IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 36-46

Automated Analysis of Freeware Installers Promoted By Download Portal

Asst. Prof. Bhavna Arora[1], Asst. Prof. Nida Parkar[2], Asst. Prof. Priti Rumao[3]
[1](Department of computer Engineering, Atharva College Of Engineering,India)
[2](Department of computer Engineering, Atharva College Of Engineering,India)
[3](Department of computer Engineering, Atharva College Of Engineering,India)

Abstract: *Freeware is proprietary software that can be used free of charge. A popular vector for distributing freeware is download portals, i.e. websites that index, categorise, and host packages. download portals can be abused to distribute doubtlessly unwanted applications (doggy) and malware. The abuse may be due to doggy and malware authors importing their ware, by means of benign freeware authors joining as affiliate publishers of pay-according to-installation (PPI) services and other associate programs, or by means of malicious down load portal owners. The authors perform a scientific observe of abuse in download portals. They build a platform to move slowly down load portals and apply it to down load 191 okay home windows freeware installers from 20 download portals. They analyse the gathered installers and execute them in a sandbox to screen their set up. They degree an general ratio of domestic dog and malware between 8% (conservative estimate) and 26% (lax estimate). In 18 of the 20 down load portals examined the amount of puppy and malware is underneath nine%. but, additionally they locate two download portals exclusively used to distribute PPI downloaders. ultimately, they detail special abusive behaviours that authors of unwanted programs use to distribute their applications thru down load portals*.

Keywords: *VirusTotal, Freeware, Softonic, Freemium, Rogueware*

I. Introduction

Freeware is proprietary software program that may be used with out economic fee. Freeware is shipped in binary shape and ought to not be stressed with open-source software that is also unfastened however affords access to its source code. A related model is shareware in which the software program is to begin with free to use, however users are expected to pay to preserve the use of it. In contrast, freeware is free to use for limitless time. even as freeware can be used freed from fee, authors may additionally want to cowl their development fees and benefit from their freeware. this may be achieved through freemium fashions wherein the consumer pays for advanced capability, voluntary user donations, advertisements, and by using imparting third-birthday party software thru commercial pay-consistent with-install (PPI) services [1, 2]. as an example, Skype makes use of a freemium version wherein customers pay for calling telephone numbers and sun's Java gives customers to additionally install the Yahoo toolbar.

A popular vector for dispensing freeware is down load portals that are web sites that index, categorise, and host applications. down load portals including cnet [3], softonic [four], or tucows [5] are a assembly point for freeware authors that need to advertise their programs and for customers searching out a specific program or functionality. customers can leverage download portals for attempting to find popular freeware in a category (e.g. video software, security equipment, and windows issues), surfing through the program metadata (e.g. version, creator, platform), reading software reviews, and in the end downloading the chosen packages. download portals enable freeware authors to distribute their applications, increasing their user base. the use of download portals, the freeware writer can save on classified ads prices required to let users realize about the freeware's life. Authors on a low budget also can avoid putting in a webpage for the freeware. The down load portals invest in advertising and have a motivation to rank extraordinarily on seek engine consequences to attract users that may be monetised via classified ads and PPI schemes.

down load portals can be abused as a distribution vector for probably undesirable packages (pup) and malware. doggy are a class of unwanted software, that whilst now not outright malicious like malware, comprise behaviours taken into consideration dangerous via many users. even as the boundary between domestic dog and malware is occasionally blurry, prior paintings has attempted to delineate what constitutes domestic dog [6–eight] and businesses together with Google [9], Microsoft [10], and MalwareBytes [eleven] have regulations for defining what behaviours make a program pup. two styles of packages which are widely considered doggy are spyware that aggressively pushes classified ads and rogueware that scares users into shopping for a software program license, regardless of its restrained capability.

Download portals can be used to distribute domestic dog and malware in three approaches. First, download portals can be abused by way of doggy and malware authors to distribute their applications, by means of uploading their undesirable software and disguising it as doubtlessly beneficial freeware. second, authors of benign freeware may also become affiliate publishers of business PPI offerings, bundling their freeware with a PPI downloader and uploading the bundle to a download portal. whilst putting in the package, customers will be offered 1/3-birthday celebration advertiser applications, which can be puppy. In reality, prior work has measured that at the least 25% of puppy is shipped thru 24 industrial PPI offerings [1]. 1/3, download portal proprietors may be untrustworthy and use their download portals to purposefully distribute unwanted software to visitors.

While numerous weblog posts factor to download portals being bloated with doggy [12–14], their conclusions are primarily based on advert hoc measurements finished on the pinnacle downloaded packages of a few download portals. in this paintings, we perform a scientific have a look at of abuse in download portals. We build a platform to move slowly download portals. We use our platform to down load all windows freeware provided via 20 down load portals. This enables analyzing domestic dog and malware incidence beyond that of the most popular downloads. Our crawling downloads 191 ok packages from the 20 down load portals, which correspond to 157 k specific files with a cumulative size of 2.5 TB. We examine the amassed freeware to pick out doggy and malware and execute the programs in a sandbox to examine what changes they perform to the system, e.g. changing a browser's homepage and installing browser adjustments. Our analysis addresses the following three main question

- what percentage of programs in down load portals are pup and malware? We use two policies to quantify undesirable (i.e. domestic dog or malware) packages in download portals. Our conservative policy identifies as undesirable any software flagged by way of more than three AV engines, while our lax coverage identifies as unwanted any application flagged by as a minimum one AV engine. We degree an universal ratio of undesirable packages throughout all download portals analysed ranging among 8% (conservative) and 26% (lax). some of the undesirable packages pup (seventy six%) dominates malware (24%). For 18 of the 20 down load portals tested, the quantity of domestic dog and malware is mild ranging among 8.five% and a low as 0.2%. these ratios are significantly lower than those suggested in previous works that best study the pinnacle downloads [12–14]. We trust our measurements are greater correct since we study all programs indexed by means of a down load portal.
- Are there down load portals which might be virtually abusive? We become aware of two download portals, run with the aid of the same company, which serve a hundred% domestic dog. the ones download portals are exclusively used to distribute a PPI downloader. Regardless what software the consumer chooses to down load, he is always furnished with a customized PPI downloader. We provide a detailed analysis of the operation leveraging those two down load portals and become aware of some other 12 comparable download portals from the equal proprietors.
- How are down load portals abused? Our analysis uncovers exclusive abusive behaviours that authors of unwanted programs employ to distribute their packages through down load portals. We examine authors uploading the identical file as one of a kind applications in the identical down load portal, in addition to throughout more than one down load portals. We pick out a few authors using external hyperlinks to bypass security assessments by using download portals. We display that the failure to pick out repetitive abusers is full-size throughout download portals, as opposed to constrained to a few careless download portals. subsequently, we observe impersonation of benign famous authors through other authors that want to leverage their reputation, e.g. to hide their undesirable programs as harmless.

II. Download Portals

Download portals index massive quantities of packages from specific authors. To allow customers finding this system they're inquisitive about, or a software that matches a selected functionality, programs are usually grouped into categories and listed the use of keywords. download portals have existed for at least two decades with famous down load portals consisting of cnet and softonic being released in 1996 and 1997, respectively. The down load portals can also host the applications themselves, i.e. the report is downloaded from domains owned by way of the download portal, may also hyperlink to the writer's web site where the program is hosted; or may provide each forms of downloads.

Most download portals be given submissions from software program developers via paperwork where a developer specifies facts about its software program inclusive of software name, model, description, and application's URL. each download portal calls for exclusive facts from the developers. some down load portals also assist submissions thru the portable software description (PAD) well known, an XML schema delivered in 1998 via the affiliation of software Publishers to standardize software metadata [15].

Download portals face challenges in figuring out that a submitted program matches its description and that it's far uploaded by its actual writer. a few down load portals may additionally take steps in the direction of decreasing abuse, e.g. analysing submitted files the use of on-line offerings consisting of VirusTotal (VT) [16].

These days, some down load portals like filehippo and softonic have stopped accepting submissions with the aid of builders. these down load portals analyse the freeware surroundings themselves to choose new packages to add.

PPI: A popular software monetisation mechanism are PPI agreements where an advertiser, i.e. a software publisher interested in dispensing its software to users, will pay a third-birthday party to assist with the distribution. PPI agreements can be bilateral between software program publishers, e.g. Oracle distributing the Ask toolbar with its Java platform [17]. They also can take the form of commercial PPI services that connect a couple of advertisers interested by dispensing their programs with more than one associate publishers.

Willing to offer those marketed applications to users that set up the affiliate's program [1, 2]. associate publishers are often freeware authors that very own packages that customers need to put in. They package deal their freeware with a PPI downloader and distribute the bundle, e.g. by means of importing the package deal to download portals, in alternate for a commission paid for each installation. when a consumer installs the bundled freeware, the PPI downloader offers to the person the advertised programs. If the user installs an advertised software, the advertiser pays the PPI service for the set up and the affiliate publisher receives a fee. some down load portals which includes cnet run their personal business PPI provider to complement their commercial income. Freeware authors importing their applications to the down load portal are invited to sign up for as publishers of the download portal's PPI provider to monetise their programs.

An alternative PPI model is for software publishers to at once recruit associates to distribute their software program without the usage of a PPI service. a few huge doggy publishers have affiliate packages together with Mindspark, Babylon, Systweak, and Spigot [1]. Freeware authors can sign up as associates of such programs, attain a package and distribute it, e.g. through uploading to download portals.

Installers: most programs want to be established before they may be finished. The set up system might also, among others, test if machine requirements and dependencies are met, setup software files into a particular folder structure, configure services that run routinely, download sources from the internet, and make this system easy to launch (e.g. adding shortcuts and entries to the start menu). To ease set up, packages are frequently allotted as installers, i.e. auxiliary applications (e.g. setup.exe) responsible for putting in a target application. most documents downloaded from down load portals correspond to installers.

Analysed down load portals: We analyse 20 down load portals that offer windows applications. the chosen portals may additionally offer packages for different structures as nicely, however we crawl simplest the windows packages they provide. we have selected down load portals that target unique geographic places and of different reputation (in keeping with their Alexa rating [18]) because there may be differences in behaviours between those sorts, e.g. how they vet publishers leading to extraordinary quantities of abuse. table 1 indicates the listing of 20 down load portals analysed in this work. For every down load portal, it shows the abbreviated call we use to consult the download portal, its URL, the Alexa rating, whether the down load portal accepts software submissions the usage of forms and PAD files, and the structures for which it indexes programs. down load portals that do not take delivery of submissions thru bureaucracy nor PAD may also receive them thru electronic mail or now not receive submissions at all.

III. Approach

Our method strategies one down load portal at a time and comprises 4 steps: instruction, crawling, report processing, and execution. all through guidance, an analyst manually generates one portal metadata record for the download portal, which includes all the wished information to crawl the down load portal. The crawling, report processing, and execution steps are automated and illustrated in Fig. 1. The crawling takes as input a portal metadata file, and mechanically navigates a selenium-based totally crawler [36] to download all of the packages the download portal offers. It outputs the downloaded files and saves all facts about the crawling right into a important database. The file processing extracts facts statically from the downloaded files (e.g. filename, filetype), collects the report file from VT [sixteen], extracts executable documents from records, and exams the authenticode virtual signature (if signed). The execution takes as enter the programs, runs them in a sandbox, generates an execution document, and saves the report records inside the crucial database. every of these four steps is detailed next in its very own subsection.

Table 1: Download portals analysed in this work, their Alexa ranking (from 10 October 2016), whether they accept software submissions through forms or PAD files, and the platforms covered (Windows, Android, MacOS, Linux). GP means it redirects to Google Play

Portal name		Submission				Platforms	
	Alexa	Form	PAD	Win.	And.	Mac	Lin.
uptodown [19]	155	✓	✗	✓	✓	✓	✓
cnet [3]	164	✓	✓	✓	✓	✓	✗
softonic [4]	200	✗	✗	✓	GP	✓	✗

filehippo [20]	615	✗	✗	✓	✗	✗	✗
softpedia [21]	1589	✓	✓	✓	✓	✗	✗
soft112 [22]	4672	✗	✓	✓	GP	✓	✓
majorgeeks [23]	6206	✗	✗	✓	GP	✗	✗
soft32 [24]	6640	✓	✓	✓	✓	✓	✗
eazel [25]	8760	✗	✗	✓	✓	✓	✗
fileforum [26]	9449	✓	✗	✓	✗	✓	✓
filehorse [27]	9980	✗	✗	✓	✗	✓	✗
portalprogramas [28]	12,171	✓	✗	✓	✓	✓	✗
freewarefiles [29]	13,556	✓	✗	✓	✗	✗	✗
tucows [5]	25,084	✓	✓	✓	✗	✓	✓
snapfiles [30]	33,545	✓	✗	✓	✗	✗	✗
filecluster [31]	56,379	✓	✓	✓	✗	✗	✗
descargarmp3 [32]	104,352	✗	✗	✓	GP	✗	✗
download3000 [33]	230,115	✓	✗	✓	✗	✗	✗
fileguru [34]	308,929	✓	✓	✓	✗	✗	✗
geardownload [35]	466,545	✓	✓	✓	✗	✗	✗

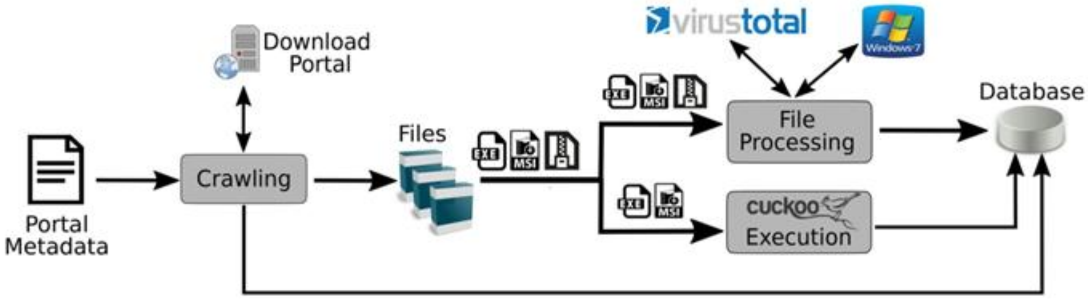


Fig. 1:*Summary of our approach for one download portal, excluding the preparation step*

III. I Practise

download portals share the purpose of allowing users to discover the software that they are inquisitive about. This creates similarities of their shape and layout. but, each download portal is special. as an instance, down load portals range of their layout and the records they acquire about the programs. This makes it tough to crawl a huge number of down load portals to analyse the software program they index. To address this difficulty, at the beginning of this mission we studied more than one down load portals to become aware of a commonplace abstraction that labored for they all and permits adding a new down load portal to be crawled with very constrained attempt.

All the down load portals we've tested percentage a simple shape in which every application within the down load portal has its personal program page, which gives the facts that the down load portal has gathered on the application. the program web page may additionally contain, amongst others, this system name, model, description, license, platform, language, length, creator, wide variety of downloads, critiques, date of book, date of last update, screenshots, previous versions, and down load links. The precise program attributes inside the program page vary across down load portals. one-of-a-kind packages within the same download portal can also have specific application attributes, e.g. if an upload shape has non-obligatory fields that a few authors fill and others do no longer. however, we had been able to identify a subset of six application attributes which might be available for all packages in all download portals analysed: call, model, platform, length, writer, and down load link.

The output of the practise step is a portal metadata report that has all the information needed for the crawling. mainly, the portal metadata record contains 3 components: the way to listing all of the software pages, how to discover the six application attributes from this system page, and the way to download the program document from the download link or button within the application page. We describe them next.

list the program pages: we can classify download portals into lessons concerning a way to listing all program pages inside the down load portal. elegance 1 down load portals permit us to list all programs immediately and sophistication 2 down load portals permit us to list all programs in each class. class 1 download portals may additionally offer an immediate link to listing all applications or may additionally enable looking with wildcard identifiers via which we can list all programs. for class 1 down load portals, we will extract a URL template including http://portal.com/software program/?page=X, in which X is a advantageous integer that may be monotonically extended to iterate over the hunt results. each page in this iteration incorporates, amongst other content material, a listing of program entries, every with a software URL that factors to the program web page. To discover this system URLs within the search outcomes pages, we use the path in the page's DOM. for sophistication 2 download portals, we need to first acquire the listing of all program classes. Then, we will outline a URL template which includes http:// portal.com/<category>/?web page=X and we are able to iterate over the search results of each category by using monotonically growing X and extracting this system URLs much like category 1 down load portals. The analyst gives within the portal metadata document: an URL template for sophistication 1 and class 2 down load portals, and a listing of all software program classes for class 2 down load portals. Figuring out the program attributes: To pick out the program attributes inside the program web page (call, model, platform, length, writer, and down load hyperlink), the analyst gives inside the portal metadata report a DOM route for every program characteristic, which uniquely identifies the position of a page element that incorporates the characteristic.

Downloading this system files: the program web page constantly carries an element to download the program, e.g. a down load hyperlink or a download button. however, that element won't link without delay to the document to down load (e.g. executable or archive). as an alternative, it may open a download web page with some classified ads and one or greater download links, e.g. from distinct mirrors. We call download URL to the URL from in which our crawler downloads this system report, i.e. the URL that factors to this system report. The analyst affords in the portal metadata record the sequence of clicks, i.e. DOM paths to buttons to be clicked, starting with the clicking on the download element in the software web page that the crawler desires to carry out to reach and click on at the download URL of the cutting-edge application model.

We advanced the process above iteratively. once we converged on this processing, generating a portal metadata document for a brand new download portal took us 2–3 h.

III. II Crawling

The crawling takes as enter the portal metadata record. It outputs this system documents downloaded from the portal and saves right into a significant database the crawling information inclusive of timestamps, URLs visited, and program attributes amassed from the down load portal. We use a crawler primarily based on selenium WebDriver [36] with Mozilla Firefox. The crawler follows the instructions within the portal metadata file to listing all program pages. For each software web page, it identifies this system attributes the use of the DOM paths in the portal metadata document and shops the attributes. If the program is a home windows freeware, it clicks at the down load link, and follows the commands in the portal metadata report to find the down load URL.

As soon as the download URL is clicked, an try and download this system report is done. If the download has now not started out after 30 s, the crawler tries again. A file down load may additionally fail for the following five motives: (i) the download hyperlink is broken; (ii) the down load has no longer finished in 5 min, a timeout we chose to restriction the most size of a downloaded document. This timeout provides sizeable garage financial savings and handiest affects documents 2 hundred MB–2 GB depending on the bandwidth of the download portal (despite this restriction the downloaded applications use 2.five TB disk garage); (iii) any internet page has now not completed loading inside 30 s; (iv) the down load link redirects our crawler to an external web page and does now not factor directly to a application file, i.e. the user is expected to discover the right download link in the publisher's web site; (v) the download portal refreshes the webpage with which our crawler is interacting, e.g. to change an commercial.

After every down load strive, whether a success or now not, the crawler outputs a tuple with: timestamp, download portal identifier, errors code, application URL, down load URL, 4 program attributes from this system web page (program name, length, version, and author), and a record identifier if the download become a hit.

we are interested by whether the down load portals host the downloaded programs onsite or absolutely redirect customers to the publisher's website. For this, we manually construct a mapping, the use of Whois and DNS records, of which domains in the down load URLs belong to each download portal. for example, cnet makes use of effective 2d-level domains for website hosting packages: cnet.com and downloadnow.com.

III.III Report processing

The record processing step statically analyses the documents and saves all data in the imperative database. It first tactics each downloaded report to gain: MD5, SHA1, SHA256, size on disk, filename, and filetype. Then, it tries to decompress documents to extract any executables inside. observe that a downloaded archive may also include other records, so that is a recursive method, which we element in segment four.1. next, the document hash of every executable, at once downloaded or extracted from an archive, is used to query VT [16], an internet provider that examines consumer-submitted documents with a large number of security equipment. on the time of scripting this paper, VT analyses submitted documents using 70 AV engines consisting of all important AV providers.

(https://www.virustotal.com/en/about/credits/). Engines are frequently updated and the listing of engines evolves over time. If the record is known to VT, a report is downloaded that contains, amongst others, the variety of AV engines that detect the report, the timestamp whilst the report turned into first submitted, and file metadata. We publish to VT all files downloaded from the download portals that are smaller than 30 MB. This threshold is because of the VT API, which has a restrict of 32 MB. We decreased the restriction to 30 MB due to the fact we determined mistakes with files close to the limit. We use maliciousness regulations: conservative and lax. The conservative policy is to take into account a application undesirable (i.e. doggy or malware) if it detected by way of extra than 3 AV engines inside the VT record. This coverage is designed to minimise false positives due to some AVs committing an error inside the detection. The lax policy considers unwanted any software detected by using at the least one AV engine. We use the lax policy as an higher bound. For programs dispensed as records, we remember them unwanted if the archive itself, or any report within the archive, satisfies the coverage. To classify an unwanted executable as either malware or pup, and to decide its circle of relatives, we use AVClass [37], a malware labelling device. AVClass takes as input the AV labels in a VT report; removes noise from AV labels by addressing label normalisation, widespread token detection, and alias detection; and outputs for each sample whether or not it's miles doggy or malware, its most possibly circle of relatives name, and a self belief issue primarily based on the agreement throughout AV engines. The final document processing step is to analyse the signed executables. Code signing is a way that embeds a digital signature in an executable, which enables verifying the program's integrity and authenticating its writer. previous work has proven that nicely signed executables detected with the aid of AVs are predominantly doggy, when you consider that it's miles difficult for malware to reap a legitimate code signing certificates from a Certification Authority (CA) [38]. The report processing issue validates the authenticode signature in executable documents. For this, it uploads all executables to a windows VM and uses the Microsoft-provided validation device to check if the executable is signed and whether or not the signature validates the usage of one-of-a-kind guidelines (e.g. default and kernel driver). Signed executables are in addition processed via our personal code to retrieve the X.509 leaf certificate and extract, among others: concern CN, company CN, PEM and DER hashes, validity duration, signing hash, virtual signature algorithm, signed record hash (known as Authentihash), and public key. For executables which are signed and whose signature validates, we will with a bit of luck discover the writer's identity.

III. IV Execution

We run downloaded executables inside the Cuckoo Sandbox [39]. Cuckoo receives an executable, assigns it to a VM for execution, and generates a behavioural file for the execution. we've got carried out a few extensions to Cuckoo to higher suit our wishes. these consist of adding a few anti-anti-evasion techniques to harden the sandbox [forty, 41], extending the Graphical person Interface (GUI) exploration, and adding signatures for precise events that we want to be notified approximately. these modifications are unique under.

The huge majority of executables downloaded from the down load portals correspond to installers (e.g. firefox_setup.exe) a good way to installation the actual program binaries at the end host (e.g. firefox.exe) upon execution. Such installers are normally GUI- based and require user interplay to finish the installation. Cuckoo affords capability to discover buttons in home windows released at some stage in the execution, and to routinely click on buttons labelled with keywords together with ‘next’ or ‘affirm’ simulating the default user behaviour of accepting all home windows to quick set up this system. but, the listing of key phrases used to identify those buttons is pretty small and restricted to English. as a consequence, we extended the keyword listing and translated the keywords into popular languages such as German and Spanish.

We also extended the signatures module of Cuckoo, which enables defining signatures for activities of hobby that Cuckoo can immediately record. This module offers overall performance improvements to perceive the ones occasions. as an example, we should test the listing of registry changes provided through Cuckoo to see if a specific key that stores internet Explorer's homepage has been modified. however, it's miles considerably extra green to construct a signature for that registry key and allow Cuckoo routinely record its modification. Our signatures include events along with whether browser extensions were installed, and whether some browser settings have been altered.

We configure Cuckoo to use 30 VirtualBox VMs on a single host strolling Ubuntu sixteen.04 LTS. every VM is configured with 1 GB of RAM and 20 GB hard disk. The VMs run home windows 7, which continues to be the most famous OS [42]. Our VM image has a large quantity of popular programs mounted such as internet Explorer, Chrome, Firefox, Opera, Java, Adobe Reader, Adobe Flash player, and the.net framework. that is finished so that we are able to take a look at changes that the performed packages may carry out on those packages.

IV. Evaluation

This phase evaluates our method. section four.1 offers the results of crawling the download portals, section four.2 examines the prevalence of unwanted applications in down load portals, section four.three compares the doggy, malware, and benign packages behaviours, phase four.4 summarises the execution outcomes, and segment four.5 gives a case have a look at on down load portals related to PPI offerings.

IV. I Download portals crawling information

on this segment, we present some general information on the crawling. We examine the safety factors along with incidence of unwanted packages and abusive behaviours within the subsequent sections.

table 2 summarises the crawling outcomes. For each download portal it presents: the date when it changed into crawled (all dates from 2016), the variety of home windows programs supplied, the variety of efficaciously downloaded packages, the quantity of particular downloaded documents by hash, the dimensions of the downloaded documents (in GB), the cut up of the precise files by using kind (EXE, ZIP, RAR, MSI, and different), and the share of unique files hosted onsite (i.e. on domains that belong to the download portal).

overall, we downloaded 191 okay applications from the 20 download portals, corresponding to 157 ok specific documents with a cumulative size on disk of 2.5 TB. The downloaded files correspond to sixty five% of the 325 k provided applications. segment 3.2 details the motives why a down load can fail. the biggest download portals are soft112 and softpedia with 107 and sixty nine okay presented programs, respectively. We downloaded the most from softpedia with 48 k specific files, accompanied through soft112 with 43 ok. The smallest download portals were download3000 and filehorse with less than one thousand applications offered each, and 275–350 unique files downloaded. word that the down load portals are sorted by way of Alexa rating (equal order as table 2), displaying that reputation does not have a right away correspondence with length. for instance, uptodown, cnet, softonic, and filehippo all have higher Alexa ranking than the two biggest download portals.

File kinds: most downloaded files are executables (forty eight%) and records (46%). extra especially, of the precise documents downloaded forty eight% are EXEs, 40% ZIP data, 3% MSI installers, 2.7% RAR data, 1.6% JAR files, any other 2% other styles of records (.gzip, .bz2, .7z, and .cab) and the final are comprised of a long tail of over 70 filetypes consisting of JPEG, textual content files, ISO images, office documents, PDFs, and source documents (e.g. personal home page, Python, C). We robotically decompress documents, finding an additional 10 M files (a hundred and seventy okay executables) internal.

Signed executables: Of the seventy five,615 downloaded executables, 39% (29,228) are signed. of these signed executables, 76% validate efficaciously on home windows, 20% have expired certificates, 1% have revoked certificates, and the remaining three% generate various validation mistakes. There are down load portals (descargarmp3 and eazel) that sign all their executables, and each of those download portals uses a unmarried code signing certificates for signing the executables. We carry out an in-intensity analysis of these two down load portals in section 4.five.

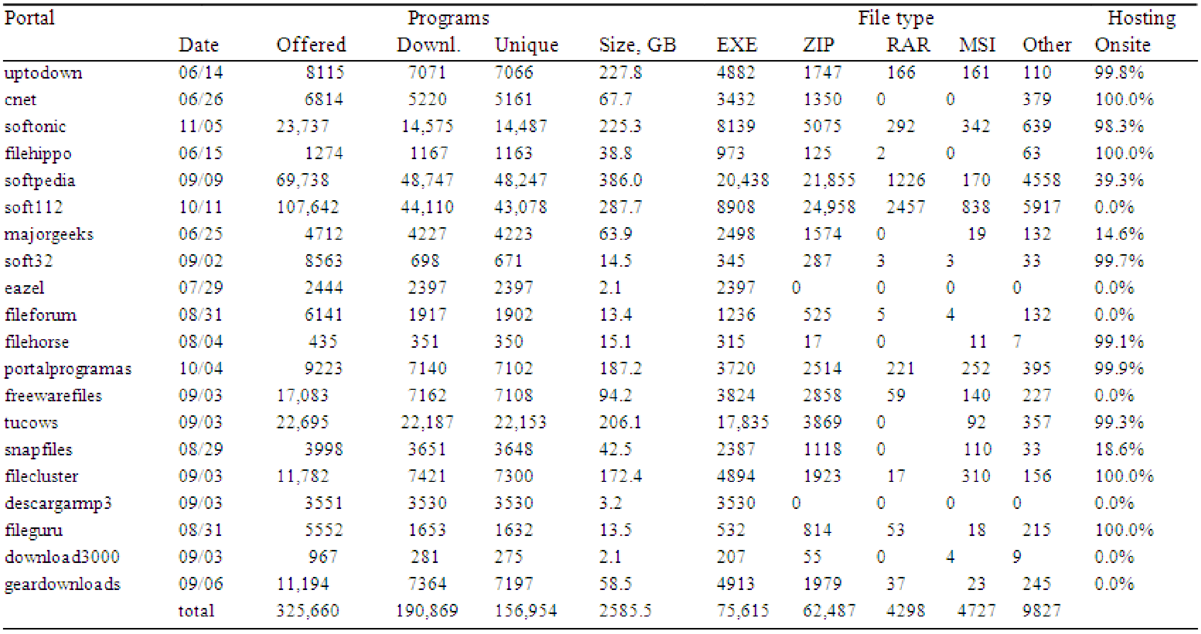
IV. II Undesirable programs in download portals

on this segment, we have a look at the superiority of unwanted packages in download portals. As explained in section three, we submit to VT all downloaded documents large than 30 MB (89% of all downloaded files). in step with the lax policy (i.e. a record is undesirable if at least one AV engine flags it), forty one,664 of the files are unwanted. in step with the conservative coverage (i.e. unwanted if flagged through greater than 3 AV engines), 12,340 documents are undesirable. accordingly, the overall ratio of unwanted packages across all down load portals tiers among eight% (conservative) and 26% (lax).

We follow AVClass at the 12,340 files flagged as unwanted via the conservative policy in order to classify them as pup/malware and to label them with a circle of relatives. of these, 9376 (seventy six%) are puppy and 2955 (24%) are malware. those numbers show that pup is more than three times extra common than malware in download portals.

Table three ranks the down load portals through percentage of unwanted programs. For each download portal, it first shows the ratio for all unwanted programs and is cut up into doggy and malware the use of the conservative coverage. Then, it indicates the overall ratio the usage of the lax coverage. download portals

Table 2: Download portals crawling results



Portal			Programs			File type					Hosting
	Date	Offered	Downl.	Unique	Size, GB	EXE	ZIP	RAR	MSI	Other	Onsite
uptodown	06/14	8115	7071	7066	227.8	4882	1747	166	161	110	99.8%
cnet	06/26	6814	5220	5161	67.7	3432	1350	0	0	379	100.0%
softonic	11/05	23,737	14,575	14,487	225.3	8139	5075	292	342	639	98.3%
filehippo	06/15	1274	1167	1163	38.8	973	125	2	0	63	100.0%
softpedia	09/09	69,738	48,747	48,247	386.0	20,438	21,855	1226	170	4558	39.3%
soft112	10/11	107,642	44,110	43,078	287.7	8908	24,958	2457	838	5917	0.0%
majorgeeks	06/25	4712	4227	4223	63.9	2498	1574	0	19	132	14.6%
soft32	09/02	8563	698	671	14.5	345	287	3	3	33	99.7%
eazel	07/29	2444	2397	2397	2.1	2397	0	0	0	0	0.0%
fileforum	08/31	6141	1917	1902	13.4	1236	525	5	4	132	0.0%
filehorse	08/04	435	351	350	15.1	315	17	0	11	7	99.1%
portalprogramas	10/04	9223	7140	7102	187.2	3720	2514	221	252	395	99.9%
freewarefiles	09/03	17,083	7162	7108	94.2	3824	2858	59	140	227	0.0%
tucows	09/03	22,695	22,187	22,153	206.1	17,835	3869	0	92	357	99.3%
snapfiles	08/29	3998	3651	3648	42.5	2387	1118	0	110	33	18.6%
filecluster	09/03	11,782	7421	7300	172.4	4894	1923	17	310	156	100.0%
descargamp3	09/03	3551	3530	3530	3.2	3530	0	0	0	0	0.0%
fileguru	08/31	5552	1653	1632	13.5	532	814	53	18	215	100.0%
download3000	09/03	967	281	275	2.1	207	55	0	4	9	0.0%
geardownloads	09/06	11,194	7364	7197	58.5	4913	1979	37	23	245	0.0%
	total	325,660	190,869	156,954	2585.5	75,615	62,487	4298	4727	9827	

All executables downloaded from those two down load portals have been unknown to VT while first downloaded, which can be predicted on the grounds that they seem to be generated on the fly because the person (or our crawler) requests them. however, as soon as submitted to VT all of them were flagged as unwanted by way of extra than three AV engines. moreover, if we run the executables from these download portals on a Cuckoo sandbox with no anti-anti-evasion techniques, they showcase a one of a kind behaviour. In this situation, executables downloaded from eazel deploy a software called Remebeher and those from descargarmp3 a software referred to as Liret. No 1/3-celebration gives are proven to the person. this transformation of behaviour suggests anti-sandboxing tests.

Each download portals nation in their phrases and situations that they belong to Vittalia internet, a regarded puppy publisher that used to run a PPI application referred to as OneInstaller [1], which appears defunct. We question the IP addresses of those portals the usage of VT and discover every other 12 Vittalia down load portals hosted on those IP addresses inclusive of solodrivers.com, filewon.com, fileprogram.internet, and downloadsoft.nl. The PPI downloader provided by using the Vittalia download portals isn't always the one from the OneInstaller PPI that Vittalia used to run. as a substitute, the AV engines identify them because the PPI downloader for InstallCore, an Israeli PPI program [1]. similarly, executables downloaded from eazel are signed via ‘FunnelOpti (Alpha standards Ltd.)’ and people downloaded from descargarmp3 are signed by way of ‘delivery Agile (New Media Holdings Ltd.)’. each New Media Holdings Ltd. and Alpha standards Ltd. are companies a part of the IronSource group, who owns the InstallCore PPI program. Finally, we observe that each one documents downloaded from eazel are hosted offsite at www.sendchucklebulk.com and those downloaded from descargarmp3 come from threedomainnames:www.sendcapitalapplication.com www.quickbundlesnew.com,www.guardmegahost.com the ones four domain names all clear up to the equal set of six IP addresses and are registered through the same privateness safety carrier in Israel. We agree with that those domains belong to InstallCore. while a user requests to download a document, the down load portals request InstallCore's API to generate on the fly a user-customised PPI downloader.

To summarise, our research shows that Vittalia has moved faraway from its very own PPI provider and instead has signed up as a writer to the greater popular InstallCore PPI service. while a user attempts to download any program from considered one of Vittalia's download portals, they are alternatively supplied an InstallCore PPI downloader generated at the fly for the user. The user may additionally determine to install a few offers from 0.33-party advertisers who pay InstallCore for distribution, and Vittalia gets a payment for every installation it enables. The person ends up installing no longer best the authentic program that it desired but also the PPI downloader, and any offers it accepts. this example examine illuminates how some down load portals are solely used to assist in the distribution of PPI downloaders and domestic dog merchandise.

Table 3: Percentage of undesirable programs in each download portal

RK	Portal		AV > 3		AV > 0
		All, %	PUP, %	Mal., %	All, %
1	eazel	100	100.0	0.0	100.0
2	descargarmp3	100	100.0	0.0	100.0
3	geardownloads	8.5	5.6	2.9	33.4
4	uptodown	8.3	5.0	3.3	32.1
5	tucows	7.0	4.7	2.3	35.4
6	download3000	5.9	4.4	1.5	30.2
7	filehorse	5.2	4.3	0.9	20.3
8	fileforum	5.1	3.0	2.1	33.4
9	softonic	4.9	1.9	3.0	30.1
10	majorgeeks	4.8	2.6	2.2	28.8
11	filehippo	4.3	3.3	1.0	21.6
12	softpedia	4.1	1.7	2.4	25.1
13	cnet	3.5	1.3	2.2	25.5
14	filecluster	3.3	2.1	1.2	25.0
15	freewarefiles	2.8	1.4	1.4	25.5
16	snapfiles	2.8	1.8	1.0	26.3
17	soft112	2.3	1.1	1.2	13.9
18	soft32	1.6	0.4	1.2	16.8
19	fileguru	1.4	0.5	0.9	15.6
20	portalprogramas	0.2	0.1	0.1	16.5

Table 4: Top ten PUP and malware families

Rank		PUP				Malware	
	Family		Files	Type	Rank	Family	Files
1	installcore		6033	PPI	15	delf	50
2	opencandy		757	PPI	17	autoit	38
3	securityxploded		202	DP	18	zbot	32
4	pswtool		148	Generic	23	joke	30
5	spigot		100	Aff.	29	scar	23
6	prefchanger		95	Generic	32	bumble	19
7	relevantknowledge		89	Marketing	34	crawler	19
8	installmonetizer		87	PPI	36	rbot	17
9	installmonster		77	PPI	37	atraps	16
10	outbrowse		72	PPI	38	ircbot	16

Table 5: Top ten undesirable authors with more than 50 files

Rank	Name	DP	Files	Signed	Pub.	All, %	PUP, %	Mal., %
1	zebnet	6	74	74	2	68	15	53
2	myplaycity	7	100	67	1	49	49	0
3	securityxploded	11	397	1	1	48	39	9
4	freeridegames	1	73	73	1	48	47	0
5	siteken	1	314	0	0	41	31	10
6	xilisoft	10	142	56	2	31	31	0
7	adobe	10	127	48	3	17	17	0
8	nirsoft	16	438	33	2	16	13	3
9	mediafreeware	4	85	9	1	15	12	4
10	microsoft	17	1930	1156	21	4	4	<1

V. Related work

Download portals: security providers have analysed the pinnacle downloads of down load portals and concluded that they're bloated with pup [12–14]. In concurrent and impartial paintings, Geniola et al. [forty three] accumulate 800 installers of promoted applications from eight down load portals. They execute them in a sandbox and discover that 1.three% of these installers drop well-known pup to the system and 10% installation a browser or a browser extension. One essential aim of this work is measuring the amount of abuse in down load portals,

i.e. the proportion of doggy and malware. the main dilemma of earlier works in the direction of that intention is that they examine most effective the top downloaded programs or the promoted programs, which won't be representative of all distributed packages. In contrast, we have downloaded all the windows packages presented by using 20 down load portals. we've got accrued 75,615 unique executables, nearly orders of importance greater than prior works. Our outcomes display an usual ratio of doggy and malware among 8 and 26%, substantially better than the 1.3% pronounced via Geniola et al. Our analysis also identifies down load portals, part of a PPI distribution carrier, which serve a hundred% domestic dog. subsequently, we have diagnosed abusive behaviours domestic dog authors employ to distribute their programs through down load portals.

Domestic dog: Early work on doggy makes a speciality of what constitutes pup [6–eight] and its deceptive methods [forty five–forty seven]. research on doggy has lately revived with some of papers inspecting pup occurrence and its distribution thru industrial PPI offerings. Thomas et al. [48] measured that ad-injectors, a sort of puppy that modifies browser sessions to inject classified ads, affect five% of specific day by day IP addresses accessing Google. Kotzias et al. [38] studied abuse in home windows Authenticode with the aid of analysing 356 okay samples from malware feeds. They found that puppy has quick elevated in so-known as malware feeds when you consider that 2010, that the extensive majority of nicely signed samples are domestic dog, and that doggy publishers use excessive record and certificate polymorphism to steer clear of protection tools and CA defenses together with identity validation and revocation. In a separate work, Kotzias et al. [1] used AV telemetry of 3.9 M actual hosts for analysing doggy incidence and its distribution via business PPI offerings. They observed pup hooked up in 54% of the hosts and diagnosed 24 industrial PPI services that distribute over 1 / 4 of all of the domestic dog of their 2013–2014 dataset. they also determined that business PPI offerings used to distribute puppy are disjoint from underground PPI offerings used to distribute malware [forty nine]. In simultaneous and impartial paintings, Thomas et al. [2] analysed the advertiser software distributed to US hosts via 4 commercial PPI services. They used SafeBrowsing data to degree that PPI services power over 60 million down load events each week inside the 2d 1/2 of 2015, nearly three instances that of malware. Nelms et al. [50] analysed web-primarily based classified ads that use social engineering to deceive users to down load pup. They found that maximum packages dispensed this manner are bundles of loose software with pup. This paintings differs from the above in that it analyses pup prevalence in download portals.

Sandboxing: Many works have proposed sandboxing platforms for malware analysis [fifty one–fifty four]. the ones might also use in-visitor tracing of home windows API calls [fifty four], emulation [51], hardware-supported virtualisation [fifty two], and naked machines [fifty three]. on this paintings, we use the open-supply Cuckoo sandbox [39].

VI. Conclusion

On this illustration, we have achieved a systematic observe of abuse in download portals, which index freeware from a couple of authors. we have built a platform to crawl download portals and feature carried out it to down load 191 ok windows freeware installers from 20 download portals. we've got analysed the amassed installers and performed them in a sandbox. We measure an common ratio of domestic dog and malware between 8% (conservative) and 26% (lax). In 18 of the 20 down load portals the amount of domestic dog and malware is slight, i.e. beneath 9%. however, we additionally find download portals solely used to distribute PPI downloaders. we've carried out a radical evaluation of those down load portals. Finally, we've distinct unique abusive behaviours that authors of unwanted packages use to distribute their packages via down load portals which include uploading the same file as unique applications, using outside links to pass protection checks, and impersonating benign famous authors.

References

[1]. Kotzias, P., Bilge, L., Caballero, J.: ‘Measuring PUP prevalence and PUP distribution through pay-per-install services’. USENIX Security Symp., Austin, TX, USA, August 2016

[2]. Thomas, K., Crespo, J.A.E., Rastil, R., *et al.*: ‘Investigating commercial pay- per-install and the distribution of unwanted software’. USENIX Security Symp., Austin, TX, USA, August 2016

[3]. CNET: http://download.cnet.com, accessed November 2017

[4]. Softonic: https://www.softonic.com/, accessed November 2017

[5]. Tucows: http://tucows.com, accessed November 2017

[6]. Bruce, J.: ‘Defining rules for acceptable adware’. Virus Bulletin Conf.,

[7]. Dublin, Ireland, 2005

[8]. McFedries, P.: ‘Technically speaking: the spyware nightmare’, *IEEE Spectr.*, 2005, **42**, (8), pp. 72–72

[9]. Pickard, C., Miladinov, S.: ‘Rogue software: protection against potentially unwanted applications’. Int. Conf. Malicious and Unwanted Software,

[10]. Fajardo, PR, USA, 2012

[11]. Google: ‘Unwanted software policy, 2018’. Available at https:// www.google.com/about/unwanted-software-policy.html

[12]. Microsoft: ‘How Microsoft antimalware products identify malware and unwanted software’, 2018. Available at https://www.microsoft.com/en-us/ wdsi/antimalware-support/malware-and-unwanted-software-evaluation- criteria

[13]. MalwareBytes: ‘PUP reconsideration information – how do we identify potentially unwanted software?’, 2018. Available at https:// www.malwarebytes.com/pup/

[14]. EmsiSoft: ‘Mind the PUP: top download portals to avoid’. Available at http:// blog.emsisoft.com/2015/03/11/mind-the-pup-top-download-portals-to-avoid/, accessed December 2016

[15]. Heddings, L.: ‘Here is what happens when you install the top 10 download.com apps’. Available at http://www.howtogeek.com/198622/heres- what-happens-when-youinstall-the-top-10-download.com-apps, accessed November 2017

[16]. Heddings, L.: ‘Yes, every freeware download site is serving Crapware’. Available at https://www.howtogeek.com/207692/yes-

every-freeware- download-site-is-serving-crapware-heres-the-proof/, accessed November 2017
[17]. ASP, The Association of Software Professionals: http://padsites.org, accessed November 2017
[18]. Virustotal: https://virustotal.com/, accessed October 2017
[19]. O. Java: ‘What are the ask toolbars?’. Available at https://www.java.com/en/ download/faq/ask_toolbar.xml
[20]. Alexa Website Ranking: http://www.alexa.com/, accessed November 2017
[21]. Uptodown: https://www.uptodown.com, accessed November 2017
[22]. FileHippo: http://filehippo.com, accessed November 2017
[23]. Softpedia: http://www.softpedia.com/, accessed November 2017
[24]. Soft112: http://soft112.com, accessed November 2017
[25]. MajorGeeks: http://majorgeeks.com, accessed November 2017
[26]. Soft32: http://soft32.com, accessed November 2017
[27]. Eazel: http://eazel.com, accessed November 2017
[28]. FileForum: http://fileforum.betanews.com, accessed November 2017
[29]. FileHorse: http://filehorse.com, accessed November 2017
[30]. PortalProgramas: http://portalprogramas.com, accessed November 2017
[31]. FreewareFiles: http://freewarefiles.com, accessed November 2017
[32]. SnapFiles: http://snapfiles.com, accessed November 2017