Three Variants of Murofet's DGA

bin.re/blog/three-variants-of-murofets-dga/

Murofet, also called LICAT, is a member of the ZeuS family [1], [2], [3]. It uses a Domain Generation Algorithm (DGA) to determine the current C2 domain names. There exist at least three different versions of Murofet's DGA, some of which I couldn't find reimplementations online. In this short blog post I list the three variants that I looked at and discuss the properties of each. Although all versions share a similar algorithm, the resulting domains are very different.

The DGAs

All DGAs share the following procedure:

1. Pick a start index based by multiplying the current minute by 17

2. Concatenate the following values:
    1. year + 0x30 (one byte)
    2. month (one byte)
    3. day rounded to the frequency *freq* (one byte)
    4. (variant 2 only) value 0 (zero or one byte)
    5. the index (four bytes); variant 2 rounds the index to multiples of 2.
3. XOR the resulting 7 to 8 bytes with an optional four byte key. Only variant 2 uses a key.
4. MD5 hash the 7 to 8 bytes.
5. Iterate over the 16 bytes of the md5 hash and determine characters of the second level domain based on the bytes.
6. Append a top level domain based on the current index.
7. Increment the index, taking the result modulo 60*17 = 1020. Repeat from step 2 up to 800 times.

The following pseudo-code summarizes the DGA:

```
SET startindex = minute*17
SET tlds to array of top level domains

FOR i in 0 TO 799
    index = (startindex + i) % (60*17)
    hash = md5( XOR_key(
                    (year & 0xFF) + 0x30 .
                    month .
                    floor(date.day/freq)*freq .
                    optionally_round(index)
                    )
              )
    domain = ""
    FOREACH byte in hash DO
        append zero or more characters to *domain* based on *byte*
    ENDFOR


    FOR index = 0 TO LEN(tlds) - 1
        tld = tlds[j]
        m = len(tlds) - index
        IF i % m = 0 THEN
            append *tld* to *domain*
            break
    ENDFOR
    PRINT domain
ENDFOR
```

The main difference between the DGA variants is how they generate second level characters based on the md5 bytes.

Variant 1

Variant 1 doesn't use a key. Domains change every week (*freq* = 7). It generates the second level characters with the following loop:

```
FOREACH byte in hash DO
    d = (byte & 0x1F) + 'a'
    c = (byte >> 3) + 'a'
    IF d != c THEN
        IF d <= 'z' THEN
            domain += d
        IF c <= 'z' THEN
            domain += c
ENDFOR
```

The top level domains are *.ru*, *.biz*, *.info*, *.org*, *.net*, and *.com*. A complete implementation in Python can be found here.

Variant 2

Variant 2 uses a key. Keys observed in the wild are for example 0xD6d7A4BE and 0xDEADC2DE. Variant 2 has a granularity of 1 day and generates the second level characters with the following loop:

```
FOREACH byte in hash DO
    tmp = (byte & 0xF) + (byte >> 4) + 'a'
    IF tmp <= 'z' THEN
        domain += tmp
ENDFOR
```

The top level domains are *.biz*, *.info*, *.org*, *.net*, and *.com*. A complete implementation in Python can be found here.

Variant 3

Variant 3 does not use a key. Domains change every 7 days. Second level characters are generated as follows:

```
FOREACH byte in hash DO
    a = (byte & 0xF) + 'a'
    b = (byte >> 4) + 'q'
    IF b > 'z' THEN
        b = b - 'J'
        c = (a % 10) + '0'
    ELSE
        c = None

    domain += a
    domain += b
    IF c THEN
        domain += c
ENDFOR
```

The top level domains are *.ru*, *.biz*, *.info*, *.org*, *.net*, and *.com*. A complete implementation in Python can be found here.

Properties

The properties of the domains of the three variants are as follows:

Property	Variant 1	Variant 2	Variant 3
Granularity	1 week	1 day	1 week
Keyed	no	yes	no
TLDs	.ru, .biz, .info, .org, .net, .com	.biz, .info, .org, .net, .com	.ru, .biz, .info, .org, .net, .com
SLD characters	a-z	a-z	a-z, 0-9
SLD length range	0 - 32	0 - 16	32 - 48
SLD expected length	25.2	15.1	38
SLD properties	uniformly distributed	*p* the most frequent, dropping of on both sides (*a* being very uncommon). Domains appear in pairs.	digits appear in pairs only, and only from 10 to 69

Examples

Variant 1

The second level characters are more or less equally distributed. The domains can't be easily attributed to Murofet simply looking at them:

```
giywswshrgxcvoqgvrkthmfa.ru
xaiqpbprgymbvrwmzgiyprgdsk.com
amgqgularpzxeapztxenbx.net
pfscijbmthyfiyjgergugtkbqyh.org
xglfcmsgorvwfilhmzlcxxvkfege.info
rcteqwkequojntibvfyfaluwh.biz
mjfqylbiaunffuaeunzdqdwscu.ru
qobeylpxgpfknlptukyddqvklztg.com
rgwgizukficdgetwsxovtcknwkfm.info
betgyaeswxorwcvsdezdupbmb.org
```

Variant 2

The second level characters are noticeably more frequent around *p*. The lengths are 14 - 16, seldom much shorter. Also domains come in pairs:

```
vxzyptrgrsndon.biz
vxzyptrgrsndon.com
gpkfnvjuvvoyqpv.net
gpkfnvjuvvoyqpv.org
imjfpqgmjtdlojn.info
imjfpqgmjtdlojn.biz
repkmqpzfilmwzf.org
repkmqpzfilmwzf.com
xugoslqquofodj.info
xugoslqquofodj.org
```

The DGA not only changes domain sets every day (compared to weekly for variant 1 and 3), but also supports virtually unlimited distinct sets of domain because of the key.

Variant 3

In contrast to the first two variants, the second level characters also contain digits. Digits always appear as tuples, and are always between 10 and 69 in value. The length of the sld is at least 32, and can go up to 48:

```
gui45gzgzmtbtp22c59fxatf62k47bvmzhwfq.ru
o21aua67pqpym49g33mqf62b28m39guixbtd50cv.com
awgyg13nva57b48p52g23j26eqkxjzl28e51nqn20.net
pscriubwkzh34fui45j56gsb48g13g53dvbyit.org
htlscwctoyf22fupvhwjvc59h64l28fvfsgs.info
brpzeym69gveyexjwoziqf42f32fqove21lt.biz
esmum69fyivbua17nsf12ase61n30dynrgzc19.ru
axbsivpxhtpskwmvnxpzevirdyfvl38dt.com
btgti35evfuc59mrgsd20c69h44o11drkwgvfw.info
bsd40g33asc29o21oyczgrfzdsjre41pqmqcz.org
```