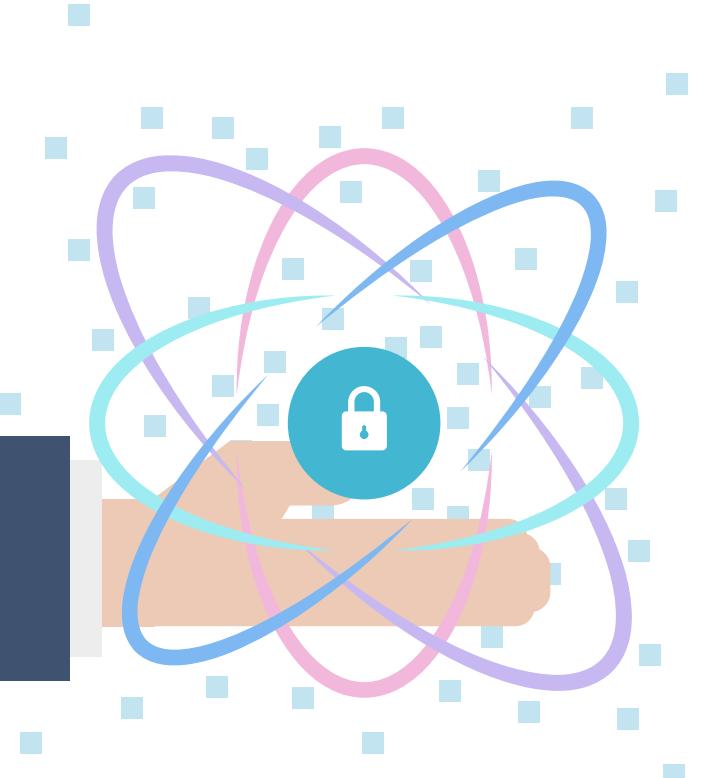


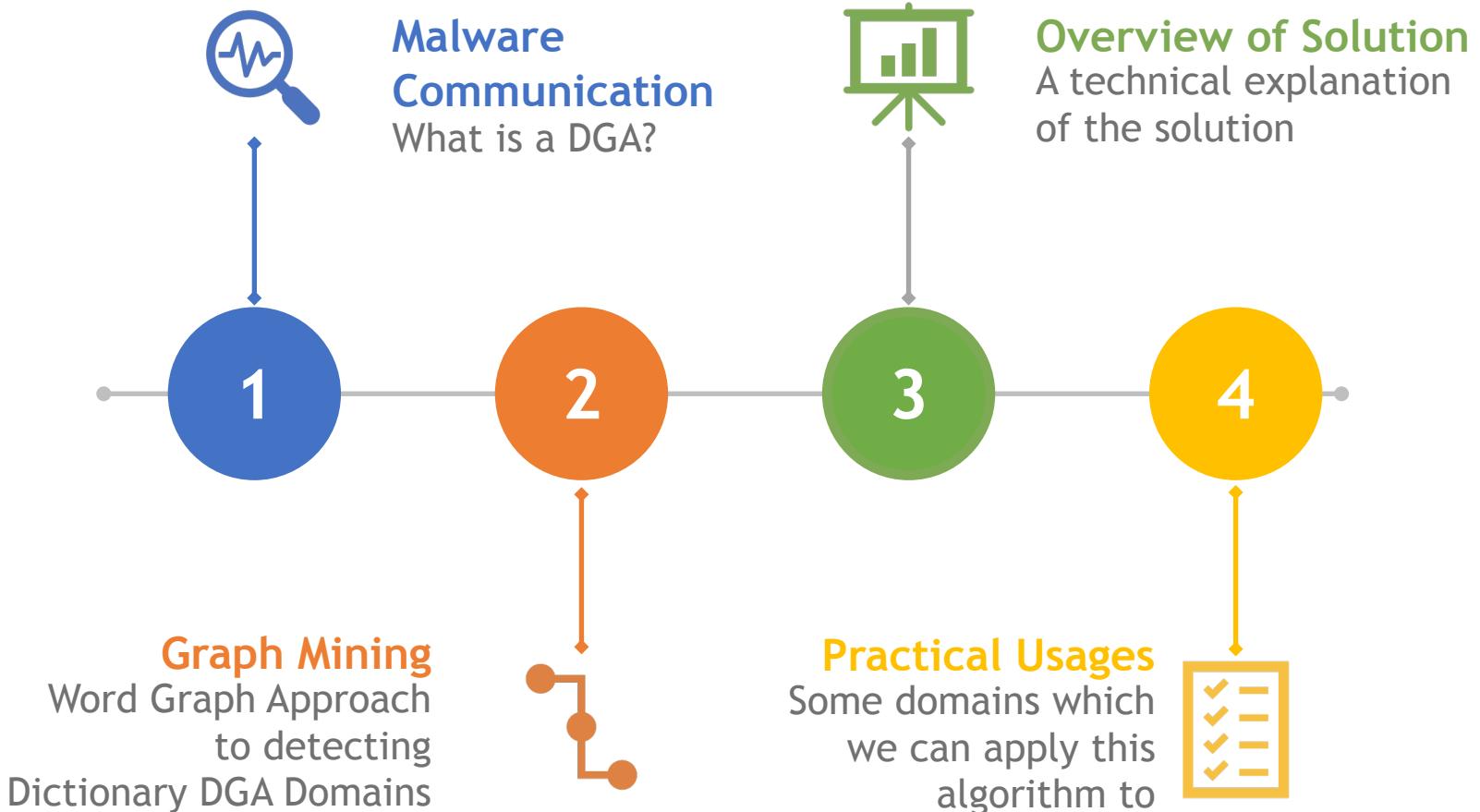
# Graph Based Machine Learning



## Detecting Wordlist Domain Generation Algorithms

Sng Yu Feng Chester @chestersng  
Lam Fhong Sheng @chrispooh007

# Roadmap



# Attacker's POV



# Hardcoded IP Address



Hey 11.22.33.44, what do  
you want me do?



Hey 11.22.33.44, here is  
the payload... {username:...,  
password:...}



11.22.33.44

# Author writes DGA

```
● ● ●  
def dga(date, magic, number):  
    seed = date.year + date.month + date.day + magic  
    r = Rand(seed)  
    ...  
    if i == 0x33:  
        r = Rand(magic)  
    v1 = r.rand()  
    ra = []  
    for i in range(10):  
        ra.append(v1 % 10)  
        v1 //= 10  
  
    domain = ""  
    for x in ra:  
        domain += LETTERS[x]  
  
    if ra[0] < len(TLDS):  
        tld = TLDS[ra[0]]  
    else:  
        tld = DEFAULTTLD  
  
    domain += tld  
    yield domain
```



# Author writes DGA

```
● ● ●  
def dga(date, magic, number):  
    seed = date.year + date.month + date.day + magic  
    r = Rand(seed)  
  
    for l in range(1, number):  
        if l == 0x33:  
            r = Rand(magic)  
        v1 = r.rand()  
        ra = []  
        for l in range(10):  
            ra.append(v1 % 10)  
            v1 /= 10  
  
        domain = ""  
        for x in ra:  
            domain += LETTERS[x]  
  
        if ra[0] < len(TLDs):  
            tld = TLDs[ra[0]]  
        else:  
            tld = DEFAULTTLD  
  
        domain += tld  
        yield domain
```



```
● ● ●  
def dga(date, magic, number):  
    seed = date.year + date.month + date.day + magic  
    r = Rand(seed)  
  
    for l in range(1, number):  
        if l == 0x33:  
            r = Rand(magic)  
        v1 = r.rand()  
        ra = []  
        for l in range(10):  
            ra.append(v1 % 10)  
            v1 /= 10  
  
        domain = ""  
        for x in ra:  
            domain += LETTERS[x]  
  
        if ra[0] < len(TLDs):  
            tld = TLDs[ra[0]]  
        else:  
            tld = DEFAULTTLD  
  
        domain += tld  
        yield domain
```

```
● ● ●  
def dga(date, magic, number):  
    seed = date.year + date.month + date.day + magic  
    r = Rand(seed)  
  
    for l in range(1, number):  
        if l == 0x33:  
            r = Rand(magic)  
        v1 = r.rand()  
        ra = []  
        for l in range(10):  
            ra.append(v1 % 10)  
            v1 /= 10  
  
        domain = ""  
        for x in ra:  
            domain += LETTERS[x]  
  
        if ra[0] < len(TLDs):  
            tld = TLDs[ra[0]]  
        else:  
            tld = DEFAULTTLD  
  
        domain += tld  
        yield domain
```

# Step 1



Run DGA in advance with time seed of tomorrow 7p.m.

Produces:

a9djc92c.com

d02kd0x.net

19cjsodos.org

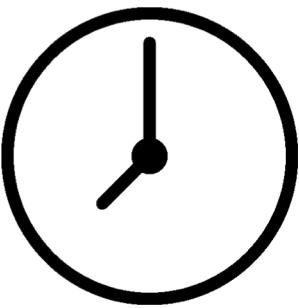
...

Registers one of them:

19cjsodos.org



# Step 2



Run DGA with time seed now

Produces:

a9djc92c.com

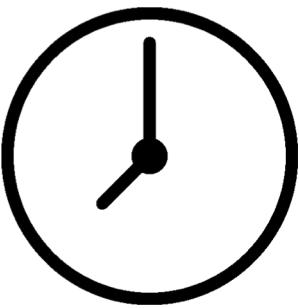
d02kd0x.net

19cjsodos.org

...



# Step 3



Attempts to communicate with all of them:

a9djc92c.com

d02kd0x.net

19cjsodos.org

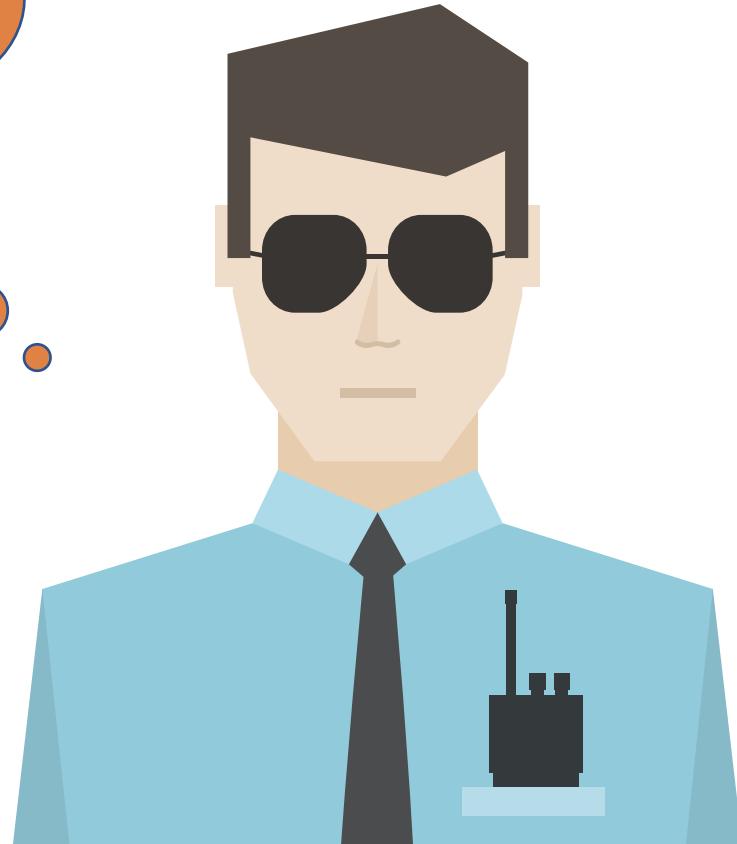
...

Successfully Established Connection with:

19cjsodos.org



# Defender's POV



# Which is the DGA Domain?

The screenshot shows the homepage of the Sounds Wonderful website. The header features the text "Sounds Wonderful" and a small logo with the words "sounds wonderful". Below the header is a navigation menu with links to HOME, ABOUT, SERVICES, EVENTS, NEWS, STORE, and CONTACT. A search icon is also present. The main content area has a dark red background with a photograph of hands playing a double bass and a guitar. Overlaid on the photo is the text "Finding Your True Voice and Expression Connection – Self-Empowerment – Presence". Below this, a smaller text block reads: "Sounds Wonderful supports people from all walks of life to find their true voice and express the power that lives inside us all." At the bottom of the page, there is a list of services: "Corporate seminars – Schools – Conferences – Voice Coaching Concerts & House Concerts – Workshops & Retreats – Choirs". A testimonial quote from Ella Rubell is displayed at the very bottom.

Finding Your True Voice and Expression  
Connection – Self-Empowerment – Presence

Sounds Wonderful supports people from all walks of life to find their true voice and express the power that lives inside us all.

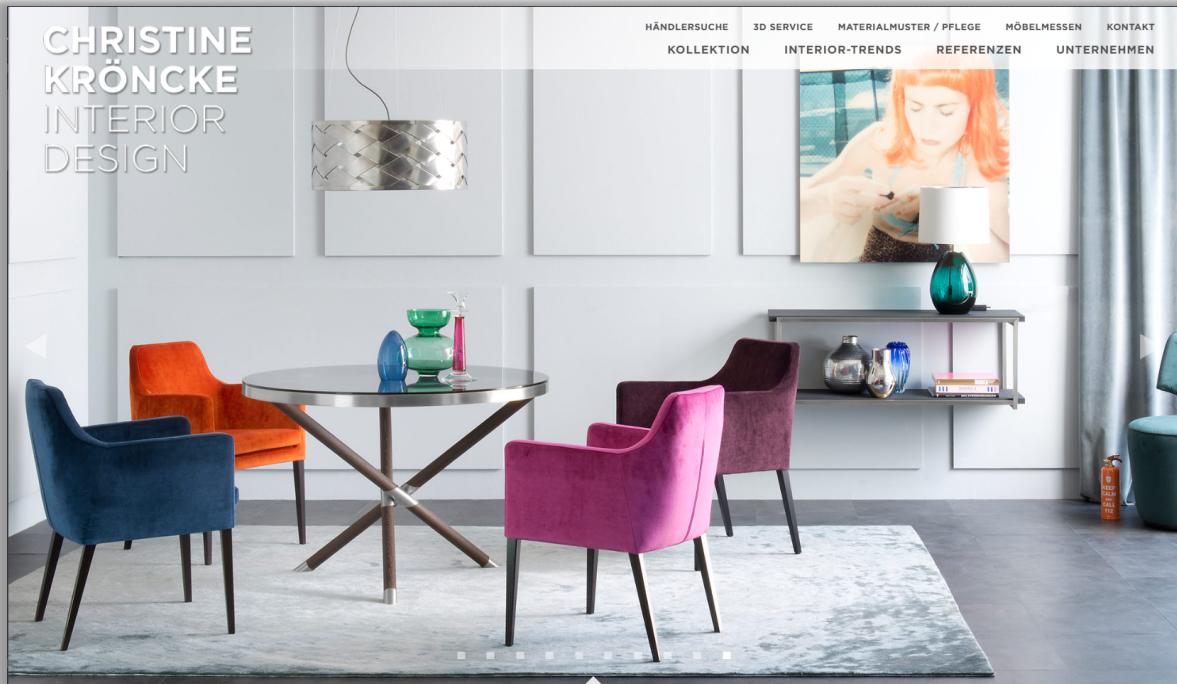
Corporate seminars – Schools – Conferences – Voice Coaching  
Concerts & House Concerts – Workshops & Retreats – Choirs

*'Chris James has an amazing way of helping people to discover their own vocal strength and potential. I was very impressed with his ability – in working with an SBS collaborator – to transform their voice-over delivery to something that carried great colour, power and energy.'*  
*Ella Rubell, Producer Online Documentaries, SBS TV & Radio*

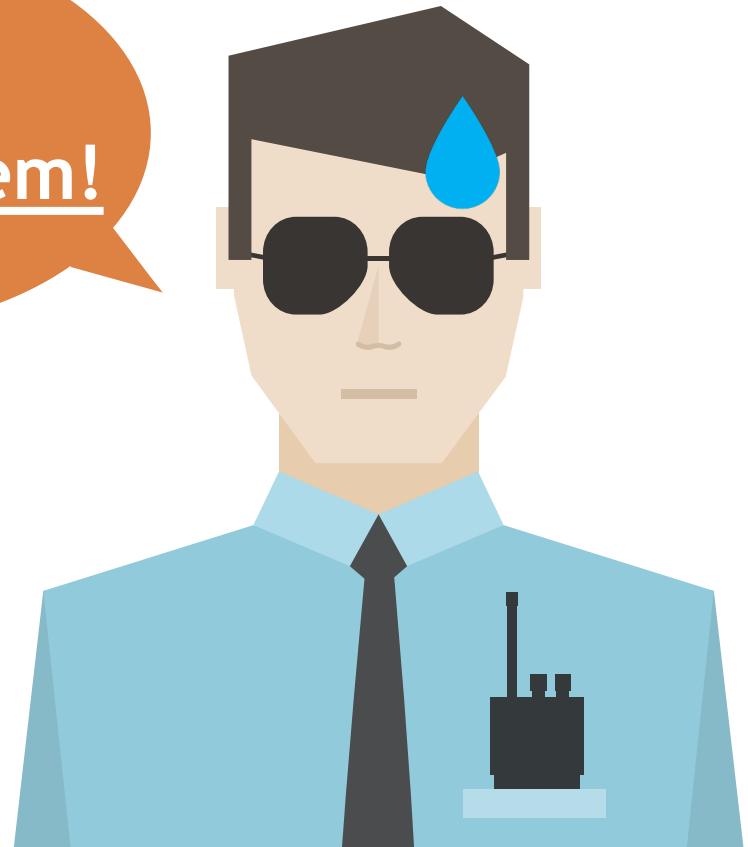
That's easy.  
It is the  
second one!



# Which is the DGA Domain?



This is a  
hard problem!



# Not just one type of DGA...

Types	Example Family	Examples Domains
Arithmetic (A)	DirCrypt	<a href="http://wejcqzbosbczzlnikyvt.com">wejcqzbosbczzlnikyvt.com</a> , <a href="http://muiccxbvkvjb.com">muiccxbvkvjb.com</a> , <a href="http://tqwpmppckhidiss.com">tqwpmppckhidiss.com</a> , <a href="http://gzredieexn.com">gzredieexn.com</a> , <a href="http://ghhcwlldtj.com">ghhcwlldtj.com</a>
Wordlist (W)	Matsnu	<a href="http://accident-require.com">accident-require.com</a> <a href="http://paintfinance.com">paintfinance.com</a> <a href="http://lawyersit-direction.com">lawyersit-direction.com</a> <a href="http://troublepace-summer.com">troublepace-summer.com</a>
Hashing (H)	Bamital	<a href="http://cd8f66549913a78c5a8004c82bcf6b01.info">cd8f66549913a78c5a8004c82bcf6b01.info</a> <a href="http://aa24603b0defd57ebfef34befde16370.cz.cc">aa24603b0defd57ebfef34befde16370.cz.cc</a> <a href="http://5e6efdd674c134ddb2a7a2e3c603cc14.org">5e6efdd674c134ddb2a7a2e3c603cc14.org</a>
Permutation	“Explosive” malware from “VolatileCedar”	<a href="http://explorerdotnt.info">explorerdotnt.info</a> <a href="http://dotnetexplorer.info">dotnetexplorer.info</a> <a href="http://dotntexplorere.info">dotntexplorere.info</a> <a href="http://xploreredotnet.info">xploreredotnet.info</a>

# 2018 Research on DDGA Detection



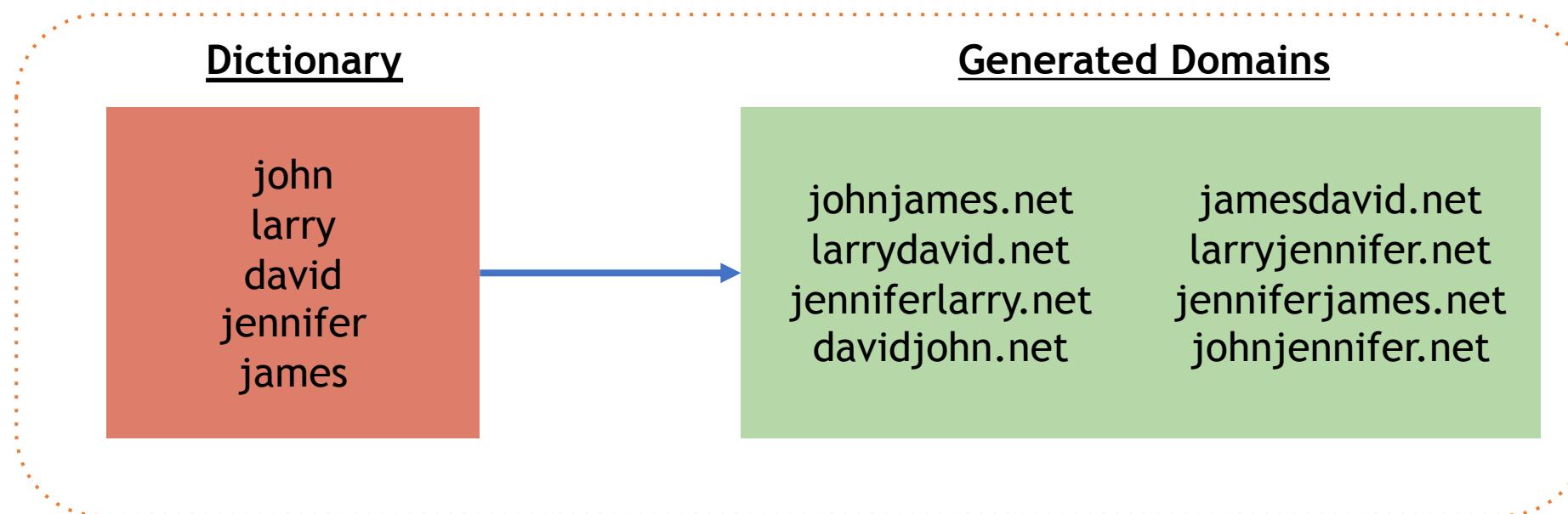
Mayana Pereira

Data Scientist  
Infoblox Inc.

# Observation 1

Malware generate their domains using a **fixed-size dictionary**. Normally, these dictionaries are **small** (hundreds of words) to reduce payload size.

Word will be **reused**, since the size of the dictionary is small.

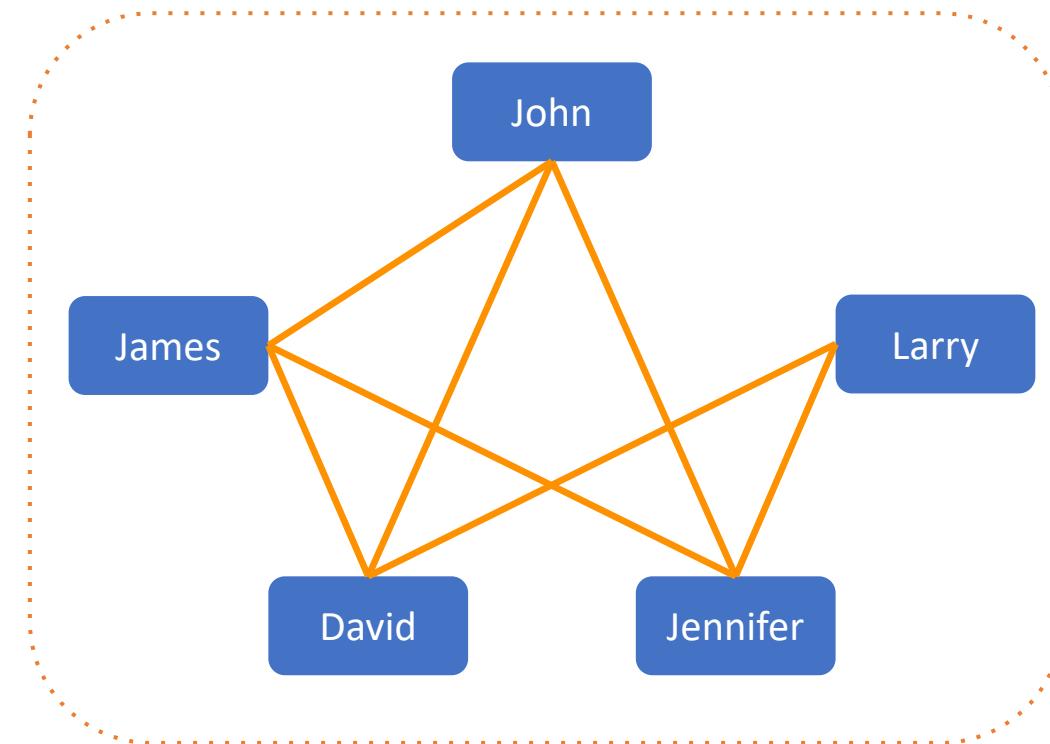


## Observation 2

We can make use of **graph theory** to form a solution for this problem.

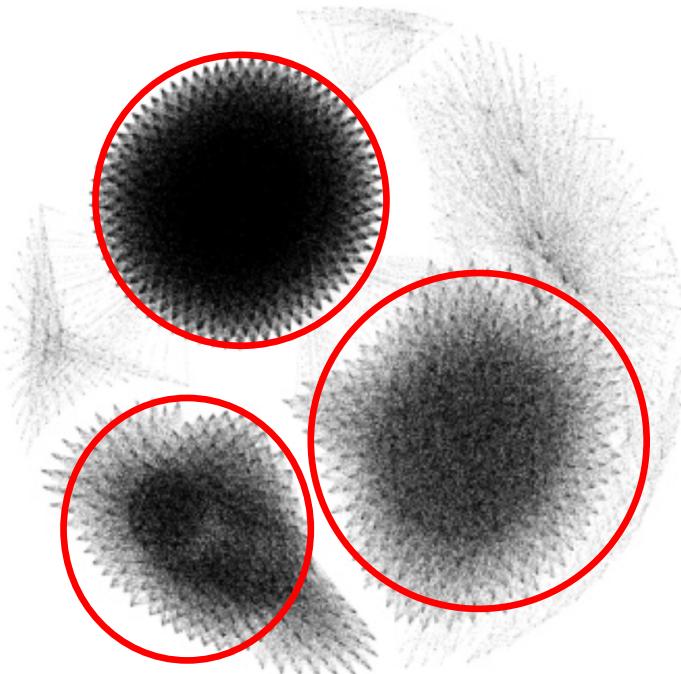
1. Let the words in the dictionary be **nodes**
2. Two nodes are linked if they appear in the same domain.

jamesdavid.net  
larryjennifer.net  
jenniferjames.net  
johnjennifer.net

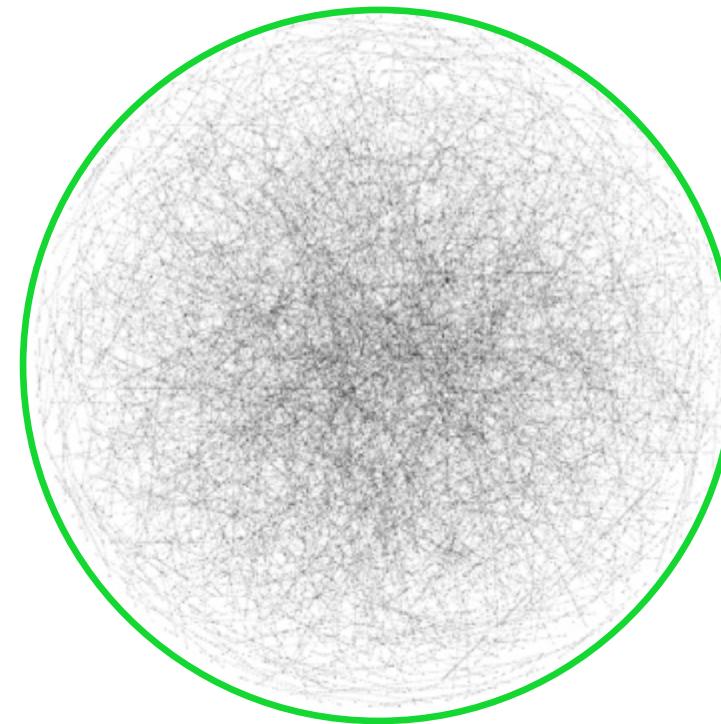


## Observation 3

Dictionary-generated domains cluster **differently** from benign domains.



Dictionary Generated Clusters



Benign Domains Clusters

# Components of the Solution

## Word Breaking Algorithm

doghouse.net

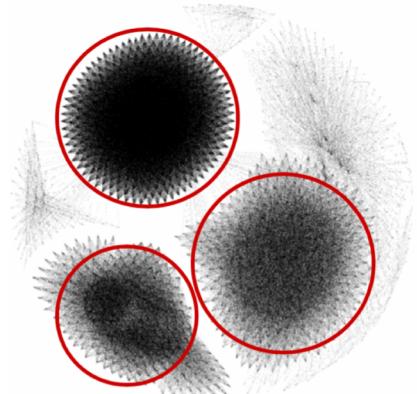


dog house

Break domains up into individual words

Implemented a new algorithm to increase efficiency

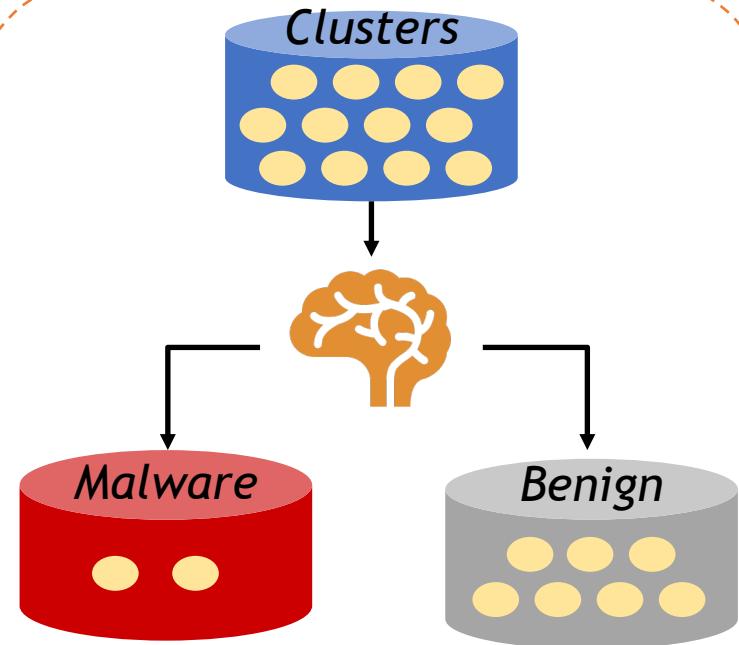
## Graph Library



Create a word graph and extract clusters

Used an additional graph library to provide clear visualisation

## Artificial Intelligence



Malware

Benign

Used a larger dataset to detect more families of malware

# Repeated words!

christianasheenagh.net

catharinesheenagh.net

christianasherilyn.net

clevelandsheridan.net

mastermansherilyn.net

sebastiansherilyn.net

grenvillesheenagh.net

clevelandshakesheave.net

sebastiansheenagh.net

christophershavonne.net

grenvilleshakesheave.net

sebastiansheridan.net

clevelandsheenagh.net

grenvillesherilyn.net

mastermansheridan.net

grenvillesheenagh.net

christophershakesheave.net

sebastianshakesheave.net

christianasheridan.net

catharinesherilyn.net



# Wordbreaking Algorithm

tomjack.net

jilljane.net

jackjill.net

tomjane.net

## Word

tom

jack

## Domains

tomjack.net

tomjack.net

## Word

## Domains



# Wordbreaking Algorithm

tomjack.net

jilljane.net

jackjill.net

tomjane.net

<u>Word</u>	<u>Domains</u>
tom	tomjack.net
jack	tomjack.net
tomj	tomjack.net
ack	tomjack.net

<u>Word</u>	<u>Domains</u>



# Wordbreaking Algorithm

tomjack.net

jilljane.net

jackjill.net

tomjane.net

<u>Word</u>	<u>Domains</u>
tom	tomjack.net
jack	tomjack.net
tomj	tomjack.net
ack	tomjack.net
jil	jilljane.net
ljane	jilljane.net

<u>Word</u>	<u>Domains</u>



# Wordbreaking Algorithm

tomjack.net

jilljane.net

jackjill.net

tomjane.net

<u>Word</u>	<u>Domains</u>
tom	tomjack.net
jack	tomjack.net
tomj	tomjack.net
ack	tomjack.net
jil	jilljane.net
ljane	jilljane.net
jill	jilljane.net
jane	jilljane.net

<u>Word</u>	<u>Domains</u>



# Wordbreaking Algorithm

tomjack.net

jilljane.net

jackjill.net

tomjane.net

<u>Word</u>	<u>Domains</u>
tom	tomjack.net
jack	tomjack.net
tomj	tomjack.net
ack	tomjack.net
jil	jilljane.net
ljane	jilljane.net
jill	jilljane.net
jane	jilljane.net
jillj	jilljane.net
ane	jilljane.net

<u>Word</u>	<u>Domains</u>



# Wordbreaking Algorithm

tomjack.net

jilljane.net

jack**kjill**.net

tomjane.net

<u>Word</u>	<u>Domains</u>
tom	tomjack.net
jack	tomjack.net
tomj	tomjack.net
ack	tomjack.net
jil	jilljane.net
ljane	jilljane.net
jill	jilljane.net
jane	jilljane.net
jillj	jilljane.net
ane	jilljane.net
jac	jackjill.net

<u>Word</u>	<u>Domains</u>
kjill	jackjill.net



# Wordbreaking Algorithm

tomjack.net

jilljane.net

jackjill.net

tomjane.net

<u>Word</u>	<u>Domains</u>
tom	tomjack.net
jack	tomjack.net, jackjill.net
tomj	tomjack.net
ack	tomjack.net
jil	jilljane.net
ljane	jilljane.net
jill	jilljane.net, jackjill.net
jane	jilljane.net
jillj	jilljane.net
ane	jilljane.net
jac	jackjill.net

<u>Word</u>	<u>Domains</u>
kjill	jackjill.net



# Wordbreaking Algorithm

tomjack.net

jilljane.net

jackjill.net

tomjane.net

<u>Word</u>	<u>Domains</u>
tom	tomjack.net
jack	tomjack.net, jackjill.net
tomj	tomjack.net
ack	tomjack.net
jil	jilljane.net
ljane	jilljane.net
jill	jilljane.net, jackjill.net
jane	jilljane.net
jillj	jilljane.net
ane	jilljane.net
jac	jackjill.net

<u>Word</u>	<u>Domains</u>
kjill	jackjill.net
jackj	jackjill.net
ill	jackjill.net



# Wordbreaking Algorithm

tomjack.net

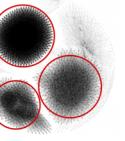
jilljane.net

jackjill.net

tomjane.net

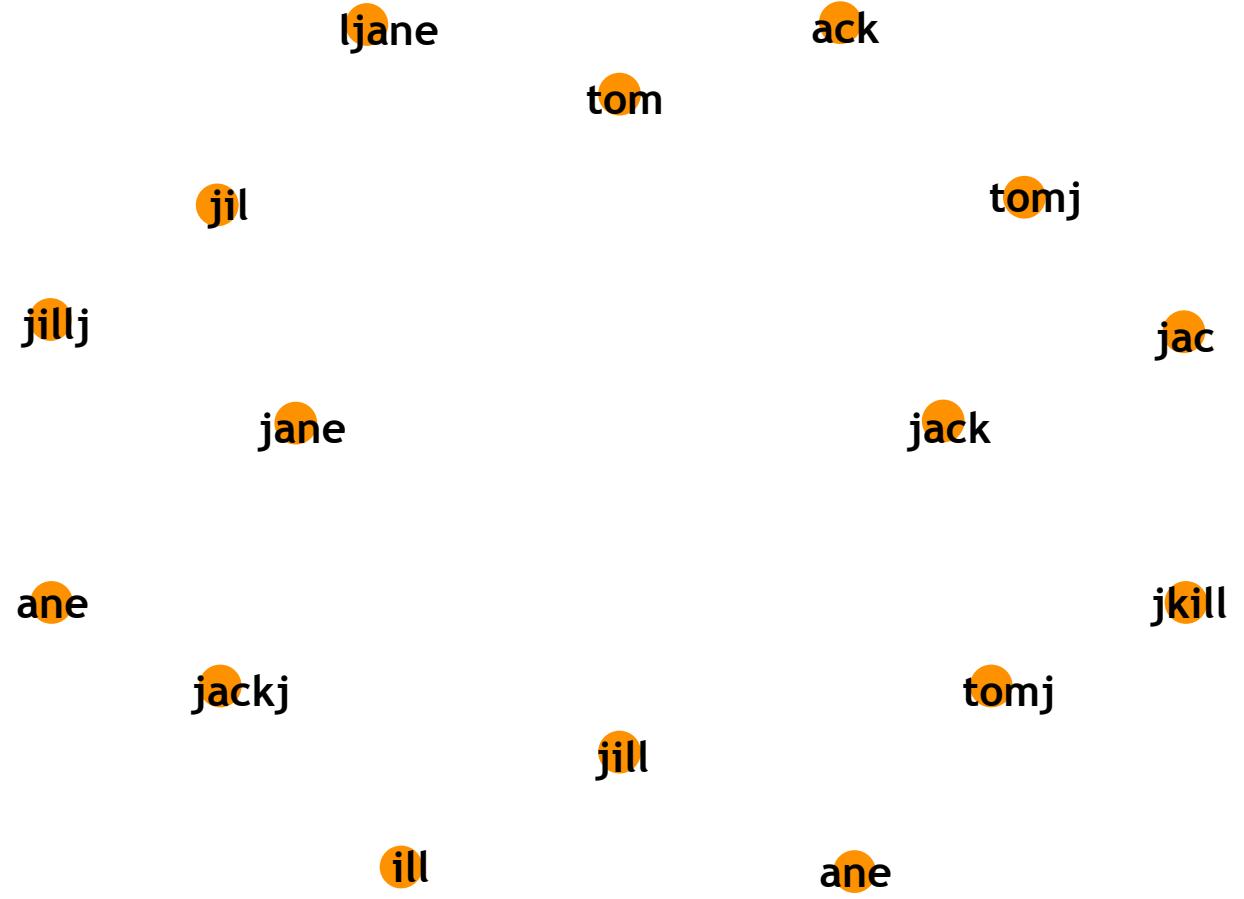
<u>Word</u>	<u>Domains</u>
tom	tomjack.net, tomjane.net
jack	tomjack.net, jackjill.net
tomj	tomjack.net
ack	tomjack.net
jil	jilljane.net
ljane	jilljane.net
jill	jilljane.net, jackjill.net
jane	jilljane.net, tomjane.net
jillj	jilljane.net
ane	jilljane.net
jac	jackjill.net

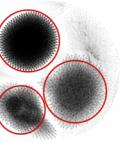
<u>Word</u>	<u>Domains</u>
kjill	jackjill.net
jackj	jackjill.net
ill	jackjill.net
tomj	tomjane.net
ane	tomjane.net



# Graphing Algorithm

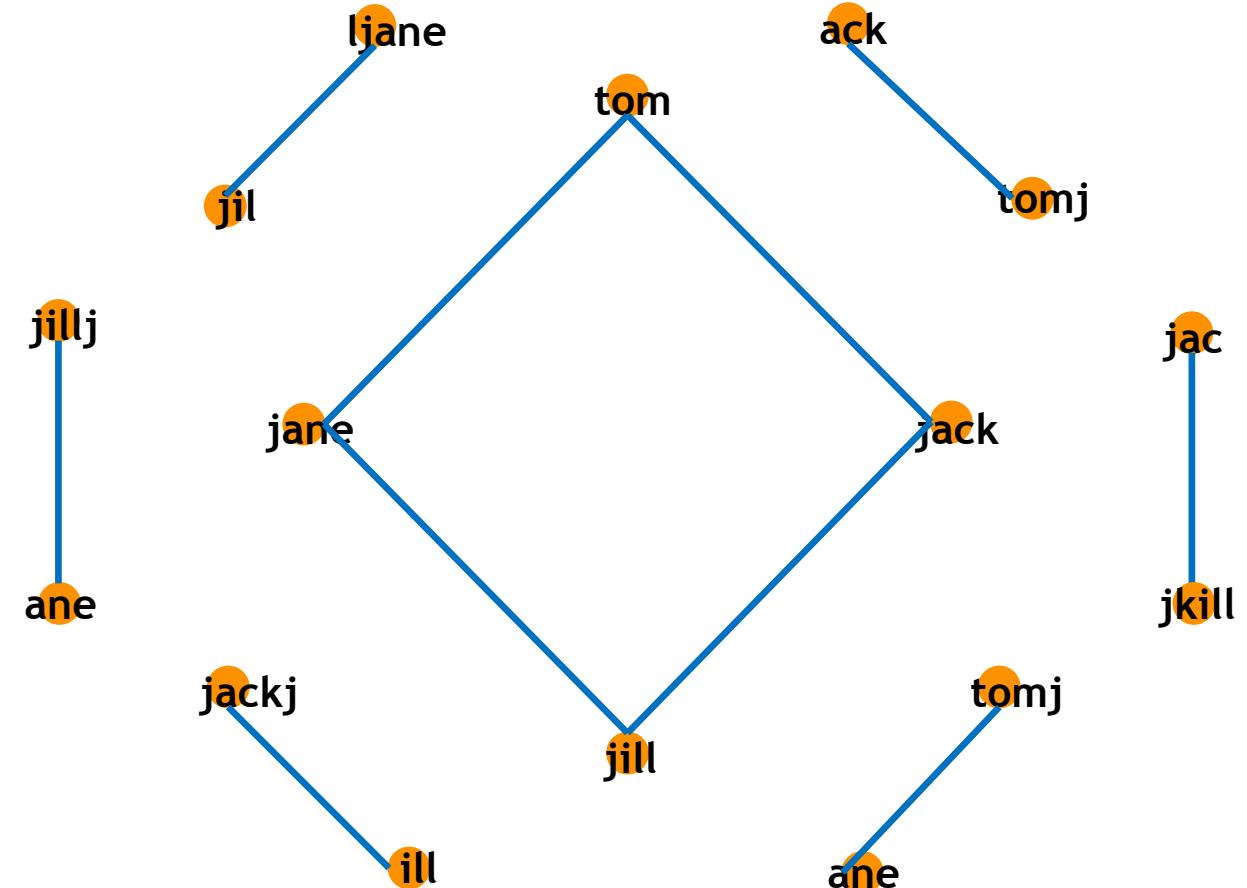
<u>Word</u>	<u>Domains</u>
tom	tomjack.net, tomjane.net
jack	tomjack.net, jackjill.net
tomj	tomjack.net
ack	tomjack.net
jil	jilljane.net
ljane	jilljane.net
jill	jilljane.net, jackjill.net
jane	jilljane.net, tomjane.net
jillj	jilljane.net
ane	jilljane.net
jac	jackjill.net
kjill	jackjill.net
jackj	jackjill.net
ill	jackjill.net
tomj	tomjane.net
ane	tomjane.net

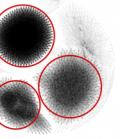




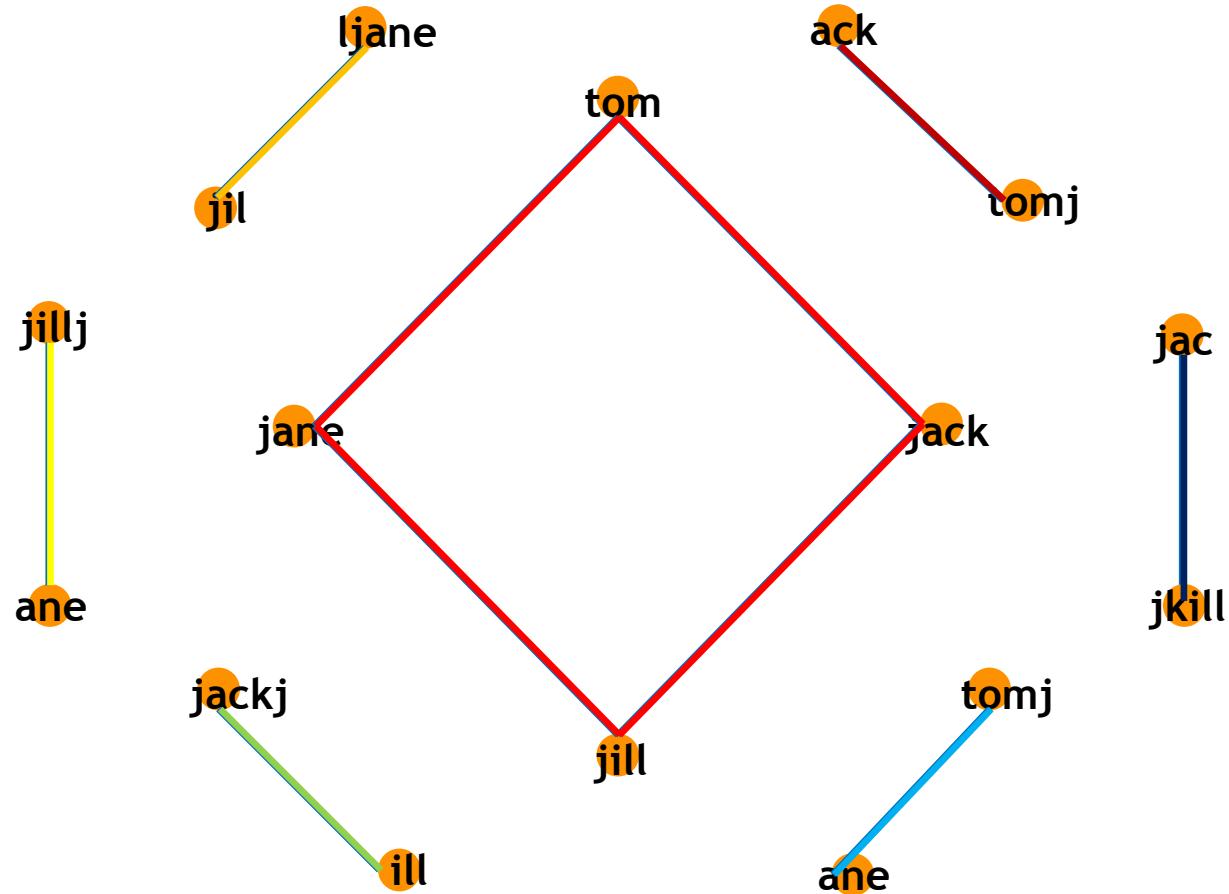
# Graphing Algorithm

Word	Domains
tom	tomjack.net, tomjane.net
jack	tomjack.net, jackjill.net
tomj	tomjack.net
ack	tomjack.net
jil	jilljane.net
ljane	jilljane.net
jill	jilljane.net, jackjill.net
jane	jilljane.net, tomjane.net
jillj	jilljane.net
ane	jilljane.net
jac	jackjill.net
kjill	jackjill.net
jackj	jackjill.net
ill	jackjill.net
tomj	tomjane.net
ane	tomjane.net



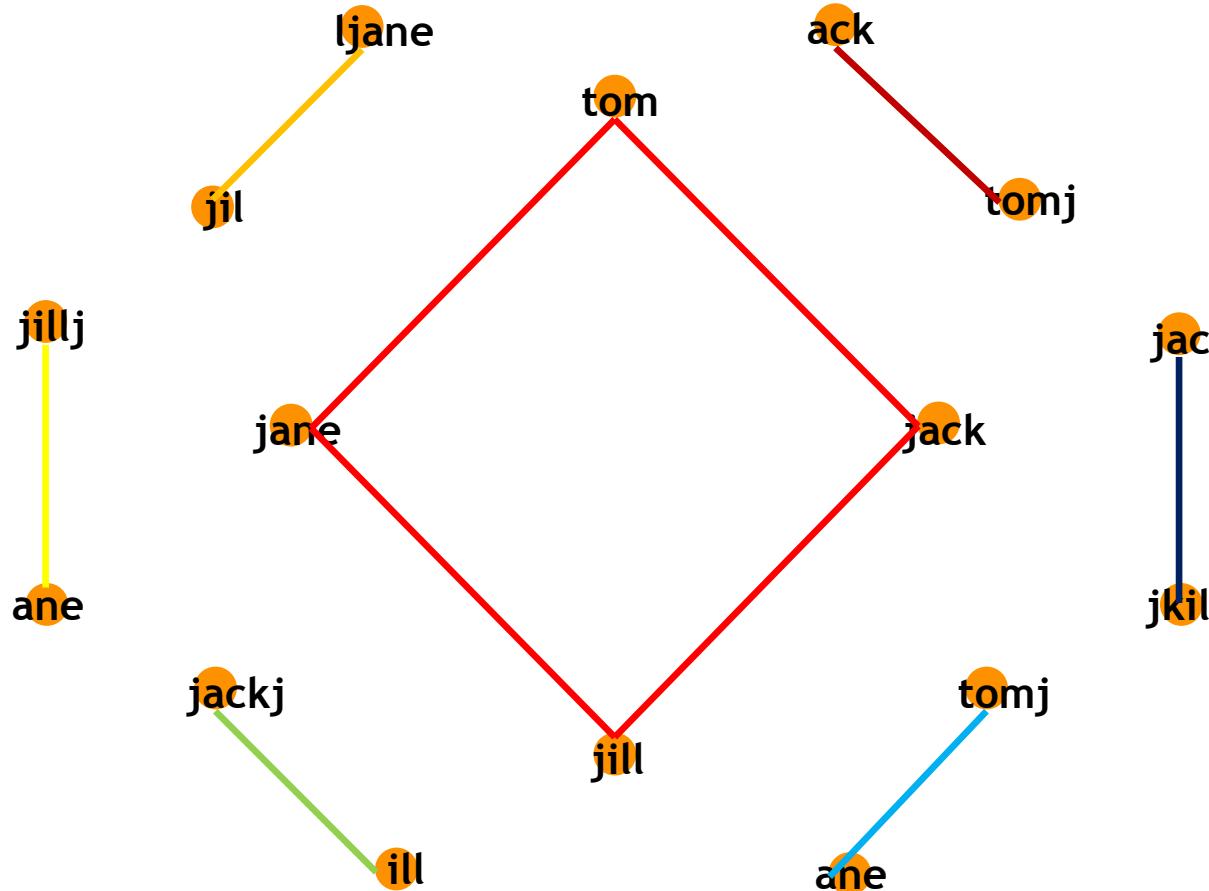


# Graphing Algorithm





# Artificial Intelligence



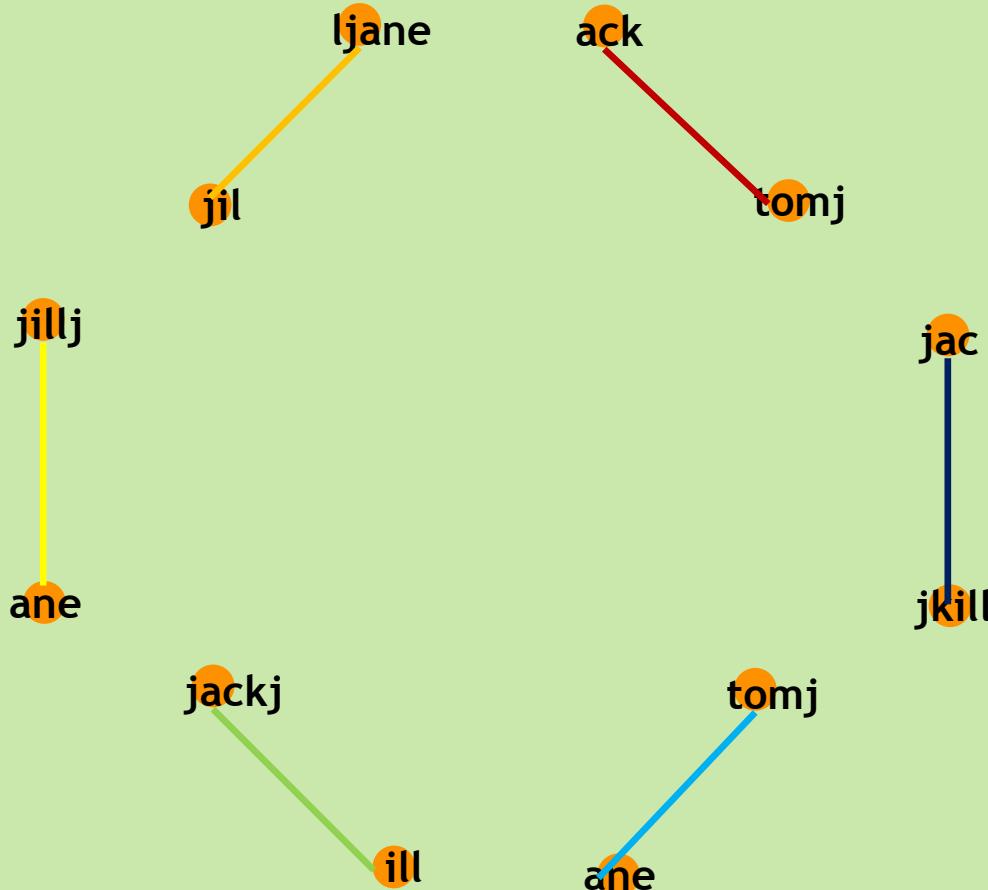
Feature Vector
Average Node Degree
Maximum Node Degree
Minimum Node Degree
Cardinality of Cycle Basis
Average Cycles Per Node
Average Clustering Per Node

Features that suggest clustering

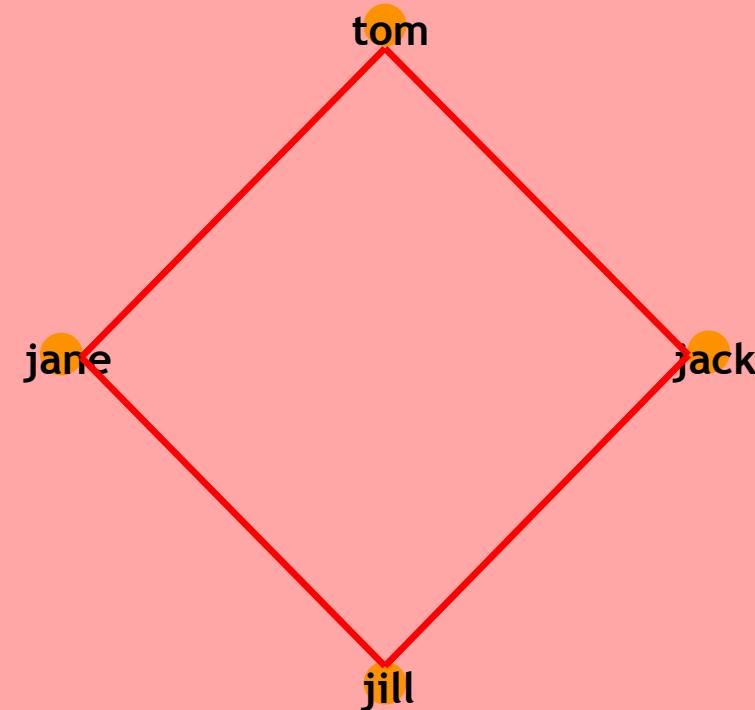


# Classification

Benign Dictionary



DGA Dictionary



# Running on REAL DATA

## How to mitigate BKDR\_BAYROB using TrendMicro Products

🕒 Updated: 24 Nov 2016   Product/Version: Deep Discovery 3.0   Platform: N/A N/A

### SUMMARY

The BAYROB/NIVDORT malware arrives via spam email as an attachment and goes through two stages of infection. The first stage uses Domain Generating Algorithm (DGA) to download its updated copy that contains backdoor capability. The second stage of infection, downloads a bitcoin-miner on an infected machine.

BAYROB was first seen in 2007 stealing only eBay accounts until it evolved and appeared again in 2015 with its backdoor and Anti-AV capabilities. A resurgence in 2016 caused high volume of infections due to its mass mailing capabilities.

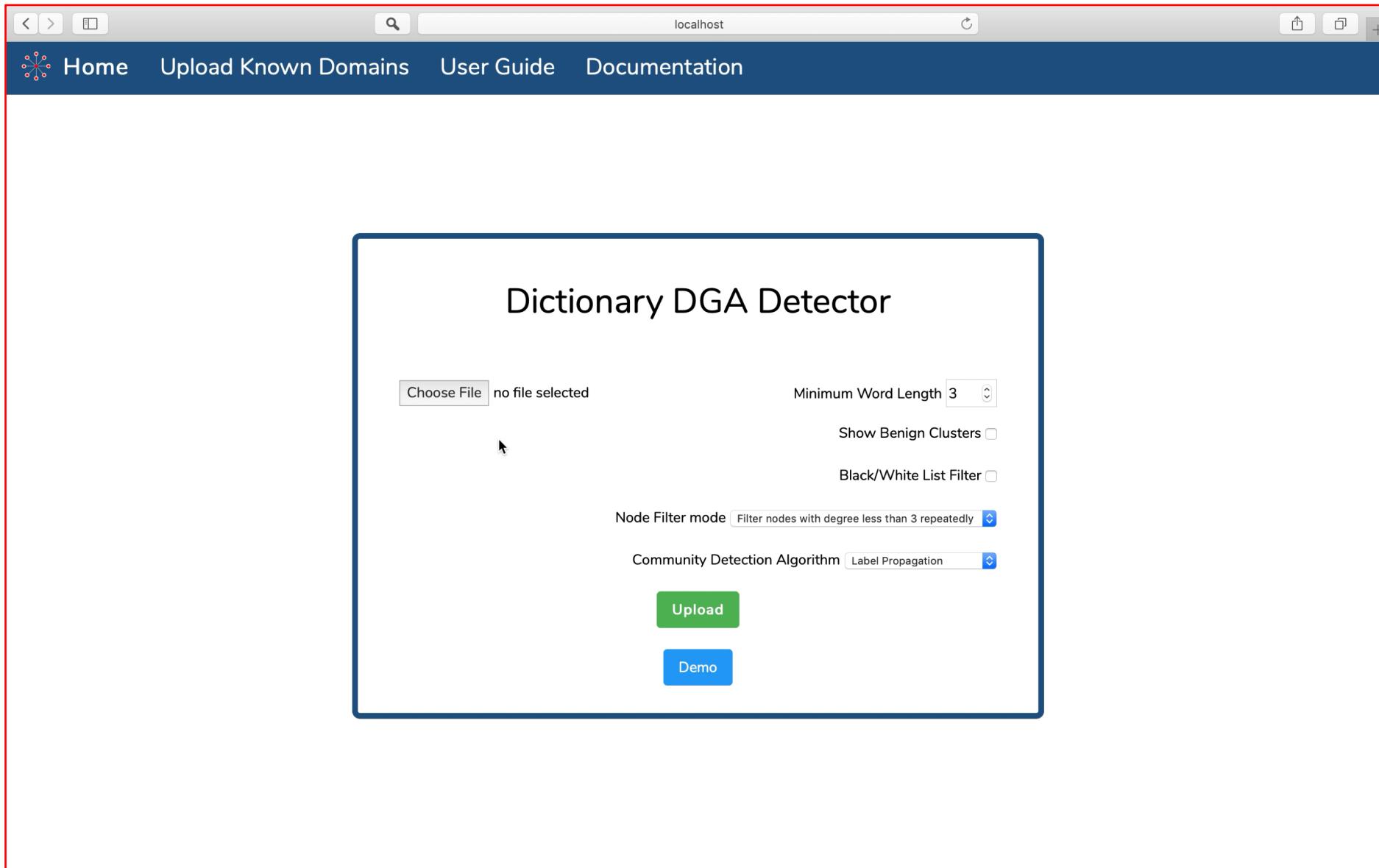
Below is a summary profile of this threat throughout the years of observation:

	2007	2015	2016
TARGET	eBay Accounts	Harvest bitcoin via coinminer	Still being determined by looking for live C&C
URL	Fixed malicious URL	Uses DGA	Uses DGA
STEALTH MECHANISM	Uses Kodak Viewer Express	Uses fake MP3 extensions and error messages	Uses names as filenames (jewell.exe) and error messages
CAPABILITY	Information theft	Information theft, backdoor capabilities, may download other malware with other functionalities, clicker capabilities	Information theft, backdoor capabilities, may download other malware with other functionalities, mass mailing capabilities
DEFENSE	None	Disables AV via registry, disables firewall, terminates AV application including watchdog	Disables AV via registry, disables firewall, terminates AV application including watchdog

- Ran the solution against 33 million .net domains (courtesy of SingCert)
- Detected hundreds of clusters
- Verified results in VirusTotal
- Domains identified in the cluster are domains used in Bayrob Malware

# Application Tour !!

# Using the Application



# Analysing Results

localhost

Home Upload Known Domains User Guide Documentation

Analysing test (50k domains).csv

DDGA Clusters detected: 14 

DDGA Domains detected: 1533

Used 62 times: window  
Used 61 times: possible  
Used 59 times: finish, mother, acter, perhaps  
Used 58 times: between  
Used 57 times: square  
Used 56 times: sweet, simple

Cluster ID	Dictionary Size	Dictionary Words	Domain
0	142	mem erstood ctricity neighbor suppose fight low mis probably bro squ brought subject prod supp experie bat several write attempt possible woman under twenty party heart simple thought ose laugh ead pleasure crowd subj winter alr report already eady arti unde meeting mother erial dent dried follow fifteen understood million ught rial begin known perhaps chance ady poss loans ience battle ween delight spb article speak mayor att summer leave fect yel between partial angry niece perfect ject borrow stu ricity experi fresh mister member water electric rstood mate ish act window fol tle sev pleas produce sum smoke possi experience material oclock adcom finish toward mee stud perf student tricity garden mill spread uce mount mountain shake happen nearly succeed alre electricity cle method asure severa eral elect gentleman square stood icity sub bly strike exper sweet action white empt ect	possiblewrite.net beginangry.net summerangry.net mountainfifteen.net possibleaction.net partyaction.net alreadymister.net partybrought.net gentlemanheart.net begindried.net finishwhite.net gentlemansquare.net possibleheart.net yellow.place laughangry.net experienceproduce.net thoughtssucceed.net studentbeans.com alreadyspeak.net mountainsuppose.net perhapsspread.net smokepleasure.net materialfifteen.net smokeangry.net subjectwhite.net knownheart.net leavearticle.net sweetdried.net waterperfect.net womanperfect.net materialproduce.net probablyneighbor.net sweetmeeting.net perhapssquare.net knownbrought.net perhapsborrow.net possiblemister.net alreadyattempt.net mountainmister.net probablystudent.net probablyunderstood.net finishspread.net materialpleasure.net waterdried.net leavemayor.net freshspeak.net gentlemanangry.net mountainsucceed.net womansquare.net laughbetween.net knowndried.net sweetdirect.net finishperfect.net winterperfect.net freshperfect.net probablysucceed.net

Search:

# Usages and Future Development



**Threat Hunting**



**Monitoring Domains  
at the National Level**



**Improve Clustering Techniques  
Improve Classification Accuracy**





Thank You



Thank You



**SingCERT**  
Singapore Computer Emergency Response Team