Prevasio

prevasio.io/blog/sunburst-backdoor-part-ii-dga-the-list-of-victims

By

Sergei Shevchenko

December 17, 2020

Sunburst Backdoor, Part II: DGA & The List of Victims

Previous Part of the analysis is available here.
Next Part of the analysis is available here.

Update from 19 December 2020:

Prevasio would like to thank Zetalytics for providing us with an updated (larger) list of passive (historic) DNS queries for the domains generated by the malware.

As described in the first part of our analysis, the DGA (Domain Generation Algorithm) of the Sunburst backdoor produces a domain name that may look like:

fivu4vjamve5vfrtn2huov[.]appsync-api.us-west-2[.]avsvmcloud[.]com

The first part of the domain name (before the first dot) consists of a 16-character random string, appended with an encoded computer's domain name. This is the domain in which the local computer is registered.

From the example string above, we can conclude that the encoded computer's domain starts from the 17th character and up until the dot (highlighted in yellow):

`fivu4vjamve5vfrtn2huov`

In order to encode a local computer's domain name, the malware uses one of 2 simple methods:

- *Method 1*: a substitution table, if the domain name consists of small letters, digits, or special characters '-', '_', '.'
- *Method 2*: base64 with a custom alphabet, in case of capital letters present in the domain name

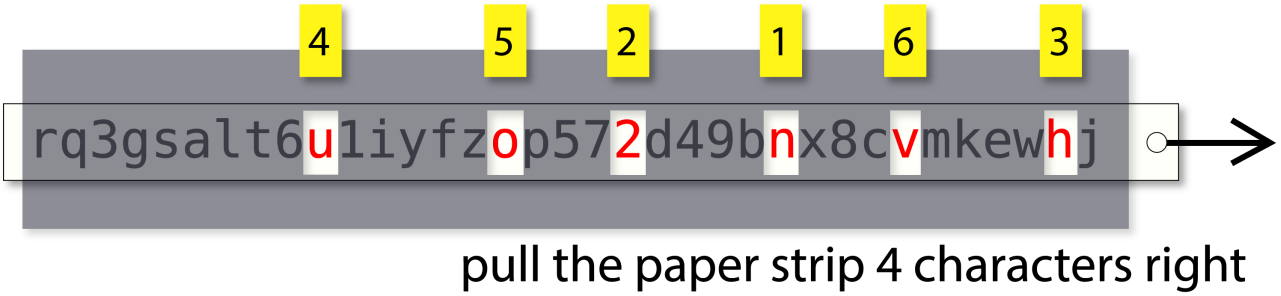
Method 1

In our example, the encoded domain name is **"n2huov"**. As it does not have any capital letters, the malware encodes it with a substitution table **"rq3gsalt6u1iyfzop572d49bnx8cvmkewhj"**.

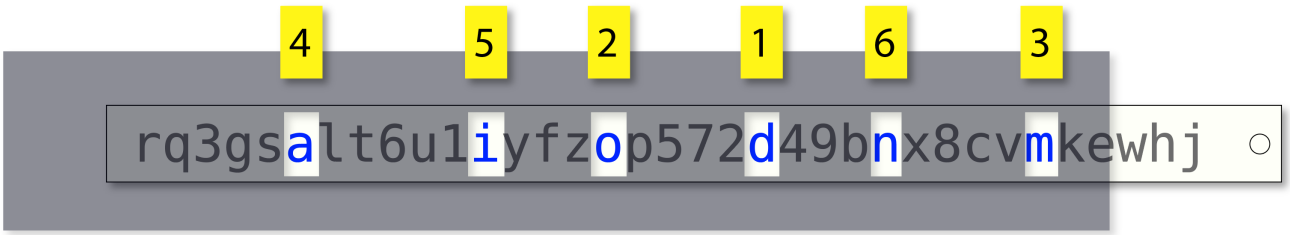
For each character in the domain name, the encoder replaces it with a character located in the substitution table four characters right from the original character.

In order to decode the name back, all we have to do is to replace each encoded character with another character, located in the substitution table four characters left from the original character.

To illustrate this method, imagine that the original substitution table is printed on a paper strip and then covered with a card with 6 perforated windows. Above each window, there is a sticker note with a number on it, to reflect the order of characters in the word **"n2huov"**, where **'n'** is #1, **'2'** is #2, **'h'** is #3 and so on:



Once the paper strip is pulled by 4 characters right, the perforated windows will reveal a different word underneath the card: **"domain"**, where **'d'** is #1, **'o'** is #2, **'m'** is #3, etc.:



A special case is reserved for such characters as **'0'**, **'-'**, **'_'**, **'.'**. These characters are encoded with **'0'**, followed with a character from the substitution table. An index of that character in the substitution table, divided by 4, provides an index within the string **"0_-."**.

The following snippet in C# illustrates how an encoded string can be decoded:

```
static string decode_domain(string s)
{
    string table = "rq3gsalt6u1iyfzop572d49bnx8cvmkewhj";
    string result = "";
    for (int i = 0; i < s.Length; i++)
    {
        if (s[i] != '0')
        {
            result += table[(table.IndexOf(s[i]) + table.Length - 4) % table.Length];
        }
        else
        {
            if (i < s.Length - 1)
            {
                if (table.Contains(s[i + 1]))
                {
                    result += "0_-."[table.IndexOf(s[i + 1]) % 4];
                }
                else
                {
                    break;
                }
            }
            i++;
        }
    }
    return result;
}
```

Method 2

This method is a standard base64 encoder with a custom alphabet **"ph2eifo3n5utg1j8d94qrvbmk0sal76c"**.

Here is a snippet in C# that provides a decoder:

```
public static string FromBase32String(string str)
{
    string table = "ph2eifo3n5utg1j8d94qrvbmk0sal76c";
    int numBytes = str.Length * 5 / 8;
    byte[] bytes = new Byte[numBytes];

    int bit_buffer;
    int currentCharIndex;
    int bits_in_buffer;
    if (str.Length < 3)
    {
        bytes[0] = (byte)(table.IndexOf(str[0]) | table.IndexOf(str[1]) << 5);
        return  System.Text.Encoding.UTF8.GetString(bytes);
    }

    bit_buffer = (table.IndexOf(str[0]) | table.IndexOf(str[1]) << 5);
    bits_in_buffer = 10;
    currentCharIndex = 2;
    for (int i = 0; i < bytes.Length; i++)
    {
        bytes[i] = (byte)bit_buffer;
        bit_buffer >>= 8;
        bits_in_buffer -= 8;
        while (bits_in_buffer < 8 && currentCharIndex < str.Length)
        {
            bit_buffer |= table.IndexOf(str[currentCharIndex++]) << bits_in_buffer;
            bits_in_buffer += 5;
        }
    }
    return System.Text.Encoding.UTF8.GetString(bytes);
}
```

When the malware encodes a domain using Method 2, it prepends the encrypted string with a double zero character: **"00"**.

Following that, extracting a domain part of an encoded domain name (long form) is as simple as:

```
static string get_domain_part(string s)
{
    int i = s.IndexOf(".appsync-api");
    if (i > 0)
    {
        s = s.Substring(0, i);
        if (s.Length > 16)
        {
            return s.Substring(16);
        }
    }
    return "";
}
```

Once the domain part is extracted, the decoded domain name can be obtained by using Method 1 or Method 2, as explained above:

```
if (domain.StartsWith("00"))
{
    decoded = FromBase32String(domain.Substring(2));
}
else
{
    decoded = decode_domain(domain);
}
```

Decrypting the Victims' Domain Names

To see the decoder in action, let's select 2 lists:

List #1

Bambenek Consulting has provided a list of observed hostnames for the DGA domain.

List #2

The second list has surfaced in a Paste bin paste, allegedly sourced from Zetalytics / Zonecruncher.

NOTE: This list is fairly 'noisy', as it has non-decodable domain names.

By feeding both lists to our decoder, we can now obtain a list of decoded domains, that could have been generated by the victims of the Sunburst backdoor.

DISCLAIMER: It is not clear if the provided lists contain valid domain names that indeed belong to the victims. It is quite possible that the encoded domain names were produced by third-party tools, sandboxes, or by researchers that investigated and analysed the backdoor.

The decoded domain names are provided purely as a reverse engineering exercise. The resulting list was manually processed to eliminate noise, and to exclude duplicate entries. Following that, we have made an attempt to map the obtained domain names to the company names, using Google search. Reader's discretion is advised as such mappings could be inaccurate.

Decoded Domain	Mapping (Could Be Inaccurate)
hgvc.com	Hilton Grand Vacations
Amerisaf	AMERISAFE, Inc.
kcpl.com	Kansas City Power and Light Company

SFBALLET	San Francisco Ballet
scif.com	State Compensation Insurance Fund
LOGOSTEC	Logostec Ventilação Industrial
ARYZTA.C	ARYZTA Food Solutions
bmrn.com	BioMarin Pharmaceutical Inc.
AHCCCS.S	Arizona Health Care Cost Containment System
nnge.org	Next Generation Global Education
cree.com	Cree, Inc (semiconductor products)
calsb.org	The State Bar of California
rbe.sk.ca	Regina Public Schools
cisco.com	Cisco Systems
pcsco.com	Professional Computer Systems
barrie.ca	City of Barrie
ripta.com	Rhode Island Public Transit Authority
uncity.dk	UN City (Building in Denmark)
bisco.int	Boambee Industrial Supplies (Bisco)
haifa.edu	University of Haifa
smsnet.pl	SMSNET, Poland
fcmat.org	Fiscal Crisis and Management Assistance Team
wiley.com	Wiley (publishing)
ciena.com	Ciena (networking systems)
belkin.com	Belkin
spsd.sk.ca	Saskatoon Public Schools
pqcorp.com	PQ Corporation
ftfcu.corp	First Tech Federal Credit Union
bop.com.pk	The Bank of Punjab

nvidia.com	NVidia
insead.org	INSEAD (non-profit, private university)
usd373.org	Newton Public Schools
agloan.ads	American AgCredit
pageaz.gov	City of Page
jarvis.lab	Erich Jarvis Lab
ch2news.tv	Channel 2 (Israeli TV channel)
bgeltd.com	Bradford / Hammacher Remote Support Software
dsh.ca.gov	California Department of State Hospitals
dotcomm.org	Douglas Omaha Technology Commission
sc.pima.gov	Arizona Superior Court in Pima County
itps.uk.net	IT Professional Services, UK
moncton.loc	City of Moncton
acmedctr.ad	Alameda Health System
csci-va.com	Computer Systems Center Incorporated
keyano.local	Keyano College
uis.kent.edu	Kent State University
alm.brand.dk	Sydbank Group (Banking, Denmark)
ironform.com	Ironform (metal fabrication)
corp.ncr.com	NCR Corporation
ap.serco.com	Serco Asia Pacific
int.sap.corp	SAP
mmhs-fla.org	Cleveland Clinic Martin Health
nswhealth.net	NSW Health
mixonhill.com	Mixon Hill (intelligent transportation systems)
bcofsa.com.ar	Banco de Formosa

ci.dublin.ca.	Dublin, City in California
siskiyous.edu	College of the Siskiyous
weioffice.com	Walton Family Foundation
ecobank.group	Ecobank Group (Africa)
corp.sana.com	Sana Biotechnology
med.ds.osd.mi	US Gov Information System
wz.hasbro.com	Hasbro (Toy company)
its.iastate.ed	Iowa State University
amr.corp.intel	Intel
cds.capilanou.	Capilano University
e-idsolutions.	IDSolutions (video conferencing)
helixwater.org	Helix Water District
detmir-group.r	Detsky Mir (Russian children's retailer)
int.lukoil-int	LUKOIL (Oil and gas company, Russia)
ad.azarthritis	Arizona Arthritis and Rheumatology Associates
net.vestfor.dk	Vestforbrænding
allegronet.co.	Allegronet (Cloud based services, Israel)
us.deloitte.co	Deloitte
central.pima.g	Pima County Government
city.kingston.	City of Kingston
staff.technion	Technion - Israel Institute of Technology
airquality.org	Sacramento Metropolitan Air Quality Management District
phabahamas.org	Public Hospitals Authority, Caribbean
parametrix.com	Parametrix (Engineering)
ad.checkpoint.	Check Point
corp.riotinto.	Rio Tinto (Mining company, Australia)

intra.rakuten.	Rakuten
us.rwbaird.com	Robert W. Baird & Co. (Financial services)
ville.terrebonn	Ville de Terrebonne
woodruff-sawyer	Woodruff-Sawyer & Co., Inc.
fisherbartoninc	Fisher Barton Group
banccentral.com	BancCentral Financial Services Corp.
taylorfarms.com	Taylor Fresh Foods
neophotonics.co	NeoPhotonics (optoelectronic devices)
gloucesterva.ne	Gloucester County
magnoliaisd.loc	Magnolia Independent School District
zippertubing.co	Zippertubing (Manufacturing)
milledgeville.l	Milledgeville (City in Georgia)
digitalreachinc	Digital Reach, Inc.
deniz.denizbank	DenizBank
thoughtspot.int	ThoughtSpot (Business intelligence)
lufkintexas.net	Lufkin (City in Texas)
digitalsense.co	Digital Sense (Cloud Services)
wrbaustralia.ad	W. R. Berkley Insurance Australia
christieclinic.	Christie Clinic Telehealth
signaturebank.l	Signature Bank
dufferincounty.	Dufferin County
mountsinai.hosp	Mount Sinai Hospital
securview.local	Securview Victory (Video Interface technology)
weber-kunststof	Weber Kunststoftechniek
parentpay.local	ParentPay (Cashless Payments)
europapier.inte	Europapier International AG

molsoncoors.com	Molson Coors Beverage Company
fujitsugeneral.	Fujitsu General
cityofsacramento	City of Sacramento
ninewellshospita	Ninewells Hospital
fortsmithlibrary	Fort Smith Public Library
dokkenengineerin	Dokken Engineering
vantagedatacente	Vantage Data Centers
friendshipstateb	Friendship State Bank
clinicasierravis	Clinica Sierra Vista
ftsillapachecasi	Apache Casino Hotel
voceracommunicat	Vocera (clinical communications)
mutualofomahaban	Mutual of Omaha Bank