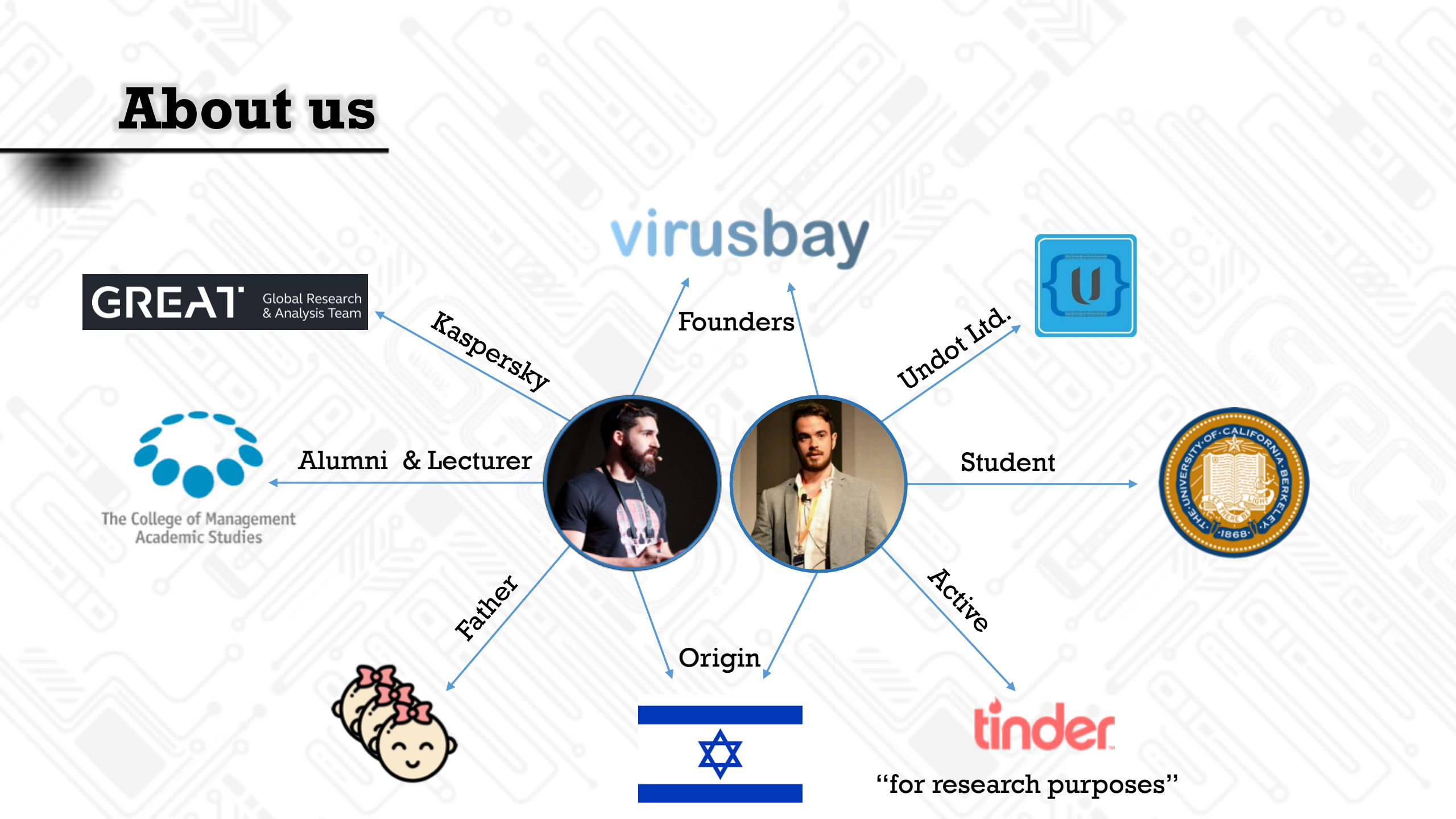
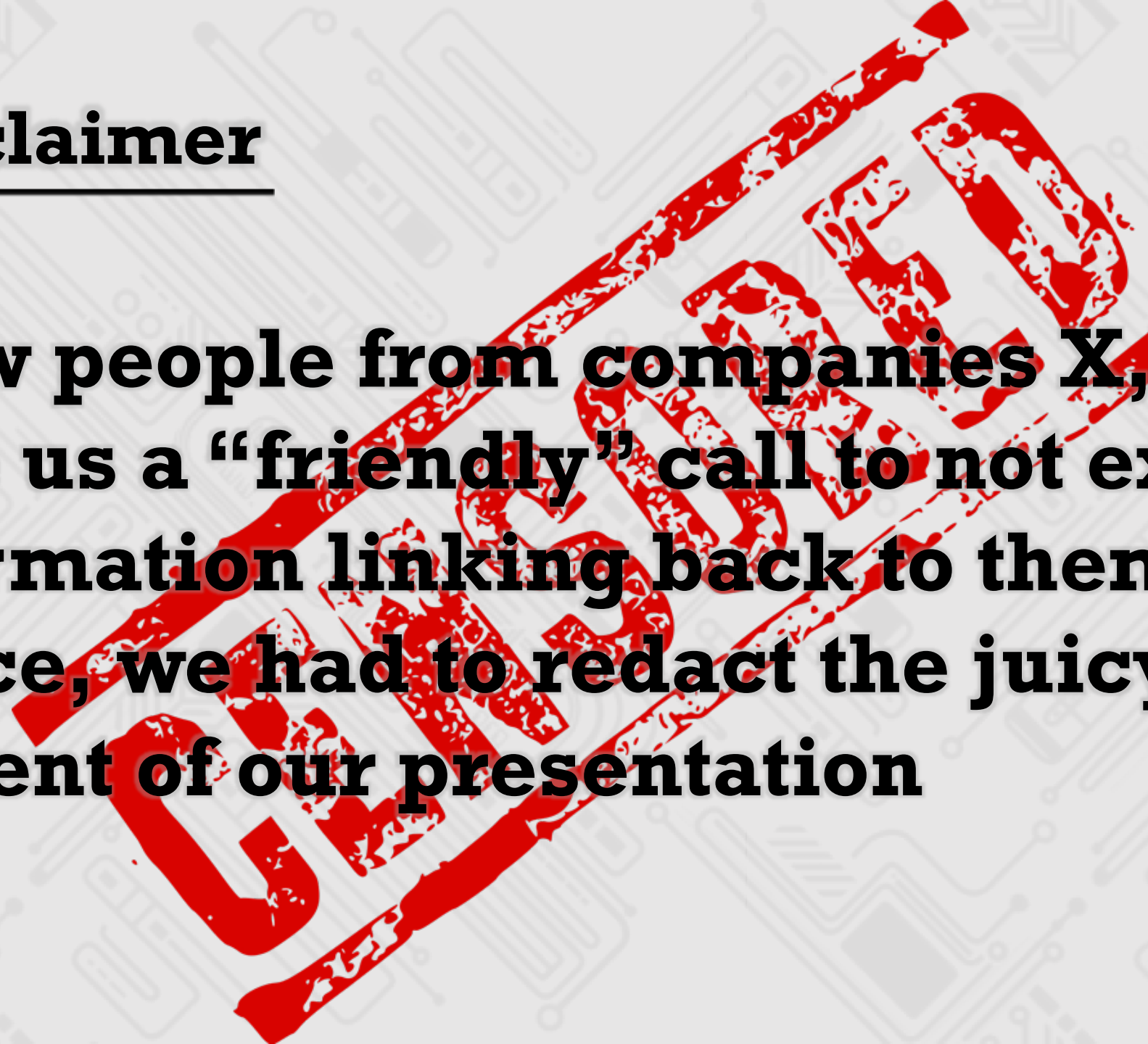
Dissecting Non-malicious Artifacts: One IP At A Time

Ido Naor (@idonaor1) **& Dani Goland** (@danigoland)

About us


virusbay
GREAT Global Research & Analysis Team
Kaspersky
Founders
Undot Ltd.
The College of Management Academic Studies
Alumni & Lecturer
Student
Father
Origin
Active
tinder
"for research purposes"

Disclaimer



A few people from companies X, Y & Z gave us a “friendly” call to not expose information linking back to them. Hence, we had to redact the juicy content of our presentation

Research Motive

```
From **********************  Tue Jul 10 03:01:02 2018
Date: Mon, 9 Jul 2018 11:30:15 +0200
MIME-Version: 1.0
Subject: New password notification
From: *********************************
To: **********************
Message-Id:
 <*************************************************************>
Return-Path: ************************
Delivered-To: ************************
Received: from localhost ([*********]) by
...
X-Orig-Recipient: ********************
X-Spam-Version: ****** Spamfilter
X-SPG-Loopback-Protection: 1
X-Halon-ID: b1eb9653-835a-11e8-9e50-005056911df2
Authorized-sender: ******************
Reply-To: *********************************
```

```
In [ ]: interesting = ['x-rejection-reason','authentication-results','x-virus-scanned','x-*********-antispam-report',
                       'x-**-*********-filtering-correlation-id','x-*********-antispam','x-*********-antispam-message-info',
                       'spamdiagnosticoutput','x-*********-prvs','x-exchange-antispam-report-test','x-******-senderhistory',
                       'x-******-esa','x-eopattributedmessage','x-**********-spam-details','x-**********-virus-version',
                        'x-*******-av','x-mimeole','x-**-**-product-ver','x-**-**as-result','x-*******-cor','x-*********-connect
                        'x-spam-status','x-*********-brts-status','auto-submitted','x-*********-spam-report','x-spam-score','x-a
                       'x-*********-spam-status','x-**********-spam-score','x-*********-spam-report','x-spam-flag','x-*****-analy
                        'x-*****-version','x-*****-result','x-messagesniffer-scan-result','x-****-scan-details']
```

Research concept

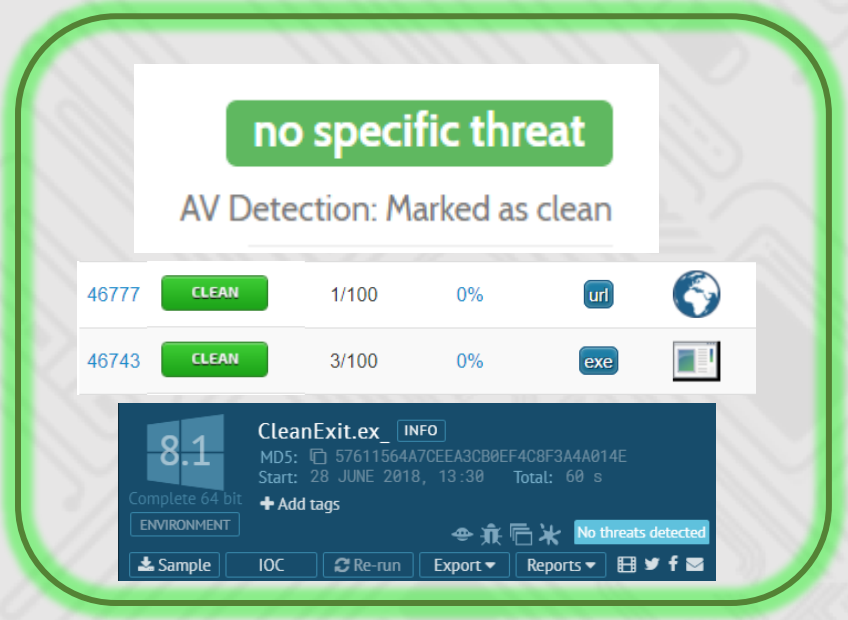
Employees use
online services
to educate themselves

Security products
ARE NOT SUPPOSED TO
use online services
to enhance detection

vendors use
data of their clients and later
have zero responsibility for it

Hackers steal data and later
publish it to online platforms

Research Scope



Garbage

Gold

Research Goal

Starting points

- Paid/non-paid services
- Repositories
- Other

Building Blocks

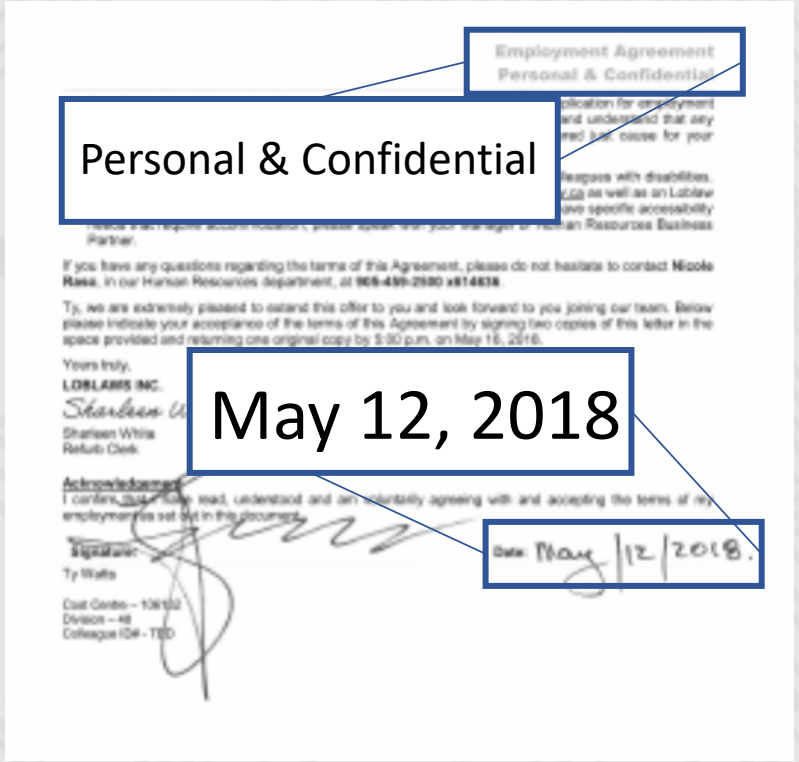
- Malware Research
- Yara
- Data Science

Goal

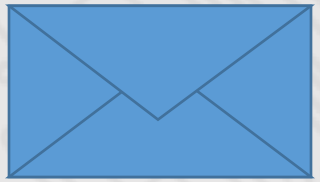
Prove that data is being unwillingly exfiltrated from organizations, and that with simple tools – it can be exposed.

Proof of Concept

- One simple search: "message"
- Mail containing non-malicious artifacts
  - Options:
    - Mail uploaded by employee
    - Mail suspected by security product
    - Mail uploaded by 3rd party
- Example of company intellectual property being leaked



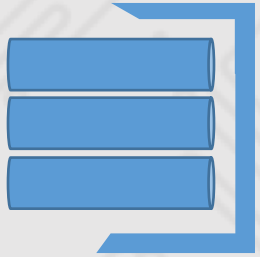
Research Subjects



Emails

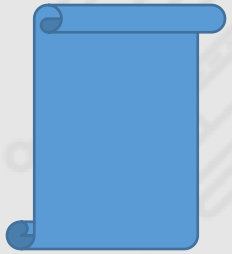
Code

Dumps



Archives

Documents



Certificates

Keys

Secrets

Graphs

Step 1: Possible Feeds

- Code repositories / Open source
- Script/Paste repositories
- Malware repositories
- Multi-scanners
- Online sandboxes
- Forums / Social platforms

Step 2: Yara rules

```
rule filter_interesting_eml {
        strings:
                $eml_01 = "From:”
                $eml_02 = "To:"
                $eml_03 = "Subject:"
                $attachment_id = "X-Attachment-Id"
                $mime_ty
        condition:
                all of (                                    and filesize > 300KB
}
```

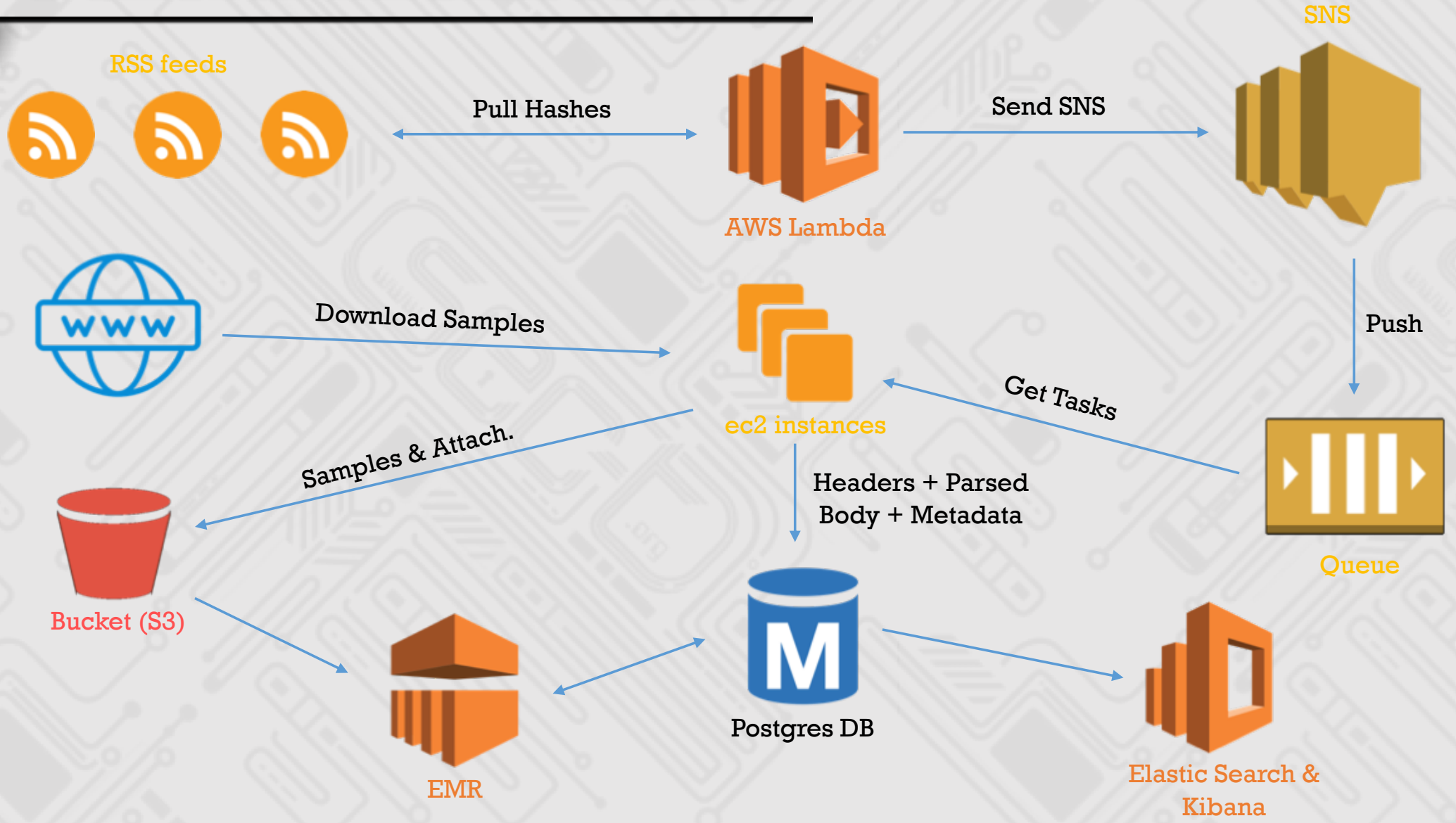
```
rule filter_interesting_outlook {
    strings:
        $a1 = {D0 CF 11 E0 A1 B1 1A E1}
        $a2 = "FROM:" nocase
        $a3 = "TO:" nocase
condition:
    all of them and filesize > 300KB
}
```

Step 3: Collection

- Searched feeds for undetected samples
- List hashes from results
- Downloaded undetected samples
- Filtered emails using Yara
- Sent matches to pipeline

```
var https = require('https');
var AWS = require('aws-sdk');
AWS.config.region = 'us-west-2';
var s3 = new AWS.S3();
let url = "https://<datasource>/...";
exports.handler = function (event,
context,callback) {
  https.get(url, res => {
  res.setEncoding("utf8");
  let body = "";
  res.on("data", data => {
    body += data;
  });
  let ids = []
  res.on("end", () => {
    body = JSON.parse(body);
    let notifs = body['notifs'];
    let filtered = {alerts:[]};
    for(let i = 0; i<notifs.length;i++)
    {
        …
        }
    }
```

General Architecture



Analysis Pipeline

#1 Feature

Spam Or Ham
- TF-IDF → Naïve Bayes
- Open Source Dataset

Email Queue

Attachment Queue

Custom Parser
- Subject & Body
- Language Detection
- Domain Names
- Geo lookup on domains
- Extract Attachments

Apache Tika

Content & Metadata Extraction

Postgres

NLP Pipeline

Preprocessing
- Stemming
- Stop word Removal
- Tokenization
- POS Tagging

scikit learn
spaCy
gensim

Named Entity Recognition
- Names Of Persons
- Organizations
- Roles(CXO,VP)
- Locations
- Monetary Values

Topic Modeling
- 10 Topics
- Top 10 words/topic

*NMF

*LDA

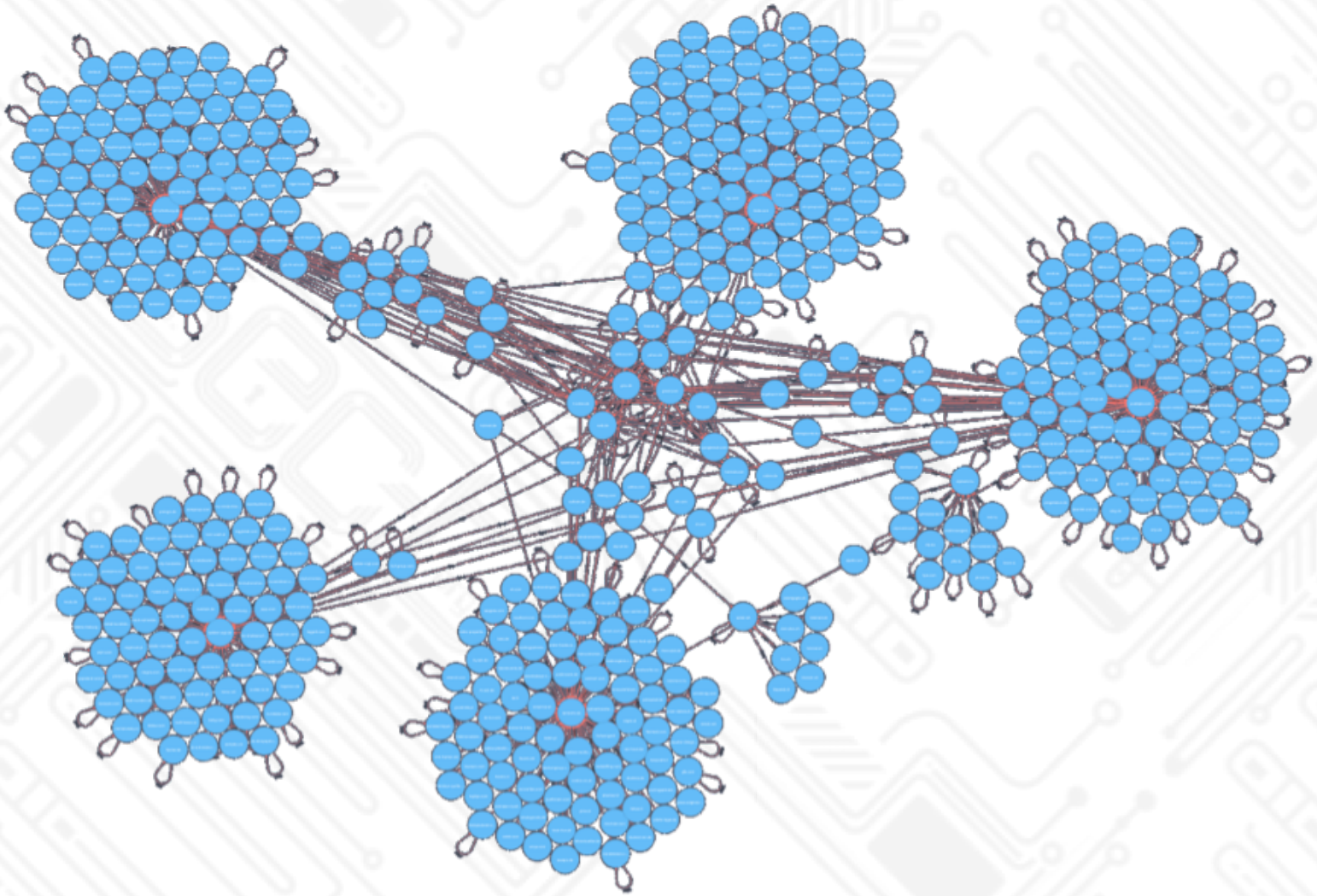
1(Casual Email) - 4(Critical/Highly Confidential)

elasticsearch

kibana

NMF = Non-Negative Matrix Factorization
LDA = Latent Dirichlet Allocation

Graph-based



Global Problem

DEMO TIME

QUESTIONS ?

(1) Please 

(2) Grab the 

(3) Stand up & speak English

VirusBay Invite Code: SECTSWEDEN2018

https://beta.virusbay.io/

THANK YOU!

virusbay

VirusBay Invite Code: SECTSWEDEN2018

https://beta.virusbay.io/