Innovative R&D by NTT

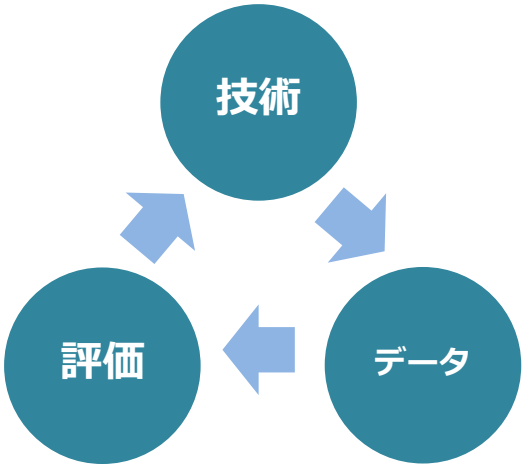
MWS Cup 2017 当日課題解説
課題１：Drive-by Download 攻撃解析

高田 雄太 (MWS 2017 企画委員)

2017年12月1日

課題１におけるこれまでの取り組み

- **2014: pcap を解析せよ！**
  - **概要：**複雑なリダイレクトの解析、クローキングの解析
  - **意図：**D3M データセット解析/活用の促進
- **2015: 改ざんされたサイトを発見せよ！**
  - **概要：**改ざんされたサイトの発見と解析
  - **意図：**独自データセット収集の促進
- **2016: 改ざんされたサイトを発見せよ！(再び)**
  - **概要：**昨年からの続きで、事前課題として出題し、解析の自動化や検知の工夫点を競争
  - **意図：**独自データセット収集の高度化および共有の促進



今年の方針

- **2017: pcap を解析せよ！(再び)**
  - **概要：**リダイレクトの解析、JavaScript 難読化の解析
  - **意図：**「独自データセット収集」から
    「データセット解析」へカムバック
    - “研究サイクル” の加速化
    - 自ら収集したデータ解析してる？塩漬けしてない？
  - 数年の間で変わらなかった攻撃特性である
    「リダイレクト」と「難読化」に関する問題を改めて出題

課題１－１

Drive-by Download 攻撃の 時系列解析

特定のある Web サイトの URL を1日ごとにおよそ半月の間アクセスした際の pcap を解析し、Drive-by Download 攻撃の時系列変化について、回答する課題

- 攻撃コードおよびマルウェアを含む通信データの分類問題
  - **悪性 Web サイトとして**振る舞う pcap
  - **悪性ではない Web サイトとして**振る舞う pcap
- 振る舞いを切り替える原因となっているWebコンテンツの解析
  - 悪性な通信データの前後日のデータ（HTML, JavaScript）を解析

【課題の意図】
- **応答Webコンテンツが変化**する悪性Webサイトの存在を知る
- 悪性Webサイトは存在し続けるのではなく、**消滅する可能性**があることを知る

課題１－１ 回答欄

一つわかると芋づる式に分かる、
ドライブバイの特徴を捉えた問題設計

ファイル名	Webコンテンツの変化内容
(F-1)	（U-1）において（U-2）を参照するHTMLタグが挿入され、また、（U-3）において使用されていたHTMLタグの参照先URLが、（U-4）から（U-2）へ変更された。 その結果、新しい参照先URL（U-2）に含まれる（U-5）を参照するJavaScriptが実行され、複数のURLへのアクセス後、最終的に攻撃コードを含むURLへのアクセスが**観測されるようになった。**
(F-2)	（U-2）にて使用されていた参照先URLが、（U-5）から（U-6）へ変更された。参照先URLが変更されたものの、（F-1）と同様の攻撃コードを含むURLへのアクセスが**観測された。**
(F-3)	（U-2）にて使用されていた参照先URLが、（U-6）から（U-7）へ変更された。そのため、（F-1）および（F-2）で観測されていた攻撃コードを含むURLへのアクセスは**観測されなくなった。**
(F-4)	（U-2）にて使用されていた参照先URLが、（U-7）から（U-8）へ変更された。その結果、（F-1）および（F-2）で観測されていた攻撃コードを含むURLへのアクセスが、再び**観測されるようになった。** しかし、翌日には（U-2）にて使用されていた参照先URLが、再度（U-7）に変更され、以降攻撃コードを含むURLへのアクセスは**観測されなくなった。**

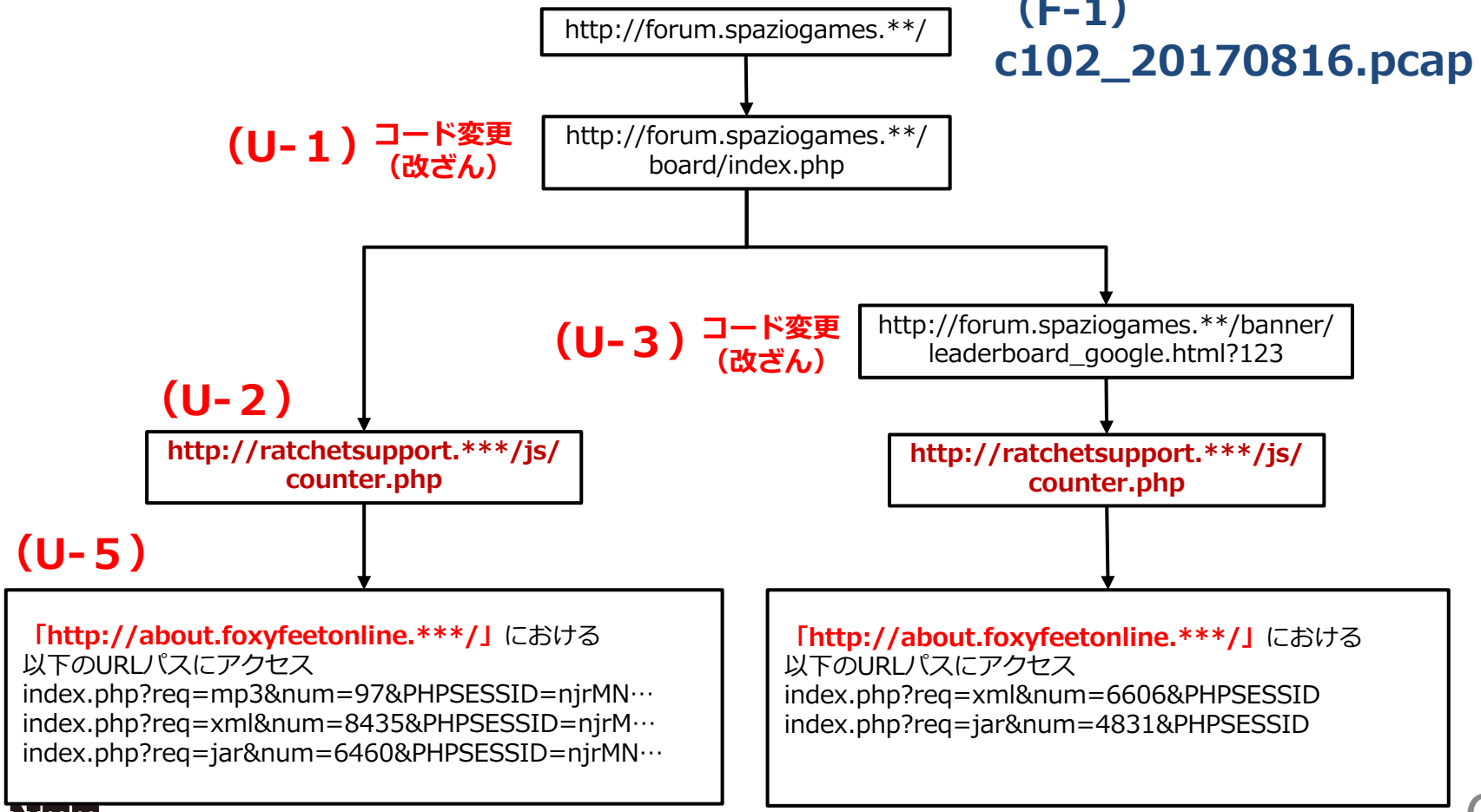
時系列解析 1/3

- **攻撃URL (攻撃コードを含むURL) の特定**
  - アクセスしたURL一覧を出力し、前後日で差分を分析することで、入口URLと異なるドメイン名や Content-Type 等を持つURLを特定
    - URL パス「?PHPSSESID=njrMNruDMh7HApz…」を含むURL
    - 差分URLの一部に「text/xml」「application/java-archive」
    - .mp3 を含む URL（実行ファイル）も重要な指標

- **アクセス先を変更する URL の特定**
  - 攻撃コードを含むURLを基にアクセス遷移を追跡
    - Referrer ヘッダやscript/iframeタグ、リダイレクトコード
  - 追跡したアクセス遷移を基に URL 変化の発生源を特定
    - **ratchetsupport.***/js/counter.php** を起点に、が転送先が変化していることに着目

時系列解析 2/3

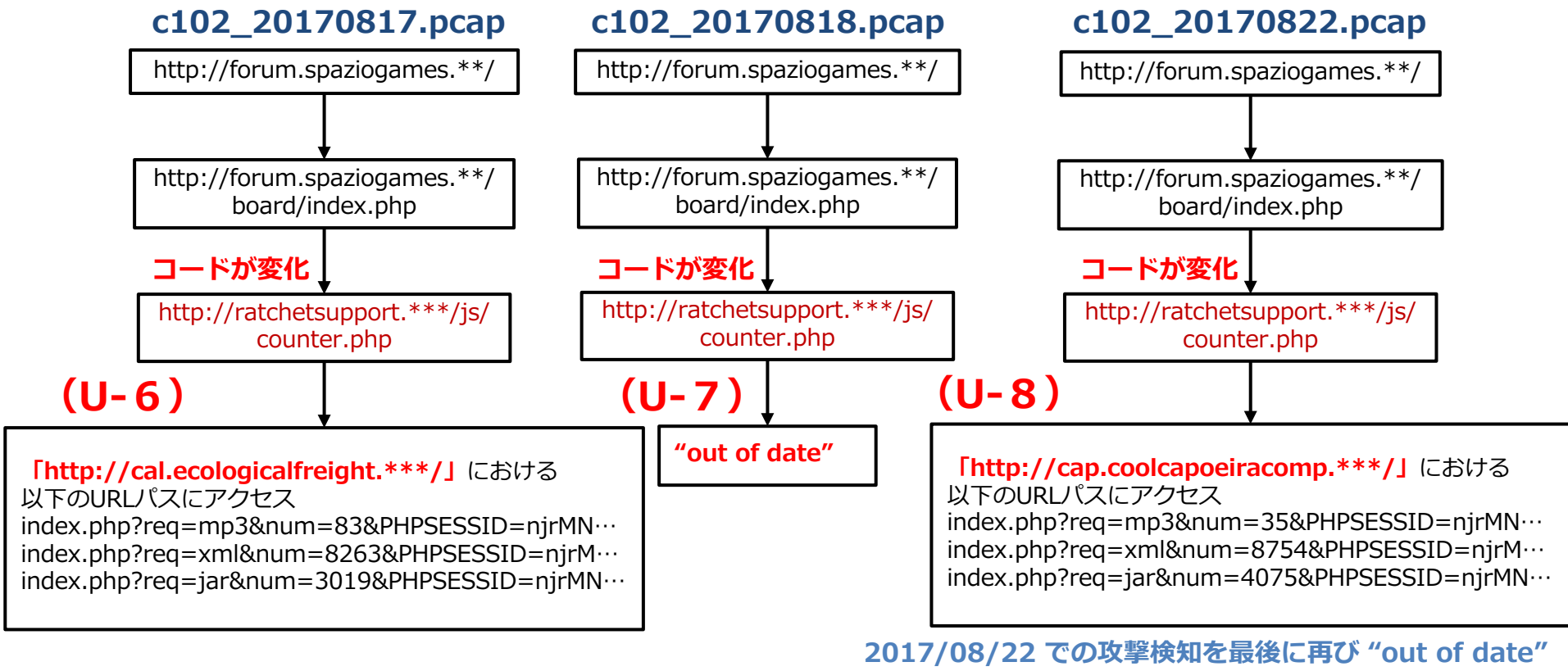
- 入口URLが読み込むファイルの内容が変更（改ざん）され、異なるドメインのURLを参照



CVE-2013-2456

時系列解析 3/3

- 転送先のドメイン名は変化し続け、同様の攻撃コードが実行される
- その後、転送先URLは “out of date” → 攻撃URL → “out of date” と変化し、最終的に攻撃は停止



課題 1 – 2

Innovative R&D by NTT

JavaScript の 難読化解除

悪性 Web サイトとの通信データを含む pcap の中から、
難読化された JavaScript を抽出・解析し、
難読化を解除したコードに含まれるコンテンツを回答する課題

- 難読化解除後に含まれるコンテンツの特定
  - 特定の関数に含まれる変数の値;
    generatePseudoRandomString() 内の変数 letters[0] を答えよ。

```
eval(function(p, a, c, k, e, r) {
  e = function(c) {
    return (c < a ? "" : e(parseInt(c / a))) + ((c = c % a) > 35 ? String.fromCharCode(c + 29) : c.toString(36));
  };
  if (!"".replace(/^/, String)) {↔}
  while (c--) {↔}
  return p;
}("D Q(f){w.3J=D(a,b){v c='';33(v i=0;i<b.F;i++){c+=K.J(a.R(i%a.F)^b.R(i))}G c};w.U=D(h){H(h.y(':'))h=h.Y(':')[0];v a
w[\"2J\"[\"s#@#n!$!s#!#!!!!$t@@r!\".l(/[^A-j-k-9\\+\\/\\=]/g,\"\")](10-4),(2-1))+\"2l\"[\"s#!u!!%b$!%@#s%!#%
1),(2-1))+\"22\"[\"s#$%$%##@u#b@#%!@@s%##!%$t@#$#r$%\".l(/[^A-j-k-9\\+\\/\\=]/g,\"\")](-7+15),(2-1))
[\"p!b#$%!%s@t$@!!$@@@@###r#$\".l(/[^A-j-k-9\\+\\/\\=]/g,\"\")](5-2),(1-0))+\"1H\"[\"s$!n#!@s$#%%@@!%
9\\+\\/\\=]/g,\"\")](0-0),(2-1))+\"1F\"[\"s$@!@%#@%n@@@#%!$!@!#!#s$@$#t@#$%###%%!r\".l(/[^A-j-k-9\\
```

【課題の意図】

- **JavaScript 難読化の解除方法**を知る
- **複雑な難読化**のケースを知る

課題１－２ 難読化解除 1/2

- 難読化された JavaScript を探索
  - ２つ存在するがいずれのファイルでもOK (common.js / browser_identificator.js)
- ２重に施されている難読化を解除
  - 一段目は **eval→console.log** へ書き換える等で解除可能 (易しい)

```
eval(function(p, a, c, k, e, r) {
  e = function(c) {
    return (c < a ? "" : e(parseInt(c / a))) + ((c = c
String.fromCharCode(c + 29) : c.toString(36));
  };
  if (!"".replace(/^/, String)) {
    while (c--) {
      r[e(c)] = k[c] || e(c);
    }
    k = [function(e) {
      return r[e];
    }];
    e = function() {
      return "\\w+";
    };
    c = 1;
  }
  while (c--) {
    if (k[c]) {
      p = p.replace(new RegExp("\\b" + e(c) + "\
    }
  }
  return p;
}("y N(f){p.2C=y(a,b){o c='';2z(o i=0;i<b.K;i++)
```

よくある packer

eval() 引数をダンプして、
難読化解除

```
function SvKxMmBSlXxAbl(f) {
  this.wTHPs = function(a, b) {
    var c = '';
    for (var i = 0; i < b.length; i++) {
      c += String.fromCharCode(a.charCodeAt(i %
b.charCodeAt(i))
    }
    return c
  };
  this.getTopHost = function(h) {
    if (h.indexOf(':')) h = h.split(':')[0];
    var a = h.split('.');
    while (a.length > 2) {
      a.shift()
    }
    return a.join('.')
  };
  this.GEwmMoSgXCGx = function() {
    return this["qRrpwLvkdOOynk"
["s!!#%%u#%b%!#s$#%$#$@!#t!!!$r@#$".repla
9\+\/\=]/g, "")](-3 + 7), (1 - 0)) + "RUzTeeqA
["s###$u%!#%!$%@#!%%##@!$b%!s!#@@t##!
-Za-z0-9\+\/\=]/g, "")](-2 + 5), (2 - 1)) + "SH
["su$#@!$##%$%$#$b$!%$#!!$#$%!%!@s#$$
[^A-Za-z0-9\+\/\=]/g, "")](-2 + 1), (1 - 0)) + "t
```

課題 1 – 2 難読化解除 2/2

Innovative R&D by NTT

- 二段目は単純な eval→console.log では解除不可能
  - 難読化解除の過程で**ホスト名** (window.location.host) が getTopHost(h) の引数とするため、pcap を参照し当該ファイルホスト名を引数 h に設定し、eval→console.log でコンテンツをダンプする (難しい！)

```
function SvKxMmBSlXxAbl(f) {
  this.wTHPs = function(a, b) {
    var c = '';
    for (var i = 0; i < b.length; i++) {
      c += String.fromCharCode(a.charCodeAt(i %
b.charCodeAt(i))
    }
    return c
  };
  this.getTopHost = function(h) {
    if (h.indexOf(':')) h = h.split(':')[0];
    var a = h.split('.');
    while (a.length > 2) {
      a.shift()
    }
    return a.join('.')
  };
  this.GEwmMoSgXCGx = function() {
    return this["qRrpwLvkdOOynk"
["s!!#%%u#%b%!#s$#%$#$@!#t!!!$r@#$".repla
9\+\/\=]/g, "")](-3 + 7), (1 - 0)) + "RUzTeeqA
["s###$u%!#%!$%@#!%%##@!$b%!s!#@@t##!
-Za-z0-9\+\/\=]/g, "")](-2 + 5), (2 - 1)) + "SHU
["su$#@!$##%$%$#$b$!%$#!!$#$%!%!@s#$$
[^A-Za-z0-9\+\/\=]/g, "")](2 - 1), (1 - 0)) + "t
```

一部関数を書き換えた後、eval() 引数をダンプして、難読化解除

どの箇所で解除に失敗するかをステップ実行で確認！

```
function createRandomNumber(r, Min, Max) {
  return Math.round((Max - Min) * r.next() + Min);
}

function generatePseudoRandomString(unix, length, zone) {
  var rand = new RandomNumberGenerator(unix);
  var subdomainlen = Math.floor(Math.random() * 32);
  var letters = "huozfexmrufmqhgnsvkehzrfrqoplpvbuaxoqeriqv
  var str = '';
  for (var i = 0; i < subdomainlen; i++) {
    str += letters[Math.floor(Math.random() * (letters.length - 1))];
  }
  str += '.'
  for (var i = 0; i < length; i++) {
    str += letters[createRandomNumber(rand, 0, letters.length - 1)];
  }

  return str + '.' + zone;
}
```

答えは “h”

まとめ

- **データセット収集と活用の促進**
  - 今後はデータセットの「共有」も活性化させたい
- **一組織による出題の限界**
  - 継続的な取り組みは重要である一方で、Drive-by に限らず、組織の偏りは課題：多様な人材・アイデアを募集！

ご興味がある方は問題作成支援委員として
ご協力 or アドバイスを是非！