JANOG的にLibrahackを解説する

京都大学
学術情報メディアセンター

上原哲太郎

本件理解のためのオススメURL

- 本家？ http://librahack.jp/
- たりき氏のまとめサイト
  岡崎市立中央図書館事件等 議論と検証のまとめ
  http://www26.atwiki.jp/librahack/
- 前田勝之氏の「サーバ管理者日記」
  http://www.nantoka.com/~kei/diary/
- 高木浩光氏の一連の記事
  http://takagi-hiromitsu.jp/diary

岡崎中央図書館事件（Librahack事件）とは

- 2つの事件に分けられる
- 事件1:「大量アクセス」事件
  - 岡崎中央図書館Webへの自動アクセスプログラムを記述・運用した男性が「偽計業務妨害」で逮捕され、22日間勾留され不起訴処分になった事件
- 事件2:「個人情報漏えい」事件
  - その岡崎中央図書館Webサーバ立ち上げ当時の内部の個人情報が、同システムを導入運用していた多数の図書館のサーバ内にばらまかれていた事件
- おまけ:Librahackの発音は「りぶらはっく」
  - 岡崎市の複合施設Libra（りぶら）に由来するから

発端としての「大量アクセス」事件

- 2010年3月半ば、岡崎市立中央図書館のWebページが度々停止するようになる
  - 原因は新着図書ページへのインターネットからの定期的・機械的アクセス（いわゆるクローラ）
  - 図書館情報システム納入業者は三菱電機インフォメーションシステムズ(MDIS)同社はこの事象について図書館に対し「ロボットによる『集中的なリンクアクセス』」「新着案内への『大量アクセス』」と説明

公開請求により開示された作業報告書より

- 3月19日「毎日18時頃にロボットによるリンクアクセスが実行されており、WEBサーバの高負荷の原因となっておりました。」
- 3月20日「ロボットによる集中的なリンクアクセスを防止するために、「CD/DVD一覧」の詳細情報の表示形式を変更いたしました。」
- 3月24日「新着案内に毎日18:00に大量アクセスがおこなわれ(3/14～)、WEBのデータベースが不調になっていた件につき対処を行いました。TOPからのアクセスするファイル名の変更を実施しています。」
- 4月1日「3/31に再度「新着資料」に大量のアクセスがあったため、ファイアウォールの設定を変更いたしました(17:00)。その後、該当のIPからのアクセスがブロックされていることを確認いたしました。」

MDISが取った対策？

- 「CD/DVD一覧」の詳細情報の表示形式を変更については不明
  - 推測だがそもそも誤認？
- TOPからのアクセスするファイル名の変更を実施はURLの「末尾に1文字付加しこれを毎日変更する」
  - …技術レベルが知れる
  - しかもパッケージ製品なのにそのレベルから変更可能
- ファイアウォールの設定を変更は、アクセスしてきたさくらインターネットのレンタルサーバのIPを遮断
- 問題は、遮断前に既に図書館は岡崎署に相談に行っている
  - 最初の相談はMDISから最初の報告を受けた3／19

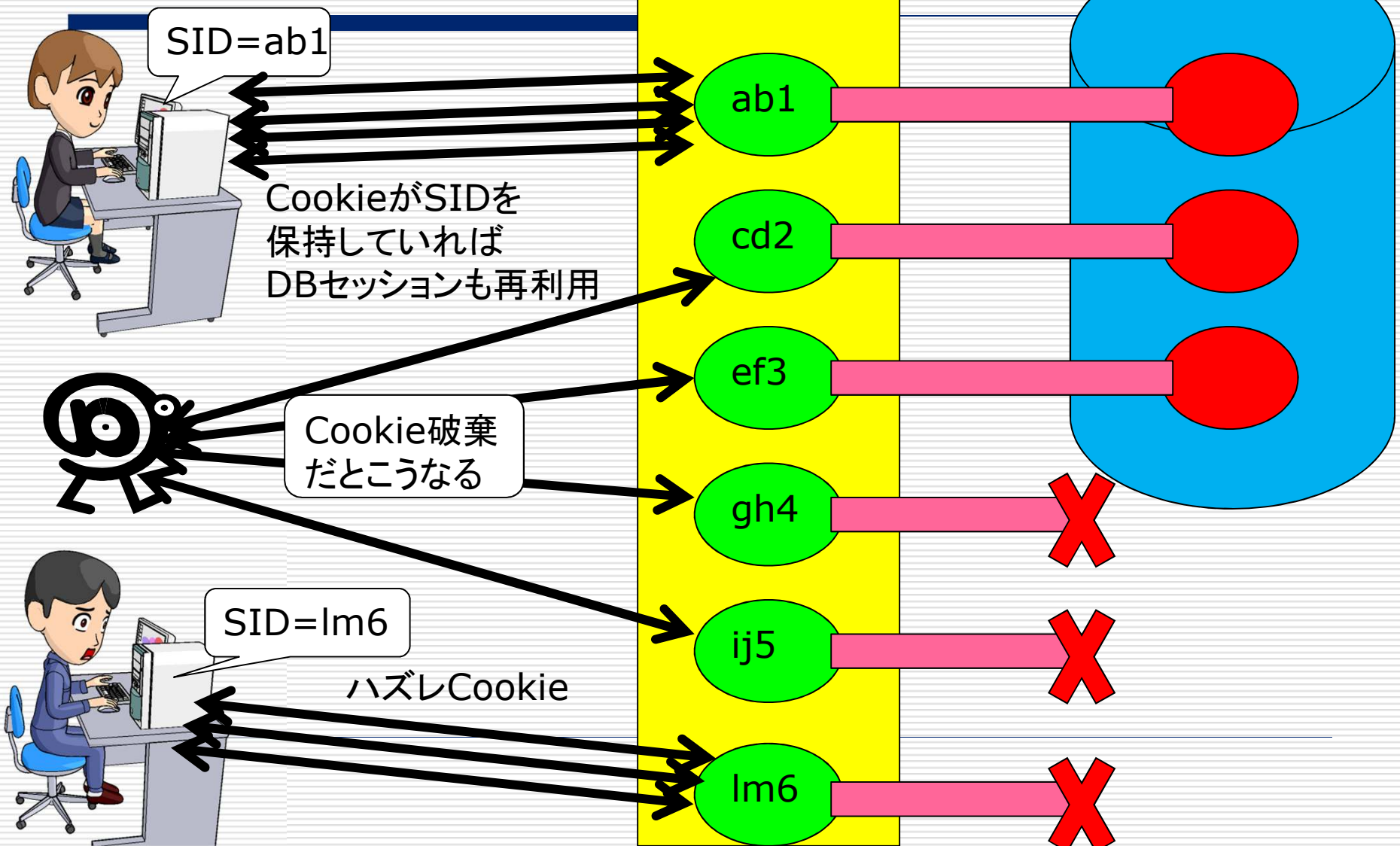
そもそも何が行われていたか

- アクセスしていたのはLibrahackこと中川氏
- 岡崎図書館の「新着図書」ページの使いにくさに業を煮やし「スクレイパー」を記述
  - 要は新着図書をDB化して差分を得ようとした
  - さくらのレンタルサーバで動作させる
  - 4月の遮断後は自宅または実家からパソコンで
- スクレイパーには元々負荷に配慮
  - シリアルアクセス、1リクエスト/秒に調整
    つまりとても「大量アクセス」と言える代物では？
  - 参考URL: http://librahack.jp/
- なのに何が起こったのか

MELIL/CSの設計ミス or バグ

- 蔵書DBはWebサーバのバックエンドのOracle
- WebブラウザのCookieをベースに作られるセッションオブジェクト毎にOracleのセッション生成
- よって、「Cookieを食わない」アクセスを連続して受けるとアクセス数分Oracleのセッションを食い潰す
- しかもマトモにエラー処理していないのでOracleセッションを握れなかったWebセッションオブジェクトがタイムアウトまで死なない
  - ハズレCookie問題
- つまりRobotが来るとそもそも必ず死ぬ構造
- 決して大量アクセスによる過負荷によるものではない

こんな感じ


Webサーバ
DBMS
SID=ab1
CookieがSIDを
保持していれば
DBセッションも再利用
Cookie破棄
だとこうなる
SID=lm6
ハズレCookie
ab1
cd2
ef3
gh4
ij5
lm6

MDISはこれをしらなかった？

- Robotが来るとシステムがおかしくなることは既知の問題だった
  - 同一システムの貝塚市等で問題発生済み
- そこでrobots.txtでロボットを避ける対応
  - これが後に国会図書館法違反になるのだがそれは別のお話…（cf.高木さんの説明）
- 岡崎市もそれで避けられていた模様なので岡崎としては初の障害事例に

これで最初の不幸が起きる

- 岡崎市はこれを警察に相談する(3/19)
  - まず民事的解決という発想はなかったのか？
  - 「図書館の自由」はどこにいった？！
- 警察は被害届提出を促し（4/2）岡崎はこれに応えて(4/15)刑事事件にする
  - この判断はどうか？
- 結果として中川氏は偽計業務妨害で逮捕(5/26)各紙で実名報道される
  - マスコミの姿勢はどうか？世間は？
  - ネット上ではこの時点から「おかしい」の声多し
- 22日間の勾留の後「起訴猶予処分」に

「大量アクセス」事件の提起した問題

- なぜMDIS、図書館、警察はシステムの問題と見ずに攻撃と見なしたのか
  - どういう行動が望ましかったか
  - 特にMDISはこの流れを止められる唯一の存在だったのではないのか？？
- 攻撃として捜査開始したとしても、警察は何故簡単に逮捕に踏み切った？
- なぜ検察は「起訴猶予処分」にした？
- プログラムを作る側からすると何をしておけば攻撃と見なされずに済むか？
- 報道被害はどうやったら救済されるのか？！

いくつかの「無知の連鎖」がある

- ネットにサービスしているはずの「図書館」がネットの常識にあまりにも無関心
  - 「クローラというものがあるというのは後で知った」
  - 「そういうことをするなら事前に教えて欲しかった」
- 警察がシステム障害というものの扱いにあまりにも無知
  - 「相性の問題ですね」
- 検察は最後までポイントを理解しなかった模様
- MDISはなぜこの連鎖を止められなかったのか

その後どうなったか

- その後、「個人情報漏えい事件」が発生する
  - MELIL/CSのWebサーバが「Anonymous ftp状態」になってたところが複数見つかり、中から…
- それを契機にMDISは「Sierとしての責任」を認めざるを得なくなる
  - 11月30日「個人情報漏えい」に関して謝罪会見
    しかし「大量アクセス事件」については謝罪せず
- これを機に岡崎図書館はMDISに対する態度を硬化、契約解除・指名停止などの措置
- MDISはプライバシーマーク一時停止処分を受ける
  - 平成23年1月24日から平成23年3月23日
- 「りぶらサポーターズクラブ」を仲介に岡崎図書館と中川氏に一定の和解(「共同宣言」)
  - しかし岡崎市は被害届の取り下げをせず(!)

ISPにとって…？

- 本件はIPAの「サービス妨害攻撃の対策等調査」報告書に反映される
  - ISPやJPCERT/CCに期待される役割を明記
- 警察への信頼が揺らぐ
  - 警察に持ち込めば逮捕されるまで止まらない？！
  - しかしISPが刑事or民事裁判に持ち込まれる前にできることは？
- サイバー犯罪対策条約に関連して法改正が予定されている:「ログ保全要請問題」
  - 刑事化のハードルは低くなる？