高速化する通信における情報セキュリティのためのデータ処理方法

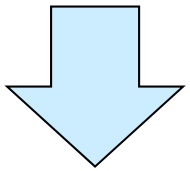
名古屋大学 情報基盤センター
情報基盤ネットワーク研究部門
嶋田 創

概要

- 背景: 近年の情報セキュリティ問題(サイバー攻撃対策)
  - 近年の高度化したサイバー攻撃の例
  - サイバー攻撃の防御とその課題
  - サイバー攻撃防御側の希望
- 高速かつ大量な通信データの処理によるセキュリティ向上
  - アノマリ検知
  - 通信データ処理のハードウェア化

Q: なぜサイバー攻撃が行われるのか

- 厳密にはサイバー攻撃を利用した犯罪(サイバー犯罪)
- 疑問
  - その攻撃手段への発想力を活かせば高収入で社会的地位の高い職業につけるのでは?



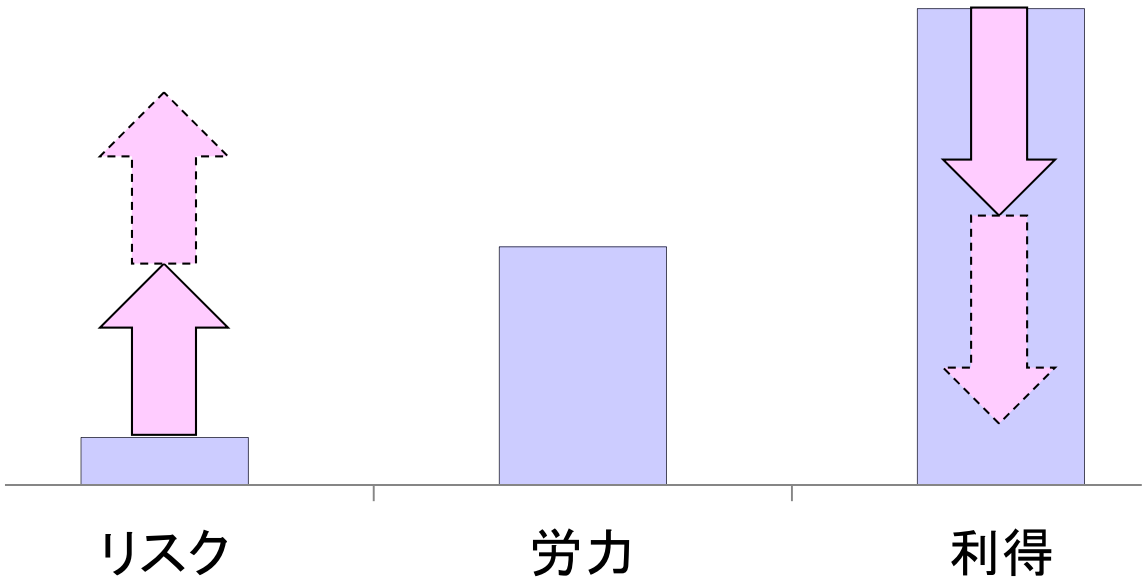
- A: 金になるから
  - 普通の職業についていたら一生かかっても稼げない金が手にはいるなら?

A: 金になるから

- クレジットカード情報
  - 悪用の他に、ブラックマーケットで売るという手も
    - フルセット揃うと1つにつき数十$程度
- 銀行(オンラインバンキング)決済情報
- 企業秘密
  - 他にも、どうしても手に入らない技術を手に入れるためとか
  - 某国は軍事技術に利用できる技術を一生懸命盗もうとしています
- 脅迫ネタ
  - 機密情報を手に入れた
  - サービス不能(DoS)攻撃をかけるぞ

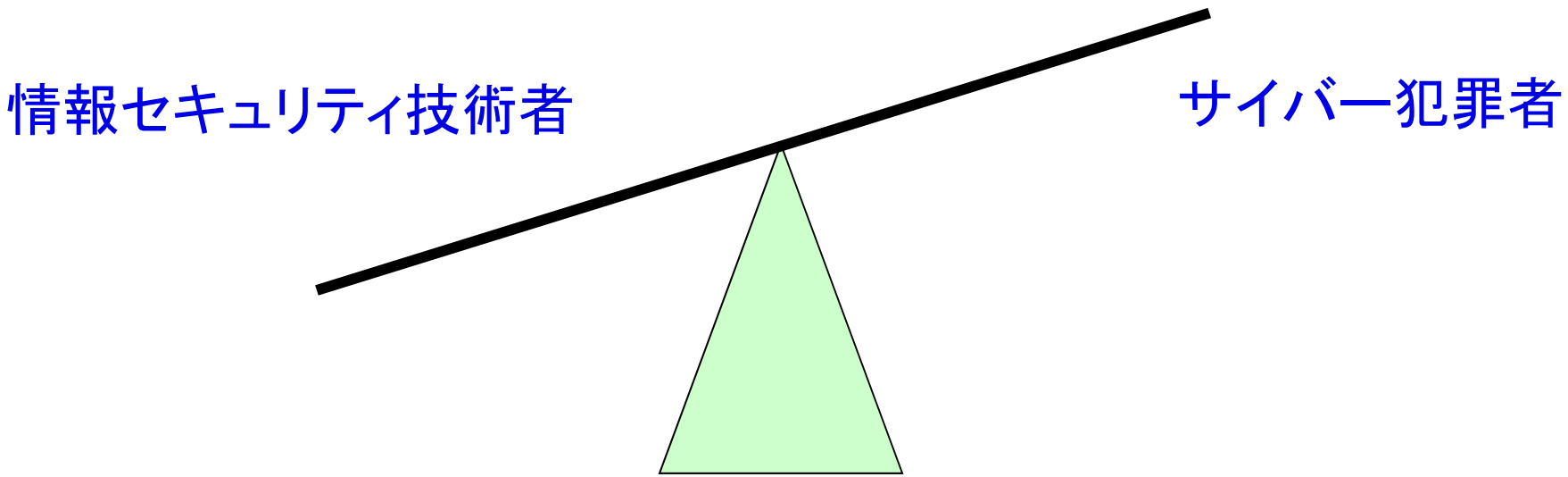
現状のサイバー犯罪

- 現在のサイバー犯罪はリスクに対して利得が大きすぎ
- リスクを上げて利得を減らす必要がある
- ただし、サイバー犯罪は世界規模なので、リスクを上げれない国家があることを考えておく必要がある



現状の勢力バランス

- 情報セキュリティ技術者の方が分が悪い
- そもそも後手後手に回ることになる
  - 犯罪者側は未発見の攻撃手段を1つ見つければ良い
  - 犯罪者は市販の情報セキュリティ技術を試せる



近年のサイバー攻撃の特徴

- 愉快犯的な物はほぼ無くなった
  - かつてのコンピュータウイルスのような拡散は少ない
  - 悪さをする人が便利なようにどんどん改良 →マルウェア
- 手間暇かけるようになった
  - 事前に目標の挙動を確認して罠をしかける
    - 目標への侵入後も即目的の活動をしない場合もあり
  - 某標的型攻撃では数ヶ月かかって目標達成
- 活動は静かに行われる
  - 発見されるのを防ぐため
  - 発見防止のために証拠隠滅までやる

マルウェアとは

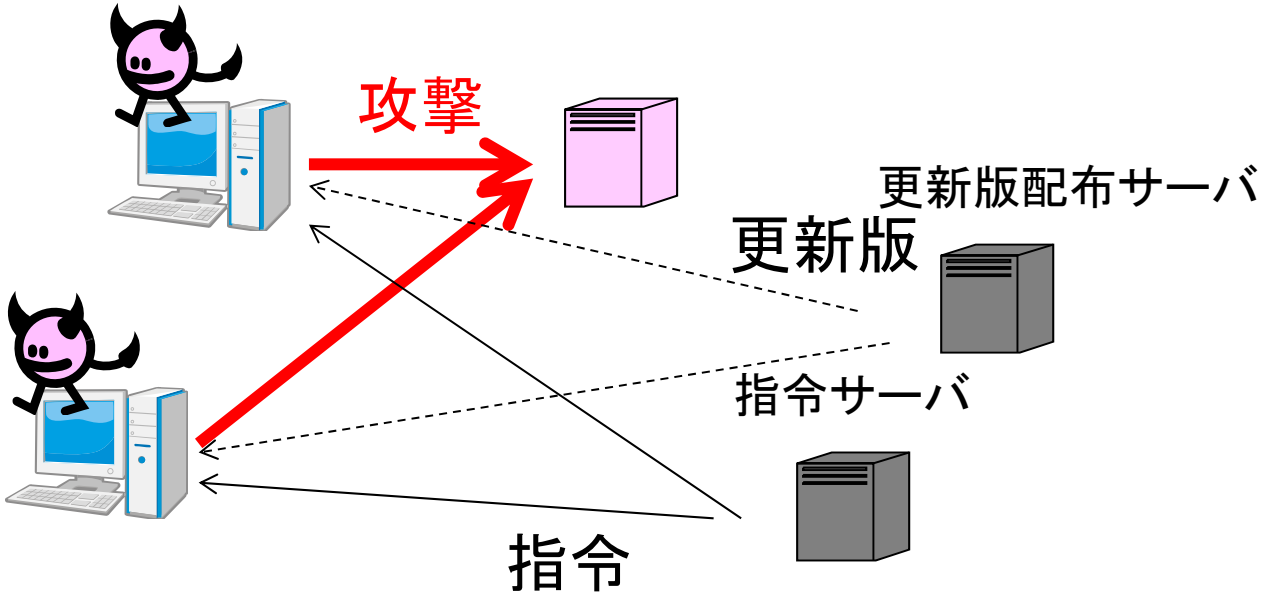
- MALicious softWARE(悪意のあるソフトウェア)の略
- かつてはよくコンピュータウイルスと呼ばれていたが、最近はマルウェアと称すことが多い
- コンピュータウイルスとの違い
  - 愉快犯や技術誇示からサイバー犯罪の道具へ
  - おおっぴらに感染/拡散しない
    - 特定のグループ/ネットワークのコンピュータにのみ感染
    - そもそも、あまりばらまくと発見される可能性が高くなる
  - おおっぴらに怪しい通信したりしない
    - 他の通信にまぎれて通信したりします
  - おおっぴらに破壊活動をしたりしない
    - 発見されると証拠隠滅することはあります

マルウェアの分類

- RAT(Remote Access Trojan)
  - 遠隔で感染したPCを操作可能な形にする
  - トロイの木馬、ボットネットクライアント、などもこれに分類
  - 自分が加害者になる点が怖い
- スパイウェア
  - 金融関係情報や各種サービス用ユーザ名/パスワードの窃取
  - キーロガーやスクリーンショット取得などの機能
- ダウンローダ
  - Drive-by Download攻撃の途中で利用
- 昔ながらのもの
  - ウイルス: 無差別に近い拡散、PCに何らかの以上を発生させる
  - ワーム: 増殖することに特化

RAT(Remote Access Trojan)

- トロイの木馬、バックドア作成、踏み台ツールの発展
- 指令を受け取って攻撃などの動作を取る
  - 昔はIRC経由が多かったが、マークされるようになったので最近はHTTPやHTTPS経由で指令受信
  - 後述するDDoS攻撃などに利用
  - 遠隔で自分自身を更新することも可能



代表的なRAT: Poison Ivy

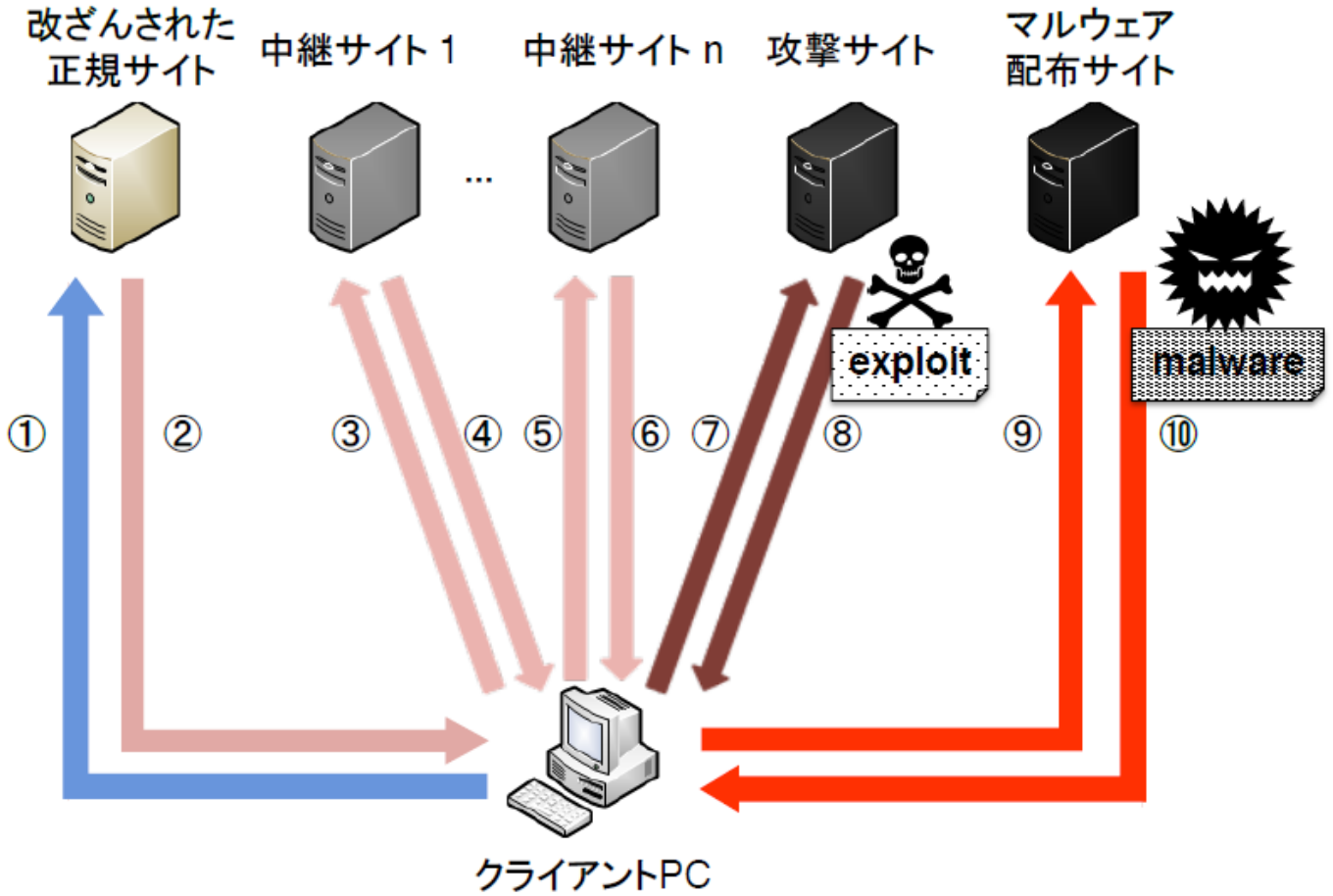
- バージョンアップしながら今も利用されている
- 機能
  - スクリーンショット、音声、Webカメラの画像の取得
  - アクティブなポートの表示
  - キー入力操作情報の収集
  - 開いているウィンドウの管理
  - パスワードの管理
  - レジストリ、プロセス、サービス、デバイス、インストールされているアプリケーションの管理
  - ファイル検索、同時に多数のファイル移動の実行
  - リモートシェルの実行
  - サーバの共有
  - 自身の更新、再起動、終了

マルウェアの送り込み方

- 昔ながらのメール
  - 本体に添付することは減ってきてウェブからのダウンロードが中心に
  - 複数のダウンロードを繰り返すDrive-by Download攻撃も
- ウェブからのダウンロード
  - 攻略されたウェブサイトから配布
  - ウェブ広告にまぎれて配布
- 標的型攻撃(APT: Advanced Persistent Threat)
  - 近年話題の攻撃
  - 標的(機密情報サーバなど)にマルウェアを入れるまでに複数の踏み台PCを経由
  - 応用: 水飲み場型攻撃
    - 特定のユーザがよく見るウェブサイトにマルウェアをしかける

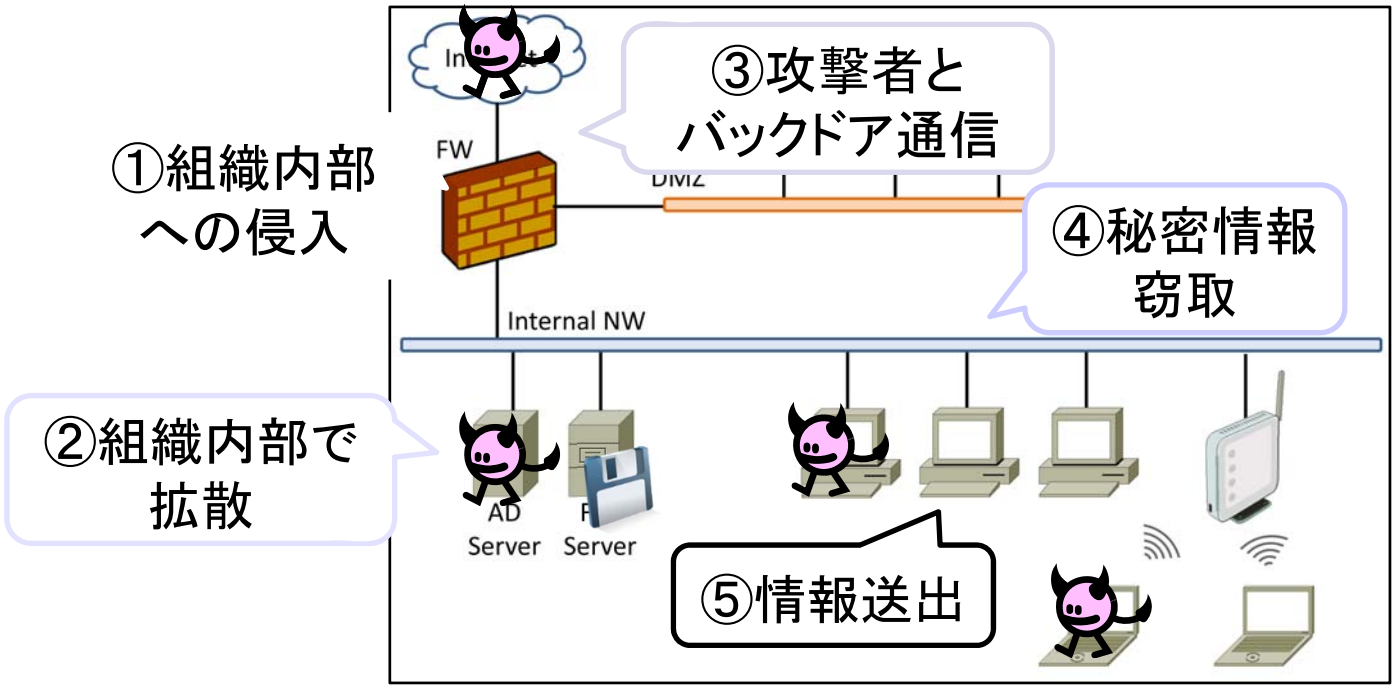
Drive-by Download攻撃

- 複数のダウンロードによってマルウェア配布を欺瞞/隠蔽



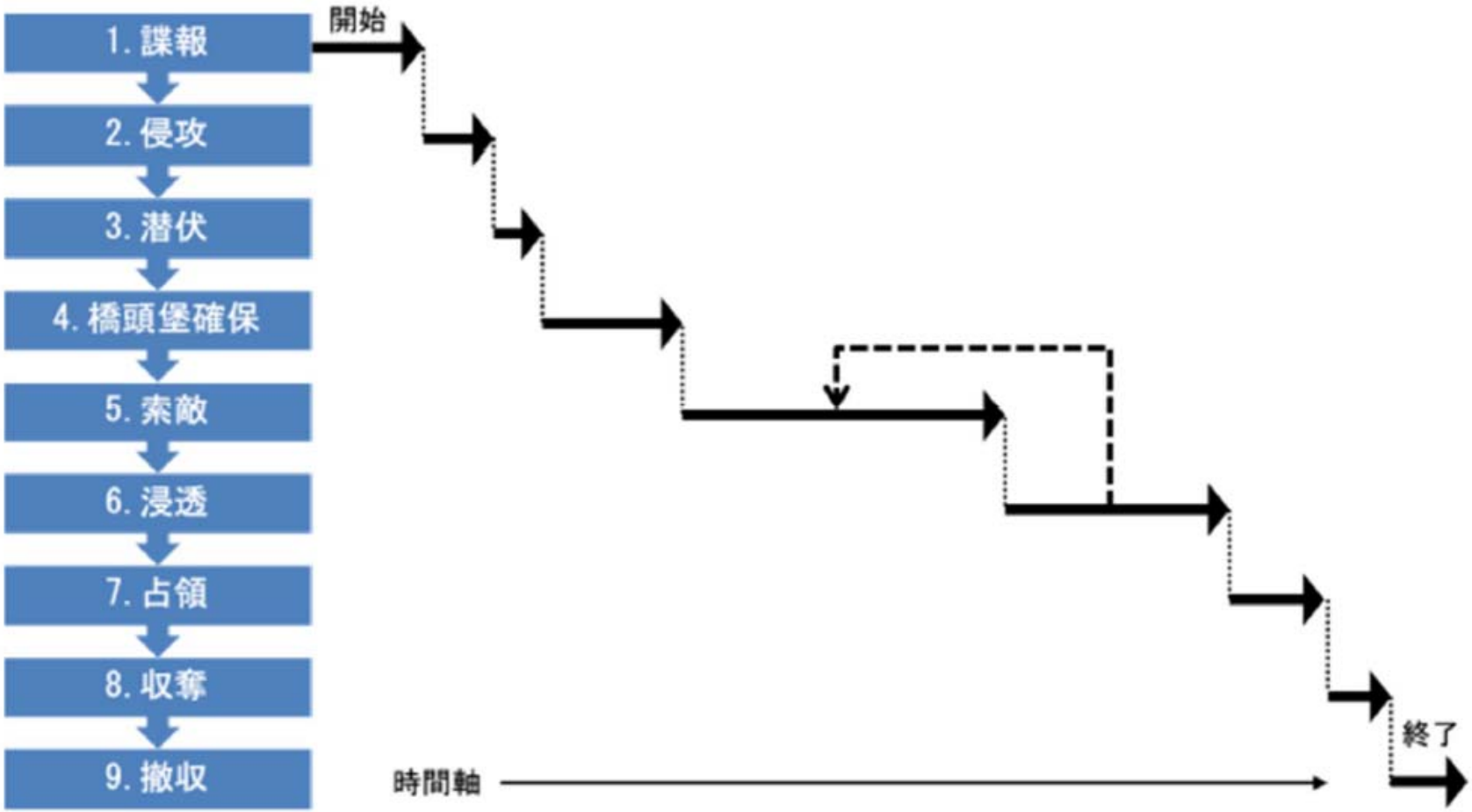
標的型攻撃の動作

- 非常にざっくり書くと5段階
  - 侵入前に組織の内部構成を調査することもある
  - 組織内での拡散において、潜伏、索敵を行うこともある



標的型攻撃の時間軸

- 長いものだと一連の攻撃に数ヶ月かける



標的型攻撃の実例(1/3)

三菱重工への標的型攻撃

- 発覚: 2011/8/11にサーバが再起動を繰り返すため
- 影響範囲
  - サーバー 45台、従業員用PC 38台
  - 8種類のウイルスを発見
  - 11の事業所から発見
- 発端: “原発のリスク整理”という添付ファイル
  - 東日本大震災(2011/3)の直後
  - Adobe Flashの脆弱性を利用
  - 送信元は内閣府実在の人物の名前、メールアドレスを騙る
  - 三菱重工は原発を作っている(いた)ので、受け取った人は疑わない

標的型攻撃の実例(2/3)

JAXAへの標的型攻撃[1]

- 特に長期間に渡った例(1年8ヶ月にも及ぶ)
- 発見: 2012/11/21
- 発端: 2011/3/17
  - なりすましメールの添付ファイル

[1] http://www.jaxa.jp/press/2013/02/20130219_security_j.html

標的型攻撃の実例(3/3)

EmEditorアップデートファイルを利用した攻撃[1]

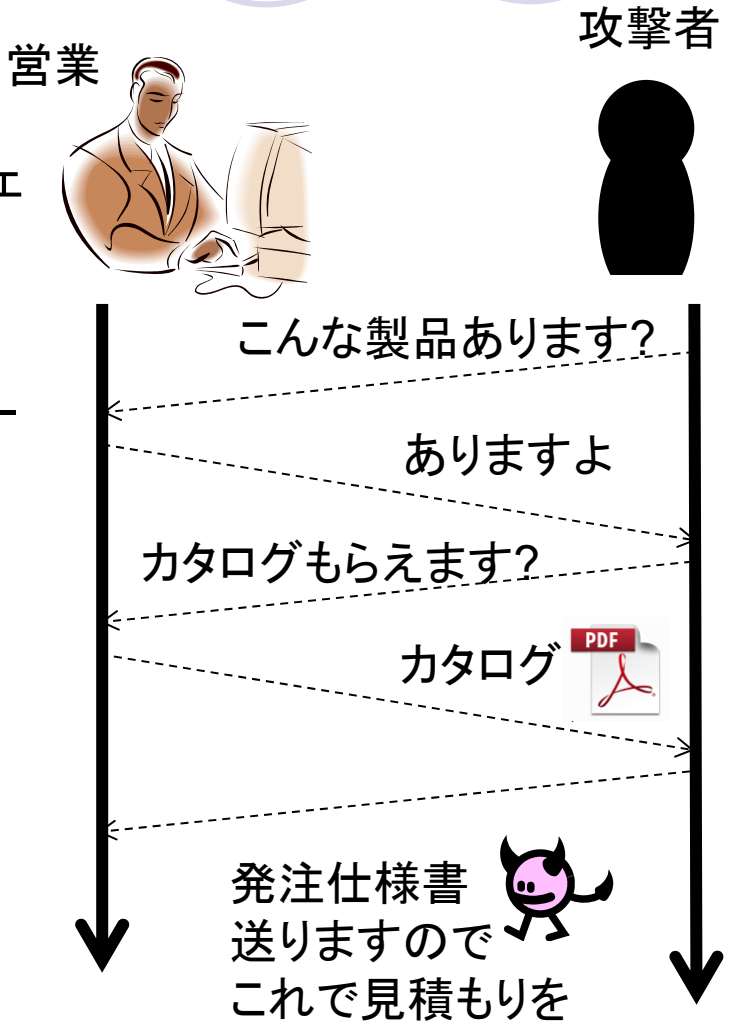
- 攻撃対象: 名古屋大学、JAXA、ISAS、朝日新聞、農林水産省など
- 以下の様な.htaccessファイルがアップデート配布ディレクトリに置いてあった
  - 指定したIPアドレスの範囲からアップデート要求があれば別ファイルを配布

```
SetEnvIf Remote_Addr “106¥.188¥.131¥.[0-9]+” install
SetEnvIf Remote_Addr “133¥.6¥.94¥.[0-9]+” install
(… 同様に70行 …)
SetEnvIf Remote_Addr “124¥.248¥.207¥.[0-9]+” install
RewriteEngine on
RewriteCond %{ENV:install} =1
RewriteRule (.*¥.txt)$ /pub/rabe/editor.txt [L]
```

[1] https://jp.emeditor.com/general/ 今回のハッカーによる攻撃の詳細について /

さらに発展した標的型攻撃

- やりとり攻撃
  - 複数回のメールのやりとりの後にマルウェア送付
- 水飲み場型攻撃
  - 「ある仕事をしている人が頻繁に見るページにマルウェアを仕掛ける」ことによる特定業種の業社への標的型攻撃
  - 例: 政府のある機関のプレスリリース、入札公告ページ
    - その機関に関連する会社に対して攻撃
    - さらにIPアドレスを制限する事例もある



標的型攻撃のマルウェア送り込み技術

- 基本的に従来の方法
- メール添付
  - 実行ファイル、ファイル実行脆弱性を利用した他のファイル
  - じゃあ、そのメールアドレス(足がつかない)の取得方法は?
- メール中でのURLの引き渡し
  - 偽ページ、マルウェア配布スクリプトを埋め込んだ(踏み台)ページへの誘導
  - じゃあ、そのURLへの仕掛け方法は?

攻撃用メールアカウント準備

- 従来だったら
  - セキュリティのゆるいフリーメールアドレスを利用する
  - ただし、あまりにも評判が悪くなると後述のブラックリストで対策される
- 近年では
  - そこそこメジャーな組織のメールアカウントを乗っ取って送信
- 無差別攻撃用
  - 従来同様にspam送信用メールサーバを(乗っ取ったPCを用いて)立てることが多い

メールアカウント乗っ取り

- メールアカウント以外にも、様々な物を乗っ取る
  - Apple ID, Google ID, FacebookなどのSNS
  - IDの共用で被害は広がる
- 近年よく見られる新しい
  - 事前に名寄せして個別サービスの同一ユーザを関連づけ
    - 例: evo0229, shimada0704, hajime1113
  - あるサービスが認証情報漏洩をした場合、他のサービスの同一ユーザアカウントで認証を試みる
- もちろん、マルウェア感染したPCを利用することも可能
  - 身に覚えの無いメール送信がいっぱい
- もちろん、標的型攻撃だけでなくspam配布にも使われます
  - 認証一発で通ってspamを数十万通送信しようとしたりする

攻撃用URL準備

- 従来なら
  - 適当なマシンでウェブサーバを走らせる
  - URLがIPアドレス直打ちなので、怪しまれたりする
- 最近では
  - 既存のウェブサーバに攻撃URL埋め込み
    - クロスサイトスクリプティング(CSS)
    - セキュリティの甘い広告サービスを利用
      - というか、Googleにも出てくることも[1]
  - ウェブサーバやコンテンツマネージメントシステム(CMS)の脆弱性を利用して、既存のウェブサーバに設置

[1] http://blog.livedoor.jp/blackwingcat/archives/1873202.html

コンテンツマネージメントシステム (CMS)

- コンテンツを登録するだけできれいなページを作成可能
  - データベースに登録されたコンテンツを、フォーマットして表示
- 代表的なCMSとシェア[1]
  - WordPress 60.6%
  - Joomla 8.0%
  - Drupal 5.2%
- バックエンドでDBを使っているので、SQLインジェクションなどの脆弱性が良く見つかる
- 本体だけでなく、プラグインにも脆弱性が見つかる
- PukiwikiなどのWikiエンジンなどもCMS
- 一番重要なこと: 最新のCMSを使いましょう

[1] http://w3techs.com/technologies/overview/content management/all

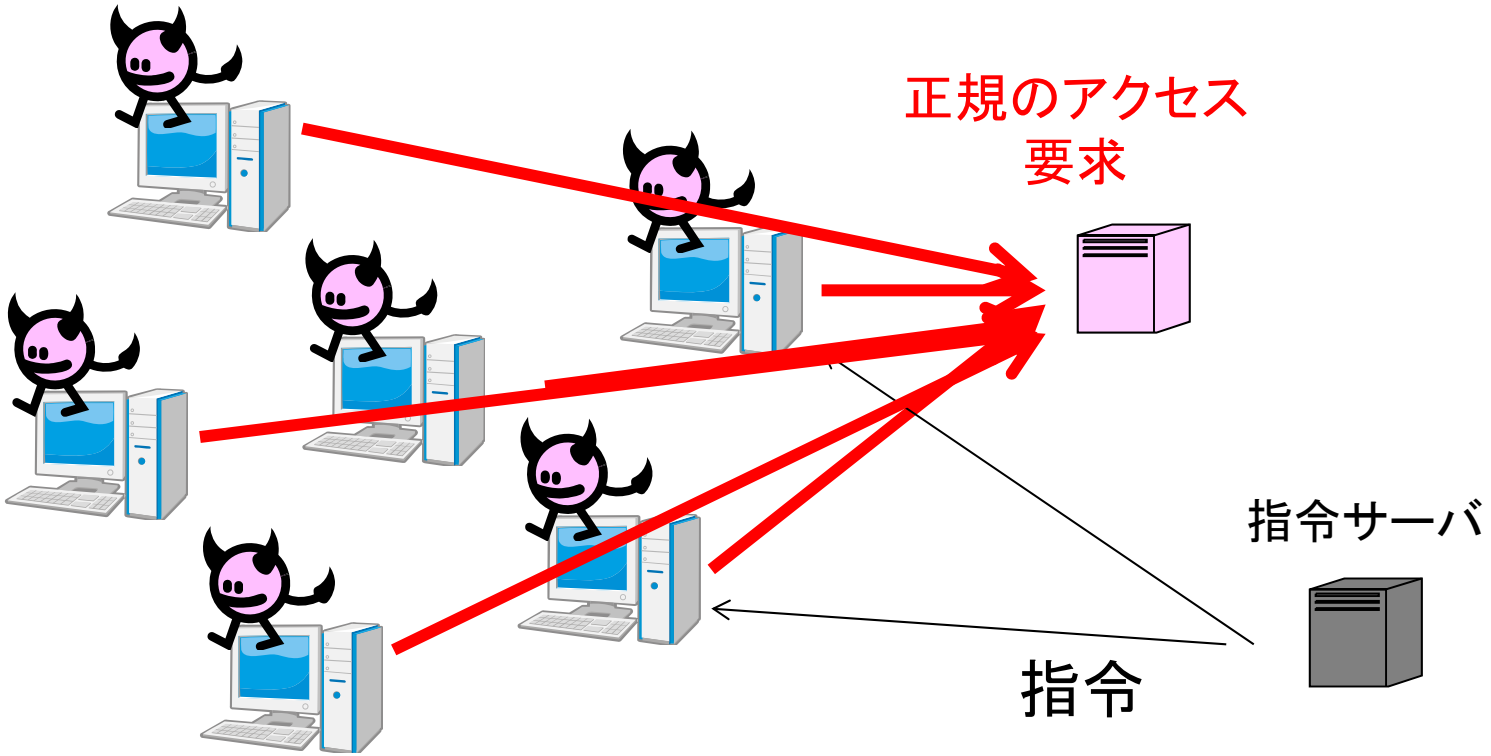
他の別URL(偽サーバ)への誘導方式

まだ「これでやられた」という実例は聞きませんが、ちょくちょく(攻撃者が実験しているのが)見られるもの

- DNSキャッシュポイゾニング
  - DNSの名前解決を毎回要求するのは無駄が多い →DNSキャッシュ
  - ここに偽の名前解決を入れて偽サーバに接続
- ネットワーク経路ハイジャック
  - 最近ではネットワーク経路解決にBorder Gateway Protocolをよく使う
    - 現在はIPv4で50万経路ぐらい
  - このBGPに偽の経路を流す
    - 別サーバに誘導
    - 自分が乗っ取ったネットワーク機器を経由するようにルート変更

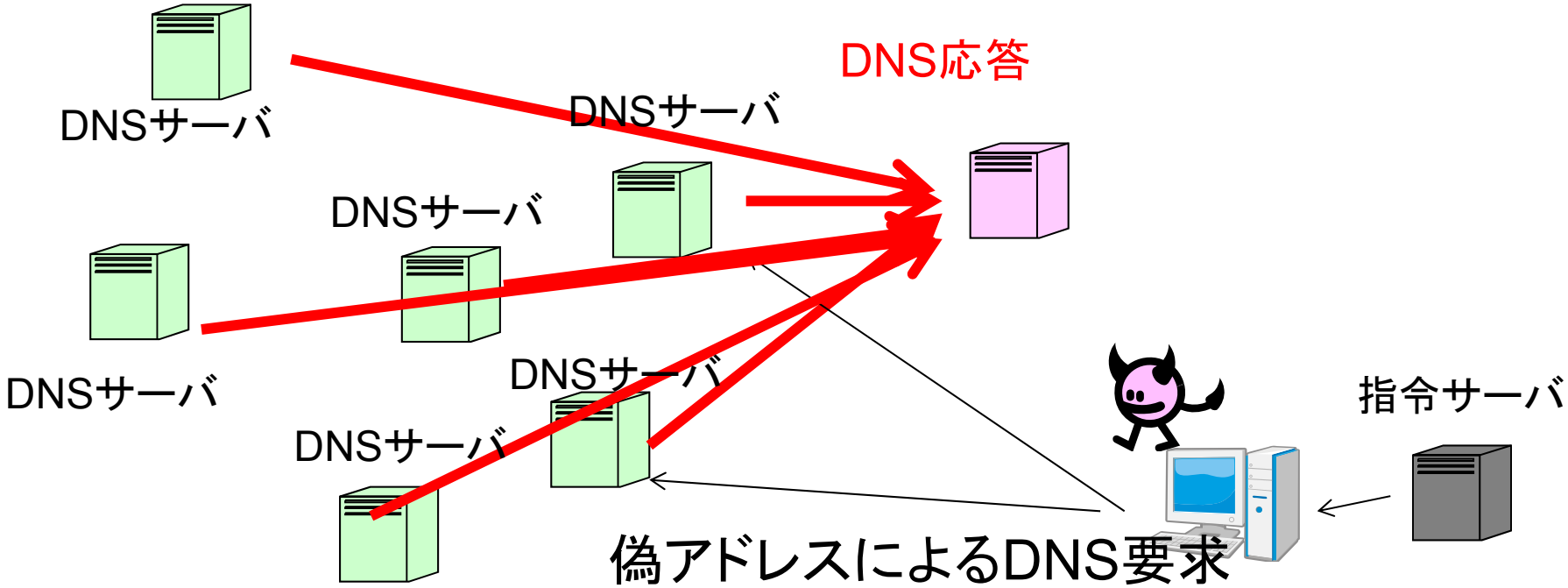
分散サービス不能攻撃(DDoS) (1/2)

- DDoS: Distributed Denial of Service
- 複数のコンピュータから正規のアクセス要求を行うことによって標的のリソースを消費
  - ネットワーク帯域、TCPのポート数、CPU負荷、など



分散サービス不能攻撃(DDoS) (2/2)

- 最近ではサーバからの応答を利用して攻撃増幅も行われる
  - DNS, NTP, SNMPなど
  - UDPはハンドシェイクが無いので送信元を偽るのは容易
- “リクエストのデータ量<応答のデータ量”ならば、少ない乗っ取り済みPCで攻撃成立可能



概要

- 背景: 近年の情報セキュリティ問題(サイバー攻撃対策)
  - 近年の高度化したサイバー攻撃の例
  - サイバー攻撃の防御とその課題
  - サイバー攻撃防御側の希望
- 高速かつ大量な通信データの処理によるセキュリティ向上
  - アノマリ検知(ビッグデータ的な不審通信検知)
  - 通信データ処理のハードウェア化

情報セキュリティ人材問題(1/2)

- じゃあ? セキュリティ技術者が増えれば問題は解決する?
  →一朝一夕には増えません
- そもそもNHKがニュースにするぐらい不足[1]

[1] http://www.nhk.or.jp/kaisetsu-blog/100/202598.html

情報セキュリティ人材問題(2/2)

- Chief Information Security Officerに月給100万クラスを準備しても、でも要求レベルの人が来ないことも
  - 法律家や警察などの論理にも精通しているのが望ましいような人
- 大学の公募も苦労しているようです
  - “情報セキュリティ技術者を育てるぜ”とコースは作ったは良いが、良い先生があつまらないという所が多々ある
  - JREC-INを見ると、某地方国立大が何度も公募を出し直していたり

→教育できる人がいないから人材が増えないという悪循環

- そもそも、情報セキュリティは情報技術の中でも若い分野
  - 当然、それに比例して人材が少ない
  - うちの研究室の教授(50直前)が最長老クラス
    - 本来ならばもっと上の人が担当する委員会の委員まで担当することになって忙しそう

セキュリティへのコスト意識の問題(1/2)

- そもそも、セキュリティ対策は、警察や消防と同じで必要無ければ嬉しい組織
  - 仕事が無いのが一番な組織
  - でも、警察や消防を不要と言う人はいないが…
  - 企業としても、直接利益を産まない所には投資しにくい
- 同様の事例として、Windows XPをまだ使う例
  - 設計が古くて情報セキュリティ的には好ましくないのだが…
  - ユーザ側としては、まだ十分に使えるOS

セキュリティへのコスト意識の問題(2/2)

- セキュリティ人材へのコスト意識
  - 実際にある募集: 薬剤師の試験雇用のパートさんより安い!
  - 人材は不足しているけど突っ込むお金はもっと不足していて、経営者のコスト感覚は致命的に不足している

求人情報

就業時間	09:00〜17:45
仕事内容	■情報セキュリティガイドラインの評価支援業務1)情報セキュリティガイドラインの評価分析 2)評価分析の結果報告書等のドキュメント作成情報漏えい防止、アクセス権の設定、改ざん防止・検出、電源対策、システムの二重化などの対策状況を確認して、評価分析、報告資料を作成します。即日〜約2ヶ月間のお仕事となりますので、ご経験を活かしたい方はぜひご応募下さい♪
雇用形態	派遣
賃金形態	時給
賃金	1,600円

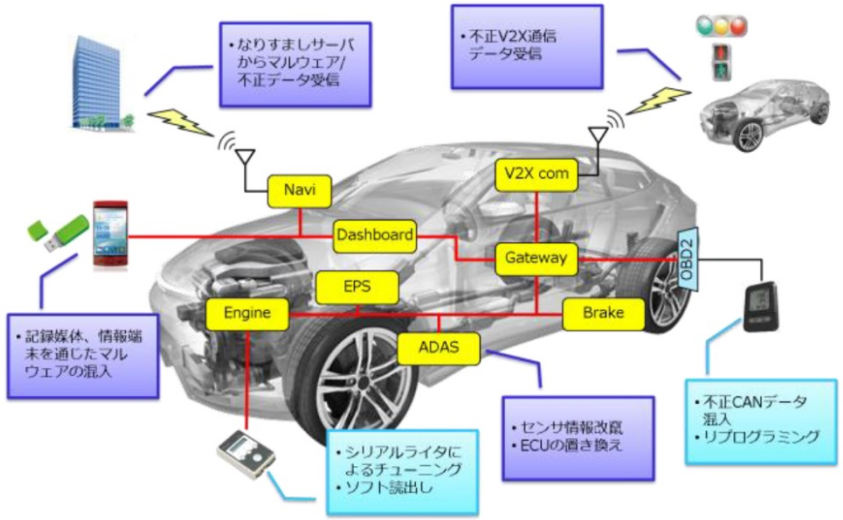
攻撃対象の増加(1/2)

- Internet of the Thing(IoT)
  - ヘルスケア用途など有望だが…

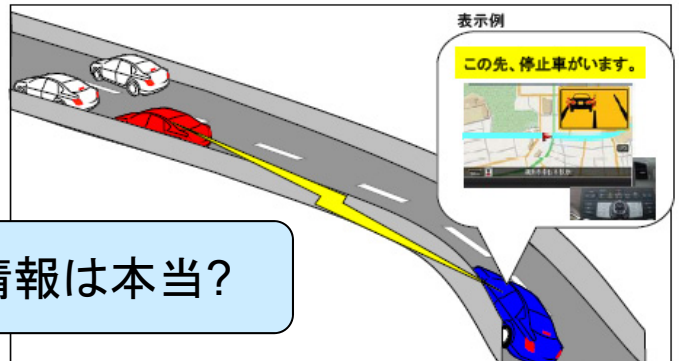
  →攻撃対象や踏み台利用の増加
- 車載ネットワーク/車間ネットワーク
  - コスト削減や交通事故削減に有望だが…
  - 車の制御システムを妨害したり
  - 他の車や信号に偽の情報を送ったり

↑ヘルスケアとIoT

↓車両制御システムへの攻撃



↓車間通信への攻撃

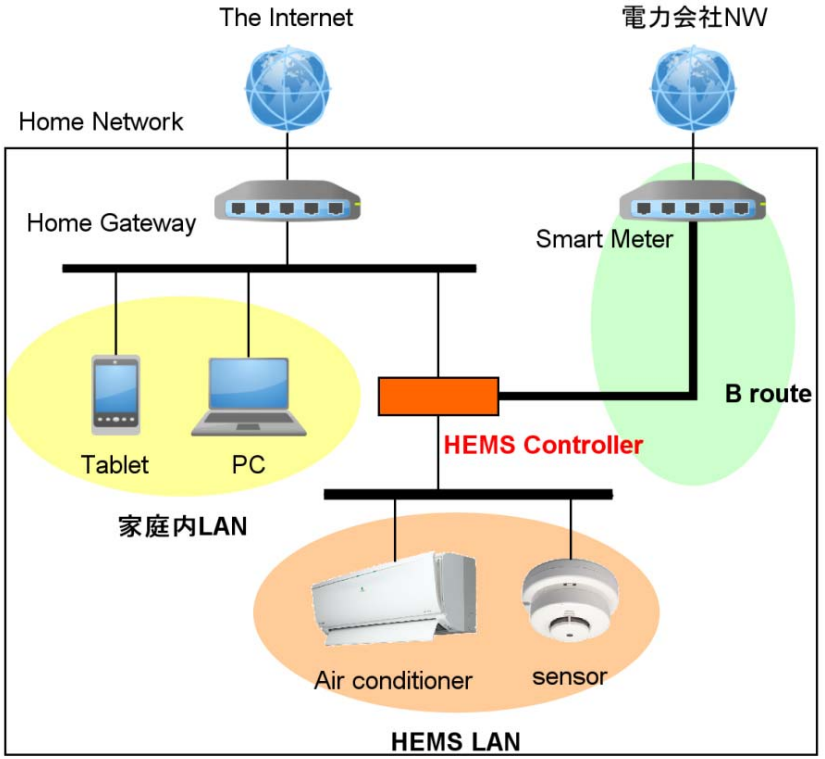


その情報は本当?

攻撃対象の増加(2/2)

- スマートグリッドの制御ネットワーク
  - HEMS (Home Energy Management Systemと連動)
  - 基本的に、家庭内LAN、HEMS LANとは分離されているはずだが…
  - 日本の住宅事情で複数サブネットのネットワーク線を通す構成できるの?

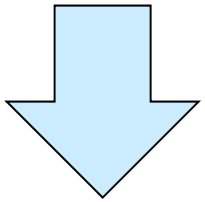
スマートグリッド普及者側が想定するネットワーク



通信量増大の問題(1/2)

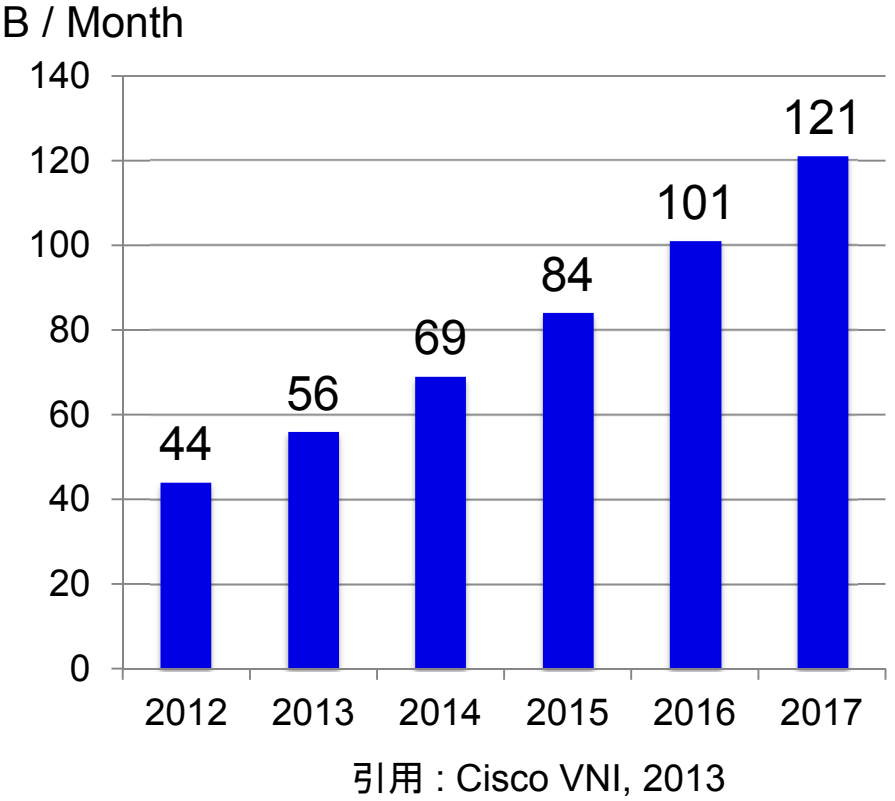
- 2017年のネットワーク接続デバイス 190億台
- 2017年の年間IPトラフィック量予測 1.4ZB
- IPトラフィック全体の年平均成長率 23%

トラフィックの増加に伴い
解析対象の増大



- 検査対象の増加
- DDoS攻撃の上限増加

→対策機器側の要性能向上

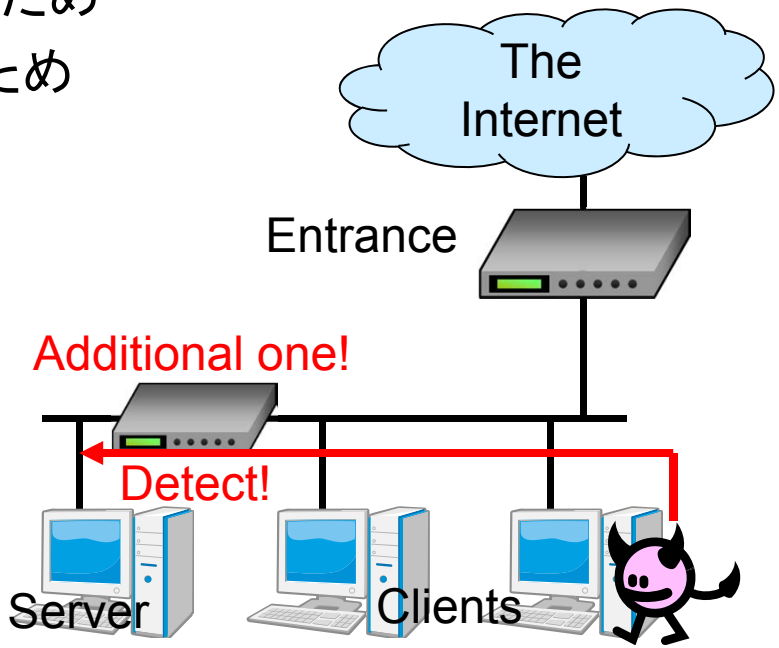


引用：Cisco VNI, 2013

通信量増大の問題(2/2)

- 近年では、対外接続部のみの不正通信は不十分
  - 標的型攻撃でセキュリティ意識の弱い部署を狙って組織内へ侵入
  - 侵入した部署から標的となる部署に攻撃をしかける
- 内部ネットワークの監視の必要性
  - 重要なマシン(e.g. サーバ)を保護するため
  - 重要な部局の仕事に影響を出さないため

対外接続部での監視に比べて
最低10倍のトラフィックをさばく必要



サイバー攻撃/犯罪対策をじゃまするもの(1/3)

内部側から(悪くないのも含む)

- セキュリティ対策への無理解
  - セキュリティ対策やEnd of Life機器の更新予算をかけてくれない
- 勝手なサイバー攻撃対策作業
  - 勝手にリカバリディスクを使ってノートPCを初期状態に戻すとか
  - ヘタすると、警察から「主犯が証拠隠滅を行った」と見られます
- 移動する無線LAN接続のクライアント
  - 外部で接続した時にマルウェアを拾ってきて内部でばらまいたりとか

サイバー攻撃/犯罪対策をじゃまするもの(2/3)

犯罪者側から

- そもそもマルウェア側に対策が行われるのを検知する機能があったりする
  - マルウェア内に偽ドメインを埋め込む →偽ドメインの名前解決があったら解析されている
  - 標的以外のIPアドレスの範囲からの通信があったら検知と判断
  - そもそも、起動時にGoogleなどのメジャーなサービスへの接続性を確認したりする
- 対策しようとするとDDoSをかけてきてじゃましようとしたりする

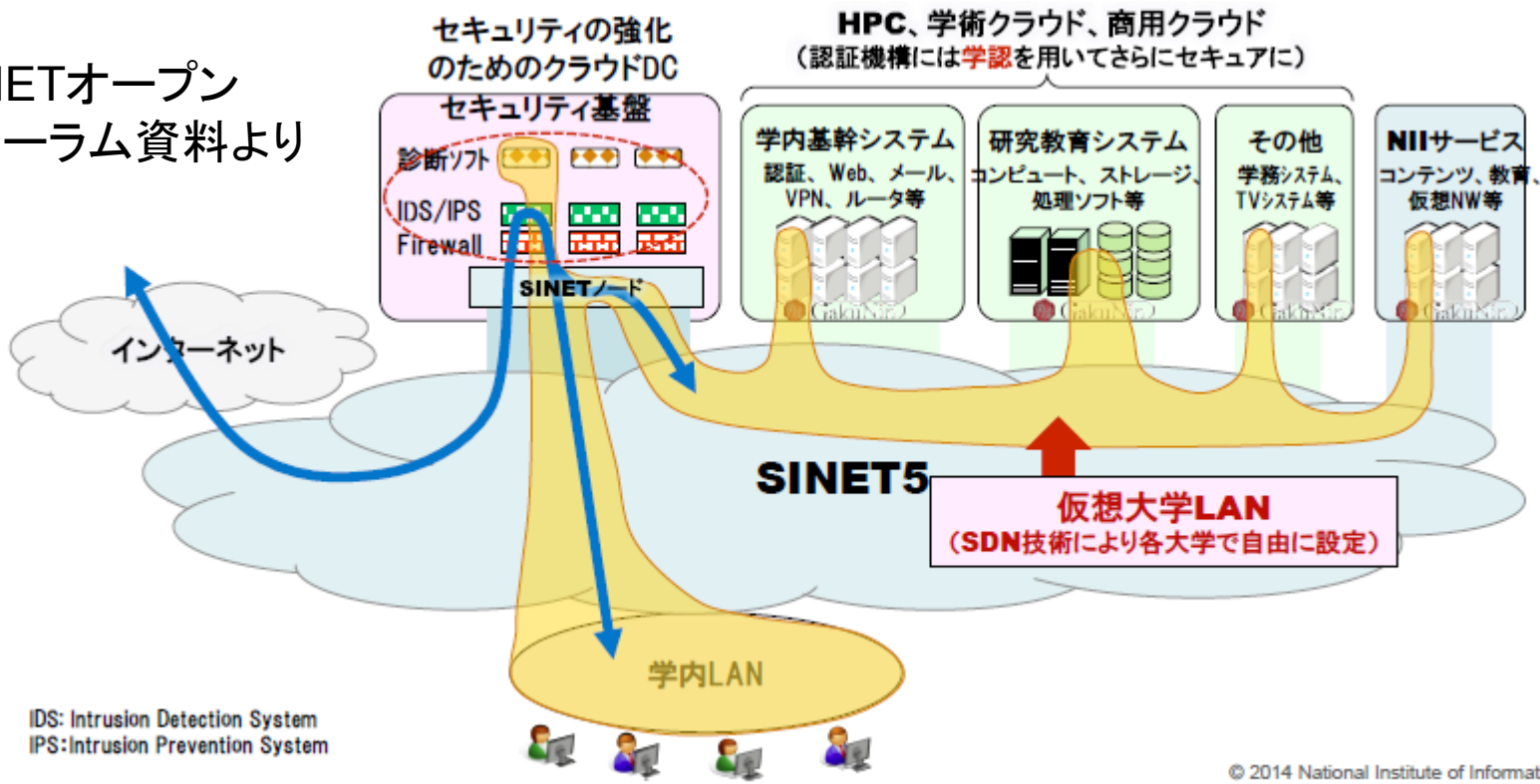
サイバー攻撃/犯罪対策をじゃまするもの(3/3)

- あまり使われない機能の追加&デフォルト有効によるセキュリティ・ホール追加
  - OpenSSLのheartbleed
  - bashのshellshock
- 右肩上がりの目標はいい加減な所でやめて欲しいのだが…
  - 新たな機能を追加すれば新たな人が無限呼び込めるとか考えているの?
  - どこかで一旦、安定に入ってもいいと思う
    - もちろん、必要性が出てきたら
  - 最近はFirefoxがその領域に

セキュリティ側から見えている希望(1/3)

- クラウドコンピューティングを利用した集中防御
  - SINETもクラウドを作成して大学の情報セキュリティを担う提案
  - ただ、運営者を信頼できるかという問題はつきまとう

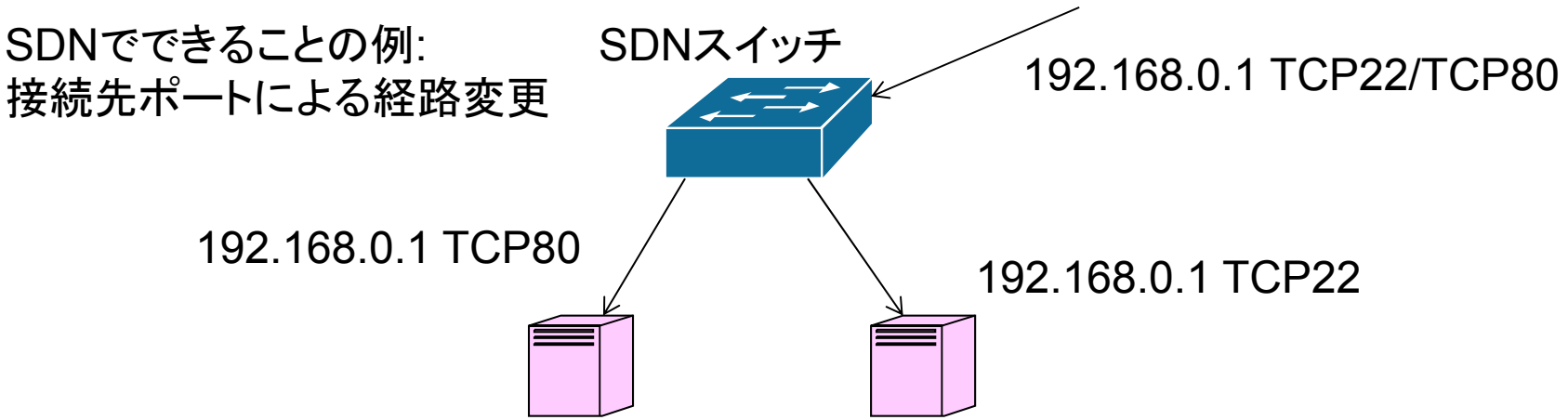
SINETオープン
フォーラム資料より



セキュリティ側から見えている希望(2/3)

- SDN(Software Defined Network)による柔軟なネットワーク
- SDNの特徴
  - ソフトウェアのような柔軟な経路選択ルール作成
    - 送信先ポート、送信元IPアドレス/ポート、など
  - 同一IPアドレスに対してTCP/UDPのポートに応じて経路選択可能

  →マルウェアの通信のみ捻じ曲げることが可能



セキュリティ側から見えている希望(3/3)

- ビッグデータ処理の応用
  - 通信解析、マルウェア分類、などへの応用
  - 異常な通信ではなく、通常の通信の定義からの情報セキュリティ適用
  - ビッグデータに向けた計算機の能力向上研究の進歩
- 人が足りないなら自動化すれば良いという目標の研究
  - 熟練情報セキュリティ技術者の知識適用の自動化
  - 別に100%を目指す必要はない
    - 自動化で80%を除外できるならば、人の負荷は1/5に

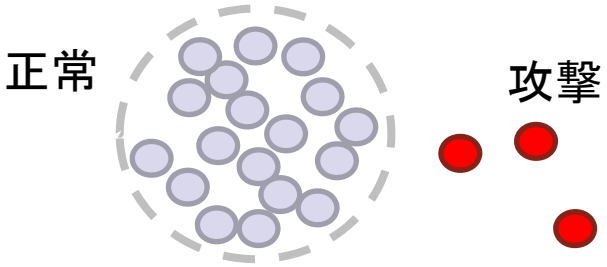
概要

- 背景: 近年の情報セキュリティ問題(サイバー攻撃対策)
  - 近年の高度化したサイバー攻撃の例
  - サイバー攻撃の防御とその課題
  - サイバー攻撃防御側の希望
- 高速かつ大量な通信データの処理によるセキュリティ向上
  - アノマリ検知
  - 通信データ処理のハードウェア化

侵入検知

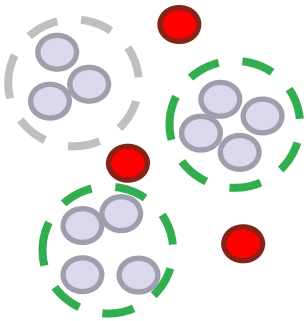
- 侵入検知システム(IDS): 攻撃を検知し管理者に警告を通知
- シグネチャ検知
  - 予め定義した攻撃の特徴(シグネチャ)と比較することにより攻撃を検知
    - ×未知攻撃を検知できない

- アノマリ検知
  - 正常トラフィックの特徴を学習しておき、これに反した異常トラフィックを攻撃として検知
    - ○未知攻撃を検知できる
  - 全通信に対して処理をかける点でビッグデータ的な検知手法

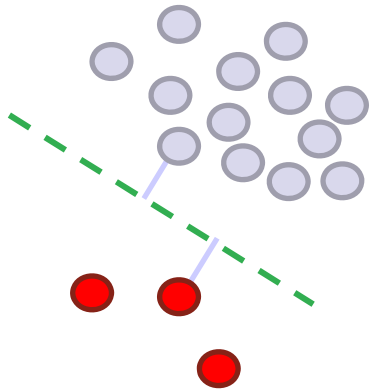


一般的なアノマリ検知手法

- セッション毎の特徴量
  - セッション: 接続が確立されてから切断されるまでの一連の通信
  - 特徴例：接続時間、接続回数、送受信バイト数、SYNエラー数など
- クラスタリングや SVM を用いた検知手法
  - セッション毎に正常か異常かを判定
  - クラスタリング
    - K-means
    - Density based
  - SVM(Support Vector Machine)
    - One-Class SVM
    - Multi-Class SVM



クラスタリング



SVM

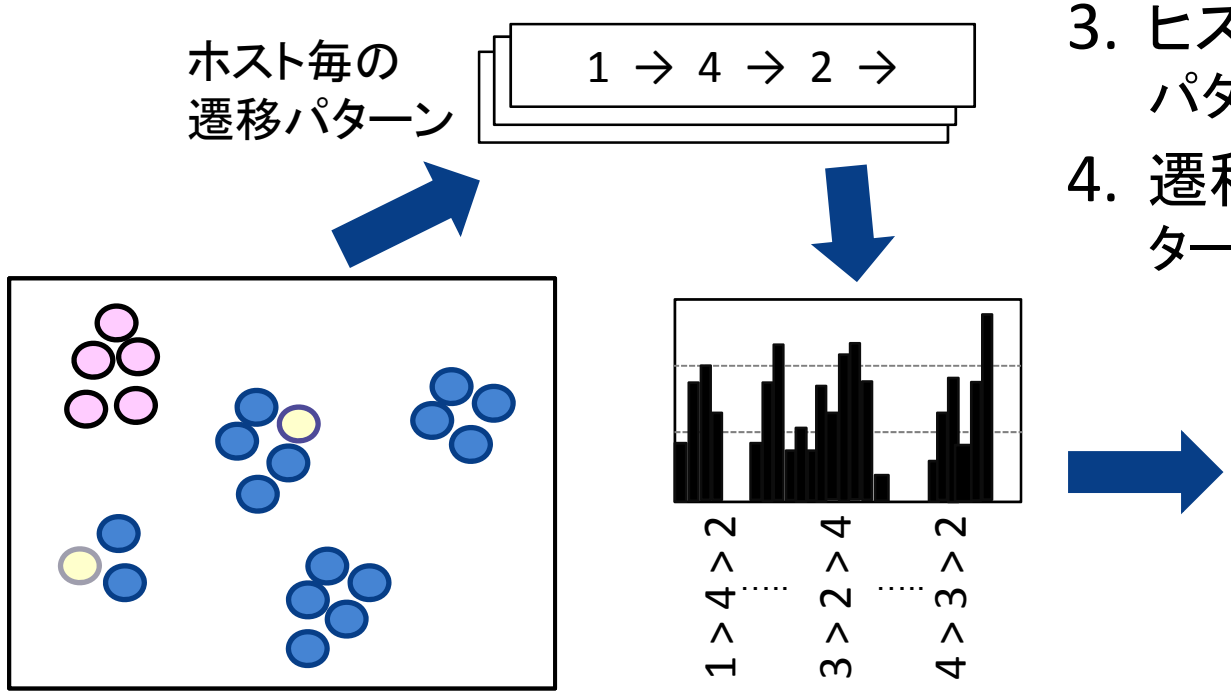
アノマリ検知を利用した未知攻撃検出

- セッション・シーケンスに着目したアノマリ検知
  - セッションのシーケンスをモデル化し、セッション単位では検知できない攻撃に対応
  - あまり見られないセッション・シーケンスに対して高得点を出しやすいので、未知攻撃の検知が可能
- 多段OC-SVMによる未知攻撃の検出

セッション・シーケンスに着目したアノマリ検知の流れ

- 学習
  1. 複数の通信状態の作成：グリッド分割を用いたクラスタリング
  2. 遷移情報の抽出：ホスト毎の遷移パターン抽出
  3. ヒストクラムの作成：各遷移パターンに関する頻度値を計算
  4. 遷移スコアの学習：遷移パターンに対するスコアの割当て

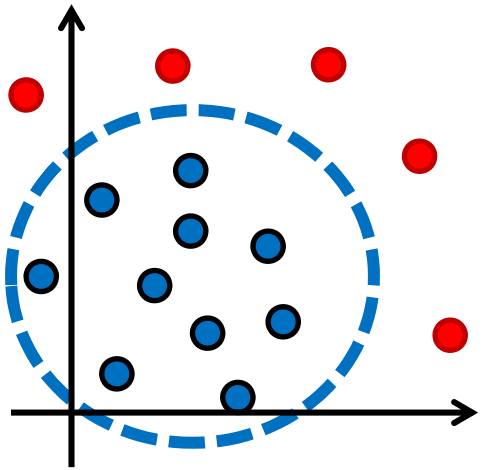
- 攻撃検知
  - 学習時に求めた遷移スコアをもとに識別スコアを計算し、閾値を超えたら攻撃と判定



遷移パターン	スコア
1 → 4 → 2	0.5
3 → 2 → 4	− 0.2
4 → 5 → 2	0.1
⋮	⋮

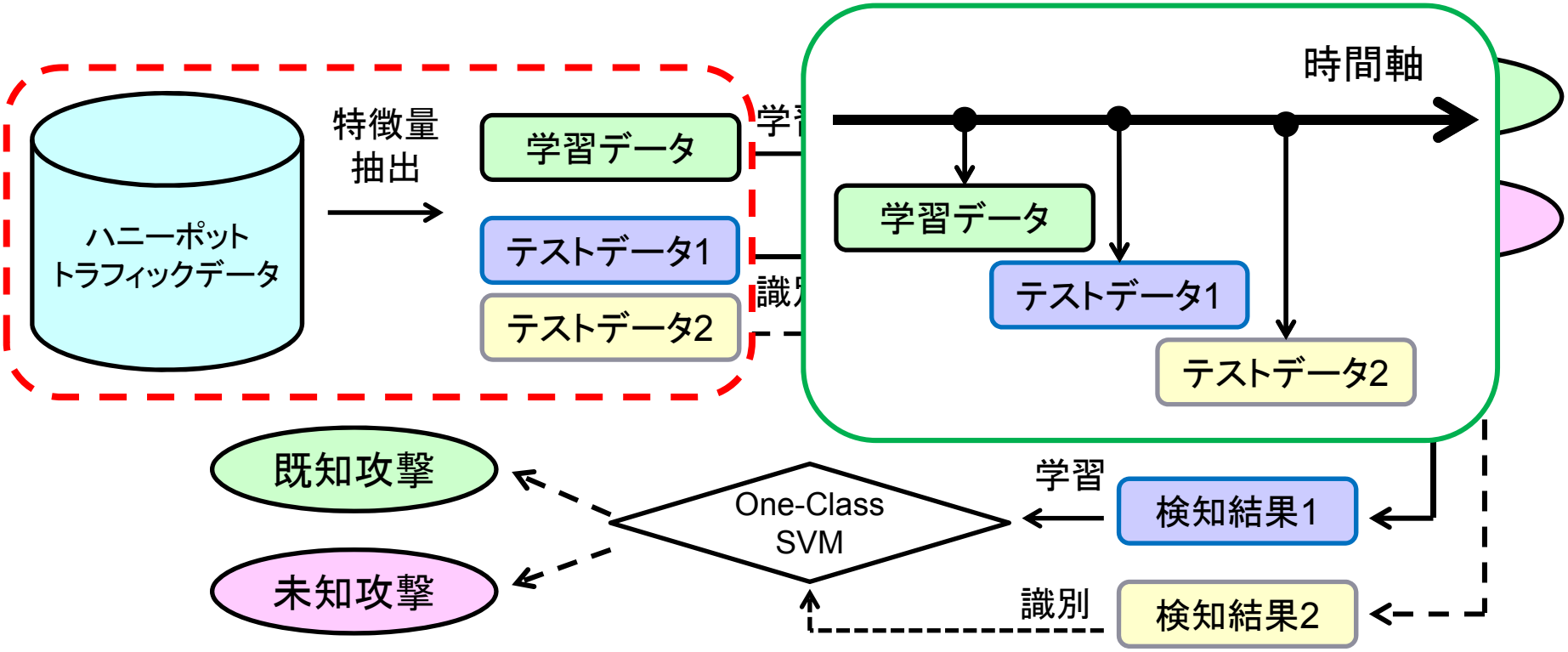
One-Class SVM(OC-SVM)

- データ集合の領域を求め、それに入っていないデータを外れ値とする
- 外れ値とデータ集合の距離が最大となる超球を求める
- パラメータvにより超球の大きさを調整
  - v=0.1なら全体の10%を除くデータにて超球を作成



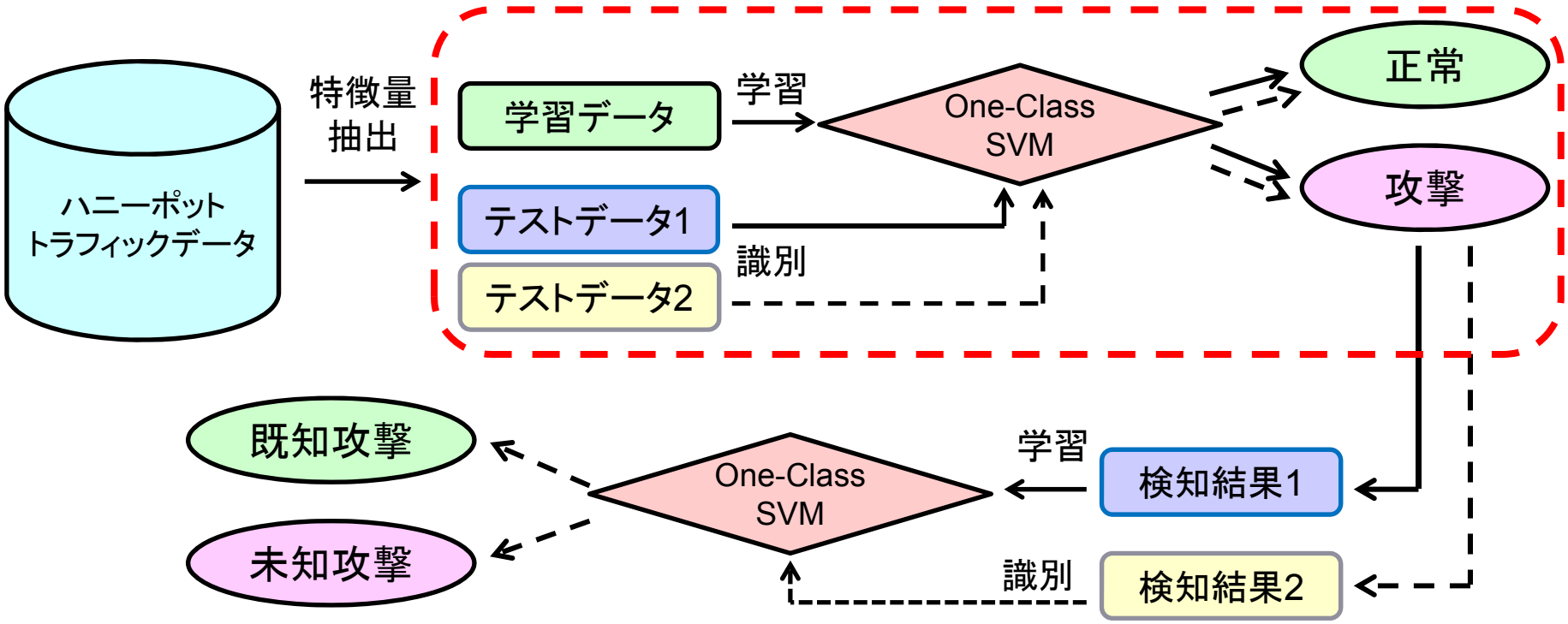
検出手法の概要

1. トラフィックデータから特徴量抽出
2. 一段目のOne-Class SVMにて攻撃検知
3. 二段目のOne-Class SVMにて未知攻撃検知



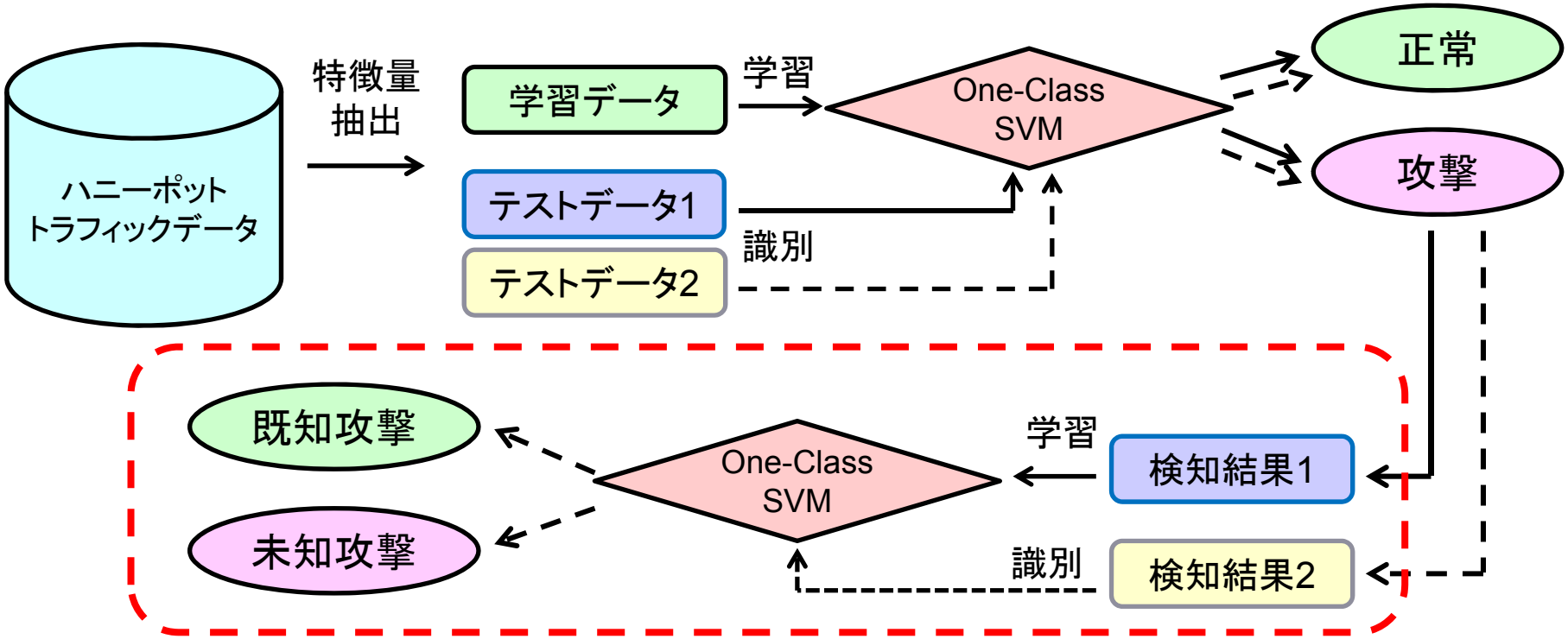
検出手法の概要

1. トラフィックデータから特徴量抽出
2. 一段目のOne-Class SVMにて攻撃検知
3. 二段目のOne-Class SVMにて未知攻撃検知



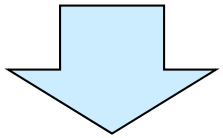
検出手法の概要

1. トラフィックデータから特徴量抽出
2. 一段目のOne-Class SVMにて攻撃検知
3. 二段目のOne-Class SVMにて未知攻撃検知



アノマリ検知とデータ処理量

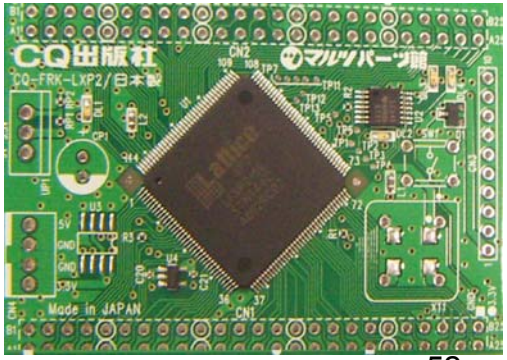
- アノマリ検知というビッグデータ的な手法で未知攻撃まで発見できそう
- しかしながら、その処理量はシグネチャ検知よりはるかに大きい
  - 全データに対して前処理をした上、類別処理を行う必要がある
  - 分散処理をやりにくい



- 通信データ処理をハードウェア化することで高速化
- ハードウェアにはFPGAを利用

FPGA: Field Programmable Gate Array

- 近年多用される再構成可能ハードウェア
- LUTを使った構成が主流
  - LUT(Look-Up Table): 任意の3-8入力の信号に対して任意の値を出力する論理素子
  - 特定用途専用回路も搭載するものも
    - ブロックSRAM
    - 乗算器
    - 組み込みプロセッサ/DSPコア
    - 高速I/O(おおむね3Gbps以上)
- プロトタイピングで多用される
  - もしくは少量生産
    - ネットワーク機器ではよくある
  - もしくはASICが来るまでのつなぎ



高速IOを持つFPGA(Altera)

- Stratix
  - Stratix IV GT(40nm): 11.3Gbps x24
  - Stratix V GX(28nm): 14.1Gbps x66
  - Stratix V GT(28nm): 28.05Gbps x4, 12.5Gbps x32
  - Stratix 10 GX(14nm): 32Gbps x?
  - Stratix 10 GT(14nm): 56Gbps x?
- Arria
  - Arria V GZ(28nm): 12.5Gbps x36
  - Arria 10 GT(20nm): 28.05Gbps x 96
- Cyclone
  - Cyclone IV GX(40nm): 3.125Gbps x8
  - Cyclone V GT(28nm): 6.144Gbps x12

アノマリ検知のハードウェア化への検討

- 様々なアルゴリズムの中からHW化に適したアルゴリズムを検討
- 使用特徴量には何が適切か
  - 特徴量にセッションを扱うのは難しい
  - セッション単位から特徴を抽出するためには，パケット，セッションを保持するバッファが大量に必要

FPGA内ブロックRAM

× 容量不足

通信量が大きい場合、SYN Flood の際にセッションを構築できない

FPGA外DRAM/SRAM

× アクセス性能がボトルネックに

セッション構築はパケットバッファに頻繁にアクセス

ハードウェア論理化にあたって

1. メモリを多量に使用しない
2. 特徴量にはシンプルなものを使用
3. ボトルネックとなる処理のハードウェア論理化が容易

PAYL[1]

- ペイロードベースのアノマリ型検知を目的としたアルゴリズム
- ペイロードに対して1- gram法を適用
- 正常な通信のみを含むトラフィックを学習データとして使用

- モデル $M_{i,j} = (\overline{p}_{00}, \overline{p}_{01}, \cdots, \overline{p}_{FF})$

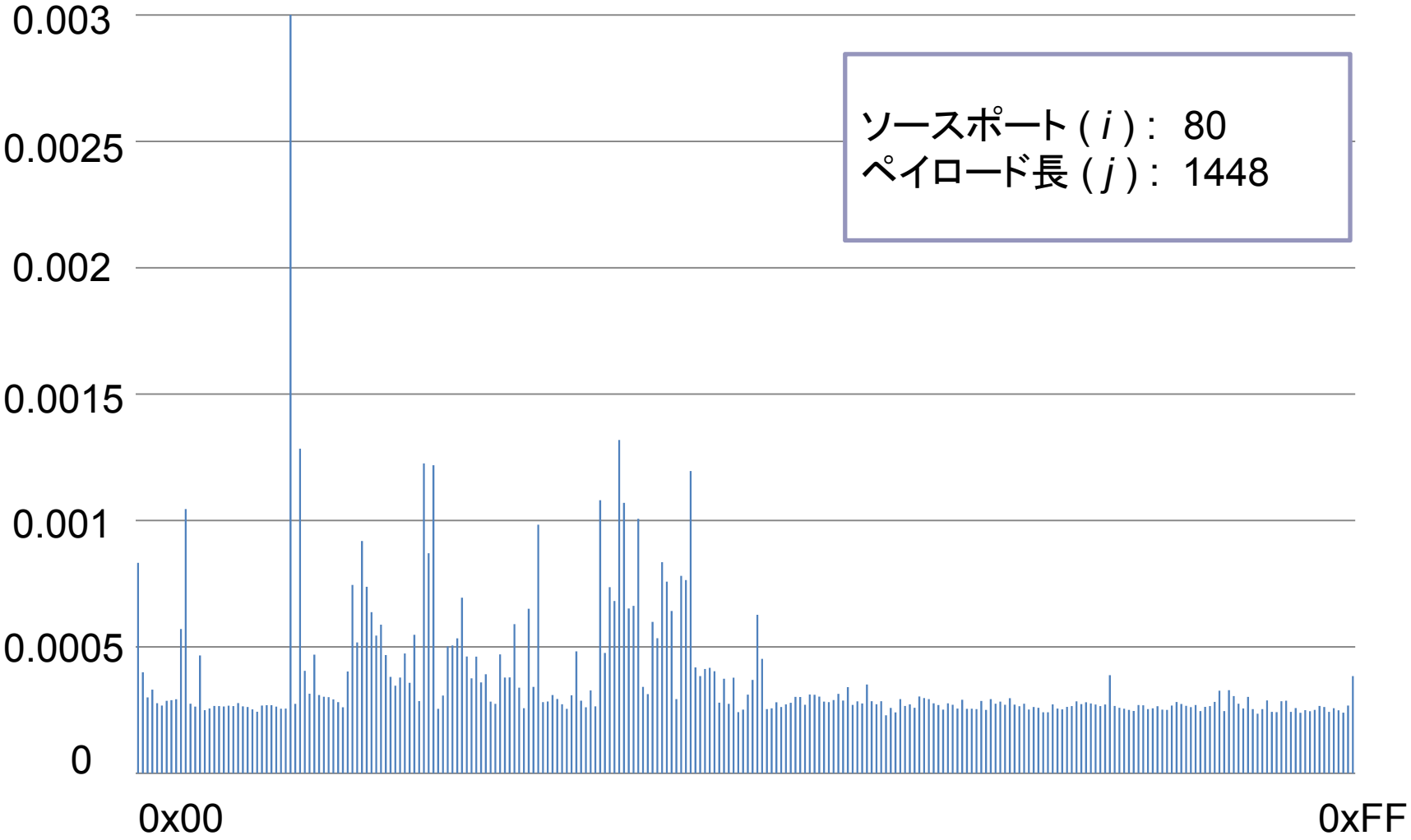
i: ペイロード長

j: ポート番号

p: ペイロード中のバイト毎(0x00 – 0xFF)の出現頻度の平均

[1] K. Wang and S. Stolfo, “Anomalous payload-based network intrusion detection,” In Proc. of 7th Intl. Symposium on Recent Advances in Intrusion Detection, pp. 203-222, Sep. 2004.

モデル $M_{i,\ j}(p)$の例


0.003
0.0025
0.002
0.0015
0.001
0.0005
0
0x00
0xFF
ソースポート (i) : 80
ペイロード長 (j) : 1448

PAYLの検知処理

- 入力パケットのペイロード長，ポート番号と一致するモデルとのマハラノビス距離を計算
- $$d(x, \overline{y}) = \sum_{i=00}^{FF} \frac{|x_i - y_i|}{\overline{\sigma_i} + \alpha}$$

x: 入力パケット
y: 学習によって生成されたモデル
σ: 標準偏差
α: 個別設定

- しきい値を超えた場合に警告
- パケット単位で 処理を行う

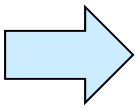
ボトルネック

- PAYLの処理をソフトウェアで実装 *1

	特徴抽出部	距離計算部
実行時間 (s) *2	0.02495	0.002328
スループット (Gbps)	0.4810	5.155
比率	0.9147	0.0853

*1 環境: CPU Intel コア i5 3.20 GHz，メモリ 8G
実装にC言語を使用，gcc4.8を用い，最適化オプション –O2 を適用
*2 1500バイトのパケット1000個を入力として用い，処理時間を計測

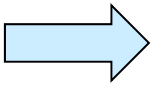
  - 特徴抽出部: パケットから特徴を抽出
  - 距離計算部: 入力パケットとのマハラノビス距離を計算
- 処理全体の約91%が特徴抽出処理



特徴抽出部のハードウェア論理化

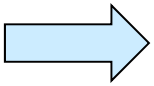
PAYLのハードウェア論理化の適性

1. メモリを多量に使用しない



パケット毎に処理を行う

2. 特徴量にはシンプルなものを使用
3. ボトルネックとなる処理のハードウェア論理化が容易

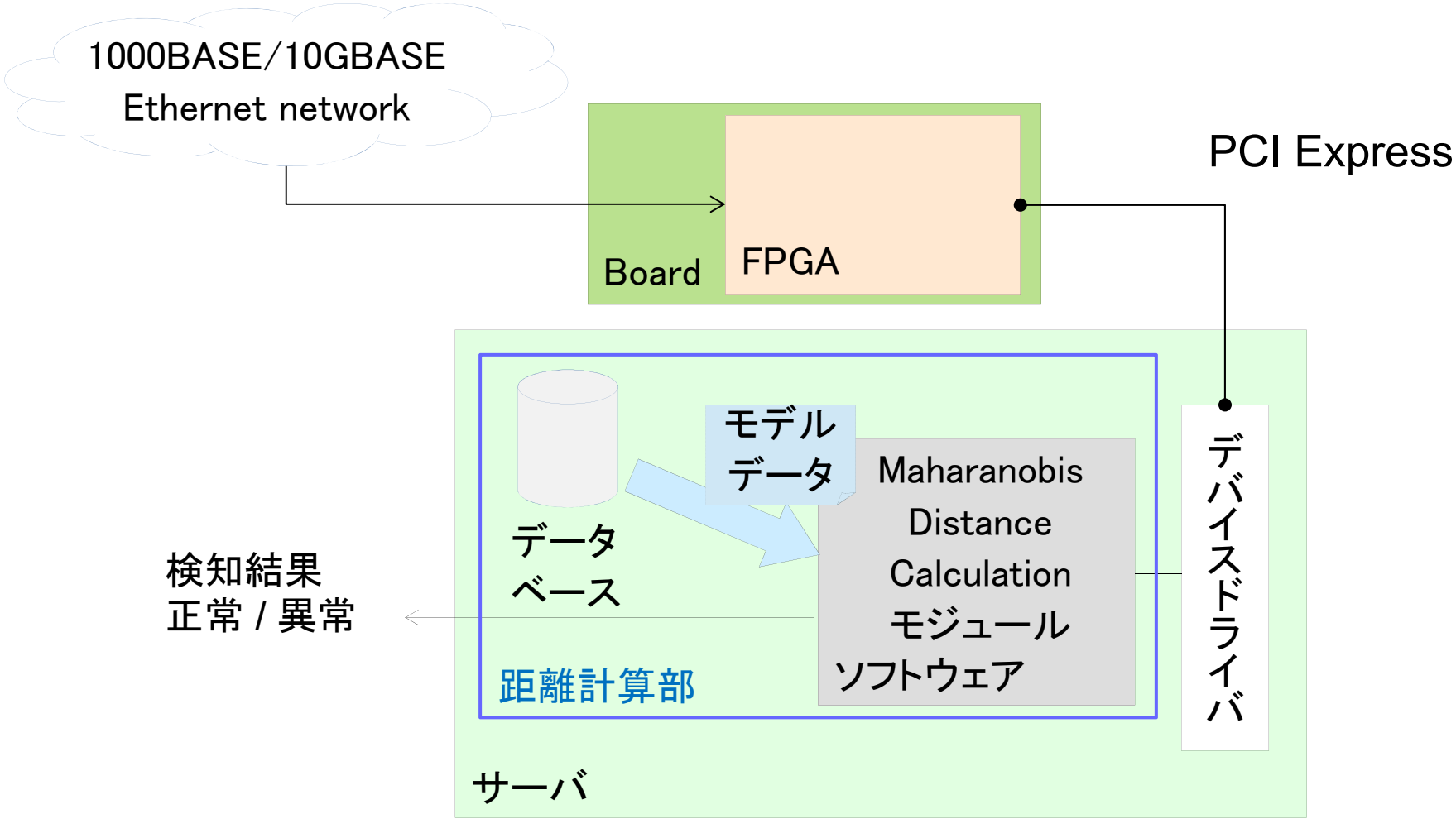


使用する特徴量はバイト毎の出現頻度
ストリングマッチによって容易に実現可能

HW/SW 部への分割と構成

FPGA で特徴抽出部を実装し，検知アルゴリズム
を行うソフトウェアと組み合わせる

システム全体の概要とサーバ



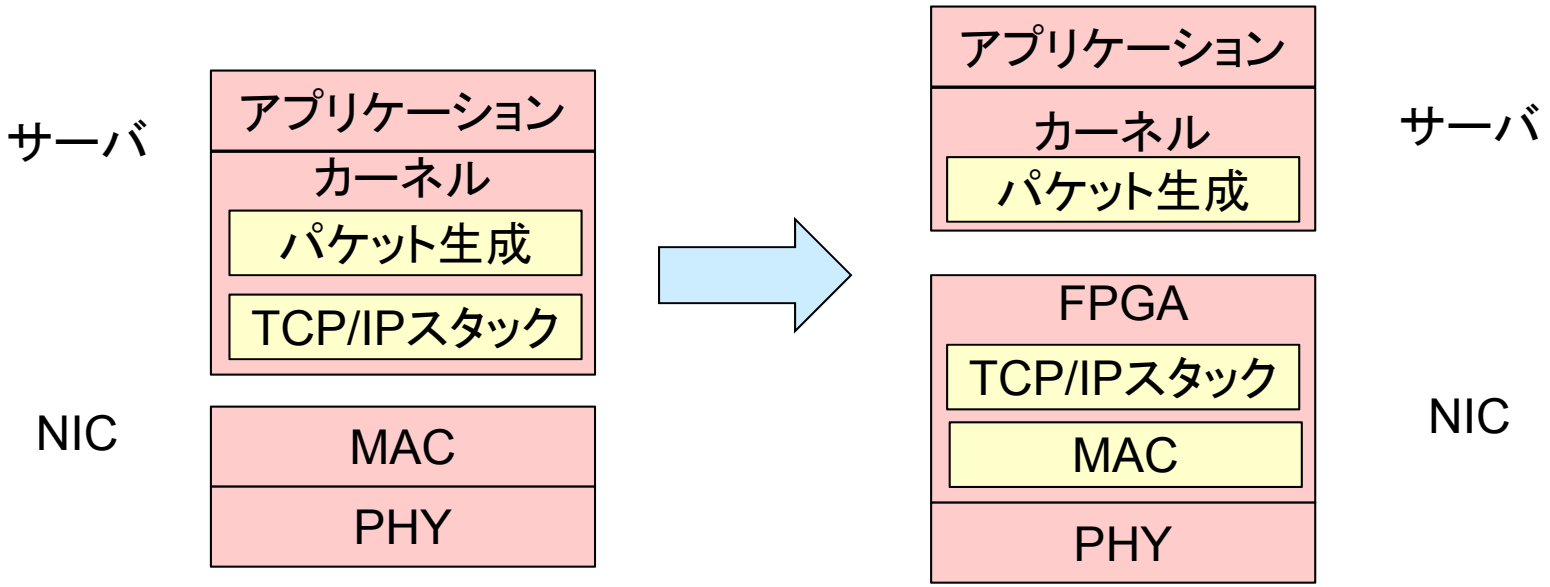
他の高速通信下のFPGA利用例1: 高頻度トレード(HFT)

- HFT: High Frequeny Trading
  - アルゴリズムによる(株式)取引方法の1つ
  - 取引時のマージンを低くするが、高頻度で取引をすることで
  - ミリ秒単位の高速(株式)取引が重要になる
    - “2005円で売り”と”2010円で買い”が出そうならば、”2006円で買って2009円で売る”という
    - 最近だとマイクロ秒とかのオーダに…
- このような取引では取引依頼の少しの遅延が大きな損失に

  →FPGAによる取引依頼部ハードウェア化
  - アルゴリズムの部分は引き続きサーバ部分

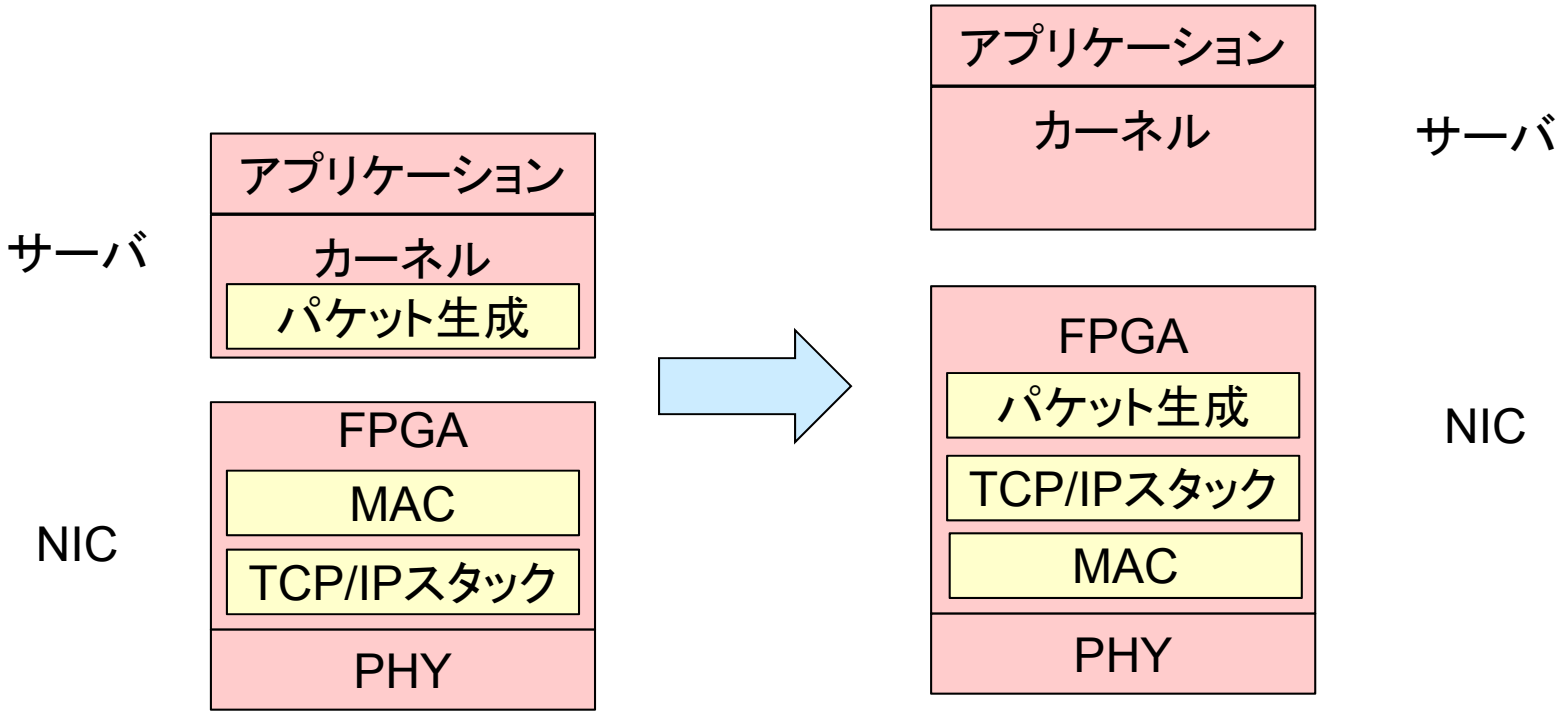
HFTのネットワークにおけるFPGA利用 (1/3)

- 初期: FPGA付きNICによるTCPオフローディング
  - TCPオフローディング: TCP/IPスタックをFPGA側で実行することでサーバ側の負荷を軽減
  - サーバで生成した取引発注の通信内容をFPGA側のTCP/IPスタックにて送信
    - FIXプロトコル:金融取引の標準プロトコル



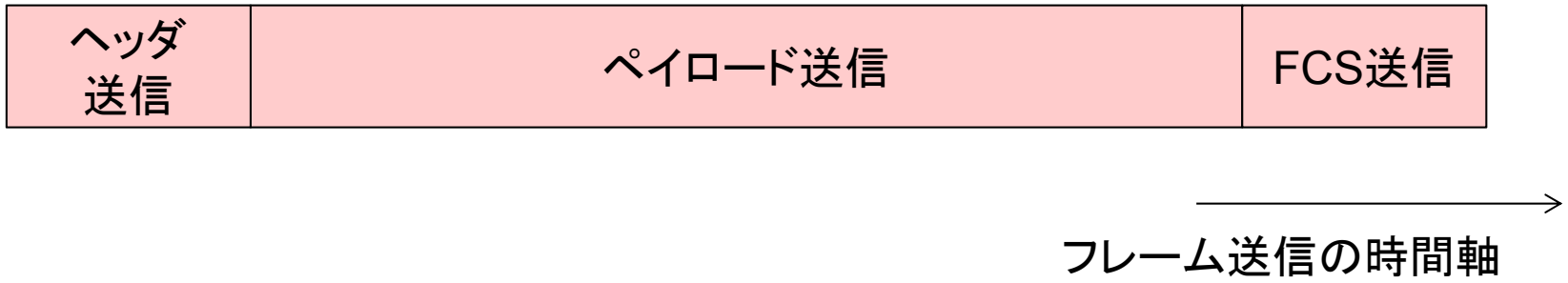
HFTのネットワークにおけるFPGA利用 (2/3)

- 中期:
  - 発注の通信をFPGA内部で生成
  - サーバ側は取引発注内容自体のリクエスト処理のみ



HFTのネットワークにおけるFPGA利用 (3/3)

- 最近: 投機的な取引リクエスト
  - 過去の値動きを元に発注すべき取引内容を予測
  - 最新の値動き結果が来る前に取引内容(のイーサネットフレーム)を送信開始
  - 予定通りの値動き: そのまま送信
  - 予定とは異なる値動き: イーサネットフレームの送信をキャンセル
    - フレーム最後のFCS(Frame Check Sequence)に誤った値を付与
    - 非常に迷惑な行為なので、当然、証券会社側の確認は取っているはず



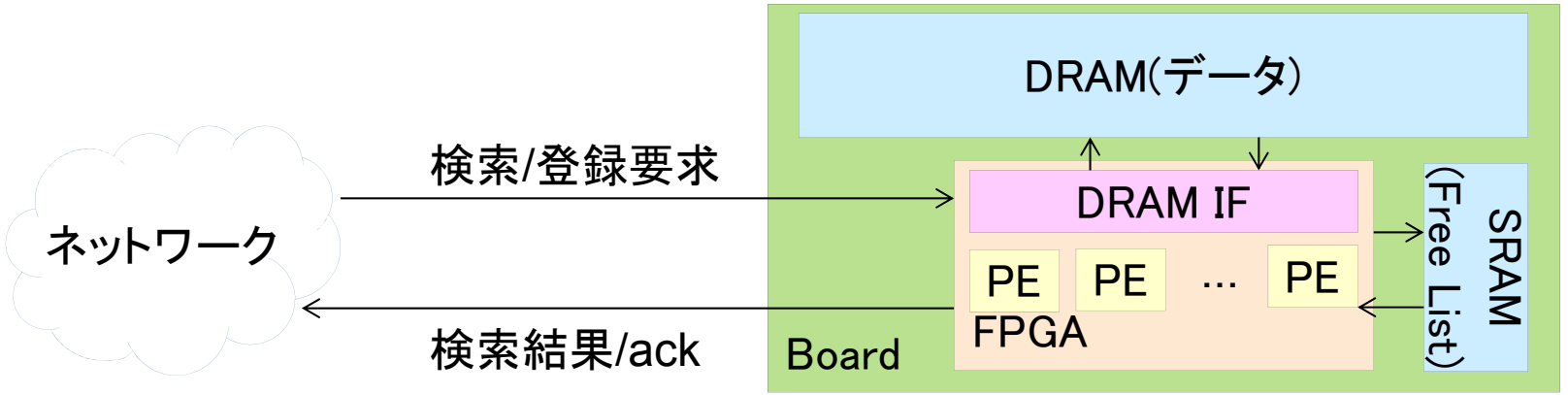
他の高速通信下のFPGA利用例2: FPGA NICによるデータ構造サーバ高速化

- ビッグデータ処理で使われるデータ構造の処理高速化
- Key-Value Store(KVS)型データ
  - Keyをインデクスとして Valueを保存
  - Keyに対して複数のValueを設定可能
  - 長所: 分散処理向き(Keyに対して別々のサーバ/スレッドを割り当て)
  - 短所: データの一貫性は保証されない
  - 代表的な実装: MapReduce, Google Big Table
- 既存のハードウェア化研究はデータベース処理を行うPE(Processing Element)をパイプライン化
- この研究は、異なるデータ型をサポートするPEを並列に動作させる

[1] 徳差ら, “FPGA NICを用いたデータ構造サーバ高速化の検討,” CPSY2014-162, 2015.

FPGA NICによるデータ構造サーバ高速化の実装

- 実装(研究会論文時)
  - FPGAボード内で完結
  - 検索データはDRAMに登録
  - 35PEで40Gbps(23.1M request per second)に対応可能
    - 1パケットは64BのKeyと64BのValue(+プロトコルヘッダ等)の組合せ
- 将来的には高速大容量の記憶装置を接続して運用?



まとめ

- 近年の情報セキュリティの目的の1つとして、金目当てで動くサイバー攻撃からいかに守るかがある
- しかしながら、複数回の攻撃で1回生功すればペイする攻撃者側の方が防御側よりも有利
  - 防御側が未知の脆弱性を利用して未知攻撃をかけて一撃必殺とか
- 未知攻撃に対して有望な防御としてアノマリ検知がある
  - 通常の通信を定義して、そこから外れたものを怪しいとする、ビッグデータ処理的な検知方法
- しかしながら、アノマリ検知は処理量が多いため、高速化の必要がある

→FPGAなどのハードウェアの活用

レポート課題

- 本講義は概論のため個々の話題の詳細は話していない
- よって、個々の話題から1つとりあげ、自身で詳細について調べてまとめること
  - 「話題から1つ」は細かなカテゴリ、おおまかなカテゴリ、いずれでもOK
  - 複数の話題を横断して調べてまとめるのもOK
  - 調べている途中で関連する話題が出てきて、そちらが興味深ければ、それをまとめるのもOK
- 1500文字以上のレポートにまとめて提出
  - 必要に応じて図を入れるのもOK