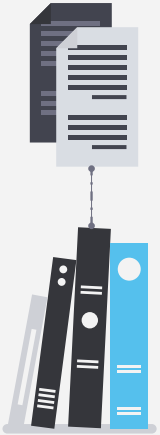
Analyzing the ecosystem of malicious URL redirection through longitudinal observation from honeypots

Computers & Security Volume 69, August 2017, Pages 155-173
Mitsuaki Akiyama, Takeshi Yagi, Takeshi Yada, Tatsuya Mori,Youki Kadobayashi

Introduction

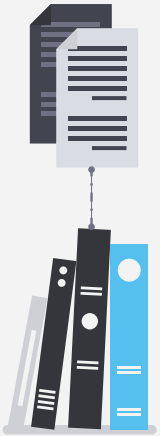


URL Redirection 被廣泛用在隱藏執行網站攻擊，攻擊者將redirect code 注入到被入侵的網站，以便將使用者導向惡意網站下載惡意軟體。

為何在如今有許多防禦方式的狀況下仍然會遇到許多活躍的惡意網站？作者認為這與 ecosystem of malicious redirection 有關。

因此作者建立 honeypot 系統，透過長期觀測來了解生態系統的演變。

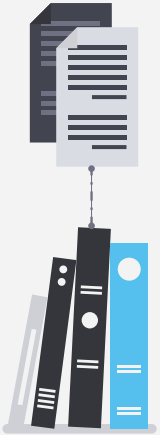
HTML injection in HTTP redirect body



http://127.0.0.1:9009?url=ws://example.com/"><script>alert(document.location)</script>

twisted.web.util.redirectTo 函數包含 HTML 注入漏洞。如果應用程式程式碼允許攻擊者控制重導向 URL，則此漏洞可能會導致在redirect response HTML body 中出現 Reflected XSS。

研究問題

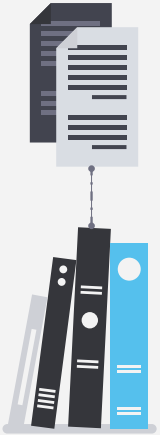


1. URL 重定向機制的關鍵特徵是什麼？
2. 隨著時間的推移，他們的目的是否改變了？

為了回答問題，作者開發一個honeypot 監控系統，專門監控URL 重導向的行為。

部屬四年，從776 個網站中提取超過10 萬個惡意重導向URL，這些網站因為被使用偷取的憑證訪問而受害。

主要發現

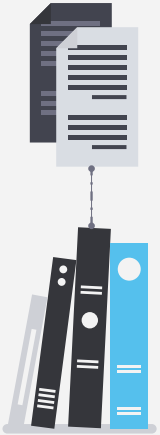


雖然透過 redirect code injection 的 URL 重導向其主要目的是引導下載惡意軟體，但近年來點擊詐欺成為新目的。

Domain Generation Algorithm (DGA)，最初用在 bot 與 C2 的通訊，現在更流行作為增加重導向 URL 的 entropy 來防止被列入黑名單的手法。

redirect chains 的中間站點應該同時使用domain-flux 和 IP-flux 來部屬，以確保穩健性。

What is URL redirection



Redirection 是自動取代存取目的地，通常透過HTTP/HTTPS 協定控制。
除此之外還有自動存取外部的網站內容，如 iframe tag，特別是對於網頁的攻擊。

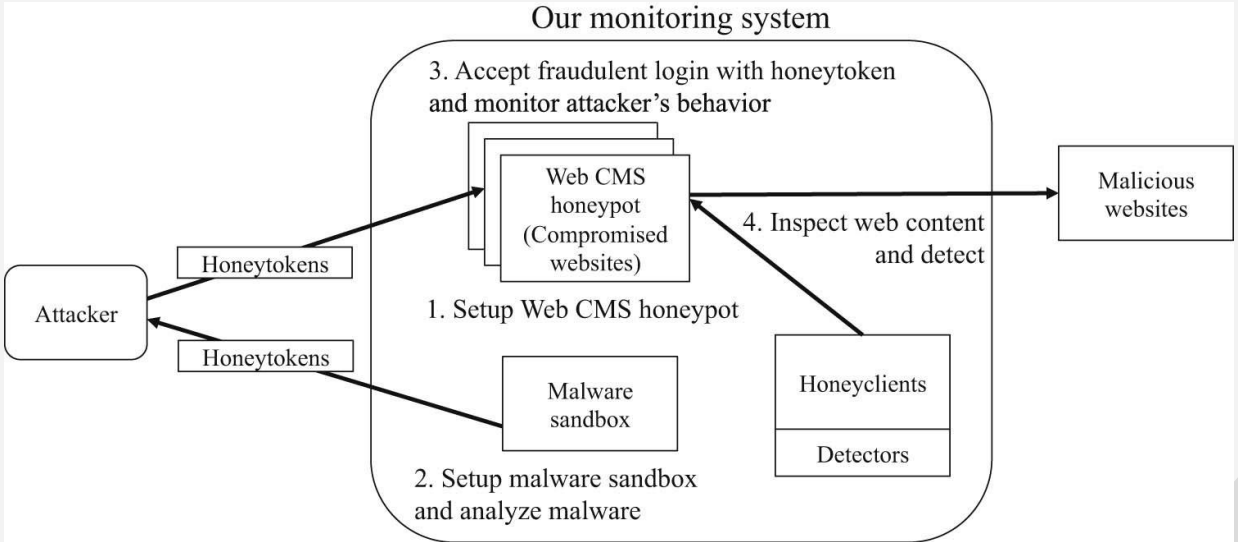
本論文的定義還包括自動發生的網頁存取(存取與初始 URL 對應的其他 URL)。
假設 URL 重導向的方法包含:

1. Tag Redirection, ex: <iframe>, <script>, <meta>
2. Script Redirection, ex: location.href, location.replace
3. HTTP Redirection, ex: HTTP-3xx

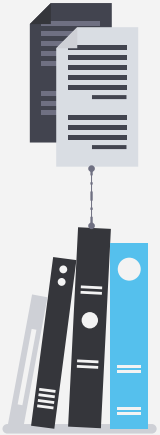
監控系統

需要先發現受感染的網站，以了解URL redirect injection。
還要使用 honeypot 來吸引攻擊者。

監控系統由多種類型的honeypot 組成，關鍵是故意向攻擊者洩漏Web CMS 的憑證(也是 honeypot)，稱為 honeytokens。



分析



分析注入URL 重導向的方法以及目標網站/URL。

監控系統有776 個網站受到入侵，其中96.5% 的網站被注入超過57K 次。
某些URL 直接使用IP，而不是domain。考慮到這種問題，作者會將這種URL 標示為網站的title。

根據統計，新出現的重導向包含11235 個網站，95.4% 使用domain，其餘使用IP。

Table 1. Summary of data.

Type	#
Period	39 months (Mar. 2012–May 2015)
Compromised websites	776 sites
Masquerader's login	59,462 logins
Content modification	57,009 logins
Inspections	323,581 times
Redirects (unique)	11,235 websites, 109,991 URLs

受感染網站上的 URL 重導向注入產生了超過 140 萬次 URL 存取，其中快一半由於 DNS 或 TCP 的錯誤而無法連線，只有 30% 的存取成功收到回應。

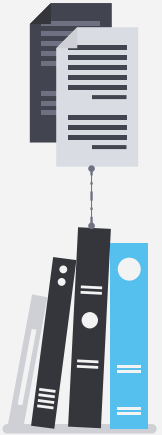
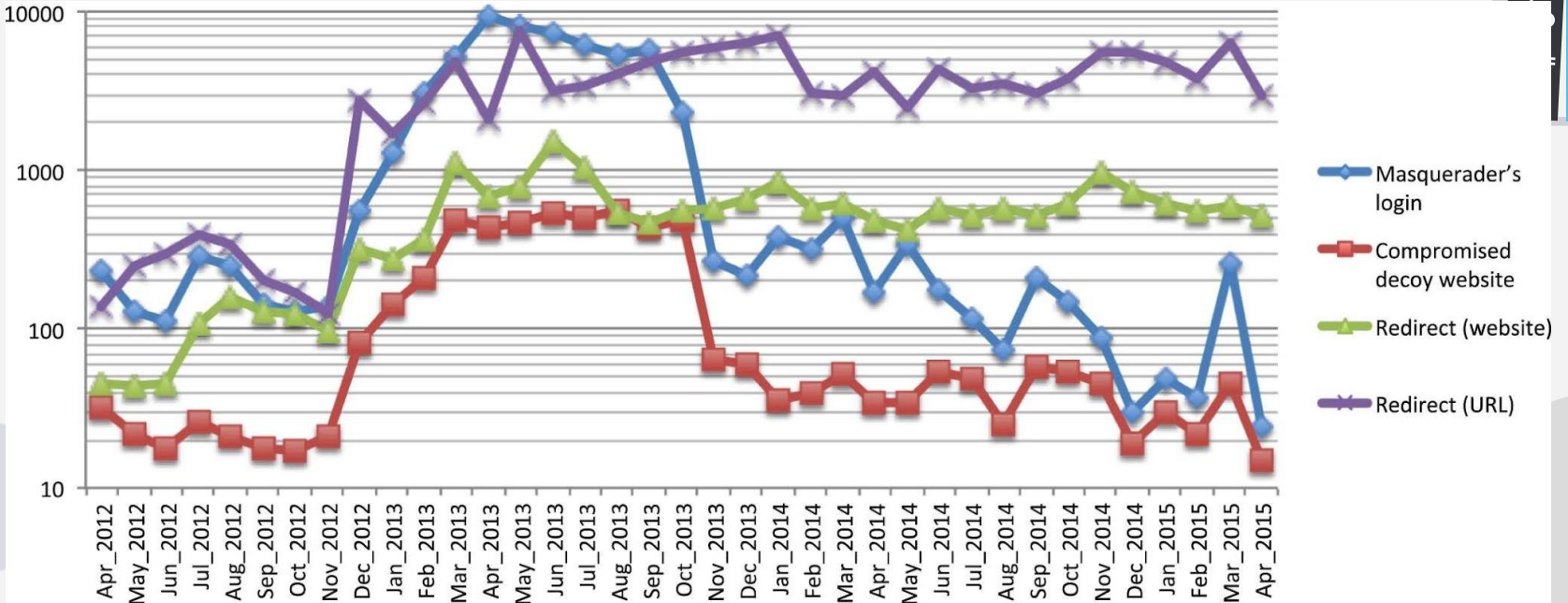


Table 2. URL status when they were accessed.

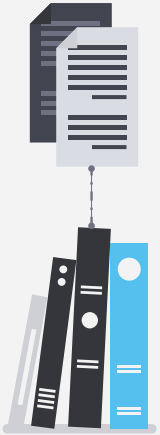
Protocol	Status	# of accesses (%)
DNS/TCP	DNS resource not found	731,272 (49.7)
	TCP connection error	56,206 (3.8)
	Other error	51,732 (3.5)
HTTP	HTTP-2xx successful	453,143 (30.8)
	HTTP-3xx redirection	83,047 (5.6)
	HTTP-4xx client error	91,371 (6.2)
	HTTP-5xx server error	4,436 (0.3)
Total		1,471,197

Injection activity

偽裝存取的數量有所減少，但重導向目標的數量仍然保持不變。
重導向目標數量持續存在的主要原因是網路廣告和自動生成的重導向目標共同造成的。



Detectors

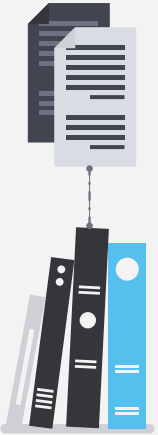


1. URL-based detector
2. Content-based detector
3. Exploit-based detector

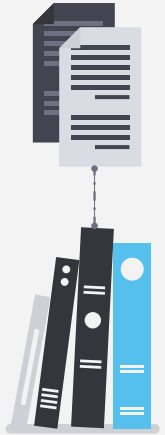
honeyclient 安裝一個有漏洞的瀏覽器以及外掛，並檢查輸入的 URL 以及使用重導向自動存取的相應 URL。
在 exploit-based 偵測，honeyclient 根據系統的異常行為來偵測。為了偵測已知漏洞，在 honeyclient 上將檢測模組放在有漏洞的地方，稱為hoenypatches。
他會檢查 data flow，並在輸入資料觸發漏洞時偵測漏洞。

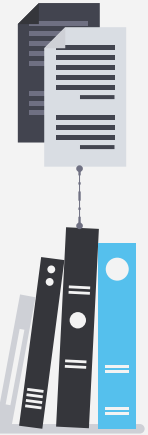
為了偵測未知漏洞，honeyclient 也對檔案系統、登錄和程序執行完整性檢查，還提取所有重導向目的地，包括檢查過程中出現的中間網站。

1. URL-based detector
   - 識別出 7296 個網站擁有的 12,841 個 URL
2. Content-based detector
   - 識別出與 3694 個網站擁有的 18,273 個 URL 相對應的 Web 內容
3. Exploit-based detector
   - 在 2841 次檢查中偵測到了偷渡式下載漏洞，其中包括 1080 個網站擁有的 6584 個 URL。懷疑被利用的網站為 74.1%，而明確試圖利用的網站為 9.61%。



標準的方法是使用script 、iframe 或 meta tag 進行標籤重導向。iframe 和 meta tag 插入率分別僅為 1.0% 和少於 0.1%。相較之下，隨著時間的推移, script tag 插入成為主要的注入方法，佔所有觀察到的檔案修改的70.8%。在這種情況下，幾乎所有script tag 插入都是沒有 src 屬性的script tag。這意味著混淆後的JavaScript 位於 scrtip tag 內，並在對其自身進行反混淆後動態輸出iframe tag 或執行位置重導向。由於這種混淆，惡意重導向程式碼隱藏了特定的重導向目的地（即URL）。



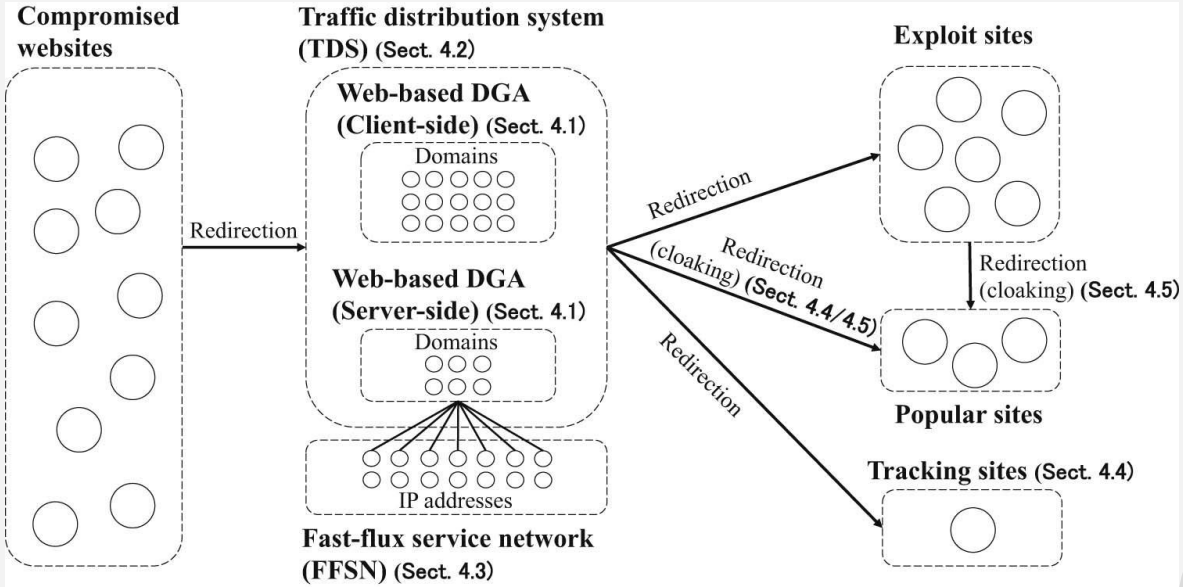


重導向目的地也不一定與惡意軟體感染有關，而是在某些情況下與網路廣告和追蹤有關。網路廣告和追蹤的目的是誘導點擊詐欺貨幣化，例如欺詐性地利用按點擊付費廣告(PPC)。
先前，攻擊者需要招募大量受惡意軟體感染的主機，例如DNSChanger、Koobface 和ZeroAccess，才能完成點擊詐欺貨幣化。相較之下，為受感染的網站注入重導向程式碼是招募一般公共網路使用者進行PPC 的替代方法。攻擊者入侵的網站越受歡迎，攻擊者無需付出太多努力就能同時招募更多的網路使用者。因此，注入重導向程式碼比招募點擊機器人更具成本效益和可擴展性。

這些 URL 通常在<path>部分中包含一次性令牌或編碼的單獨數據，例如時間戳、客戶端資訊和原始URL（引用URL），以便為每次訪問重新產生唯一的URL。

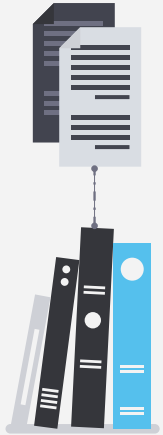
redirection techniques

1. domain-flux by web-based DGAs
2. IP-flux by fast-flux service networks (FFSNs)
3. redirection controlling by traffic distribution networks (TDSs)
4. target profiling by tracking services.



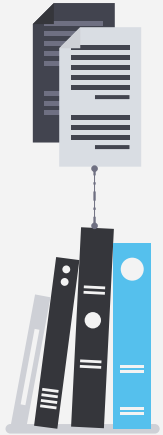
DGA 的使用自 2008 年開始變得流行。最初, DGA 用於感染後階段的C&C, 例如, 受惡意軟體感染的主機僅在很短的時間內與特定產生的網域(即C&C 伺服器)進行通訊。在觀察到的重定向URL 中發現了許多可疑的AGD。我們檢查 DGA 機制如何在URL 重導向生態系統中使用。我們的研究表明, DGA 也被用作重導向機制的關鍵組成部分。觀察到的URL 重定向 DGA 的使用大致分為兩類:

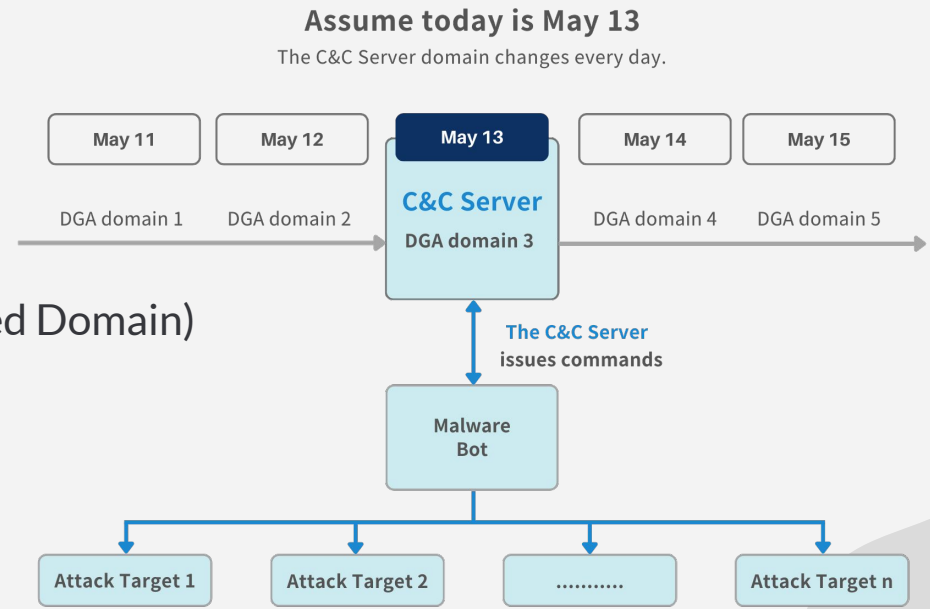
1. client-side domain generation (CDG)
2. server-side domain generation (SDG)



DGA and AGD

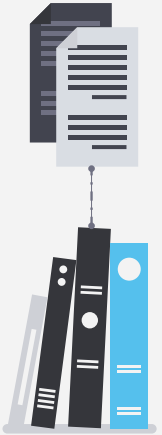
域名生成演算法(DGA, Domain Generation Algorithm)被用在一些惡意軟體或殭屍網路中，定期生成大量域名，作為其指揮控制伺服器的伺服器名字。每個 DGA 域名都是新生成的，因此對於傳統的惡意網站數據庫來說是未知的。這些域名通常很長，且不包含常見的英文單字。





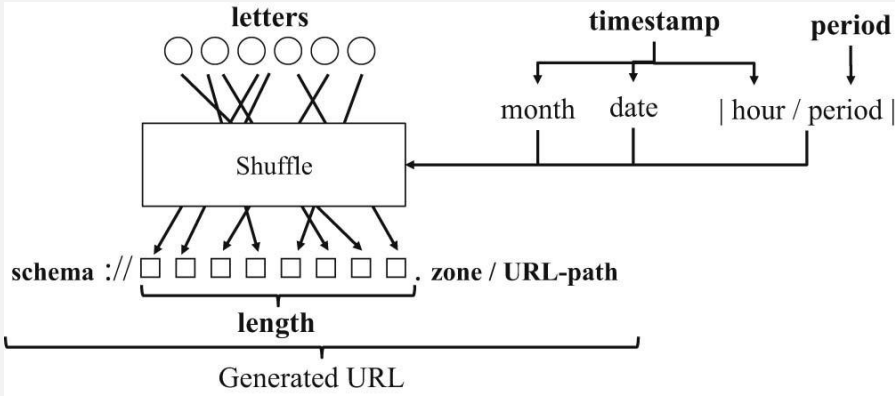
AGD (Algorithmically Generated Domain)

Client-side domain generation

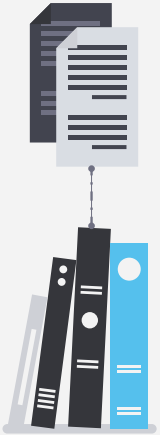


CDG 以 JavaScript 形式實作。當 Web 瀏覽器存取包含 JavaScript DGA 的頁面時，它會在瀏覽器上執行並偽隨機產生網域及其 URL。之後，JavaScript DGA 輸出重定向程式碼，例如 iframe tag 設定 src 屬性作為產生的 URL。

JavaScript DGA 高度混淆以阻礙分析。我們使用瀏覽器模擬器來完全執行混淆的 JavaScript，並提取 eval() 和 document.write() 的輸入/輸出值作為去混淆的人類可讀程式碼的候選值。然後，我們手動分析了提取的JavaScript，並確定了用於建立重導向 URL 的以下參數，包括偽隨機生成的網域。



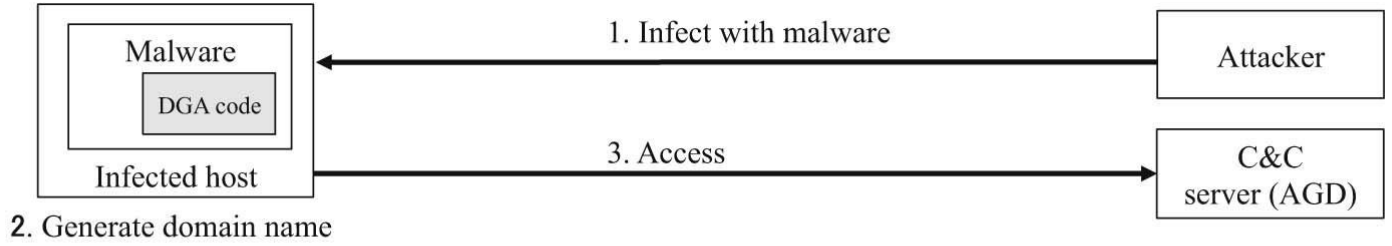
Server-side domain generation



在觀察到的重導向中，特徵URL 集合具有固定長度的隨機字元主機名稱和類似的URL 路徑（即 count[1-9][0-9]?\.php ）

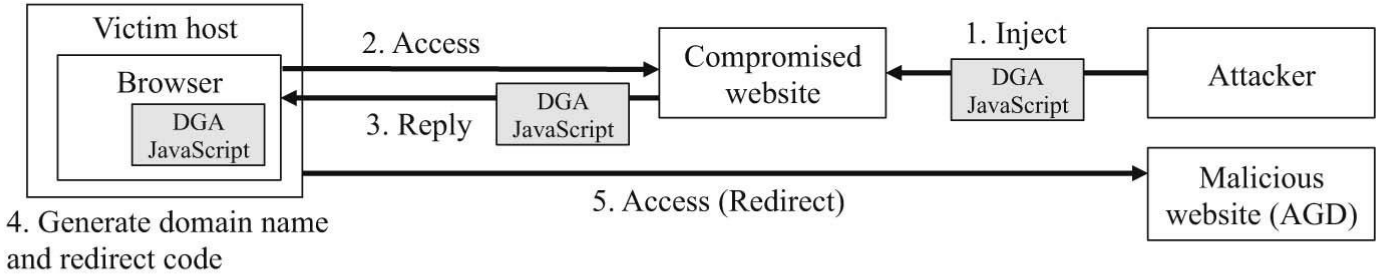
在注入的重導向程式碼中僅設定了硬編碼URL
假設攻擊者之前執行DGA 在伺服器端生成域名，並將帶有生成域名的重導向程式碼注入到他或她自己的受感染網站中。

C&C-based DGA

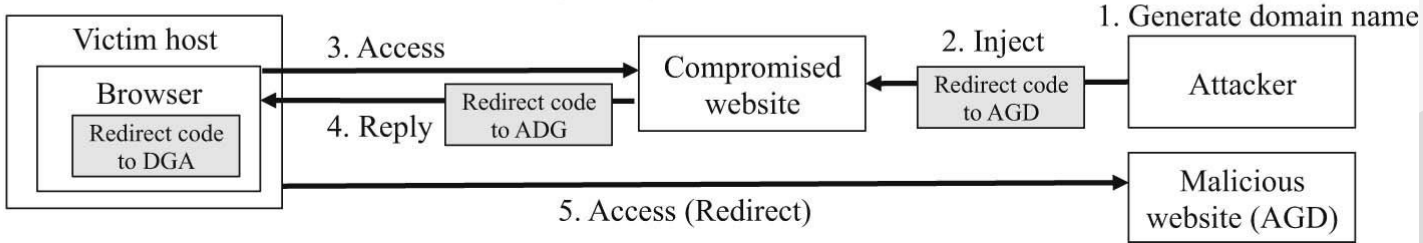


Web-based DGA

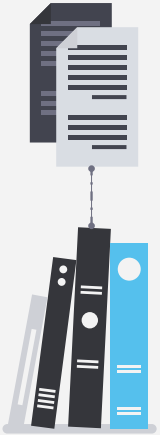
Client-side Domain Generation (CDG)



Server-side Domain Generation (SDG)



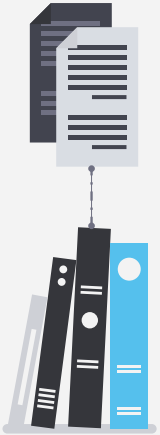
Traffic distribution systems



觀察到的 AGD 在 URL 路徑中有兩種特徵字串模式：count[1-9][0-9]?\.php 和 in.cgi\?[1-9][0-9]? 。其中一個（in.cgi ）在先前的論文中被稱為Sutra-TDS，它是一個用於構建流量分發系統（TDS）的工具包。

TDS 是放置在初始網站（即受損網站）和最終目標網站之間的中間網站。主要目的是控制最終重導向目的地以掩蓋它們。隨著時間的推移，最終網站會經常被TDS 更改。

Fast-flux service network (FFSN)

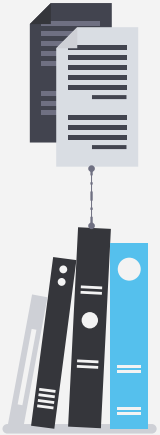


攻擊者用來隱藏惡意活動基礎設施（例如釣魚網站、惡意軟件分發或指揮與控制伺服器）的技術。它通過快速更改與域名相關聯的IP 地址來實現，這些IP 地址通常屬於受感染的設備（例如殭屍網絡中的主機）。這種技術使得追蹤和關閉惡意活動變得更加困難。

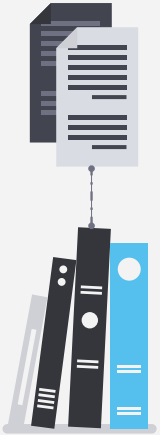
攻擊者會快速變更與該網域名稱相關聯的DNS 記錄，將多個 IP 位址與一個網域名稱相關聯。在註冊一個IP 位址幾分鐘或幾秒鐘後，就會取消註冊並替換為一個新的IP 位址。攻擊者可利用稱為循環DNS 的負載平衡技術，並為每個IP 位址設定極短的存留時間 (TTL) 來達成此目的。通常，使用的部分或全部IP 位址將是攻擊者入侵的Web 主機。在這些 IP 位址的機器將充當攻擊者原始伺服器的代理。

Mitigation

1. 對抗 DGAs
2. 透過 Sinkhole HTTP 請求發現未知的受感染網站
3. 禁用攻擊者的廣告和追蹤ID
4. 基於偽裝行為的可疑重導向路徑偵測



結論



1. RQ1：URL 重導向機制的關鍵特徵是什麼？
   a. URL 重導向機製表現出內在的變化和新趨勢
   b. 基於 Web 的 DGA 作為增加重導向URL 熵的手段已變得流行
   c. domain-flux 和 IP-flux 同時用於部署重導向鏈的中間站點，以確保重導向的穩健性
2. RQ2：隨著時間的推移，他們的目的是否改變了？
   a. 除了惡意軟體感染之外，點擊詐欺貨幣化最近已成為攻擊者的新目的
   b. 有趣的是，我們發現追蹤訪客（即受害者）統計資料的網路追蹤服務，被安裝到重導向URL 上