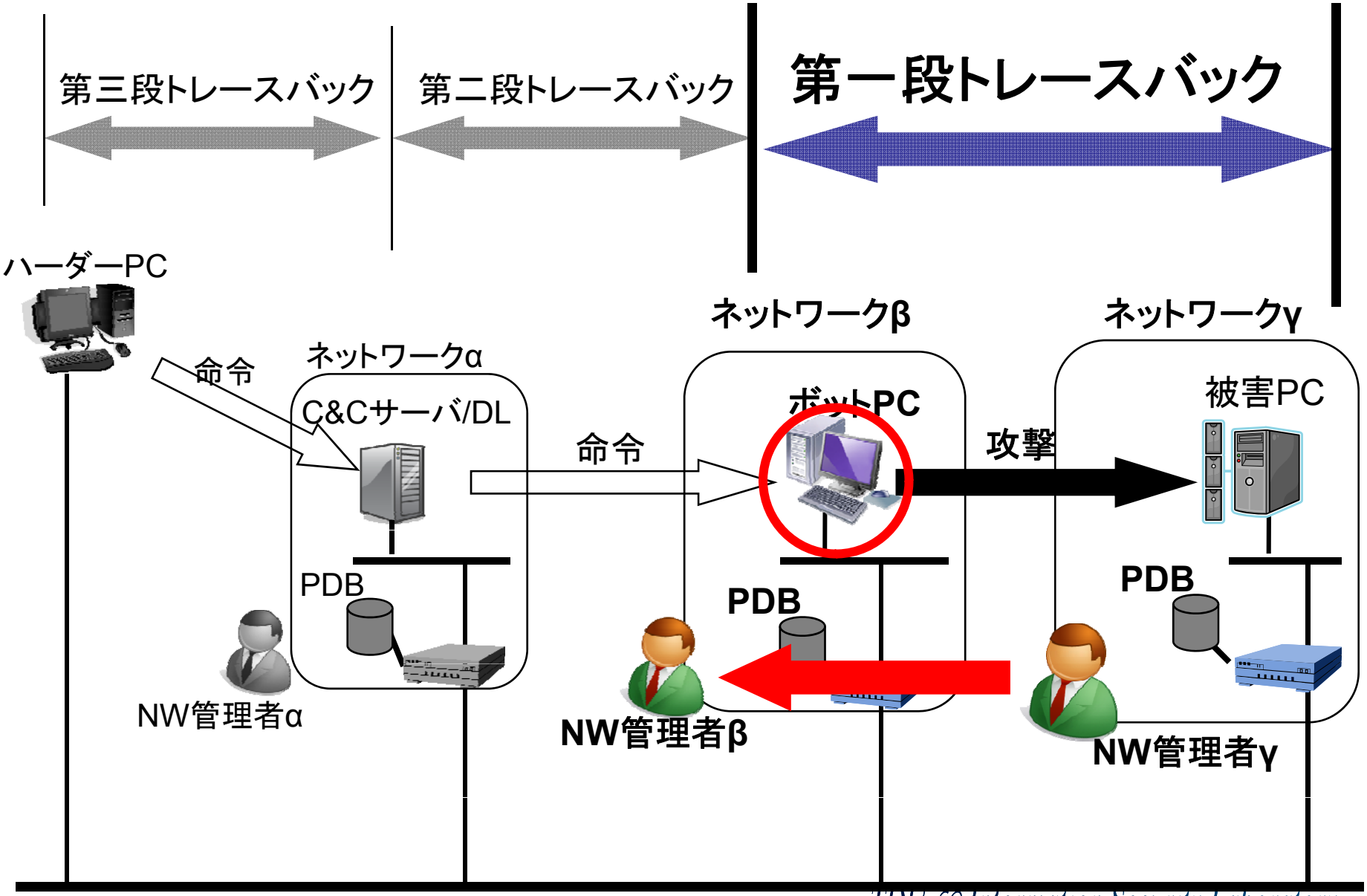
ボットネットの多段追跡システムを構想

ボットネットの根本的な解決を目指す

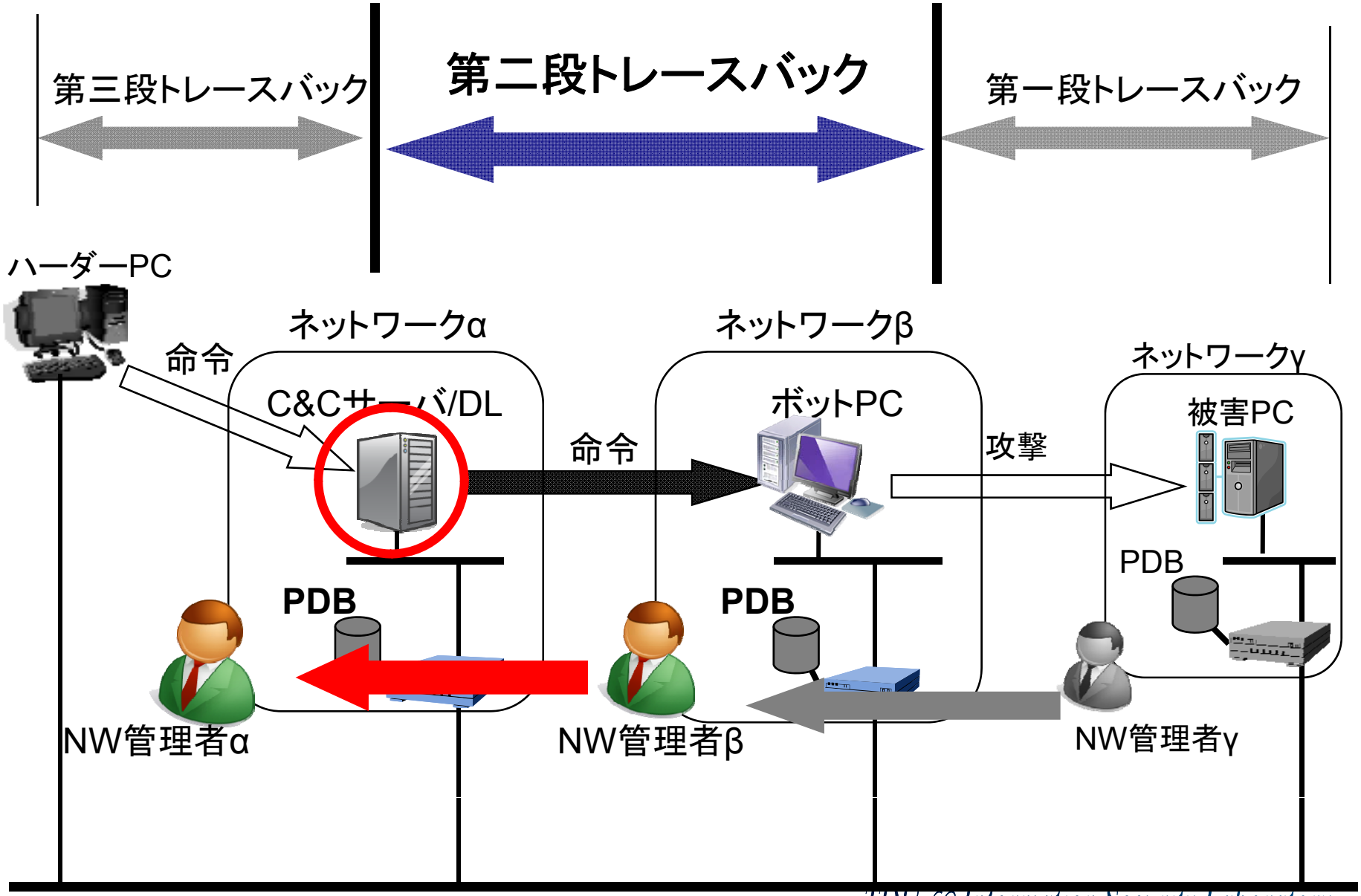
1.2 多段追跡システム概要



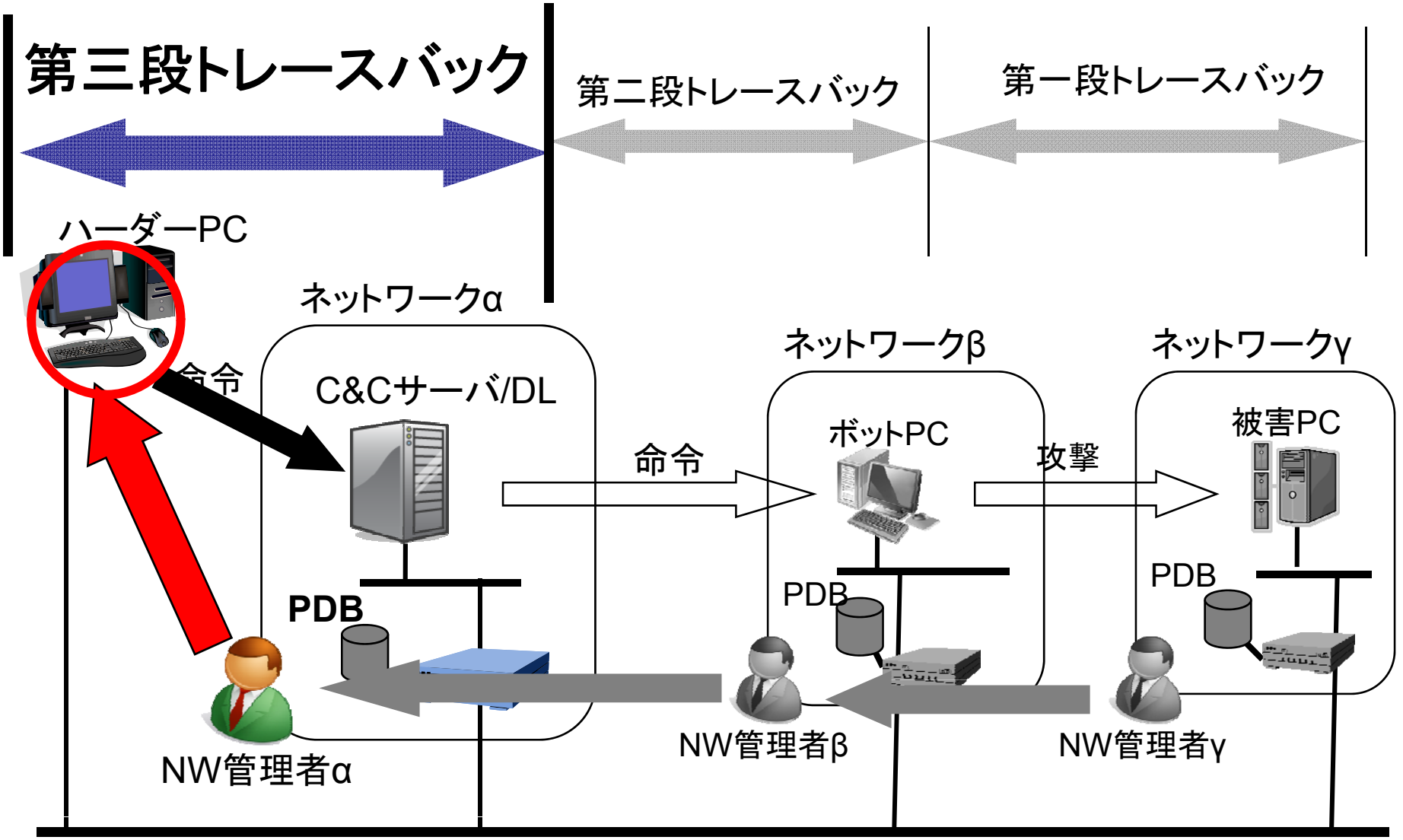
1.2 多段追跡システム概要



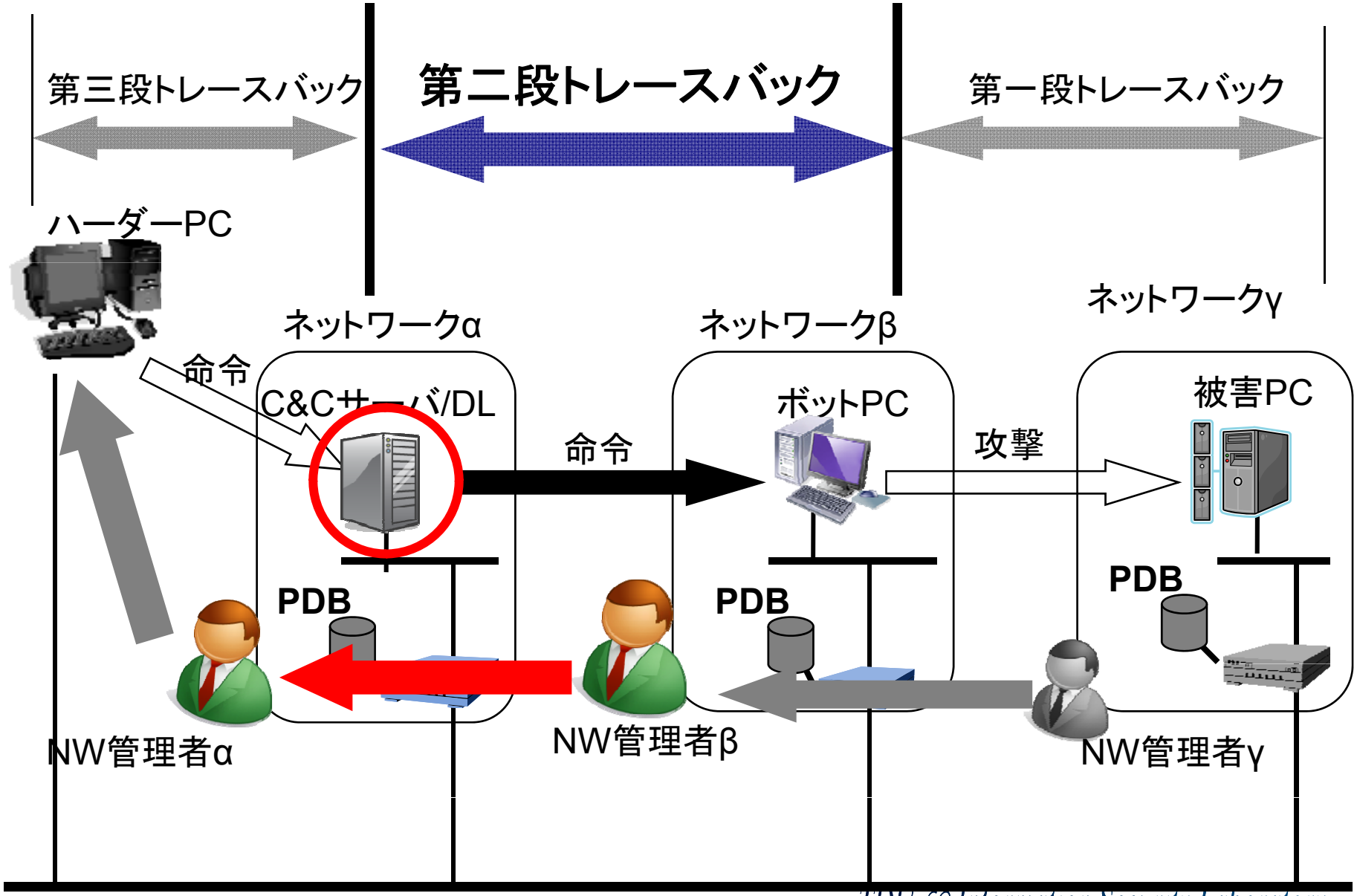
1.2 多段追跡システム概要



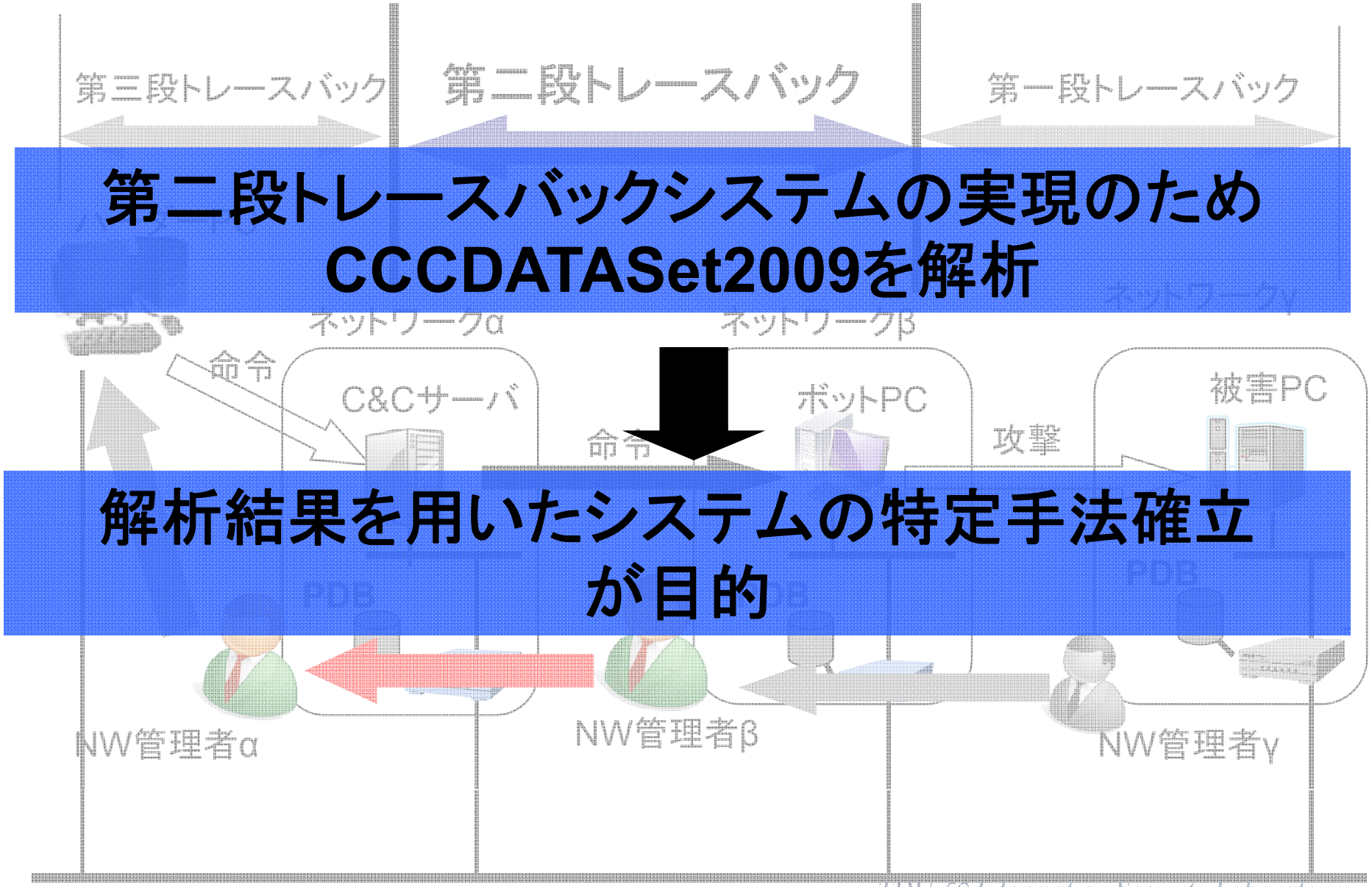
1.2 多段追跡システム概要



1.2 多段追跡システム概要



1.2 目的



第二段トレースバックシステムの実現のため CCCDATASet2009を解析

解析結果を用いたシステムの特定手法確立 が目的

目次

2. CCCDATAset2009の解析方針

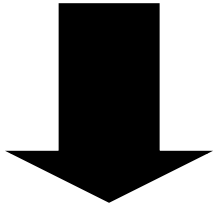
今回著者らはC&Cサーバの「ドメイン情報」に着目

調査対象データ

・ボットネットに関するドメイン情報　　：ボットネットドメイン
・ボットネットに関係しないドメイン情報：ノーマルドメイン

調査対象項目

・既存研究より5項目を選別



各調査項目ごとに2つのデータを比較
ボットネットの特徴を掴む

2. CCCDATAset2009の解析結果

- 調査対象データ
  - ボットネットドメイン
    - CCCDATAset2009の通信データ中DNS通信から　　取得したドメイン
    - 個数：24個
  - ノーマルドメイン
    - 本研究室の通信データ中DNS通信から取得したドメイン
    - 個数：50個

PC台数	20台（OS:WindowsXP）
通信データ取得時間	24時間
パケット数	約50万パケット

2. CCCDATAset2009の解析結果

調査対象項目

逆引き

DNSサーバに対する「IPアドレス→ドメイン名」の問い合わせ調査

SOAレコード

WHOIS

mail.wwwサーバの有無

同一ドメイン上にMail,Webサーバが登録されているかの調査

TTL値

DNSサーバの各管理ドメインに設定されるTTL値

参考文献:
「フィールド調査によるボットネットの挙動解析」高橋正和，村上純一，須藤年章，
平原伸昭，佐々木良一，情報処理学会論文誌，Vol,47,No.8,2007
「DNS通信の挙動から見たボット感染検知方式の検討」東角芳樹，鳥居悟
CSS2008

2. SOAレコード 1/2

- SOA(Start Of Authority)レコード
  - 各DNSサーバが管理
  - ドメインに関する設定情報
  - SOAレコードの構成
    - ゾーン名
    - ネームサーバホスト名
    - 管理者メールアドレス
    - Refresh値
    - Retry値
    - Expire値
    - Minimum値

2. SOAレコード 2/2

- SOAレコード
  - ゾーン名
  - ネームサーバホスト名
  - 管理者メールアドレス
  - Refresh値
  - Retry値
  - Expire値
  - Minimum値（ネガティブTTL値）

    ➡ Minimum値を調査対象とした

2. WHOIS 1/2

- WHOIS情報
  - JPNIC等のレジストリが管理
  - 各ドメインに関する管理者情報
  - WHOIS情報の主な構成
    - 登録ドメイン名
    - レジストラ名
    - ドメインが登録されているDNSサーバ名
    - ドメイン登録者の名前・住所
    - ドメインの登録年月日
    - ドメインの登録有効期限

2. WHOIS 2/2

- WHOIS情報
  - 登録ドメイン名
  - レジストラ名
  - ドメインが登録されているDNSサーバ名
  - ドメイン登録者の名前・住所
  - ドメインの登録年月日
  - ドメインの登録有効期限

→ 上記2つの他，2つのデータから判明する「ドメイン登録期間」の3つを調査

2. CCCDATAset2009の解析結果

- 調査結果
  - 逆引き
  - SOAレコード
  - WHOIS
  - mail.wwwサーバの有無
  - TTL値

2. 逆引きの調査結果

逆引きの調査結果

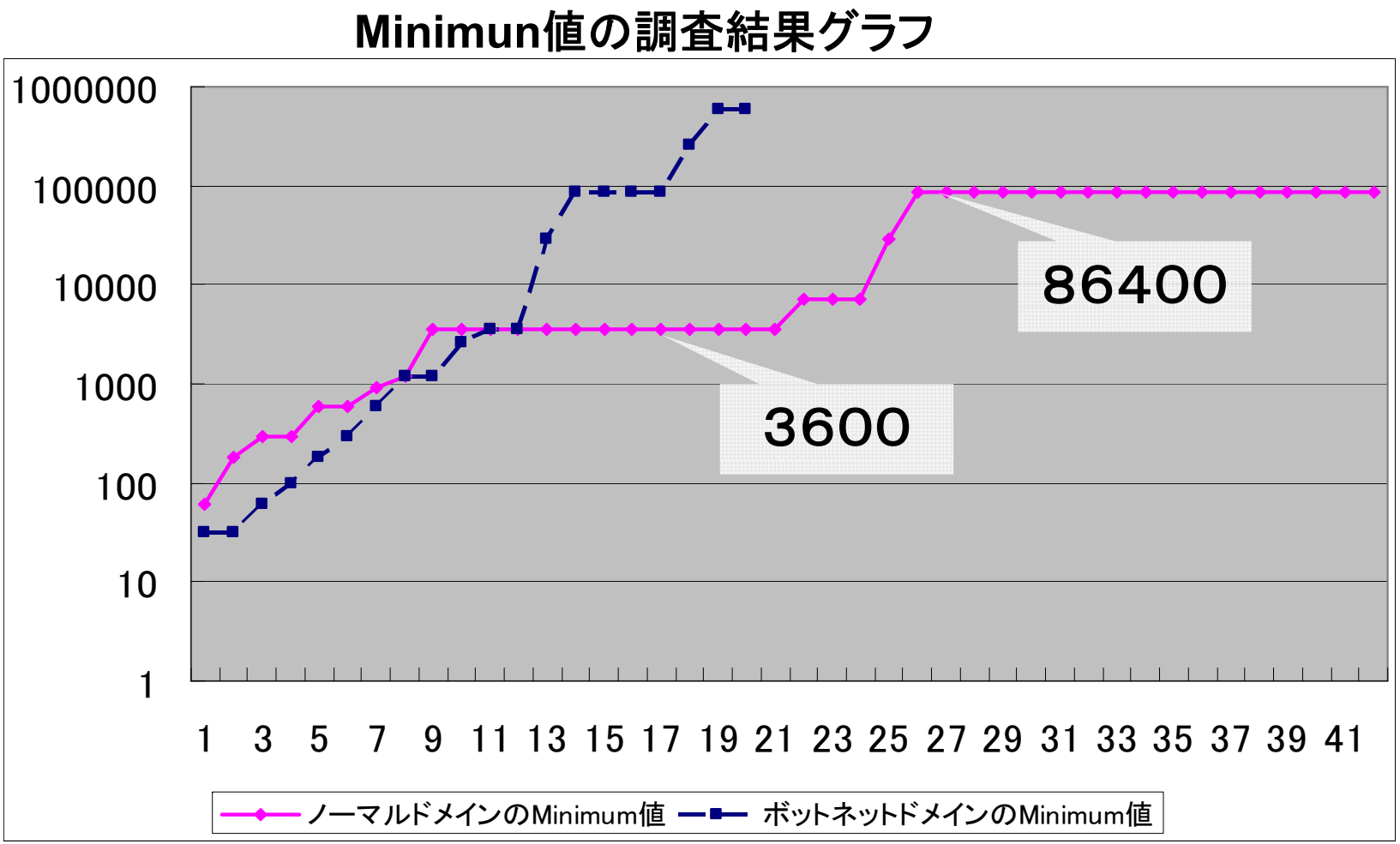
	ノーマルドメイン	ボットネットドメイン
逆引き結果が正しい	9個 （18%）	2個 （8%）
逆引き結果が正しくない	34個 （68%）	5個 （21%）
返答なし	7個 （14%）	17個 （71%）
合計	50個 （100%）	24個 （100%）

2つのデータに明らかな差がみられた

2. CCCDATAset2009の解析結果

- 調査対象項目
  - 逆引き
  - Minimun値(SOAレコード)
  - WHOIS
  - mail.wwwサーバの有無
  - TTL値

2. Minimun値（SOAレコード）の調査結果



ボットネットドメイン：　特徴は見られなかった

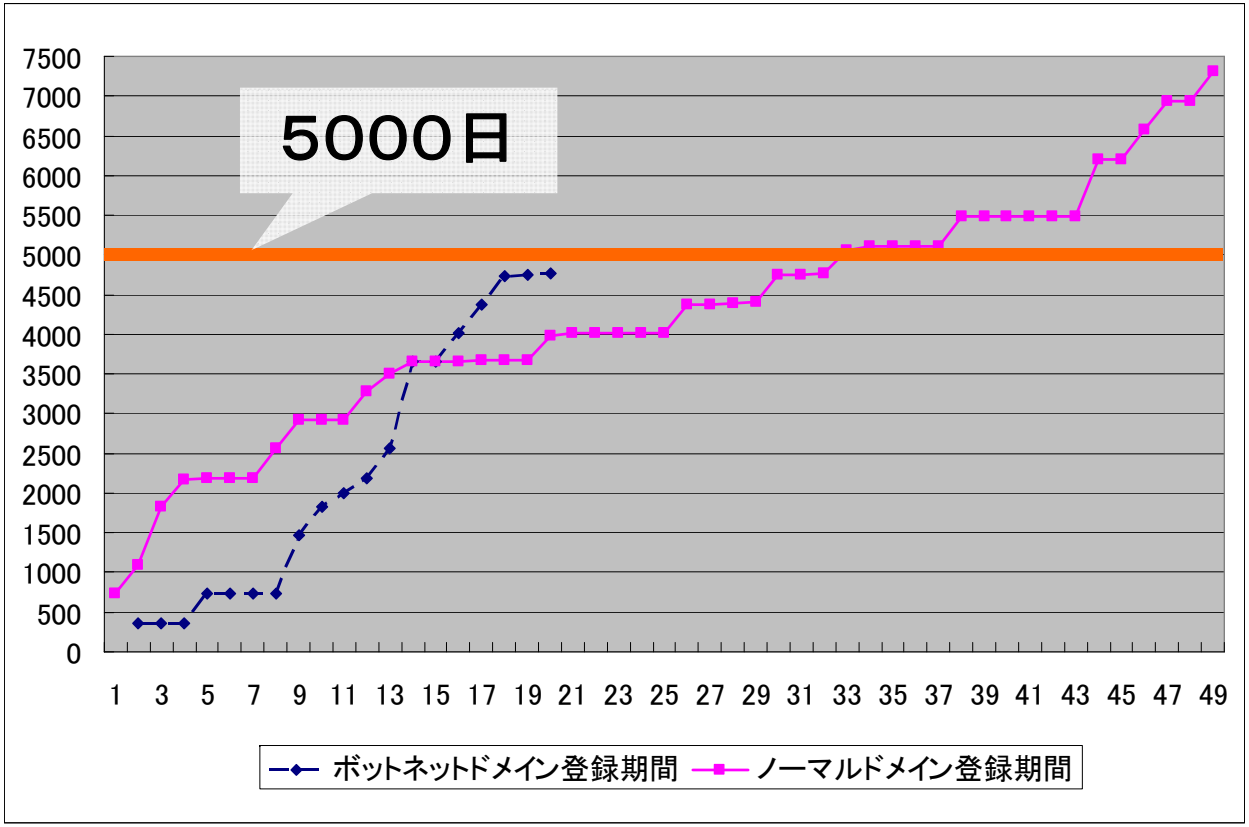
ノーマルドメイン　　：　特徴が見つかった

2. CCCDATAset2009の解析結果

- 調査対象項目
  - 逆引き
  - SOAレコード
  - WHOIS
    - ドメイン登録年月日
    - ドメイン登録終了年月日
    - ドメイン登録期間
  - mail.wwwサーバの有無
  - TTL値

2. ドメイン登録期間（WHOIS）の調査結果 3/3

ドメイン登録期間の調査結果グラフ



ボットネットドメインの登録期間が比較的短い

2. CCCDATAset2009の解析結果

- 調査対象項目
  - 逆引き
  - SOAレコード
  - WHOIS
  - mail.wwwサーバの有無
  - TTL値

2. mail.wwwサーバの有無

	ノーマルドメイン	ボットネットドメイン
Mailサーバ有	36個 （72%）	14個 （59%）
Mailサーバ無し	10個 （20%）	3個 （12%）
返答なし	4個 （8%）	7個 （29%）
合計	50個 （100%）	24個 （100%）

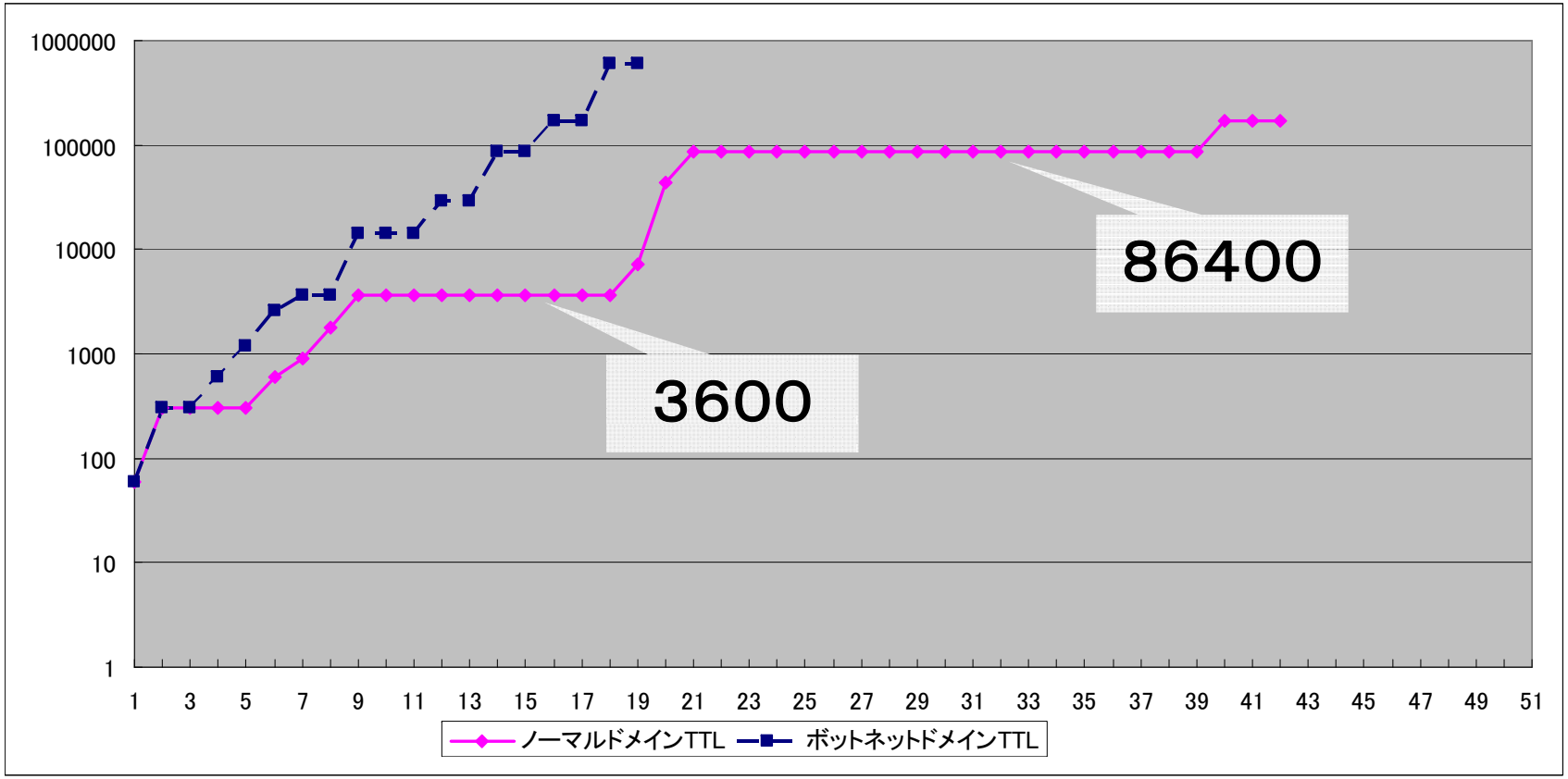
	ノーマルドメイン	ボットネットドメイン
ウェブサーバ有	45個 （90%）	13個 （54%）
ウェブサーバ無し	2個 （4%）	4個 （16%）
返答なし	3個 （6%）	7個 （30%）
合計	50個 （100%）	24個 （100%）

2つのデータに特長は見られなかった

2. CCCDATAset2009の解析結果

- 調査対象項目
  - 逆引き
  - SOAレコード
  - WHOIS
  - mail.wwwサーバの有無
  - TTL値

2. TTL値の調査結果



ノーマルドメインのTTL値が特定の値に集中しているという傾向が見つかった

2. CCCDATAset2009の解析結果

調査結果をまとめると

- 逆引き
- Minimun値(SOAレコード)
- ドメイン登録期間(WHOIS)
- mail.wwwサーバの有無
- TTL値

C&Cサーバ等の特定が可能となるような
特徴的データは得られなかった

2. CCCDATAset2009の解析結果

調査結果をまとめると

- 逆引き
- Minimun値(SOAレコード)
- ドメイン登録期間(WHOIS)
- mail.wwwサーバの有無
- TTL値

C&Cサーバ等の特定が可能となるような
特徴的データは得られなかった

そこで. . .

2. CCCDATAset2009の解析結果

調査結果データを入力データとし，数式処理を試行する

- 逆引き
- SOAレコード
- ドメイン登録期間（WHOIS）
- mail.wwwサーバの有無
- TTL値

数式処理によりC&Cサーバ等の判別が可能か実験
実験では，数式処理に**数量化理論2類**を使用

2. 数量理論2類とは

各対象に対し，パラメータにより求められる判別式，判別値を境界値と比較し，2つのグループに分ける

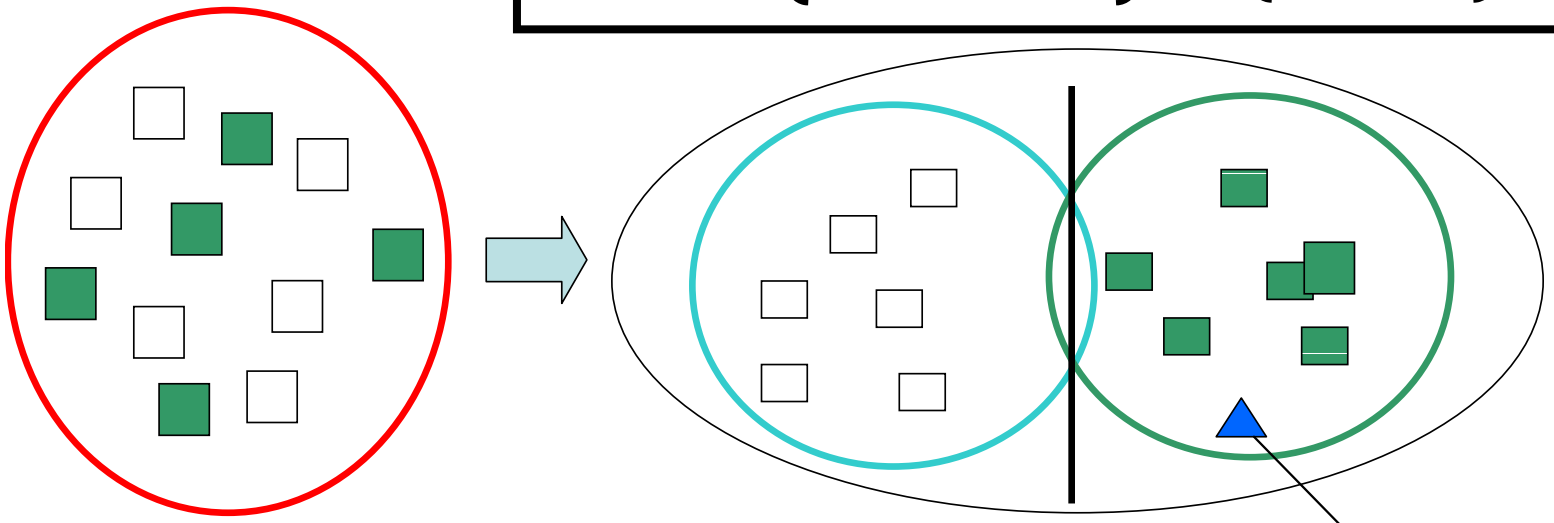
例：喫煙者と非喫煙者の判別

□ タイプA：喫煙者

■ タイプB：非喫煙者

判別式：　（年齢）　＋　（飲酒歴）　＝「判別値」

パラメータ　（20才以上：1，20才未満：0）　（ある：1，ない：0）



2. 数量化理論2類の使用に必要となるデータ

喫煙者の判別には喫煙者のデータが必要

判別式: （年齢） ＋ （飲酒歴），「境界値」

パラメータ
- 20才以上：1
- 20才未満：0

- ある：1
- ない：0

C&Cサーバ等の判別にもC＆Cサー等のデータが必要

例

判別式: （TTL値） ＋ （逆引き結果），「境界値」

パラメータ
- 900以上：1
- 900未満：0

- 返答有：1
- 返答無：0

- パラメータに5つの調査結果を使用
- 実験により，調査結果データ用いた最適パラメータを選択

目次

1. はじめに
2. CCCDATAset2009の解析結果
3. **実験**
4. システム概要
5. まとめと今後

3．7つのパラメータ(5つの調査結果)

調査結果5項目を以下のように数量化理論2類のパラメータとして設定

ドメイン登録期間	設定値
1-2500 (日)	1
2501-5000 (日)	2
5001-8000 (日)	3
NA	4

ドメイン登録終了年月日	設定値
-2010.09.13	1
2010.09.14-	2
NA	3

逆引き	設定値
返答無し	1
返答が正しくない	2
返答が正しい	3

TTL値	設定値
1-1000	1
1001-100000	2
100001-1000000	3
NA	4

Minimum値	設定値
1-100	1
101-1000	2
1001-100000	3
NA	4

Mailサーバ	設定値
あり	1
なし	2

ウェブサーバ	設定値
あり	1
なし	2

3. 実験概要

パラメータ設定実験

目的

7つのパラメータを数量化理論2類に適応
最も検知精度の高いパラメータの最小の組合わせを決定

検証実験

目的

パラメータ設定実験で決定したパラメータ使用
設定実験と別のドメインデータ入力し，判別式と境界値の検知精度を検証

3. 実験に使用するドメインデータ

パラメータ設定実験

ボットネットに関するドメイン
➡CCCDATASet2009より**10個**

ボットネットに無関係のドメイン
➡本研修室の通信データより**20個**

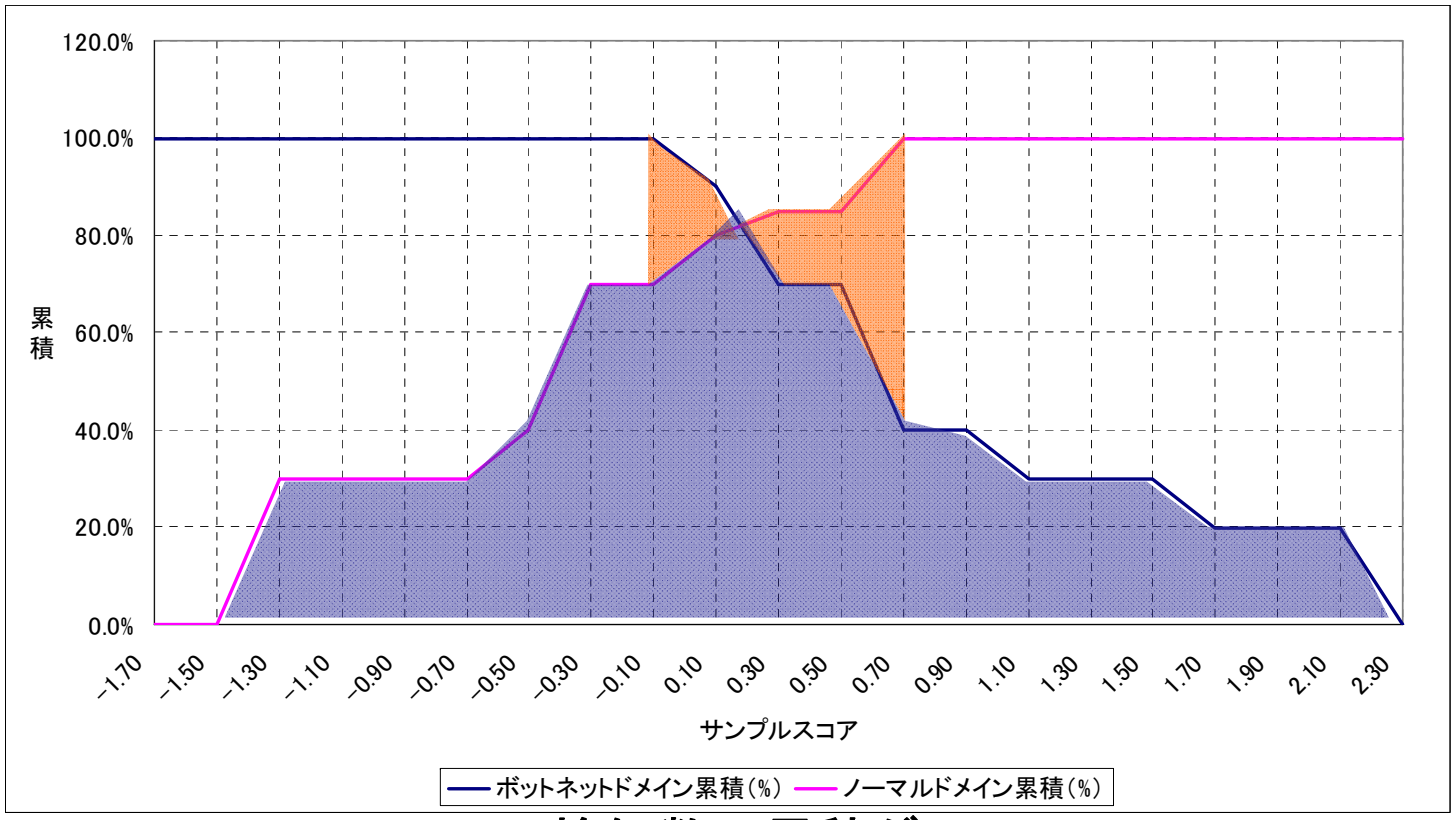
検証実験

ボットネットに関するドメイン
➡CCCDATASet2009より**10個**

ボットネットに無関係のドメイン
➡本研修室の通信データより**20個**

3. 実験結果

- 最適なパラメータの組み合わせを選択した
  - ドメイン登録期間(WHOIS)
  - 逆引き



検知数の累積グラフ

誤検知

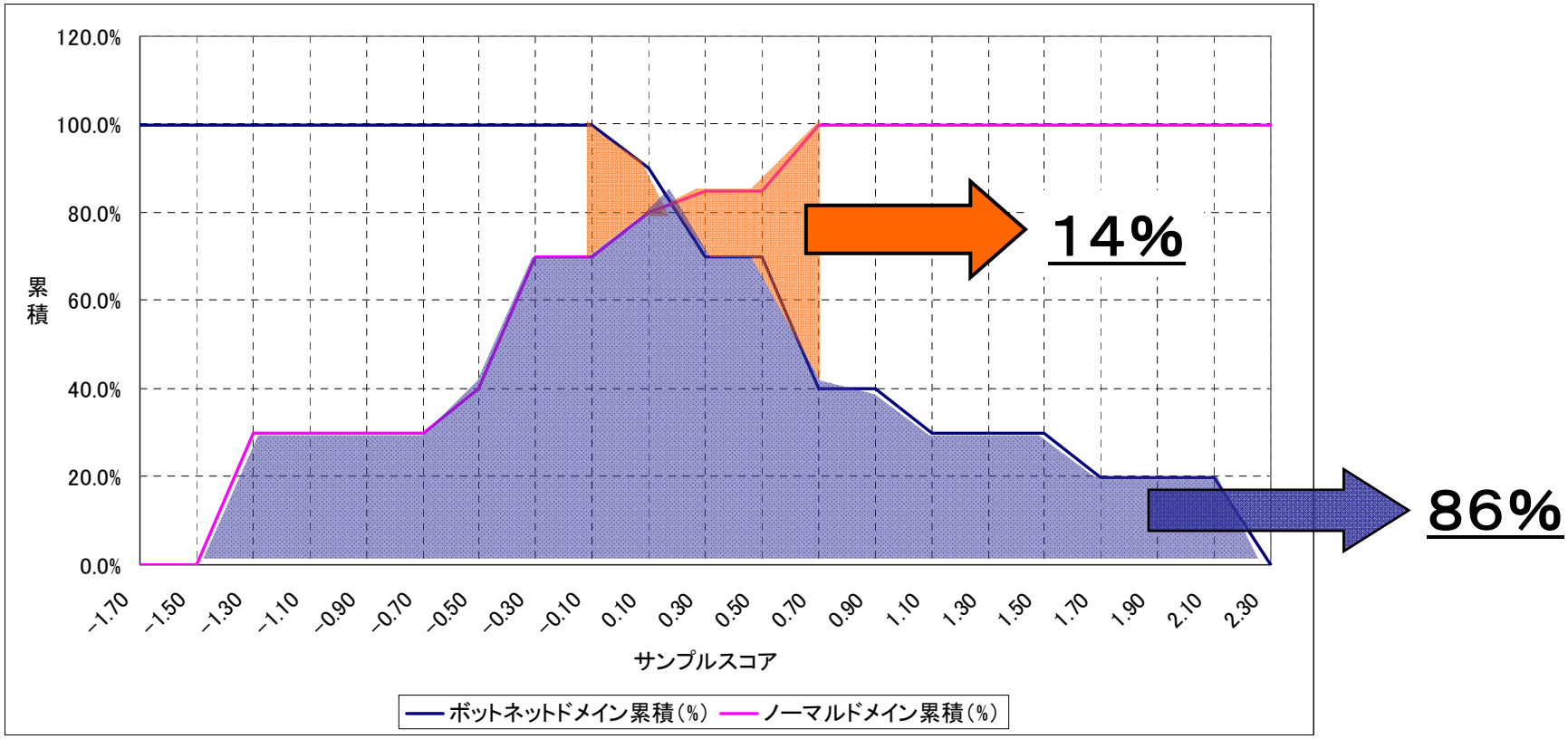
14%

検知

86%

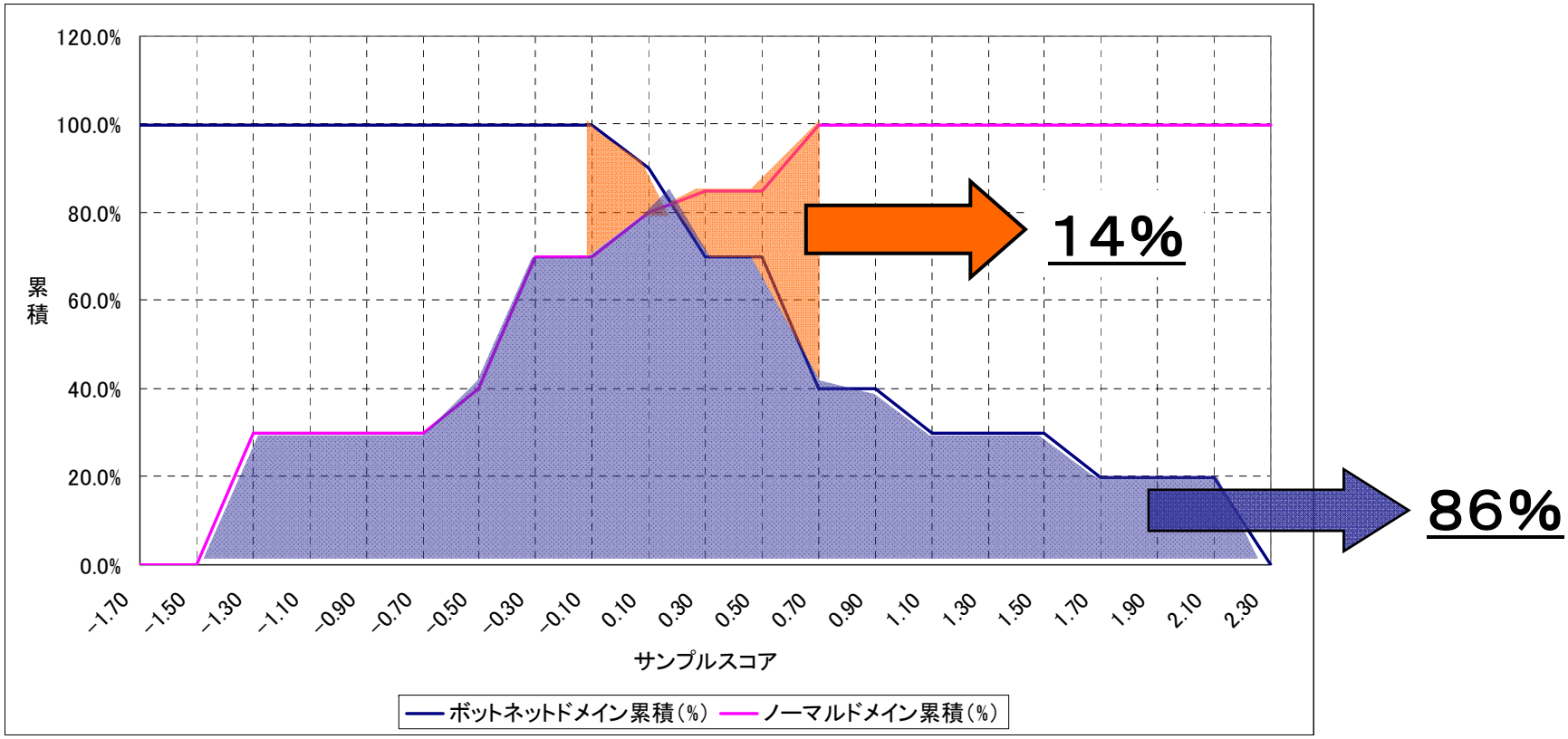
3. 実験結果

単独でのC&Cサーバ等特定には信頼性が低い



3. 実験結果

単独でのC&Cサーバ等特定には信頼性が低い



ブラックリスト方式と併用
特定精度の向上を狙う

3. ブラックリスト方式

ブラックリスト方式：概要

C&Cサーバ等ボットネットに関するホストのドメイン名を取得
通信データ中ホストのドメイン名と照合し検知を行う

ブラックリスト方式：検証

検証対象ドメイン

CCCDATASet2009の攻撃通信データから取得した
ドメイン24個

ブラックリスト取得サイト(2009年6月20日取得)

SRI Malware Threat Center

ttp://www.mtc.sri.com/

DNS-BH

ttp://www.malwaredomains.com/

ブラックリスト方式：検証結果

ブラックリストドメインとの一致数

	一致	不一致	合計
ドメイン数（%）	20（80%）	4（20%）	24（100%）

数量化理論2類を用いた検知方式での誤検知ドメイン

zonetech.info, www.getmyip.org

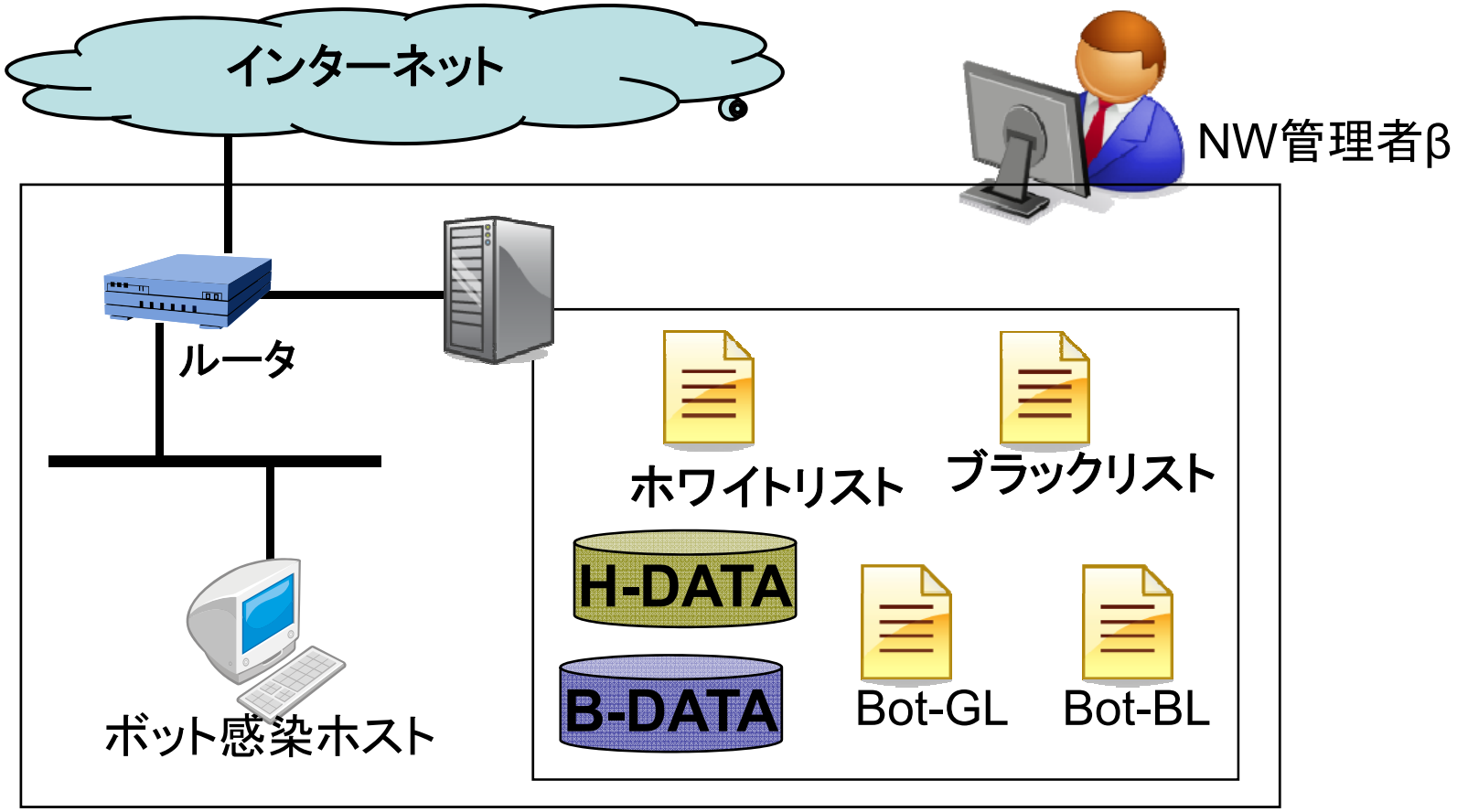
ブラックリスト方式での誤検知ドメイン

ftp.newaol.com, ftp.scarlet.be, ftp.icq.com
www.if.ee,

ブラックリスト方式と数量化理論2類を用いた方式の併用で全てのドメインを検知することができた

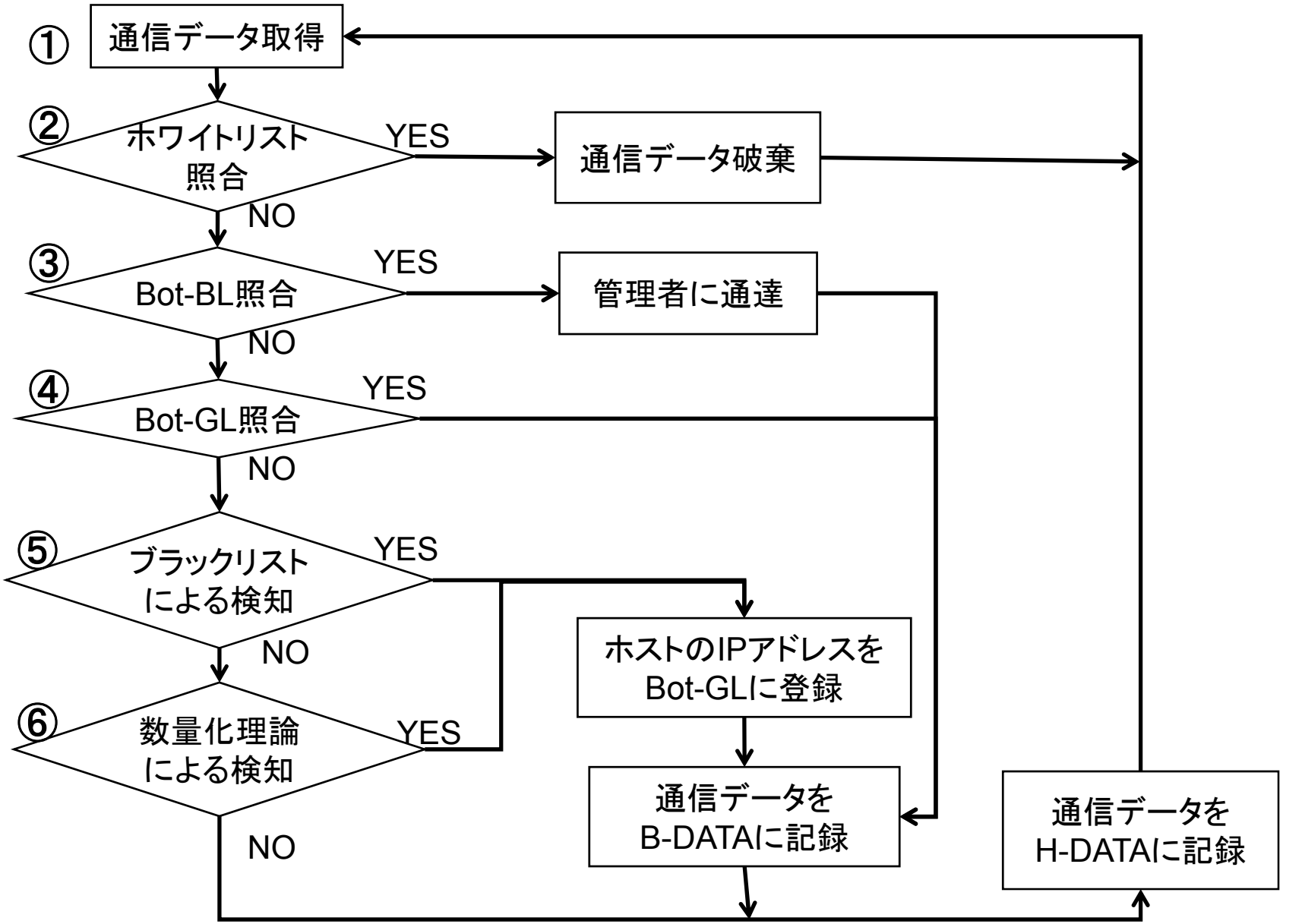
目次

4. 第二段トレースバックシステム構成

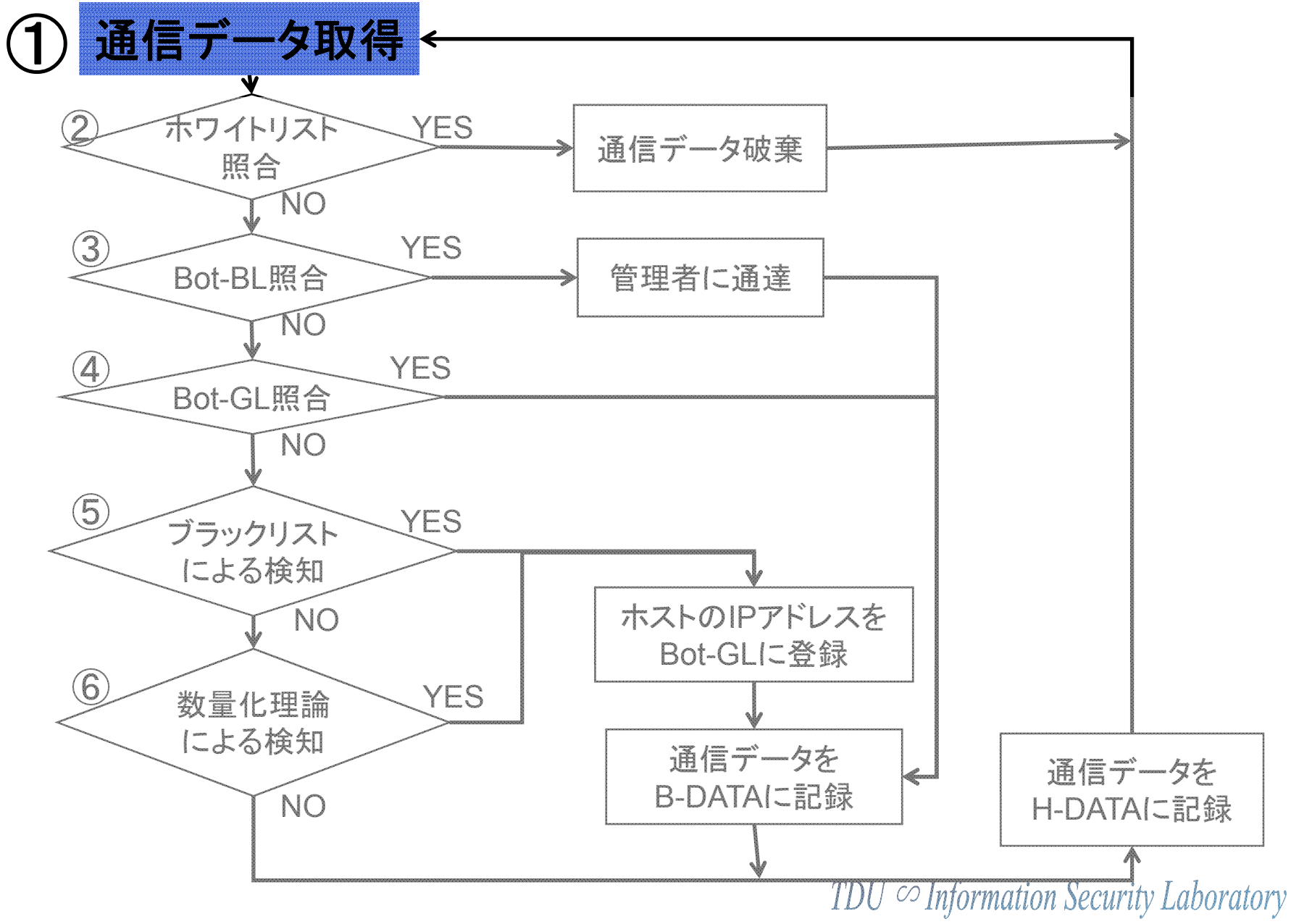


・IPアドレス，ドメイン名が記載されたリスト4種
・通信データを記録するデータベース2種
計6種で構成

4. 第二段トレースバックシステム フロー



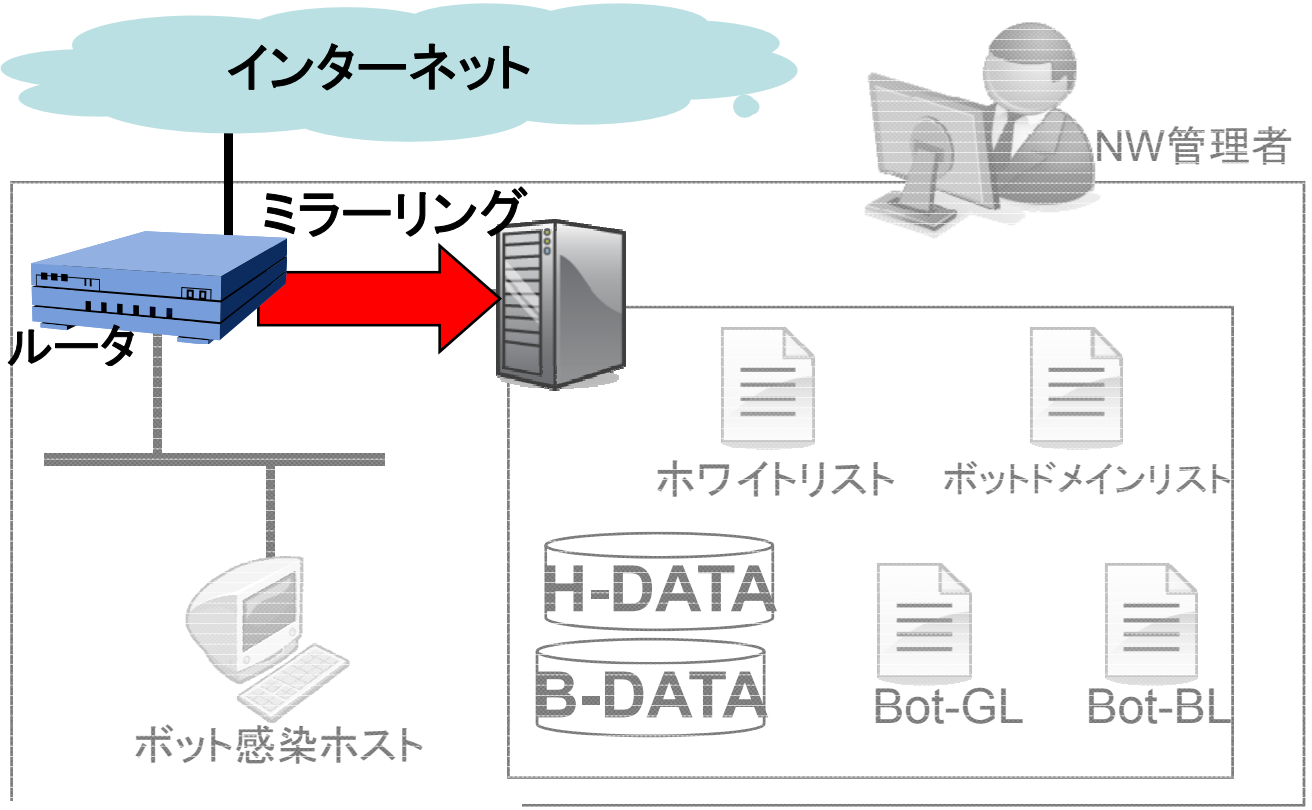
4. 第二段トレースバックシステム フロー



4. 第二段トレースバックシステム構成

1. ネットワークを通過する通信データを取得

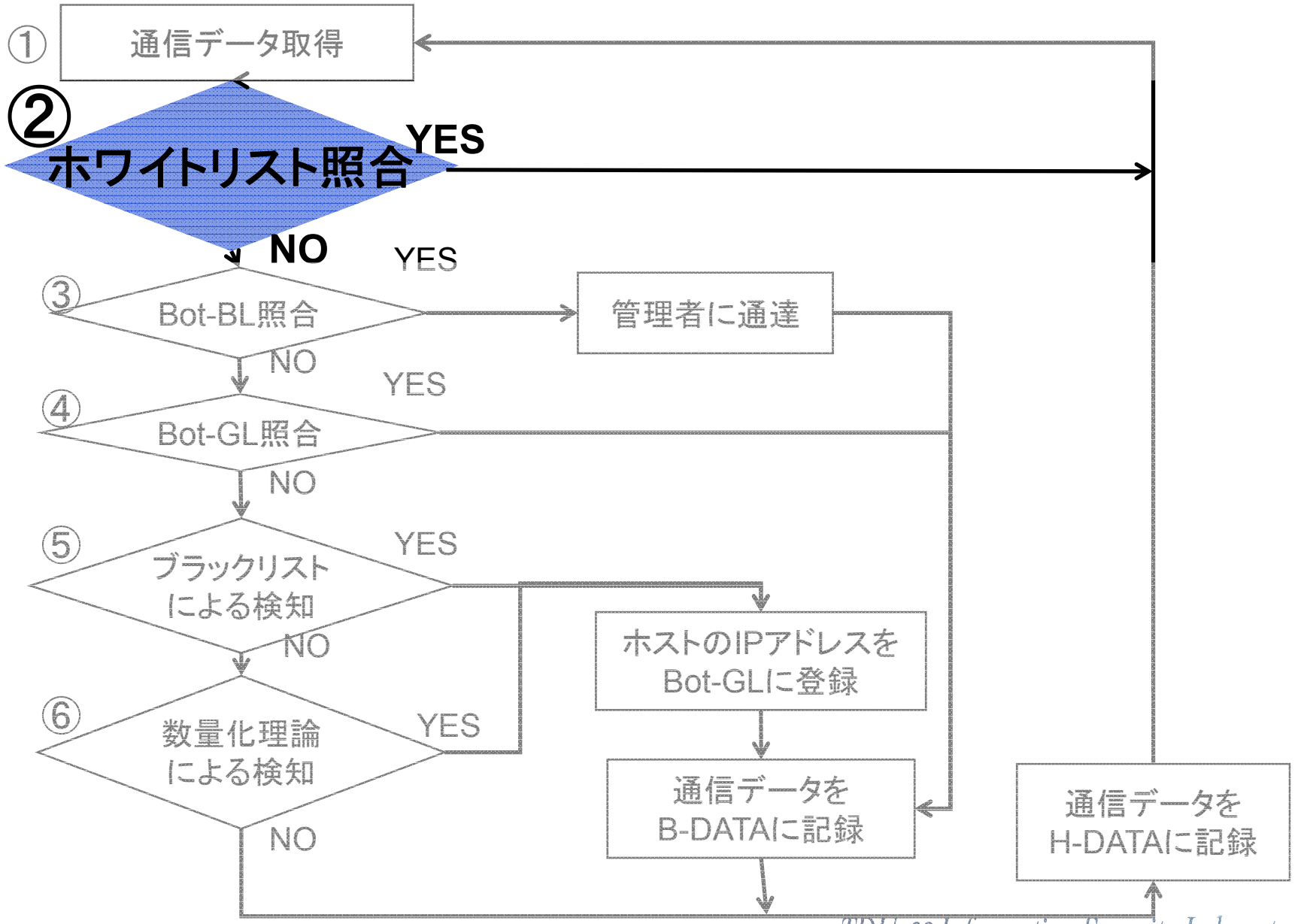
通信先IPアドレス，DNS通信中のドメイン名を抽出



通信データの取得方法

ルータからのポートミラーリング等によりコピーされた通信データを使用

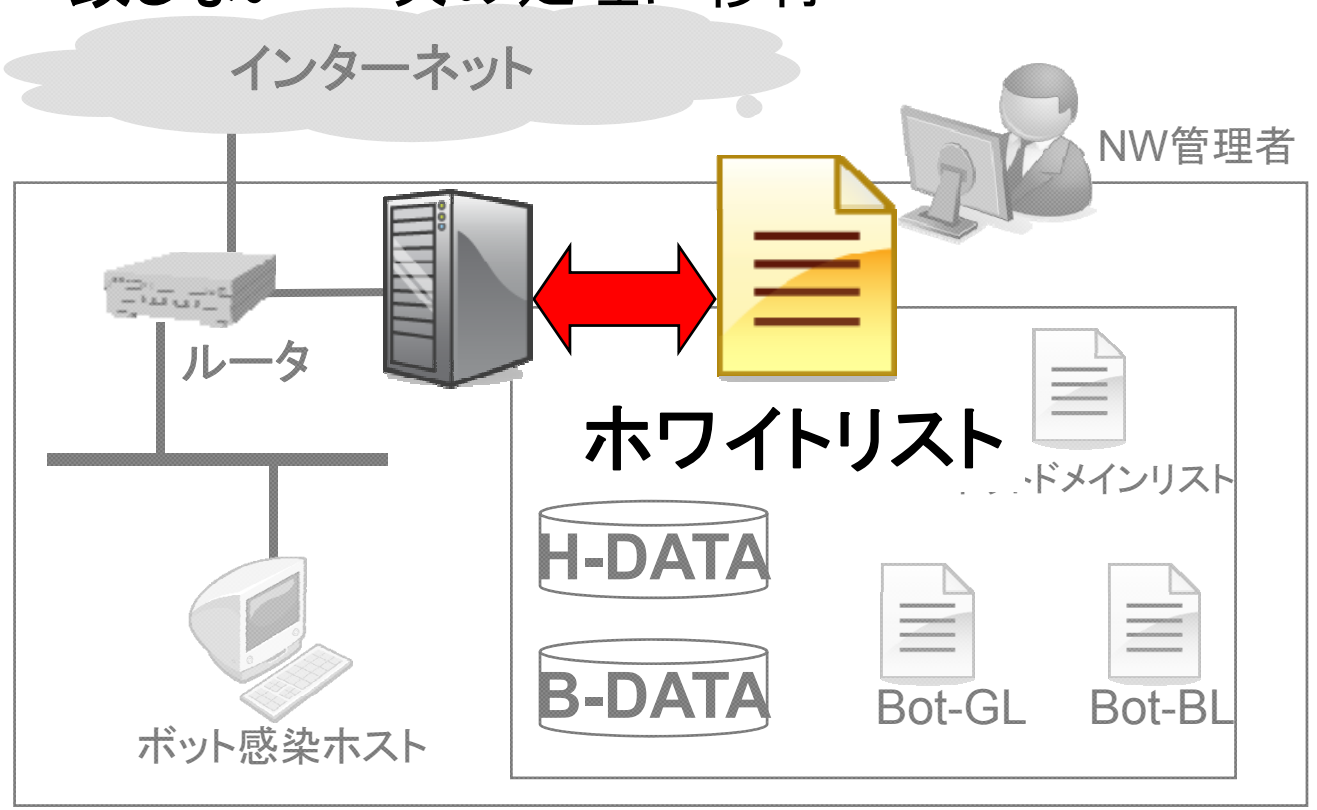
4. 第二段トレースバックシステム フロー



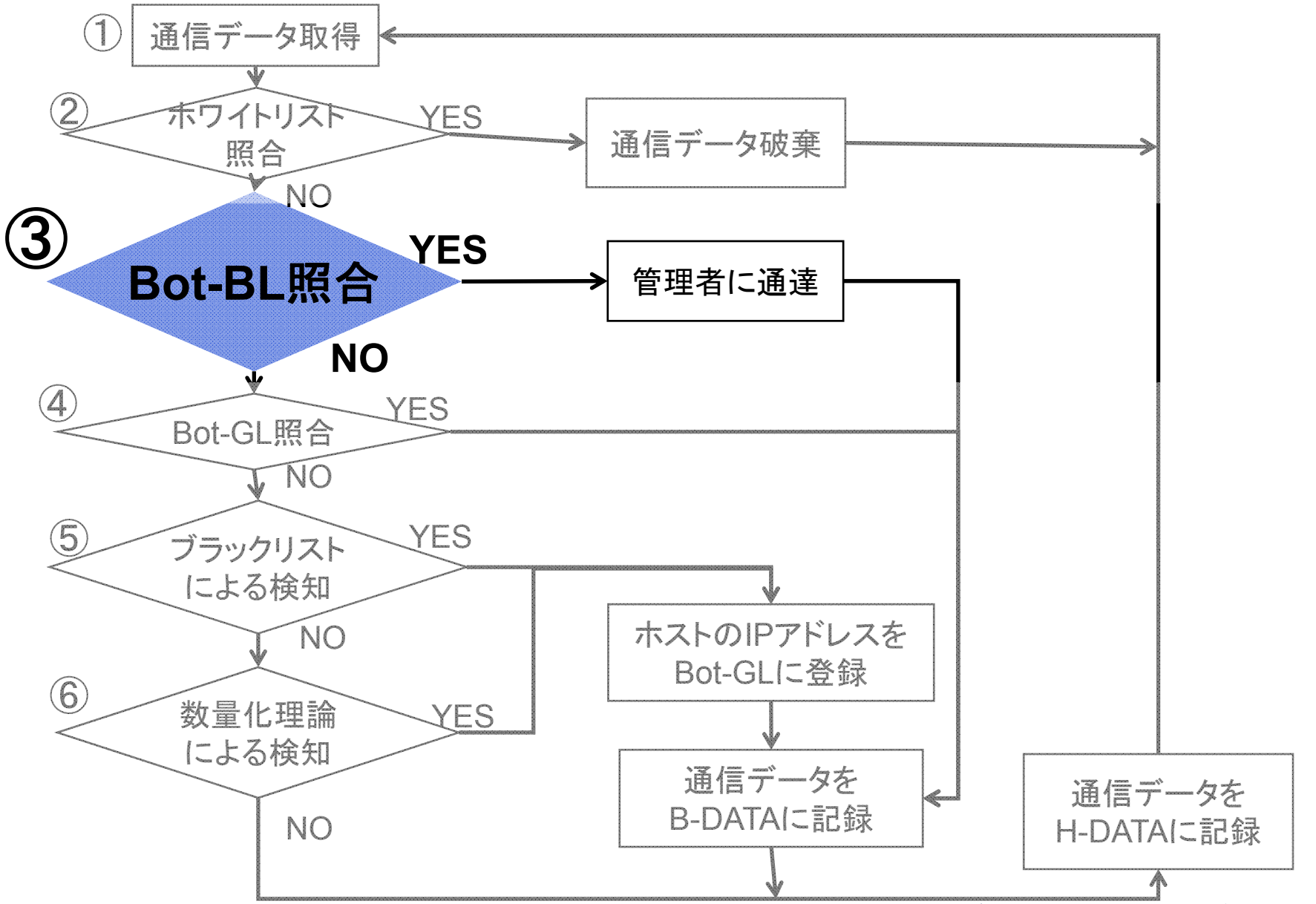
4. 第二段トレースバックシステム構成

2. 取得したIPアドレスとホワイトリストを照合

- 一致する　　安全な通信であるとみなす
- 一致しない　　次の処理に移行



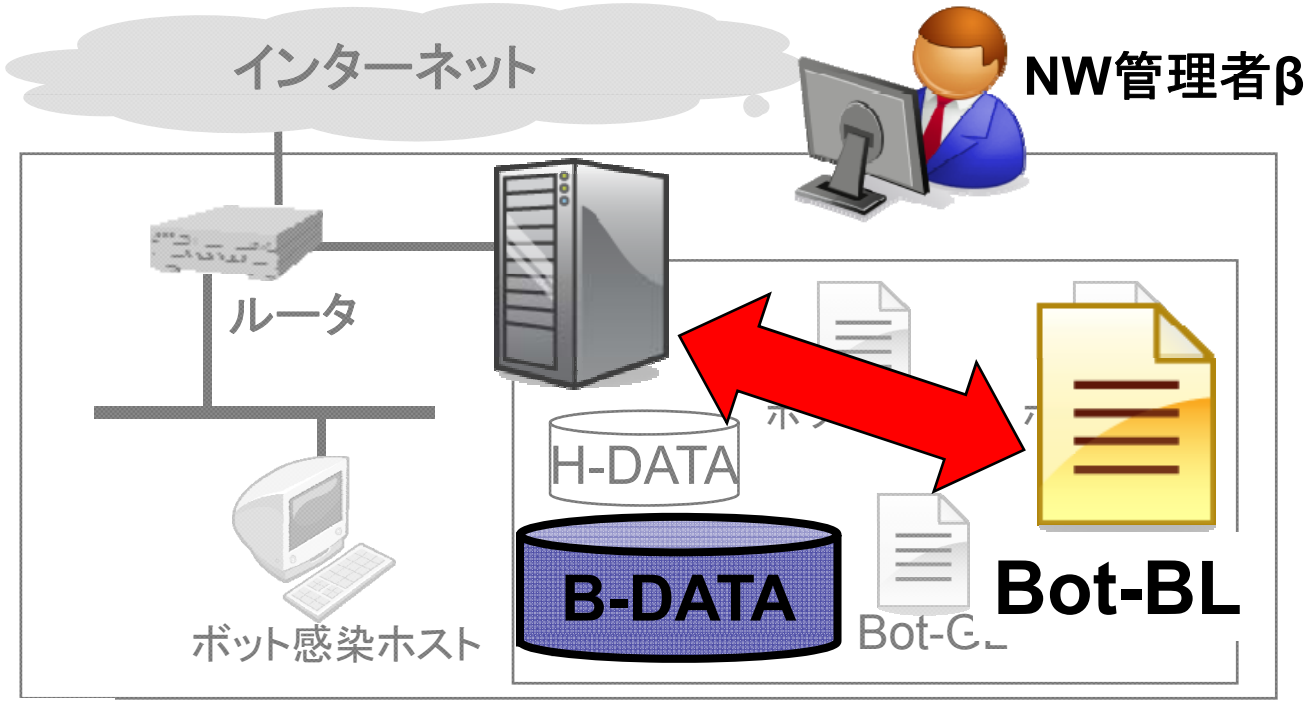
4. 第二段トレースバックシステム フロー



4. 第二段トレースバックシステム構成

3. 取得したIPアドレスとBot-BLを照合

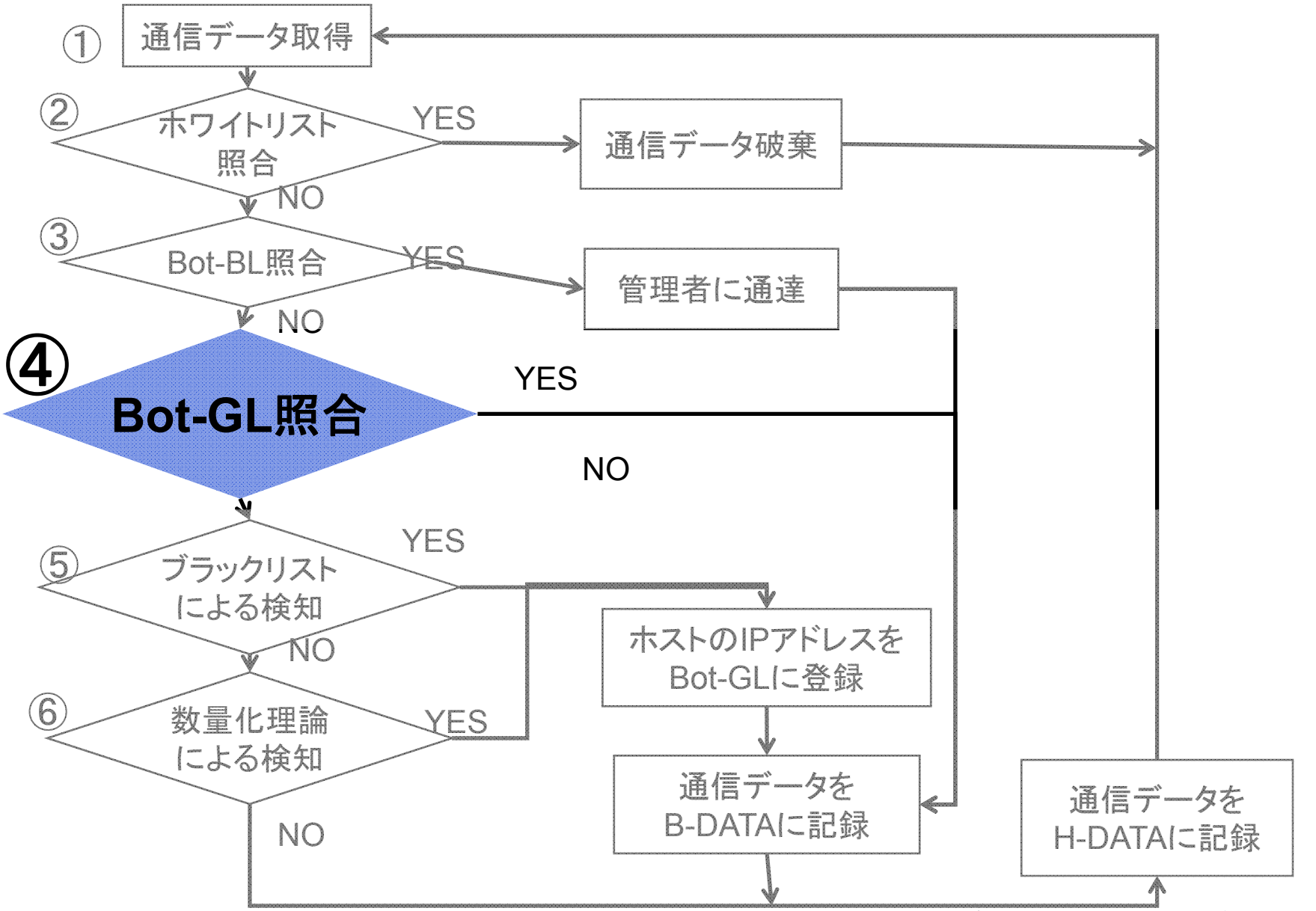
- 一致する　　通信データをB-DATAに記録・管理者通達
- 一致しない　次の処理に移行



Bot-BL

第二段トレースバックシステムがC&Cサーバ/DLであると判別したIPアドレスリスト

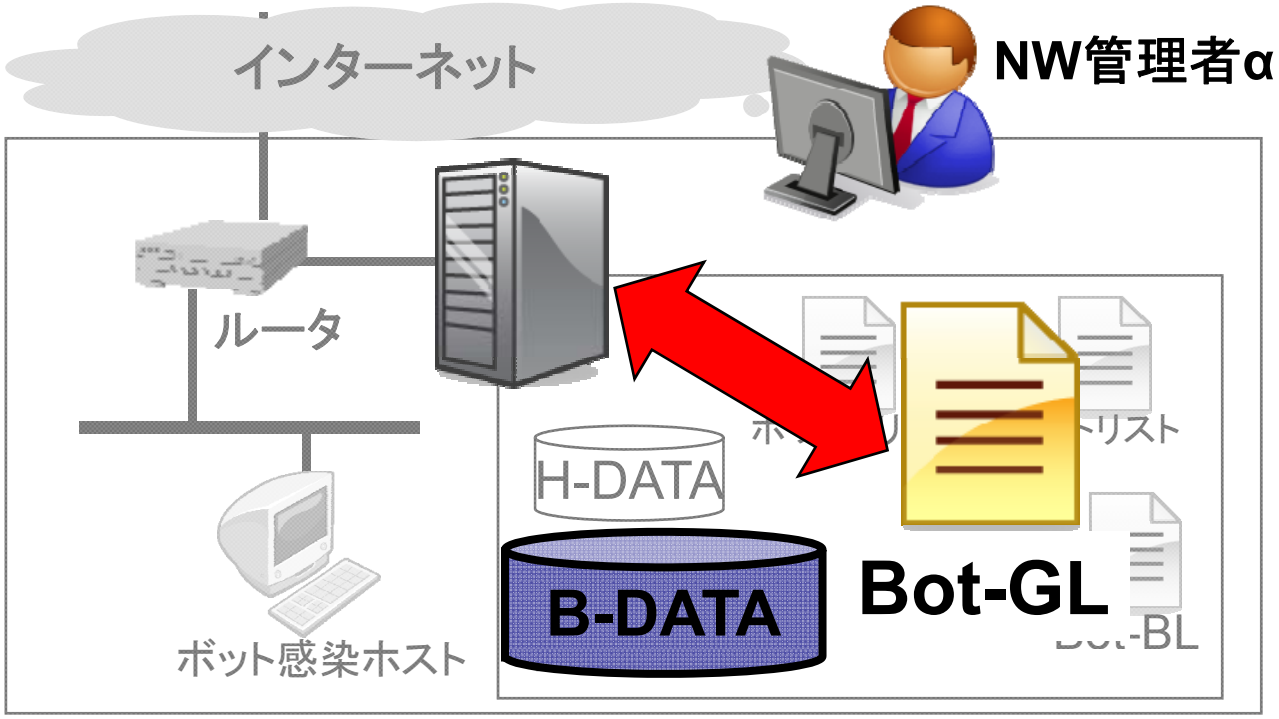
4. 第二段トレースバックシステム フロー



4.　第二段トレースバックシステム構成

4. 取得したIPアドレスとBot-GLを照合

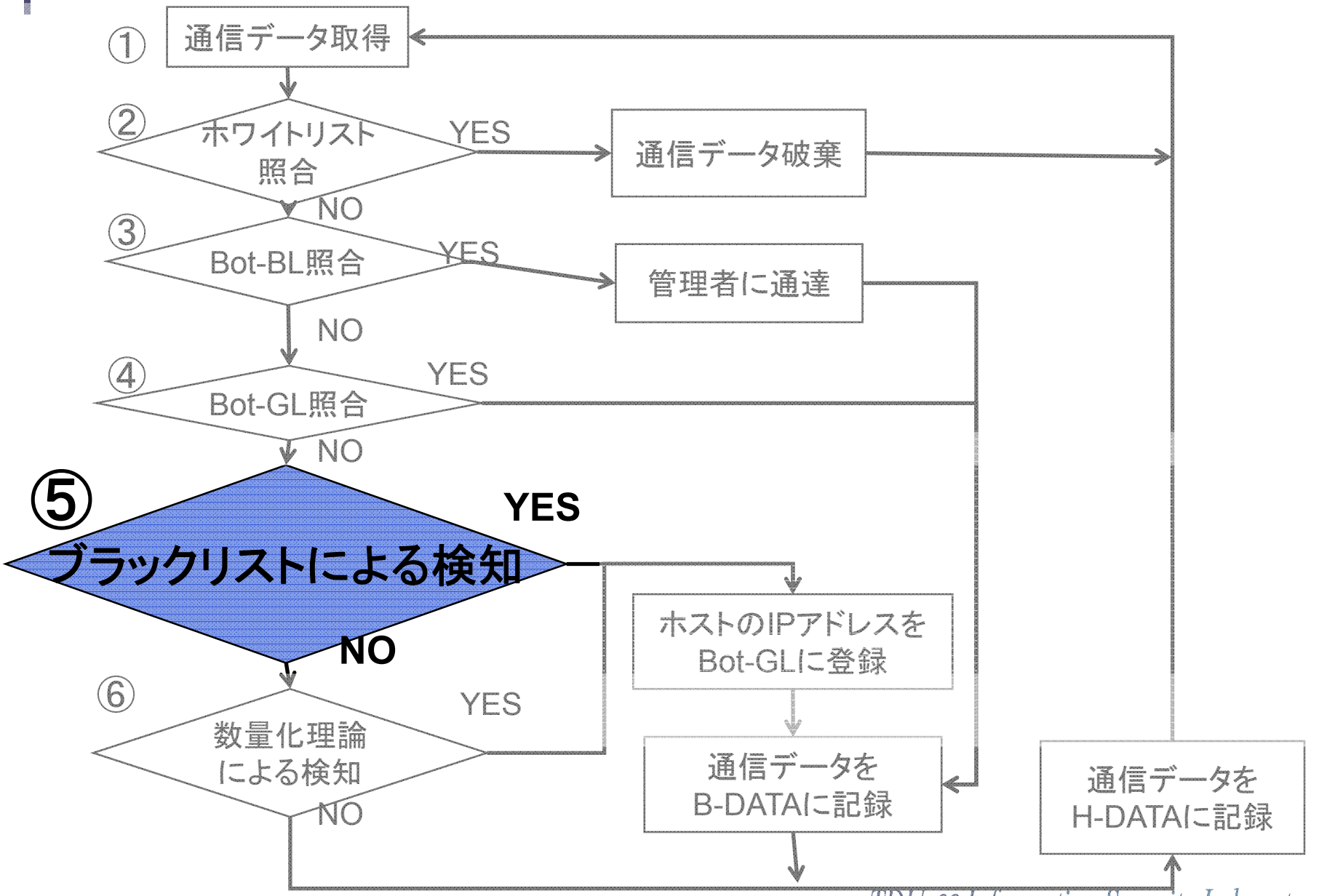
- 一致する　　通信をB-DATAに記録・管理者通達
- 一致しない　　次の処理に移行



Bot-GL

第二段トレースバックシステムが疑わしいと判別した
ホストのIPアドレスリスト

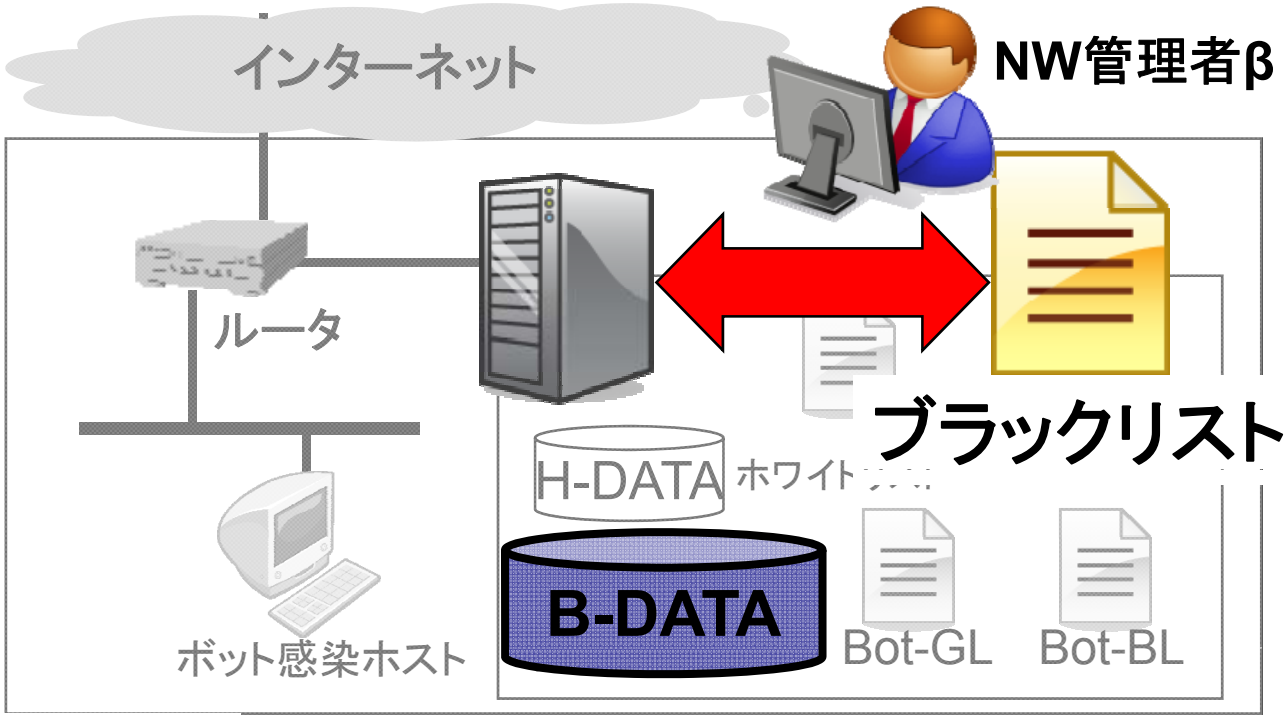
4. 第二段トレースバックシステム フロー



4. 第二段トレースバックシステム構成

5. ボットリストを用いたブラックリスト検知処理

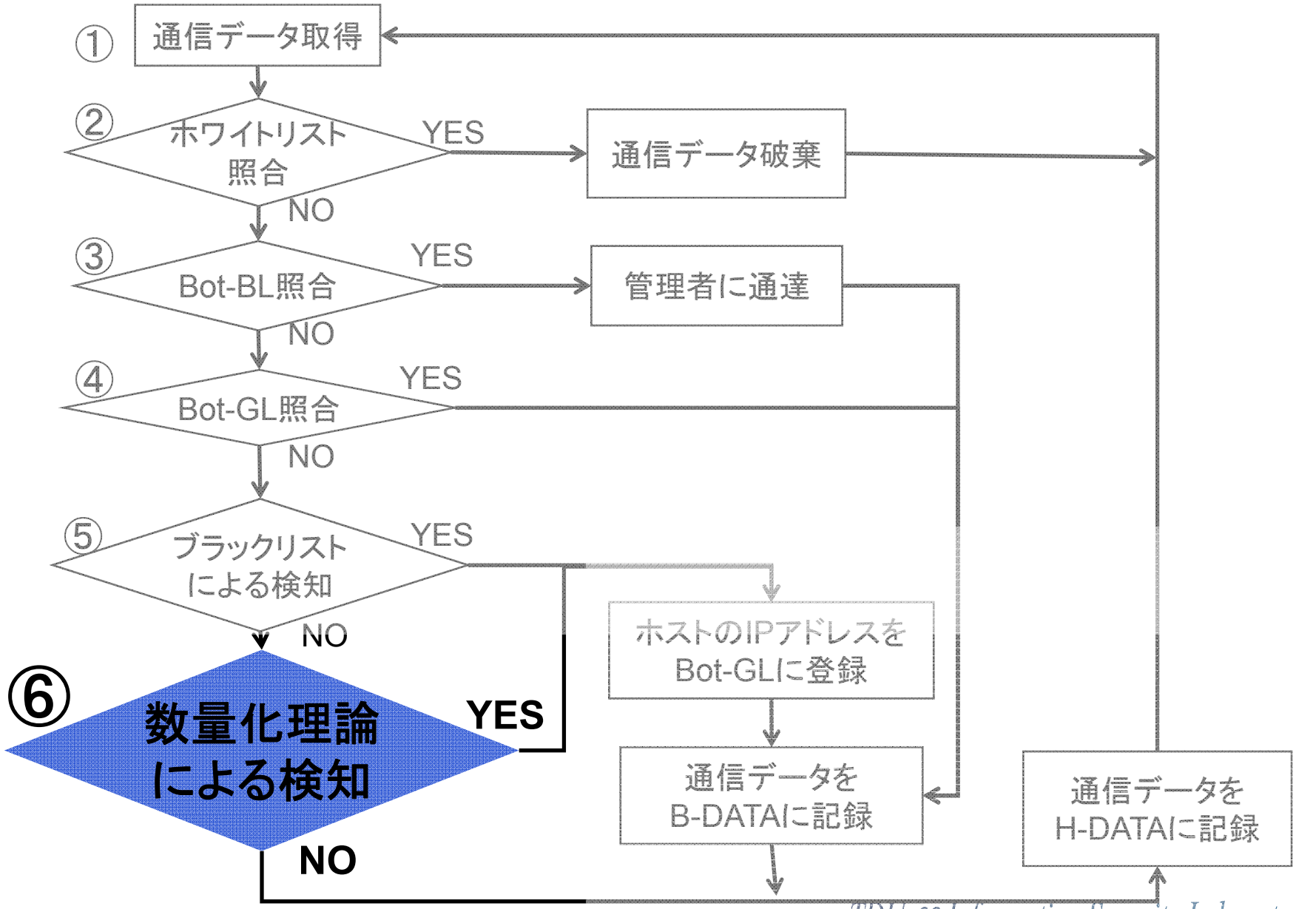
- 検知　通信をB-DATAに記録/IPをBot-GLに記録/ 管理者通達
- 無検知　　次の処理に移行



ブラックリスト

インターネット上から取得したボットネットに関するドメインのリスト

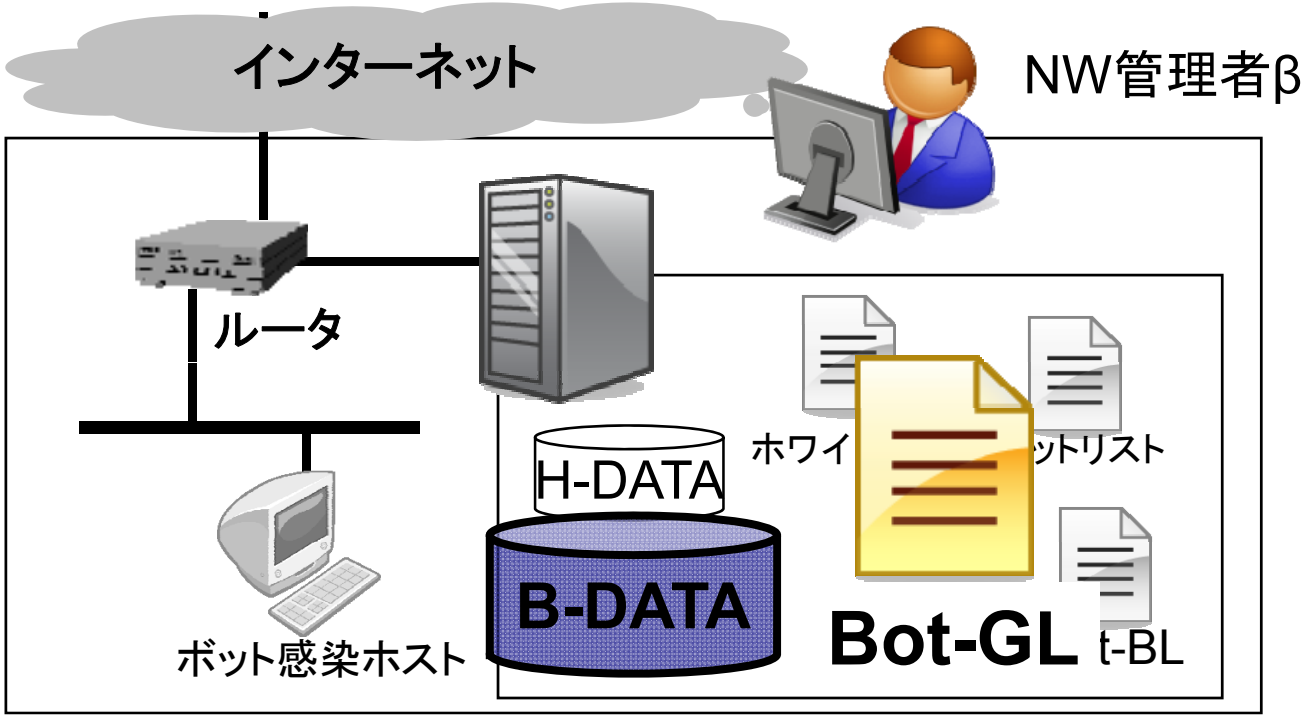
4. 第二段トレースバックシステム フロー



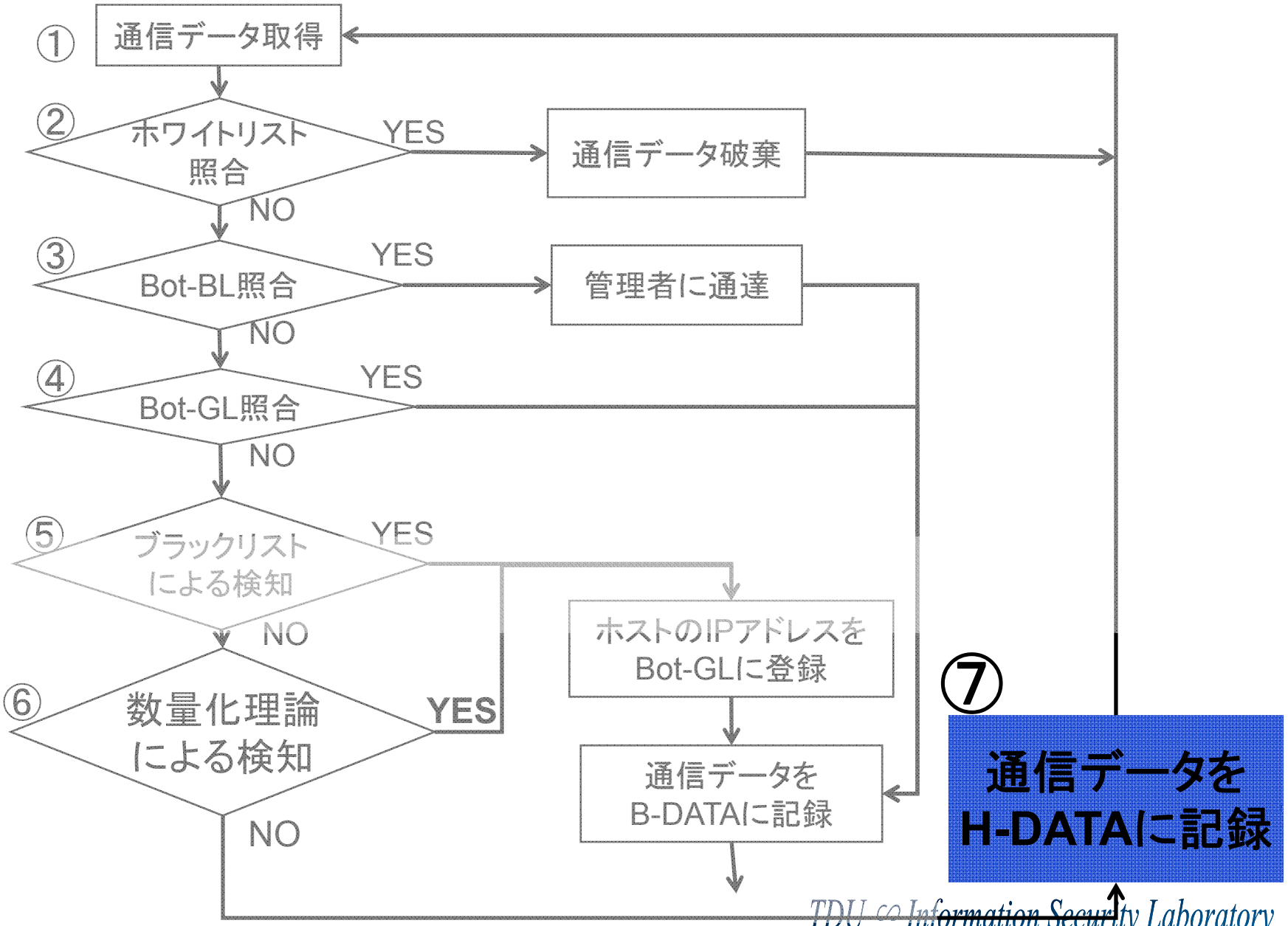
4. 第二段トレースバックシステム構成

6. 数量化理論2類を用いた検知処理

- 検知　　IPアドレスをBot-GLに登録/ 管理者通達
  通信データをB-DATAに記録
- 無検知　次の処理に移行



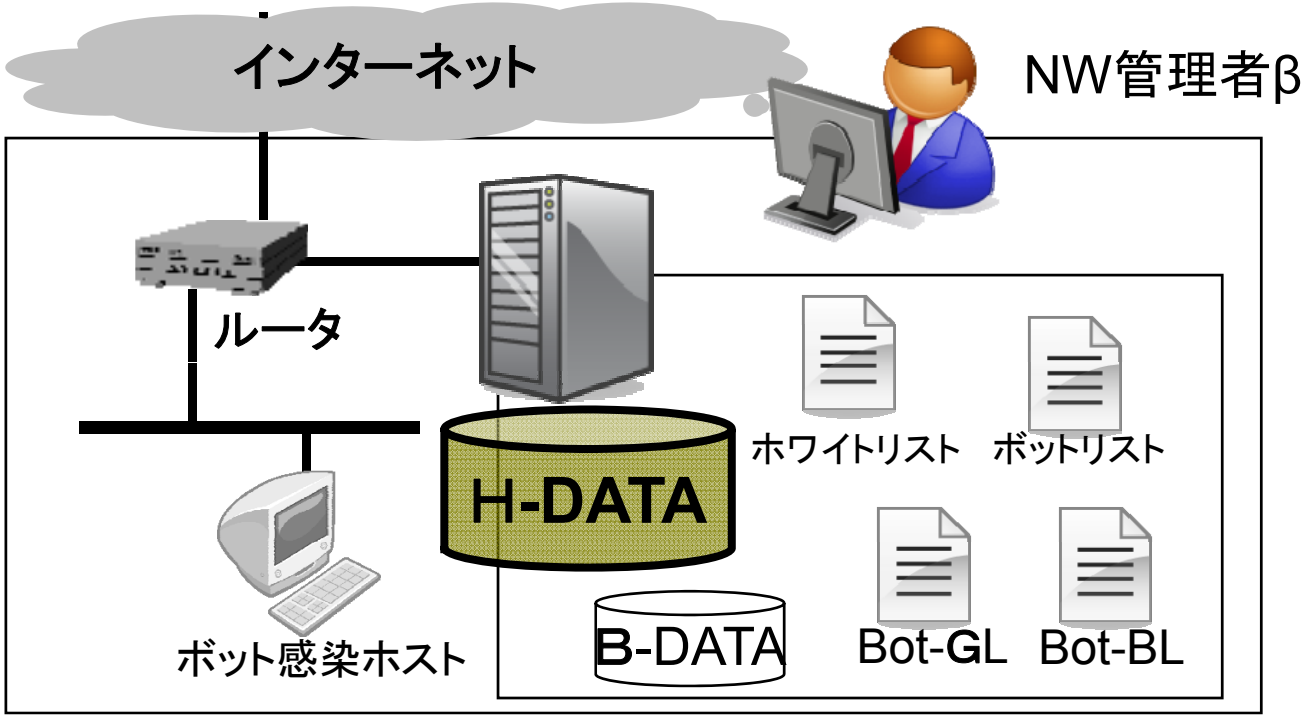
4. 第二段トレースバックシステム フロー



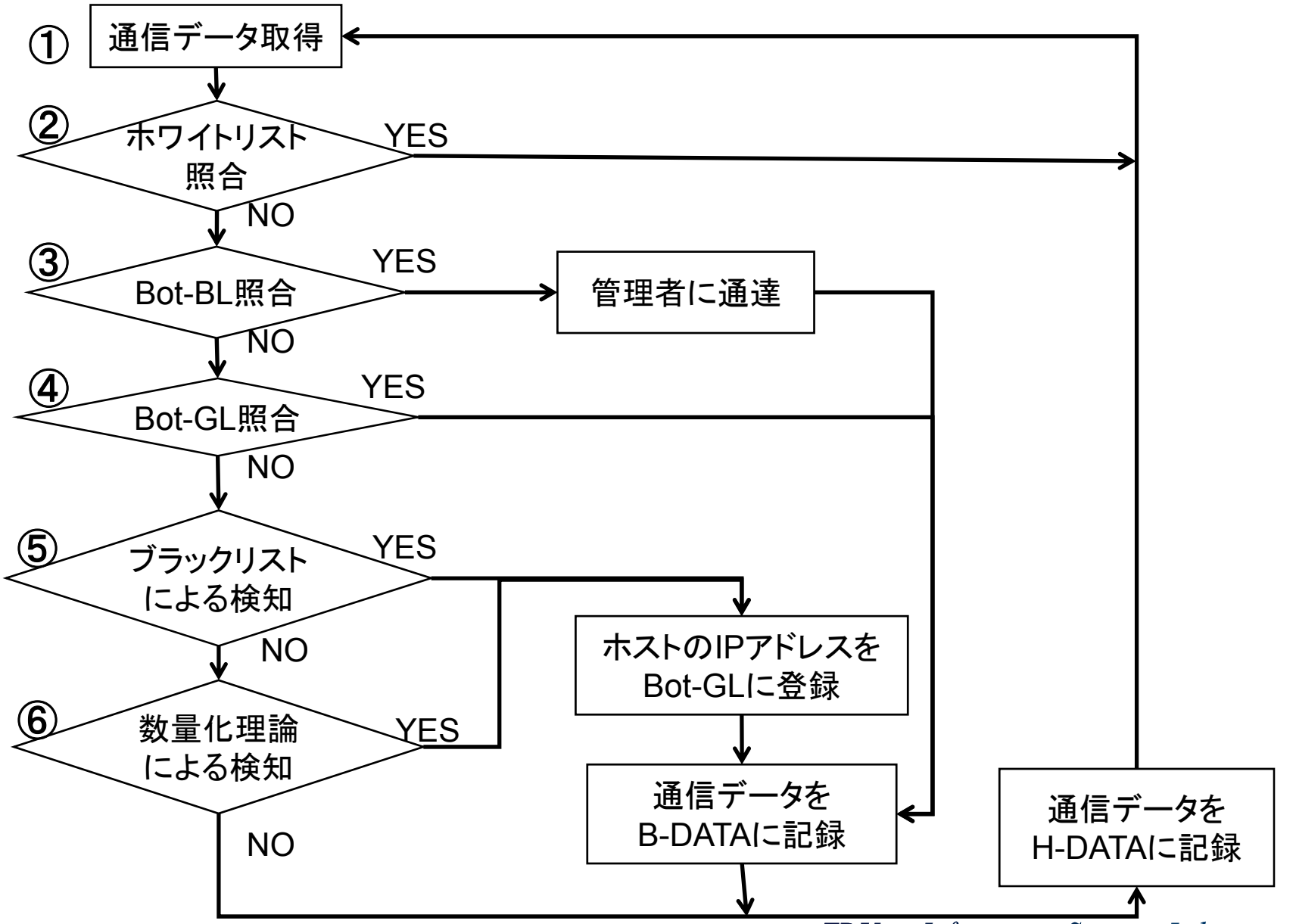
4. 第二段トレースバックシステム構成

7. 全ての通信データの記録

H-DATAに全ての通信記録を一定時間記録
第一段からの攻撃連絡時に対応



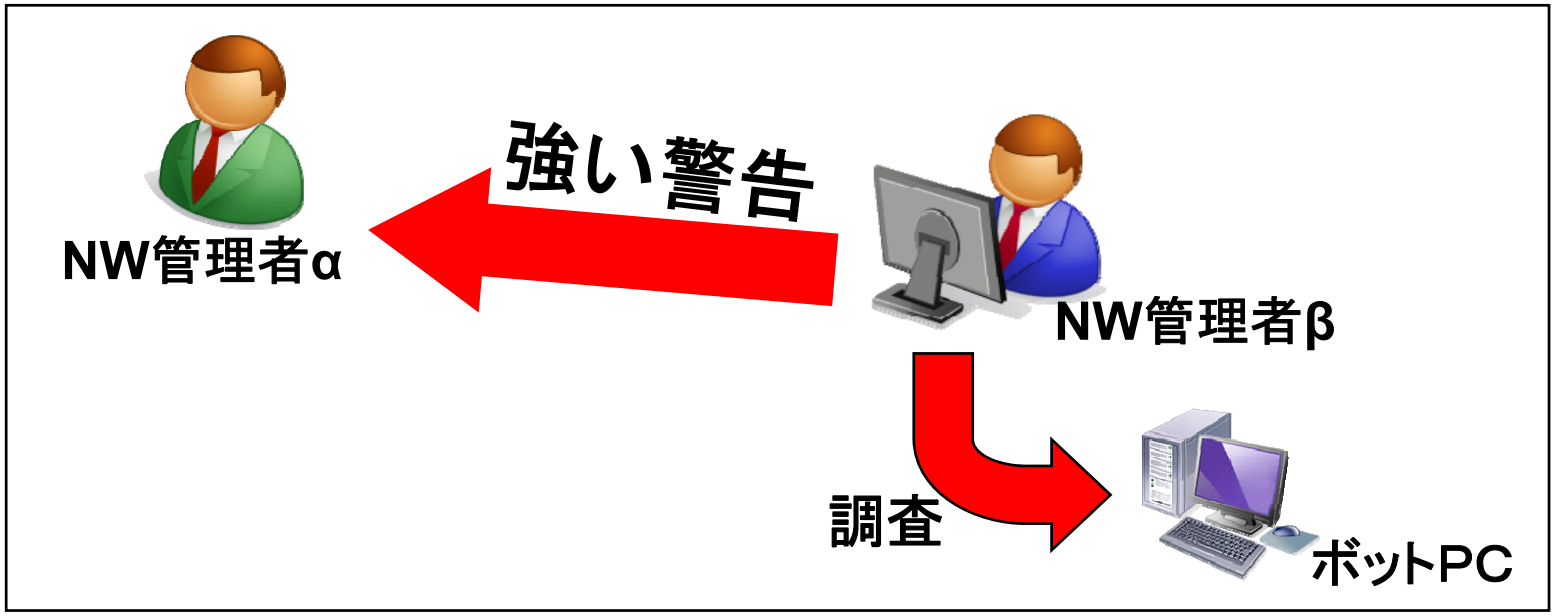
4. 第二段トレースバックシステム フロー



4. 第二段トレースバックシステム構成

第三段トレースバックとの連携①

Bot-BLに登録されたホスト処理



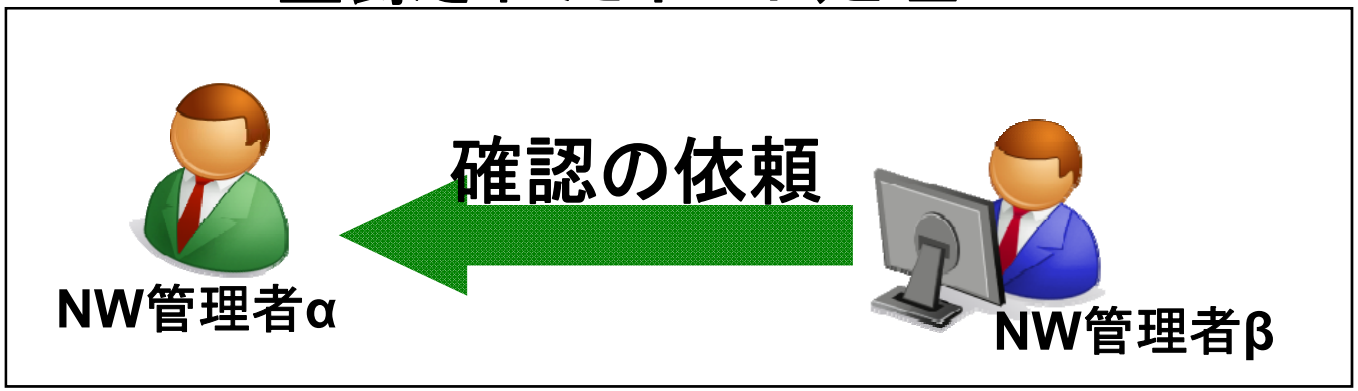
・NW管理者βはNW管理者αに強い警告を行う

・NW管理者βは自NW内のBot-BLと通信したホストの調査を行う

4. 第二段トレースバックシステム構成

第三段トレースバックとの連携②

Bot-GLに登録されたホスト処理



NW管理者αからの確認の返答

・該当ホストは正常(誤検知)

→ 該当するホストをBot-GLから削除

・該当ホストはC&Cサーバ/DLである

→ 該当するホストをBot-BLに登録
該当ホストと通信を行った自NW内ホストの調査

4. 第二段トレースバックシステム構成

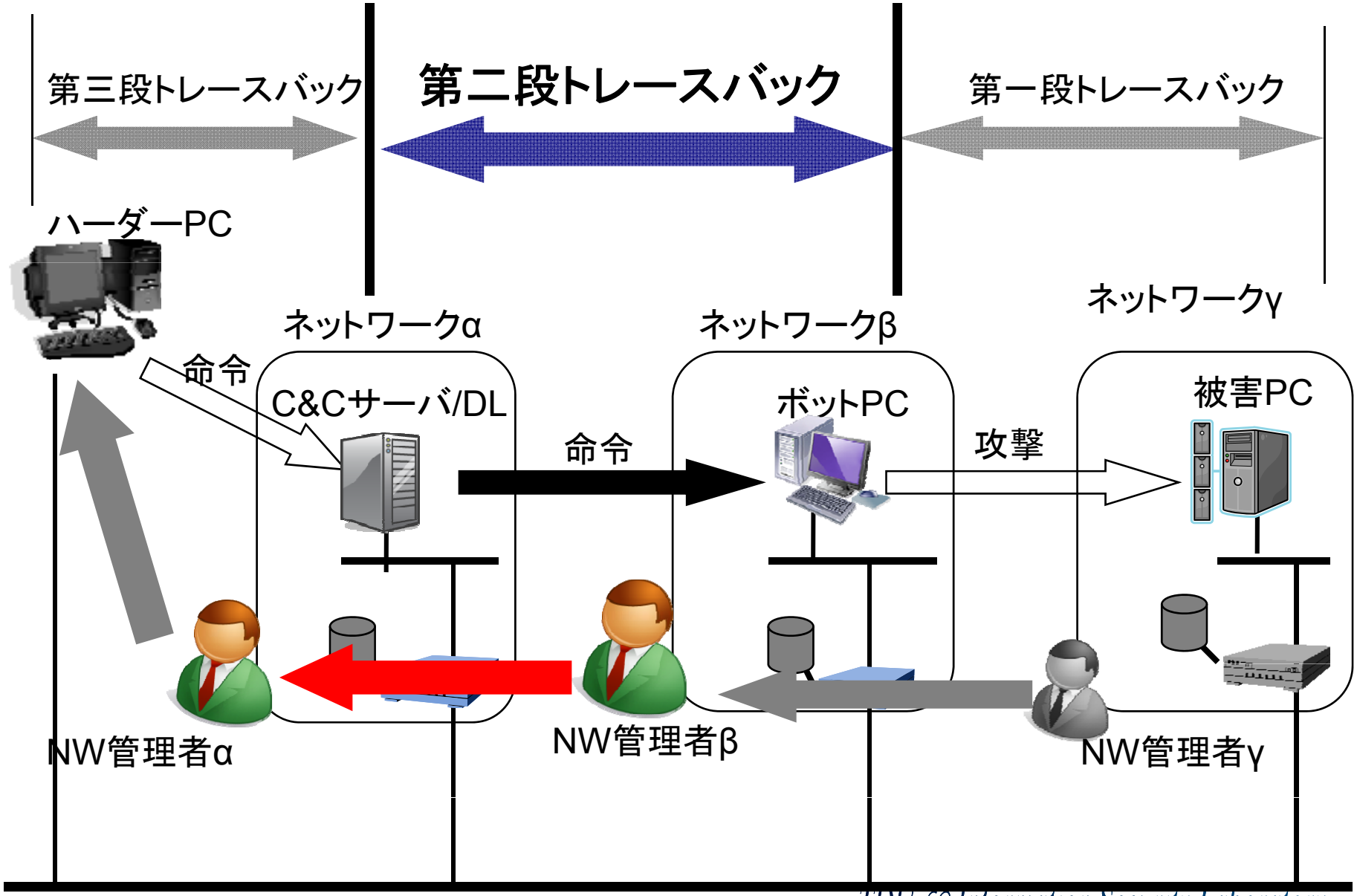
第一段トレースバックとの連携

攻撃ホスト連絡時の処理

・NW管理者γから攻撃の通達を受けたNW管理者βは
通達を受けた時点で調査を行う

➡ B-DATA及びH-DATAの通信記録にて対応

4. 多段追跡システム概要



目次

1. はじめに
2. CCCDATAset2009の解析結果
3. 実験
4. 提案システム概要
5. まとめと今後

5. まとめと今後

- まとめ
  - 多段追跡システムの，第二段トレースバックシステムを提案
  - 数量化理論2類に用いる最適なパラメータの選定，その有用性を確認
- 今後
  - 第二段トレースバックシステムの実装と評価
  - 数量化理論2類以外の方法を用いた結果の比較

御清聴ありがとうございました