Review Webserver Metafiles for Information Leakage (Bilgi Sızıntısı için Web Sunucusu Meta Dosyalarını İnceleyin)

Summary (Özet)

Bu bölüm, web uygulamasının yolunun veya işlevselliğinin bilgi sızıntısı için çeşitli meta veri dosyalarının nasıl test edileceğini açıklar. Ayrıca, Örümcekler, Robotlar veya Crawlers tarafından
kaçınılması gereken dizinlerin listesi, uygulama yoluyla Harita yürütme yolları için bir bağımlılık olarak da oluşturulabilir. Diğer bilgiler ayrıca saldırı yüzeyini, teknoloji ayrıntılarını tanımlamak veya sosyal mühendislik katılımında kullanılmak üzere toplanabilir.

Test Objectives (Test Hedefleri)

- Meta veri dosyalarının analizi yoluyla gizli veya bulanık yolları ve işlevselliği belirleyin.
- Eldeki sistemlerin daha iyi anlaşılmasına yol açabilecek diğer bilgileri özümsünün ve haritası.

How to Test (Nasıl Test Edilir)

> Aşağıda gerçekleştirilen eylemlerden herhangi biri wgetAyrıca yapılabilir. curl. . ZAP ve Burp Suite gibi birçok Dinamik Uygulama Güvenliği Testi (DAST) aracı, örümcek / kreş işlevselliğinin bir parçası olarak bu kaynaklar için kontroller veya ayrıştırma içerir. Ayrıca çeşitli Google Dorks kullanılarak veya gelişmiş arama özelliklerinden yararlanmak için de tanımlanabilirler. inurl:. .

Robots (Robotlar)

Web Spiders, Robots veya Crawlers bir web sayfasını alır ve daha fazla web içeriğini almak için daha fazla satır bağlantısını tekrarlarca geçer. Kabul edilen davranışları, web kök dizinindeki robots.txt dosyasının Robotlar Dışlama Protokolü ile belirtilmiştir.

Örnek olarak, başlangıcı `robots.txt` Google'dan 5 Mayıs'ta örneklenen dosya aşağıda aktarılır:

```
User-agent: *
Disallow: /search
Allow: /search/about
Allow: /search/static
Allow: /search/howsearchworks
Disallow: /sdch
...
```

Kullanıcı-Ajan direktifi, belirli web örümcek / robot / Crawler'ı ifade eder. Örneğin, The `User-Agent: Googlebot` Google'dan gelen örümceği ifade ederken `User-Agent: bingbot` Microsoft'tan bir tarayıcı anlamına gelir. `User-Agent: *` Yukarıdaki örnekte tüm ağ örümcekleri / robotlar / kazaklar için geçerlidir.

The (İngilizce) `Disallow` Direktif, hangi kaynakların örümcekler / robotlar / çürükler tarafından yasaklandığını belirtir. Yukarıdaki örnekte aşağıdakiler yasaktır:

```
...
Disallow: /search
...
Disallow: /sdch
...
```

Web örümcekleri / robotlar / yaralayıcılar kasıtlı olarak görmezden gelebilir `Disallow` Birinde belirtilen direktifler `robots.txt` Paylaşılan bağlantıların hala geçerli olmasını sağlamak için Sosyal Ağlardan gelenler gibi dosya. Bu nedenle, `robots.txt` Web içeriğine üçüncü taraflarca nasıl erişildiğine, saklandığı veya yeniden yayınlandığına dair kısıtlamaları uygulamak için bir mekanizma olarak düşünülmemelidir.

The (İngilizce) `robots.txt` Dosya, web sunucusunun web kök dizininden alınır. Örneğin, onu almak için `robots.txt` dan `www.google.com` kullanmak `wget` ya da `curl` : :

```
$ curl -O -Ss http://www.google.com/robots.txt && head -n5 robots.txt
User-agent: *
Disallow: /search
Allow: /search/about
Allow: /search/static
Allow: /search/howsearchworks
...
```

Analyze robots.txt Using Google Webmaster Tools (Google Webmaster Araçlarını Kullanarak robots.txt'i analiz edin)

Web sitesi sahipleri, web sitesini Google Web Yöneticisi Araçlarının bir parçası olarak analiz etmek için Google "Analize robots.txt" işlevini kullanabilirler. Bu araç teste yardımcı olabilir ve prosedür aşağıdaki gibidir:

1. Google Webmaster Araçları ile bir Google hesabıyla giriş yapın.
2. Gösterge panelinde, sitenin analiz edilmesi için URL girin.
3. Mevcut yöntemler arasında seçim yapın ve ekrandaki talimatı takip edin.

META Tags (META Etiketleri)

`<META>` Etiketler içinde yer almaktadır `HEAD` Her HTML belgesinin bölümü ve robot / örümcek / cırcırboş başlangıç noktasının web kökü, yani derin bir bağlantı dışındaki bir belge bağlantısından başlamaması durumunda bir web sitesinde tutarlı olmalıdır. Robotlar direktifi, belirli bir META etiketi kullanılarak da belirtilebilir.

Robots META Tag (Robotlar META Etiketi)

Eğer bir şey yoksa `<META NAME="ROBOTS" ... >` Giriş daha sonra "Robotlar Dışlama Protokolü" varsayılandır `INDEX,FOLLOW` Sırasıyla. Bu nedenle, "Robotlar Dışlama Protokolü" tarafından tanımlanan diğer iki geçerli giriş önyüklenmiştir. `NO...` - Yani. `NOINDEX` ve `NOFOLLOW` . .

İçinde listelenen Disallow direktifi(ler) temeline göre `robots.txt` Webroot dosyası, düzenli bir ifade araması `<META NAME="ROBOTS"` Her web sayfasından çıkarılır ve sonuç ile karşılaştırıldığında `robots.txt` Webroot'ta dosya.

Miscellaneous META Information Tags (Çeşitli META Bilgi Etiketleri)

Kuruluşlar genellikle ekran okuyucuları, sosyal ağ önizlemeleri, arama motoru indeksleme vb. Gibi çeşitli teknolojileri desteklemek için web içeriğine bilgilendirici META etiketleri yerleştirir. Bu tür
meta-bilgi bilgileri, kullanılan teknolojileri tanımlamada test edenlere ve keşfetmek ve test etmek için ek yollar / işlevsellik açısından değerli olabilir. Aşağıdaki meta bilgiler alındı
`www.whitehouse.gov` 2020 Mayıs 05 tarihinde Sayfalık Kaynak üzerinden:

```
...
<meta property="og:locale" content="en_US" />
<meta property="og:type" content="website" />
<meta property="og:title" content="The White House" />
<meta property="og:description" content="We, the citizens of America, are n
ow joined in a great national effort to rebuild our country and to restore its pro
mise for all. – President Donald Trump." />
<meta property="og:url" content="https://www.whitehouse.gov/" />
<meta property="og:site_name" content="The White House" />
<meta property="fb:app_id" content="1790466490985150" />
<meta property="og:image" content="https://www.whitehouse.gov/wp-conte
nt/uploads/2017/12/wh.gov-share-img_03-1024×538.png" />
<meta property="og:image:secure_url" content="https://www.whitehouse.go
v/wp-content/uploads/2017/12/wh.gov-share-img_03-1024×538.png" />
<meta name="twitter:card" content="summary_large_image" />
<meta name="twitter:description" content="We, the citizens of America, are n
ow joined in a great national effort to rebuild our country and to restore its pro
mise for all. – President Donald Trump." />
<meta name="twitter:title" content="The White House" />
<meta name="twitter:site" content="@whitehouse" />
<meta name="twitter:image" content="https://www.whitehouse.gov/wp-cont
ent/uploads/2017/12/wh.gov-share-img_03-1024×538.png" />
<meta name="twitter:creator" content="@whitehouse" />
...
<meta name="apple-mobile-web-app-title" content="The White House">
<meta name="application-name" content="The White House">
```

```
<meta name="msapplication-TileColor" content="#0c2644">
<meta name="theme-color" content="#f5f5f5">
...
```

Sitemaps (Site Haritaları)

Site haritası, bir geliştiricinin veya kuruluşun site veya uygulama tarafından sunulan sayfalar, videolar ve diğer dosyalar ve aralarındaki ilişki hakkında bilgi verebileceği bir dosyadır. Arama motorları sitenizi daha akıllıca keşfetmek için bu dosyayı kullanabilir. Testçiler kullanabilir `sitemap.xml` Site veya uygulama hakkında daha eksiksiz bir şekilde keşfetmek için daha fazla bilgi edinmek için dosyalar.

Aşağıdaki alıntı, Google'ın birincil site haritasından 2020 Mayıs 05 Mayıs'ta.

```
$ wget --no-verbose https://www.google.com/sitemap.xml && head -n8 sitem
ap.xml
2020-05-05 12:23:30 URL:https://www.google.com/sitemap.xml [2049] -> "si
temap.xml" [1]

<?xml version="1.0" encoding="UTF-8"?>
<sitemapindex xmlns="http://www.google.com/schemas/sitemap/0.84">
  <sitemap>
    <loc>https://www.google.com/gmail/sitemap.xml</loc>
  </sitemap>
  <sitemap>
    <loc>https://www.google.com/forms/sitemaps.xml</loc>
  </sitemap>
...
```

Oradan bir test cihazını keşfetmek, gmail site haritasını almak isteyebilir `https://www.google.com/gmail/sitemap.xml` : :

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9" xmlns:xhtml
="http://www.w3.org/1999/xhtml">
  <url>
```

```
    <loc>https://www.google.com/intl/am/gmail/about/</loc>
    <xhtml:link href="https://www.google.com/gmail/about/" hreflang="x-defau
lt" rel="alternate"/>
    <xhtml:link href="https://www.google.com/intl/el/gmail/about/" hreflang="e
l" rel="alternate"/>
    <xhtml:link href="https://www.google.com/intl/it/gmail/about/" hreflang="i
t" rel="alternate"/>
    <xhtml:link href="https://www.google.com/intl/ar/gmail/about/" hreflang="a
r" rel="alternate"/>
...
```

Security TXT (Güvenlik TXT)

`security.txt` Web sitelerinin güvenlik politikalarını ve iletişim bilgilerini tanımlamasına izin veren önerilen bir standarttır. Bunun, aşağıdakiler de dahil olmak üzere ancak bunlarla sınırlı olmamak üzere test senaryolarına ilgi duymasının birçok nedeni vardır:

- Keşif / analize dahil edilecek başka yol veya kaynakları tanımlamak.
- Açık Kaynak istihbarat toplantısı.
- Böcek Böbrekleri vb. Hakkında bilgi bulmak.
- Sosyal Mühendislik.

Dosya, web sunucusunun kökünde veya içinde bulunabilir. `.well-known/` Dizin. Eski:

- `https://example.com/security.txt`
- `https://example.com/.well-known/security.txt`

İşte LinkedIn 2020 05 Mayıs'tan alınan gerçek bir dünya örneği:

```
$ wget --no-verbose https://www.linkedin.com/.well-known/security.txt && ca
t security.txt
2020-05-07 12:56:51 URL:https://www.linkedin.com/.well-known/security.txt
[333/333] -> "security.txt" [1]
# Conforms to IETF `draft-foudil-securitytxt-07`
Contact: mailto:security@linkedin.com
```

```
Contact: https://www.linkedin.com/help/linkedin/answer/62924
Encryption: https://www.linkedin.com/help/linkedin/answer/79676
Canonical: https://www.linkedin.com/.well-known/security.txt
Policy: https://www.linkedin.com/help/linkedin/answer/62924
```

Humans TXT (İnsanlar TXT)

`humans.txt` Bir web sitesinin arkasındaki insanları tanımak için bir girişimdir. Web sitesini oluşturmaya katkıda bulunan farklı kişiler hakkında bilgi içeren bir metin dosyası şeklini alır. Daha fazla bilgi için daha sonra insanlara bakın. Bu dosya genellikle (her zaman olmasa da) kariyer veya şema siteleri / yollar için bilgi içerir.

Aşağıdaki örnek Google 2020 05 Mayıs'tan alınmıştır:

```
$ wget --no-verbose  https://www.google.com/humans.txt && cat humans.txt
2020-05-07 12:57:52 URL:https://www.google.com/humans.txt [286/286] -> "humans.txt" [1]
Google is built by a large team of engineers, designers, researchers, robots, a
nd others in many different sites across the globe. It is updated continuously,
and built with more tools and technologies than we can shake a stick at. If yo
u'd like to help us out, see careers.google.com.
```

Other .well-known Information Sources (Diğer . İyi Bilinen Bilgi Kaynakları)

Dosyaların standartlaştırılmış kullanımlarını öneren başka RFC'ler ve İnternet taslakları da var `.well-known/` Dizin. Listeleri burada veya burada bulunabilir.

Bir test cihazının RFC / taslakları gözden geçirmesi, bu tür dosyaların varlığını veya içeriğini doğrulamak için bir tarayıcıya veya buzzer'a tedarik edilecek bir liste oluşturması oldukça basit olacaktır.

Tools (Araçlar)

- Tarayıcı (Kaynak veya Dev Araçları işlevselliğini görüntüle)
- Kıvırma

- wget
- Burp Suite'in
- ZAP