ネットワーク・フォレンジックス

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

ネットワーク・フォレンジックスとは

- フォレンジックス(forensics)
  - forensic: 法廷の、法廷で用いる、弁論の
  - 法廷で使える証拠と考えてもらえばOK
- ネットワーク・フォレンジックス: ネットワークを介して得られる法廷で使える証拠
  - サーバ内部の情報なども含む
- 様々なフォレンジックス
  - デジタル・フォレンジックス: デジタルデータ(主にファイルシステム内)による証拠
  - コンピュータ・フォレンジックス: 計算機内部からの証拠
    - メモリ・フォレンジックス: メモリ(主記憶)からの証拠

→境界はけっこうあいまい

以外とネットワーク・フォレンジックスをやる局面はあるかも?

- 例: 研究室内LANにつながる無線LANからマルウェア感染端末の同定
  - マルウェア検知は研究室内PCのアンチウィルスソフトウェアが発見したものとする(怪しい通信が来た)
  - 無線LANアクセスポイント等のログを調査
    - 無線LAN APがルータ動作ならばAP内部のDHCPログや接続MACアドレスログを見る
    - 無線LAN APがブリッジ動作ならば、研究室LANのDHCPサーバのログとAPの接続MACアドレスログを見る
- 例: 研究室メールサーバからspamメールが送信された
  - 誰のアカウントを使って送信されたか、どのぐらい送信されたか
  - 使われたアカウントの所有者の持つ他のアカウントの悪用の可否(研究用サーバへのログインの可否など)
- 例: 行方不明になった無線LAN接続端末の捜索

ネットワーク・フォレンジックスで重要なこと(1/2)

- 複数の機器に分散して証拠は残る
  - どこにどのような証拠が残るかを(事前に)考えておく
- 揮発性の証拠が多いので素早く動く必要がある
  - 場合によっては、証拠集めの順序を考えておく
- 機器の不注意な操作が証拠を汚染することもある
  - 証拠を汚染しない操作を事前に考えておく
- 止めることができない機器で作業することもある
  - 作業による影響は最小限に留める方法を考えておく

ネットワーク・フォレンジックスで重要なこと(2/2)

- 書き込み保護されたコピーで行うことが望ましい場合がある
  - 不用意な操作による証拠の破壊を防ぐことができる
  - ただし、正確なコピーかどうかを(司法において)保証できるかどうかが難しい
- 単に証拠集めだけでなく、何らかのアクションにつなげる必要も起こりうる
  - 例: マルウェアに感染した端末の通信を遮断
  - 例: マルウェア感染拡大行動の試みが見られたサブネットの遮断
  - 例: 行方不明になった端末の保護

フォレンジックス全体で重要なこと

- 対象はサイバー攻撃者のみではない
  - セキュリティ意識の薄い者
    - ネットワークを使いたいから適当なIPアドレスを設定して接続を試みる
  - 内部情報を売ろうとする者
    - ファイルサーバなどから本来は不要な情報を読み出して保存
- 証拠を提示する先が、それを証拠として認識してくれるか?
  - 組織の上司とか幹部とかは?
  - 警察とか裁判官とか外部は?
  - 2つのエビデンスを意識する
    - 意見や判断が妥当であるかどうかを示す情報や痕跡のこと
    - 法的捜査における事実を立証するのに用いられる情報、および、法廷で供述として適格である情報

フォレンジックスで重要な点

- 情報の収集(Obtain Information)
- 戦略(Strategize)
- 証拠の収集(Collect Evidence)
- 解析(Analyze)
- 報告(Report)

OSCARという頭文字語に

情報の収集(Obtain Information)

- 何が起こったか
- インシデントが発見された日付、時間
- 関わった人物
- 関わったシステム、データ
- 発見以降取られた対応

細かな調査の戦略立てに必要な情報を収集(予備調査的)

戦略(Strategize)

- 揮発性の証拠を確保するための素早い戦略の決定
- 各調査にかかる概算時間を理解する
- 持っている資源の(人員、時間、設備など)のリスト化
- 有力な証拠を持つ機器等を見極める
- 機器から証拠を得るためのコストを推定する
- 入手する証拠の優先順位をつける
- 第一段階の証拠の入手/解析計画を立てる
- (初期の解析を終えたあと、「戦略」に立ち戻ってさらに証拠を入手)

証拠の収集(Collect Evidence)

- 優先順位に従って証拠の収集
  - できるだけ早く取得する(合法的に)
  - 信頼出来るコピー作る(必要に応じて暗号化)
  - オリジナルを隔離し、管理とアクセスを制限する
- 証拠入手時の注意点
  - 証拠を集めるにあたってシステムへのアクセスや取った行動の全てに対して注意深く記録を取る
    - 証拠収集作業のミスを検証
  - 証拠自体となる物を保存し、余計な物を保存しない
  - 証拠管理の連鎖が保存されるようにする
- 解析は基本的に作成したコピーに対して行う

解析(Analyze)

- 複数のソースからの
  - 情報の相関を確認(タイムスタンプなど)
  - 時系列の構築
  - 重要イベントの確定
- 複数の証拠により信頼度(確証)を高める
- 仮説を立て、証拠の意味の評価(解釈)し、追加の潜在的な証拠を探索
- 収集→解析の繰り返しによる証拠を強化
- 解析には評判の良い信頼できるツールを使うこと

報告(Report)

- 誰に対しての報告かで、重視する点を考える
  - 例: 上司、管理者、幹部、警察、裁判官
  - 自分も相手も報告にかける時間を無制限に持っているわけではない
- 必要に応じて、調査結果を専門でない人に向けて、(自然科学的な厳密さを備えた上で)明確に説明する準備も
  - 例え話は諸刃の剣

証拠のカテゴリ

- 伝統的な物: 直接、目撃、状況、業務記録
  - 直接: USBデバイス、HDD、他のPC構成要素の中身
  - 目撃: デバイスの持ち出し、デバイスの接続
  - 状況: 関連を匂わせるチャットやSNSのログ
  - 業務記録: 電子メールなどを含む業務文書
- 電子システム上の物: デジタル、ネットワーク経由のデジタル
  - デジタル: 通信のセッション、サーバ/PC内のログ
  - ネットワーク経由のログ: IDS、ファイアウォール、Webプロキシ、認証サーバ、その他のネットワーク関係サーバのログ
    - プライバシーの問題との兼ね合いが難しい
- 「伝聞」は証拠の補強にはなるが単体では証拠にならない

フォレンジックの例1

ユーザID単位での802.1x認証の認証VLANでP2P検知されたのでユーザID利用者に通知

1. IDS等のログからNAPT出口のグローバルIPアドレス、ポート、時間を抽出
2. 当該グローバルIPアドレスに対するNAPTテーブルから当該時間にポートを割り当てたプライベートIPアドレスを同定
3. 802.1x認証+DHCPのログから接続時に使用されたユーザIDを抽出
4. ユーザIDに対応する利用者の詳細(コンタクト先)を得る

- IPv4 + NAPT + DHCP + 802.1x認証なシステムが重なると、個人にたどり着くまでがめんどくさい

フォレンジックスの例2(1/2)

大学内でノートPCが無くなった

- ノートPCがなくなったのは正確にはいつだろうか?
  - 利用者に最後に使った時間と無くなったのに気づいた時間を確認
  - 無くなったのに気づいた時間近辺の、無線LANアクセスポイント、ネットワークスイッチのARPテーブル、認証サーバのログ、などを探る
- 無くなった後に(別の場所で)学内ネットワークに接続されているか?
  - もし接続履歴があった場合は接続された場所を特定
- ノートPC上には重要データが保存されているか?
  - 成績情報などの重要データは保存していないか確認
  - ノートPCに保存されている可能性がある機密性が高いメールはあるか?
    - メールサーバ側のデータと突き合わせ

フォレンジックスの例2(2/2)

- 盗まれたと仮定して、盗人がさらなる悪用とアクセス権を利用していないか?
  - 学内の認証サーバやサービスに何か操作された記録はあるか?
    - ログを利用した犯人の絞り込み
    - さらなる情報漏えいが起きた可能性も考える
    - 単なる物品目的の窃盗ではなく情報も目当てだと考えられる

ネットワーク機器からのフォレンジックス

- ネットワークの証拠はたくさんのところにある
  - 侵入検知/防御システム(IDS/IPS)
  - ファイアウォール
  - DHCPサーバ
  - DNS(キャッシュorクエリログ)
  - Webプロキシ
  - (シングルサインオン)認証サーバ
  - メールサーバ
  - ネットワークスイッチ
- 揮発性が高いものが多いため、システムが動いている間に証拠を集める
- オンラインでデバイスと対話しなくてはならないことも
  - オンラインでの情報収集は環境を変更するので、影響を最小限に抑える

ネットワーク機器からの情報収集の戦略(1/3)

- システム時間を記録する
  - 常にデバイスと信頼出来るソースの時間のズレをチェック
    - 補正しなければ、証拠を関連づけることが難しくなる
  - ツールに時間のズレを補正する機能が無い場合、手動でログを比較するか、スクリプトを使用して補正
  - 長時間デバイスを利用していると時間のずれが変わるので、事前に定期的に時間を補正する設定(NTPの設定)をしておく
- 揮発性のレベルに応じて証拠を集める
  - ARPテーブルなどは数時間で消えるし、ログもローテートの(デフォルト)設定しだいでは1週間ぐらいしか持たないことも
  - ちゃんと順序をつけた方が全ての証拠を集めやすい
  - 揮発性の高い証拠は集めるのが難しいものが多いので、不必要なときはこの限りではない

ネットワーク機器からの情報収集の戦略(2/3)

- 環境全体に残す足跡を最小限に抑えながら証拠を集めなければならない
- デバイスの再起動やシャットダウンを控える
  - ネットワークベースの証拠は揮発性のメモリの中に存在していることが多い
    - 特にMACアドレスやARPのテーブルはリブート時に保存されないことが多い
  - ディスク領域が限られていて上書きされる設定がされている場合、シャットダウン/ブートログで上書きが入ることも
- ネットワークからではなく(シリアル)コンソールから接続して収集することも
  - ネットワークからだとトラフィックを生成し状態を変更してしまう
  - 攻撃者がいる場合、攻撃者に見つかってしまう可能性あり

ネットワーク機器からの収集の戦略 (3/3)

- 調査活動を記録する
  - 最近だとフォレンジック収集/分析ツールがいくつもあり、その中には調査活動の記録機能も
    - ただし、ツールを正しく使えていたか(抜けが無いか)の記録は別途あった方が良い
  - 別途、すべての活動を映像撮影で記録しておくのが良い
    - 最近だと映像からの文字認識して抽出するツールもある
  - CLIの場合、scriptコマンドやGNU screenの機能などで記録できる
  - GUIでは記録は難しいが、可能な限り画面のキャプチャ、写真、グラフィカルな接続記録を取る
- 調査には常に足跡を残すことになる点に注意
  - 証拠を得れば得るほど足跡は増えていく

ファイアウォール(+IDS/IPS)からのフォレンジックス(1/2)

- 通信元IPアドレス/ポート、通信先IPアドレス/ポート、プロトコル(TCP/UDP)は最低でも記録されている
  - 5タプル情報と呼ぶ
  - 他の情報: 接続試行, 開始時刻, 終了時刻, 総データ転送量, プロトコル, 使用したアプリケーション, パケットの内容など
  - このあたりをまとめて取得するなら、NetFlowやIPFIXなどのフローデータをネットワークスイッチから取得もあり
- パケットの冒頭部を利用したアプリケーションの判別も可能
  - 最近のIDSなどでは著名なサービスに関する通信は判別できる
    - Facebook, Twitterなど
  - ただし、プライバシーと機器の負荷上昇の問題がある
  - URLなどの特定の文字列とのマッチングもあり

ファイアウォール(+IDS/IPS)からのフォレンジックス(2/2)

- ファイアウォール等の設定によっては、サービスやデータが攻撃者に晒されたかどうかを明らかにすることもできる
  - ただし、監視ポイントの増加などで機器負荷に影響が出る
- 調査員が証拠を集めたり、システムにアクセスするためにファイアウォールの設定を変更する必要が出ることもある
  - ただし、設定変更中のログが不完全になる
- ファイアウォールに問題がある可能性も考える
  - 攻撃者によるログの消去(侵入された、ログを流すために大量のデータを送りつけた)
  - ファイアウォールの設定ミスによるロスト

DNS(キャッシュ)サーバ(1/2)

- DNSキャッシュサーバはTime To Liveで指示されている時間だけ名前解決のログをキャッシュとして保持される
  - 揮発する前にコピーすること
- ブラックリスト入りしているドメイン名の名前解決の有無は?
  - マルウェア配布とかC&C(C2)サーバの実績のあるドメイン
  - DNSBL(http://www.dnsbl.info/)にまとめられている
  - 最近だとVirusTotalの情報も良い
- 変なドメイン名の名前解決をしていないか?
  - やたらと長いドメイン名は無いか?
  - ドメイン名のエントロピーが高い物はあるか?
    →ランダムに生成したドメインの特徴
  - 変なデータを載せて(大量名前解決で)データ送信とか無いか?
  - やたらとTTLが短いドメインは無いか(IPアドレス単位でブロックされても移動させやすい)

DNS(キャッシュ)サーバ(2/2)

- 最初からDNSクエリログを保存するのもあり
  - 昔よりもディスク容量問題は緩和されている
- より多くのデータを保持すれば、新たな挙動が見えてきたりする
  - 変なクライアントからの名前解決結果は無いか?
    - 変に一定間隔(ビーコン)、変に大量送信
- 多くのDNSサーバソフトウェアでDNSクエリログ保存機能はある
  - Amazon Route 53サービスとかクラウドサービスでも機能はある

無線LANアクセスポイント

- 基本的に、接続端末のMACアドレスとDHCP/NAPTで与えたIPアドレスやポート番号のログを利用
- ハイエンド機種だと以下のようなログやログ関係機能もある
  - VPN接続機能
  - ルーティングログ
  - システムログ
  - SNMPやsyslogによるログの転送

    →ログの長期保存のために、備えていることが望ましい

802.1X認証サーバ

- 802.1XはLANの拡張認証フレームワークの規格
  - 接続時にユーザ名とパスワードで認証
    - nuwnet1xを考えてもらえばOK
  - 有線/無線を問わない
  - アクセスログが認証システム内に保存されている
- 802.1Xの認証システムのバックエンドからログを取得
  - RADIUSサーバ
  - Active Directoryサーバ
  - LDAPサーバ
- もちろん、802.1Xではなくネットワーク接続のcaptive portalのログもあり

他のサーバ

- DHCPサーバ
  - 接続端末のMACアドレスとIPアドレスの払い出し/更新ログ
- Webプロキシ
  - クライアントのIPアドレスと、そのクライアントがアクセスしたURLのログ
- メールサーバ
  - メールの送受信時間
  - クライアントによるメール到着の確認(POP, IMAP)時間
- グループウェアのサーバ
  - active/inactiveの記録が残っているかも

ネットワークスイッチ

- ARPテーブル(L2)
  - IPアドレスとMACアドレスの関係の解決の記録
- ルーティングテーブル(L3)
  - どのサブネットやVLANに通信を試みたかの記録

- いずれもかなり短時間で消える点に注意
  - CiscoのスイッチだとARPキャッシュはデフォルトで4時間

ログ取得(1/2)

- 調査員はデバイスログを次のように考えること
  - 事件に関係のある証拠を宝庫
  - 新しい証拠を集めるもの
  - 自分自身の活動の記録を残すもの
- 機器のローカルログ
  - デフォルトで様々なログをローカルメディアに保存
  - しかし、普通は高揮発性だし限られた容量しか保存しない
  - イベントが起こるたびにコンソールを確認しない限り情報は失われてしまうことも
    - コンソールログはキャプチャするにはカメラを使うという手もある
  - 無関係なログを生成して攻撃を紛らわすのはからよくある攻撃方法
- ターミナルログ
  - script等のツールでログを保存

ログ取得(2/2)

- SNMP
  - SNMP trapを使ってイベントを出力できる
  - ネットワークに流れるSNMP通信もキャプチャ可能
    - 通常は非暗号化UDPで送信するので簡単
    - リアルタイムデータを検索ならSNMP trapの盗聴がよい
- syslog
  - 古から広く利用されているログシステムの1つ
  - ログのローテートの設定は適切に(特に古いシステムは短いことが多い)
- Windows Event Log Monitoring
  - PowershellのGet-WinEventコマンドレットを使った検索と出力
- 認証などのアカウントに関するログ

機器へのアクセス無しでのフォレンジックス

- ポートスキャン
  - 空いているポートやポートを開けているソフトウェアのバージョン調べる効率のいい方法
  - nmap等のツールを利用
  - トラフィックを生成するので、ターゲットのデバイスの状態を変更してしまう可能性
- ペネトレーションテストによるターゲットのシステムの既知の脆弱性のチェック
  - トラフィックを作りデバイスの状態を変更
  - デバイスがクラッシュすることもあるので注意

パケット解析によるフォレンジックス

- トラフィックをキャプチャしてどうする?
  - 調査の性質に応じて文字列を探したり、ファイルを抽出したり
- パケット解析は次のような事象で役立つ
  - 疑わしいトラフィックに関するIDSからの警告を受け、その原因を特定したい
  - 組織の人員が機密データをエクスポートしているので、外向き通信を特定のキーワードでサーチしたい
  - 原因が特定できない謎のトラフィックを発見した
- パケットを見る時の注意
  - オクテット単位で転送されます
  - 基本はビッグエンディアンです
- ツールとしてはWiresharkが有名だか、シンプルなtcpdumpも使い方を覚えておくと良い

パケットからのプロトコル解析

- 理想: 仕様書があり、仕様通りプロトコルが実装されてる
  - 現実: 仕様書が無かったり、いい加減に実装されていたり
  - 知的財産の保護、競争の阻止、セキュリティの目的のため故意に機密にされていることも
  - プロトコルはシンプルだが文章化されてない事例(MSとか)
- いくつかのプロトコルはIETF指定規格で文章化
  - それでもベンダが適切に実装するとは限らない(MSとか)
  - プロトコルが標準化される前に実装されたり、部分的にしか実装しなかったりすることも
  - 規格を完璧に尊守していることは稀な事例も多い
- 攻撃者がこのことを悪用することは多い
  - 侵入検知システムやFirewallを回避
  - データの密輸
  - システム障害などの騒ぎを引き起こす

最初からパケット/プロトコル解析結果を保存しておくのもあり

- (内部向け)IDSも各種alartのログを保存してくれるが、反応していない通信の記録は無い
  →全トラフィックを解析保存しておけば?
  - もちろん、解析システムにもセンサノードにもお金はかかる
- Security Onion Project[1]
  - 内部的には、Snort/SuricataなどのOSSのIDSや、Zeekなどのパケット/プロトコル解析器を呼んで記録している
  - 次スライドにあるElastic Stackで可視化も

[1] https://securityonionsolutions.com/

Security Information and Event Manager (SIEM)

- ここまでのネットワークフォレンジックのためのデータの収集や分析を自動化する試み(製品)
- もちろん、複数のソースの相関をもとにした分析も実施
- 自分でこの手の物を作る場合、最近はElastics Stackを使うことが多い
  - OSSだったが、AWSがタダ乗りしている(組み込んだクラウドサーバを有償提供)のに怒って商用利用(外部提供)に制限された
  - Logstash: データ収集
  - Elasticsearch: データ検索と分析
  - Kibana: データの可視化
  - Packetbeat(Bests): パケット/プロトコル解析

フォレンジックの状況管理ツール

- ネットワーク以外も含めたフォレンジックの状況管理について、オープンソースのものも含めていくつも出ている
- Request Tracker for Incident Response (RTIR) [1]
- Fast Incident Response (FIR) [2]
- TheHive Project [3]

[1] https://bestpractical.com/rtir
[2] https://github.com/certsocietegenerale/FIR
[3] https://thehive-project.org/

その他のネットワークフォレンジックに関して事前に知っておくと良いもの

- 質の良い脅威インテリジェンス
  - MISP(Malware Information Sharing Platform)
  - The AlienVault Open Threat Exchange (OTX)
- Cheat Sheet(カンニングペーパー)と呼ばれるフォレンジックに向けたまとめ資料の質の良い物
  - (ある程度、同じ形式でまとまっている方がなお好ましい)
  - SANS InstituteのCheat Sheet[1]
  - Malware ArchaeologyのCheat Sheet[2]

[1] https://www.sans.org/blog/the-ultimate-list-of-sans-cheat-sheets/
[2] https://www.malwarearchaeology.com/cheat-sheets

産総研への標的型攻撃におけるフォレンジック

2019/7のひろみちゅ先生@産総研による招待講演の内容の一部

- 報告書[1]は力作なのでぜひ読んで欲しいとのこと
- パスワード攻撃の(フォレンジックの)リアルケース
  - 被害範囲の特定方法
  - パスワードスプレー攻撃の可視化
    - キーボード配列のパスワードを利用
  - ユーザID特定と攻撃成功の関係

[1] https://www.aist.go.jp/pdf/aist_j/topics/to2018/to20180720/20180720aist.pdf

初動

- タイムライン
  - (1/30にphishingメール)
  - 2/6にあるSEが自IDが他大学等からの接続に気づく
- 初動での判明
  - 41アカウントから不正ログイン(最終的に143)
  - 2箇所(国内、香港)から不正ログイン
  - 利用者のパスワードは比較的強固
- 不可解: ユーザIDは独自かつ独特なのに…、全IDがやられているのではない
- 疑問: 他の被害は? いつから? 単一案件か複合案件か?

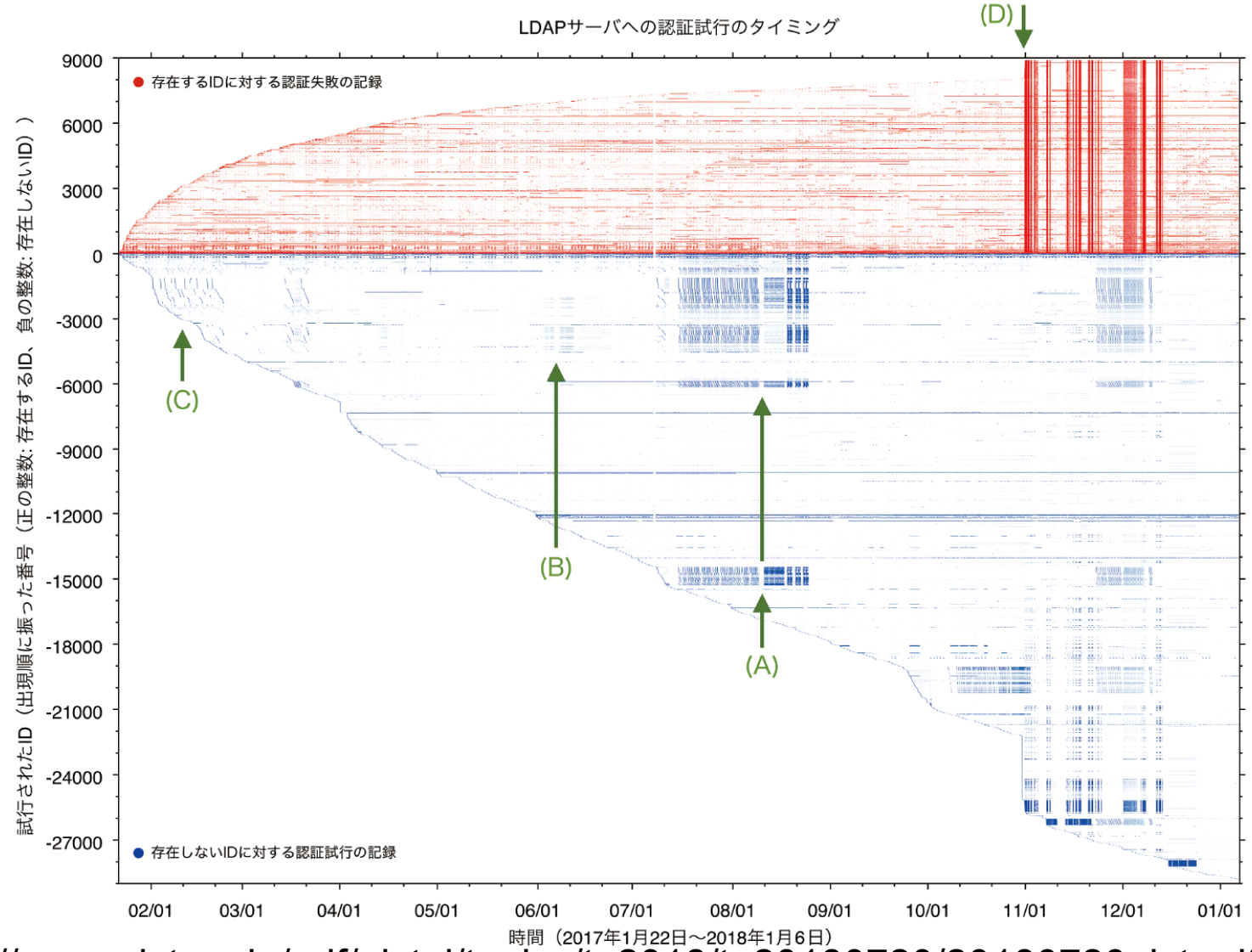
不正ログインの特定方法

1. Tor出口、既知の不正利用用VPN出口を不正IPアドレスとみなす
   - 他にも、ログインが短時間で1000km単位で動いたら不正とみなす
2. 不正IPアドレスからのログインは被害とみなす
3. 被害IDの全ログイン元IPアドレスを精査
   - 目視で問題なさそうならIPアドレスをホワイトリストに
   - 目視で問題ありそうならIPアドレスをブラックリストに
     - 海外出張などは個別に本人に確認
4. 新たに見つかったブラックリストのIPアドレスを不正IPアドレスとして2へ →繰り返し

パスワードスプレー攻撃の可視化

- 試行したIDとミス数をY軸、時間をX軸で点をプロット
  - 試行してミスした数で色が濃くなる
  - 存在しないアカウントへの攻撃が多いと色が濃くなる
- 可視化してわかったこと
  - 前年2月などにも色が濃い部分がみつかった
  - 土日に来ない、UTC+1で8:30-17:30に来た
- 最終的なID不正利用
  - 10/27に1人、11月に100人ぐらい、2月に40人ぐらい

可視化結果([1]のp.26より引用)



[1] https://www.aist.go.jp/pdf/aist_j/topics/to2018/to20180720/20180720aist.pdf

結果

- 最終的に、破られた人にパスワードを聞いて解析
  - パスワードを聞くのはあまりよろしくない(本人の傾向分かるので)

  →キーボード配列ベースのパスワードがほとんど
- 世の中の約1%はキーボード配列のパスワードの可能性
  - zxcvbnなる評価ツールは出ている
    - D. L. Wheeler, "zxcbb: Low-Budget password Sgreangth Estimation" 25th USENIX, 2016.

  →もっとキーボード配列ベースのパスワードの危険性の啓発を!

- パスワードスプレー攻撃と合わせると、IDが特定されていると1%のIDはすぐに破れることになる

フォレンジックに興味がある人へ(1/2)

- デジタル・フォレンジック研究会なる特定非営利活動法人が良い資料の刊行を行っている
  - https://digitalforensic.jp/
  - (少なくとも現時点で)学術団体ではない
    - 個人および団体で会員となることが可能
  - 講習会、イベント、分科会活動、書籍刊行、公開資料刊行、コラム刊行などの活動を行っている
  - 「証拠保全ガイドライン」は一般公開されている
    https://digitalforensic.jp/home/act/products/df-guideline-10th/
- IPAも良い資料を刊行している
  - 「インシデント対応へのフォレンジック技法の統合に関するガイド」
    https://www.ipa.go.jp/files/000025351.pdf
    - 「セキュリティ関連NIST文書」の下にある(一連の文書も
      https://www.ipa.go.jp/security/publications/nist/index.html

フォレンジックに興味がある人へ(2/2)

- 2023年の輪講で使った本が良書だった
  - オライリーブックス，"詳解 インシデントレスポンス　-現代のサイバー攻撃に対処するデジタルフォレンジックの基礎から実践まで-", 2022年1月.
    https://www.oreilly.co.jp/books/9784873119748/
  - 様々なおすすめ有償無償ツールへのポインタもあって良かった
  - 研究室図書も図書館カウンターで貸出申請して研究室側がOK出せば借りれます