Dealing with Big Data in Cyber Intelligence

Greg Day

Security CTO, EMEA, Symantec

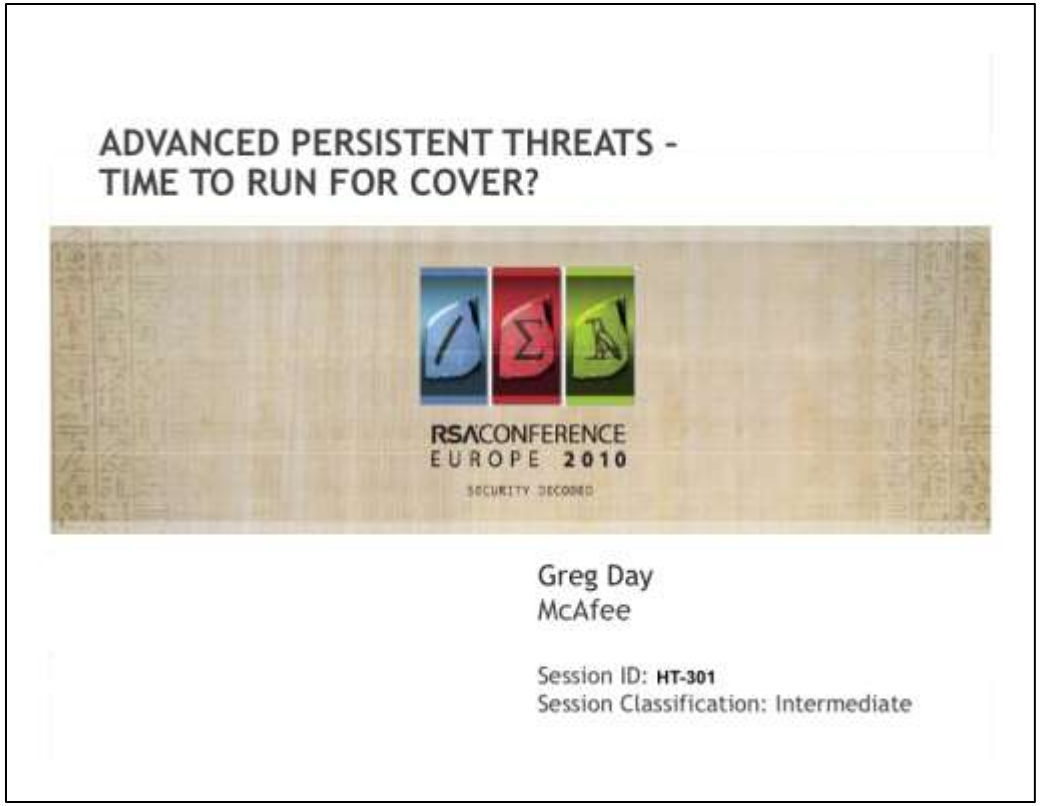
Session ID: **HT-303**
Session Classification: **General Interest**

RSA CONFERENCE EUROPE 2012

What will I take away from this session?

- What is driving big data in cyber?
- How is cyber intelligence evolving to create big data?
- How can or should I use this in my company?
- Should I be gathering my own data?

APT's - since 2010



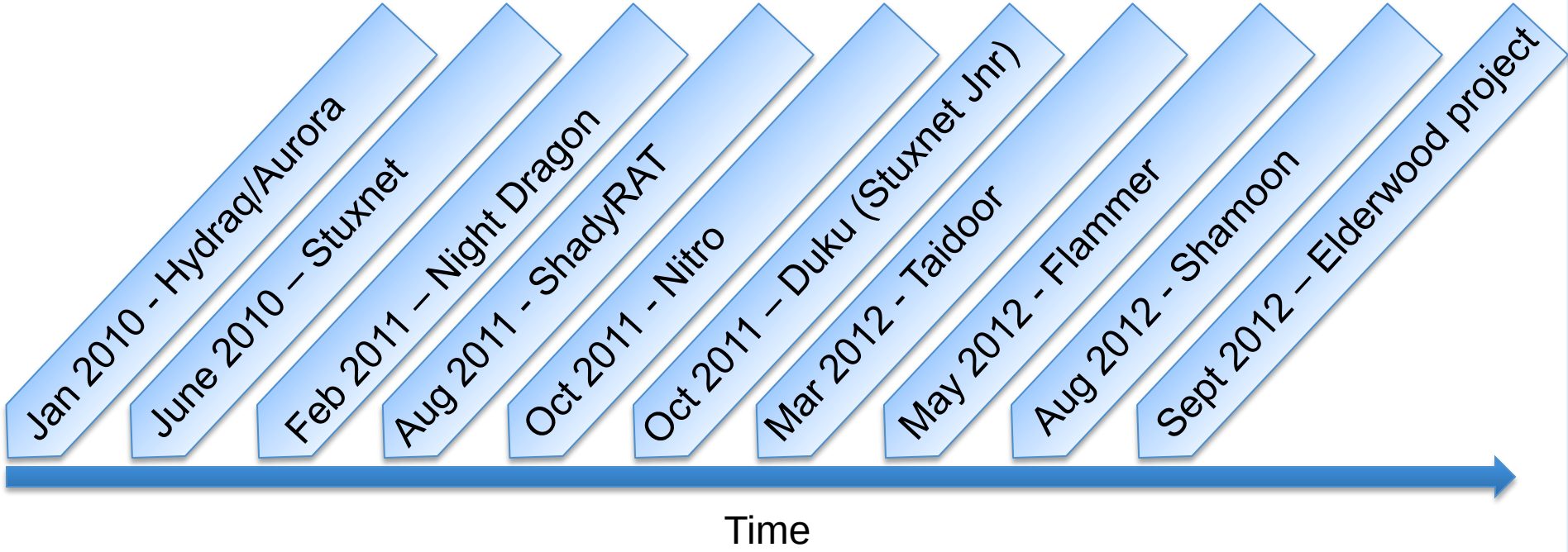
ADVANCED PERSISTENT THREATS -
TIME TO RUN FOR COVER?

RSACONFERENCE
EUROPE 2010
SECURITY DECODED

Greg Day
McAfee

Session ID: HT-301
Session Classification: Intermediate

(Advanced) Persistent Targeted attacks



> there are unconfirmed reports of infections dating back to 2007 as well.

Symantec Security Response Blog

Flamer – 29 May 2012

Attack Evolution (motive and methods evolve)

Cybercrime

ANYONE

1980's

2000's

PTTs

BUSINESS

APTs

GOV, MILTARY,
GLOBAL BUSINESS

2010

2011

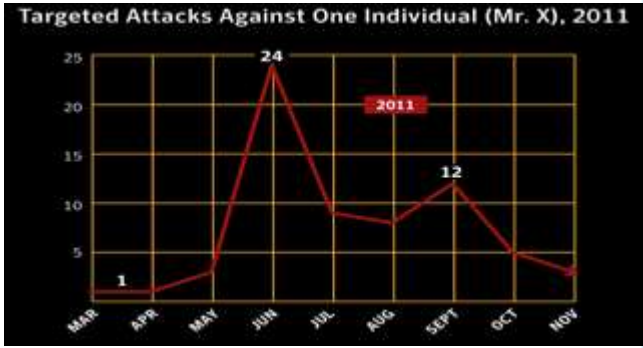
2012

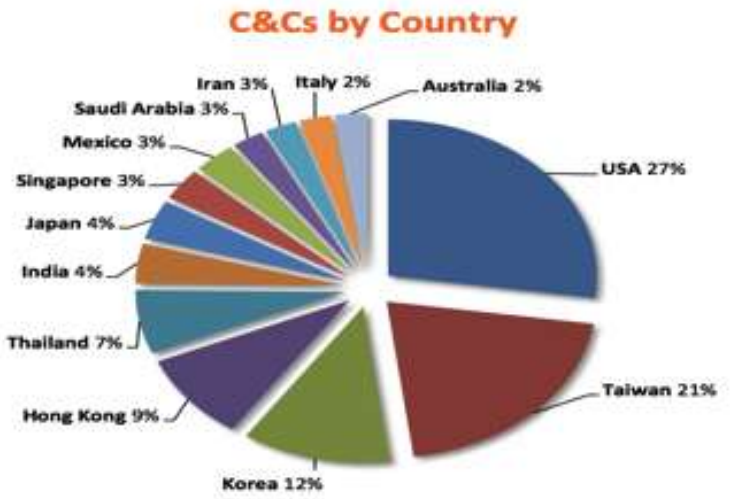
Not so advanced - Taidoor (14+ variants)

- Negotiations - US and Taiwan modernization of Taiwan‘s air- force.
- Targets: primarily private industry and influential international think tanks involved

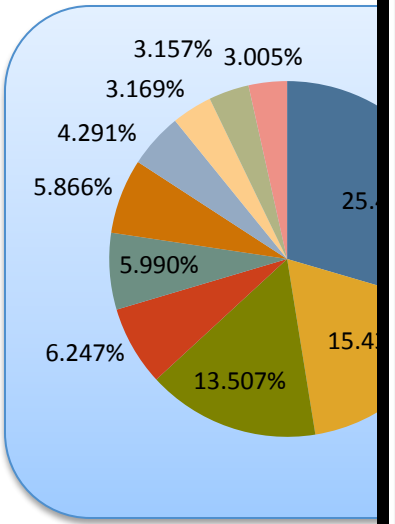
  Specifically those who have expertise in South Asia and South-East Asia policy and military strategy

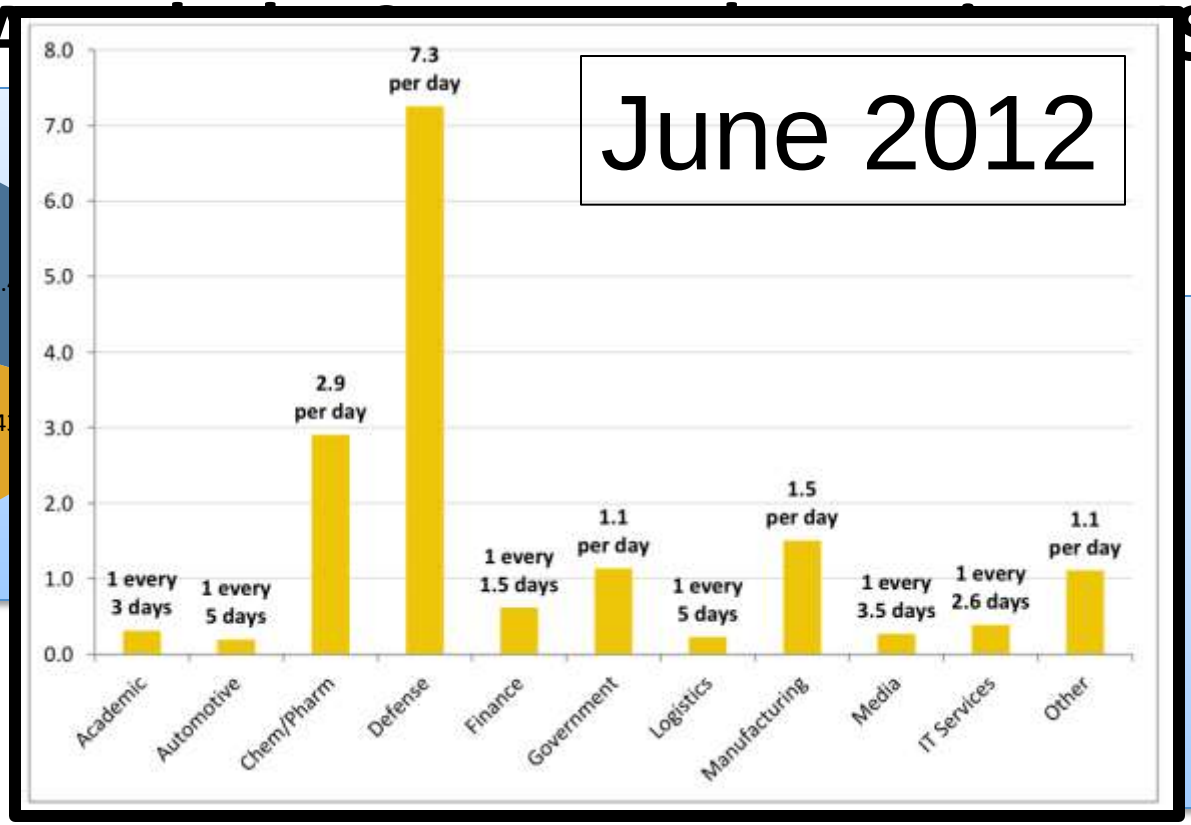
    peaked during the US-Taiwan Defense Industry Conference‘ held on September 18th-20th



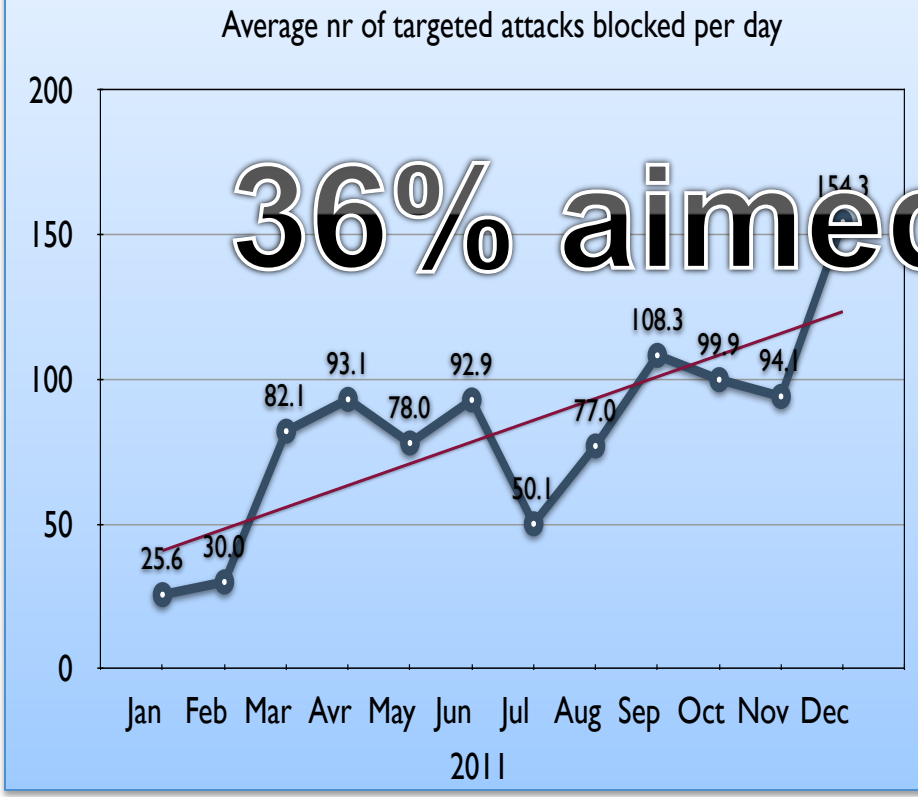


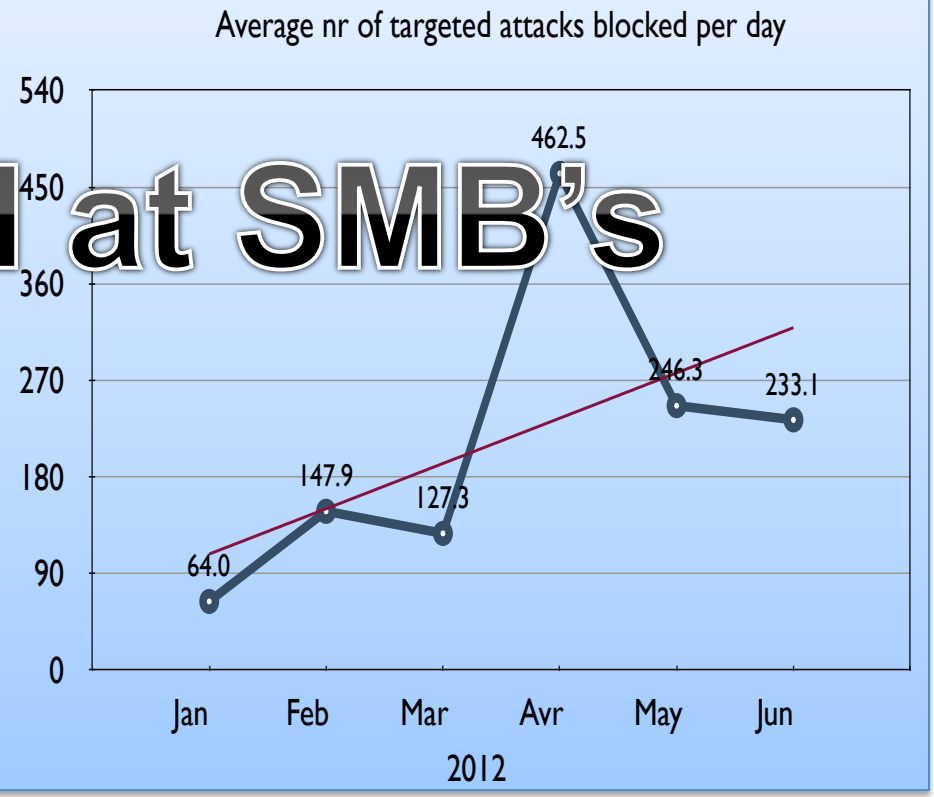
Targeted A[illegible]STR 2011





Rise in Targeted Attack Activity over Time



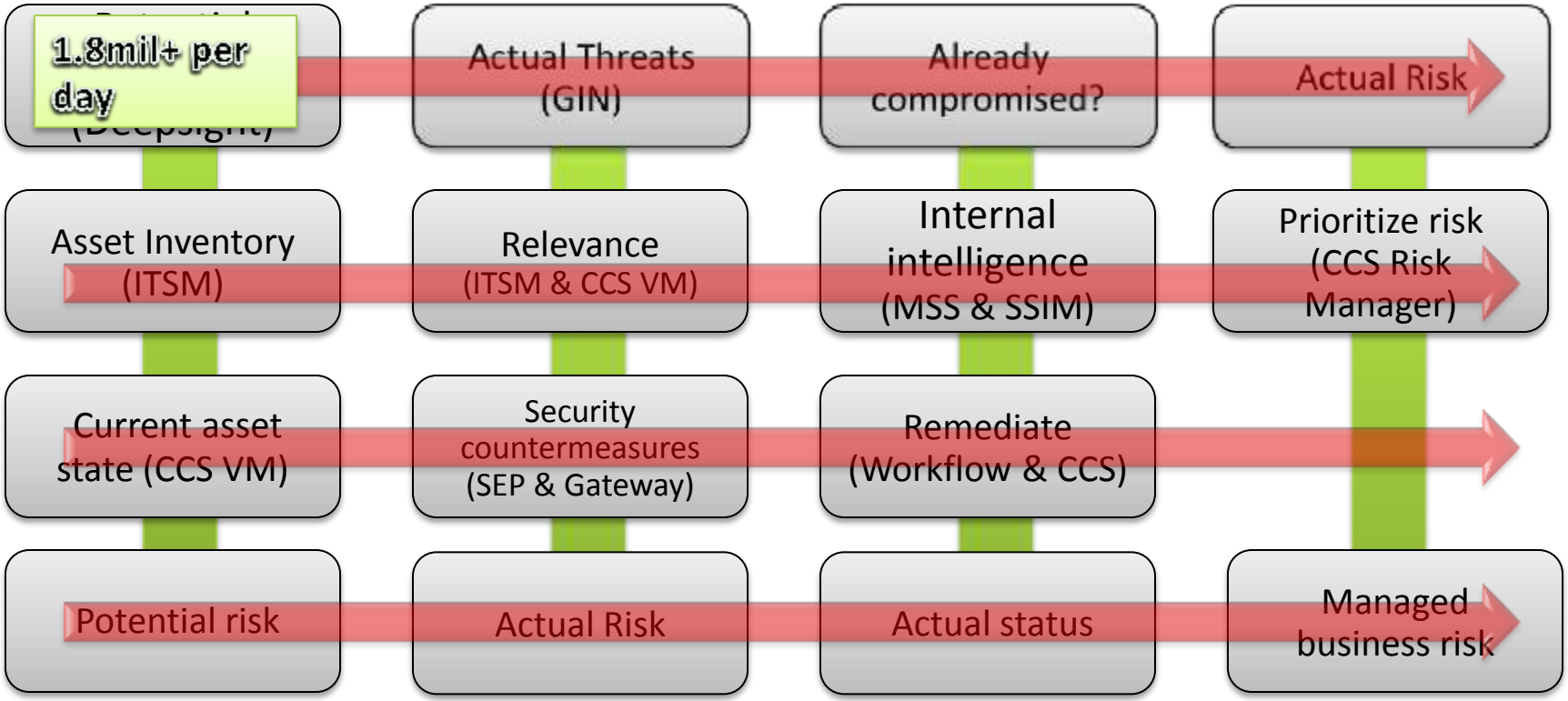


36% aimed at SMB's

So are we catching these today?

We hear about it from the masses!

Process - Can you turn 2 million+ decisions into 10?



Time?

Realistically we wait for alarm bells to ring?

So how to solve the problem?

Rethink: Anatomy of an attack

1. Research
2. Incursion
3. Discovery
4. Capture
5. Exfiltrate

Risk Posture and Policies

Strong security awareness, counter intelligence

Continuous enforcement of controls according to risk policy (mgmt and protection)

Actively monitor infrastructure (endpoint to perimeter), information and users

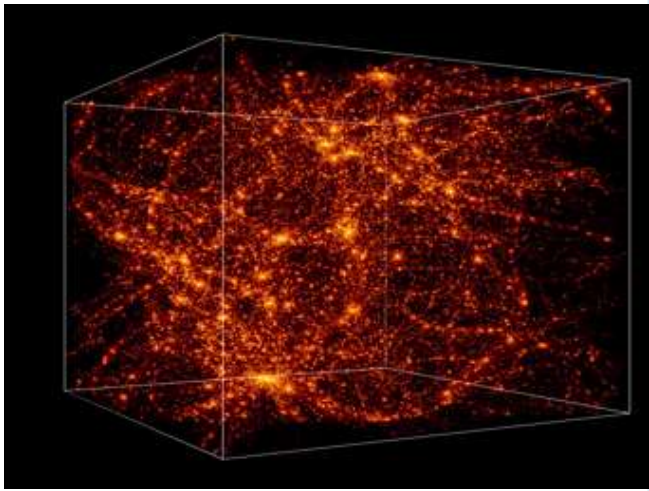
Control unusual internal movement and access of sensitive data

Counter intelligence, forensics, damage mitigations and information recovery

Reconnaissance → Weaponization → Delivery → Exploitation → C2 → Exfiltration

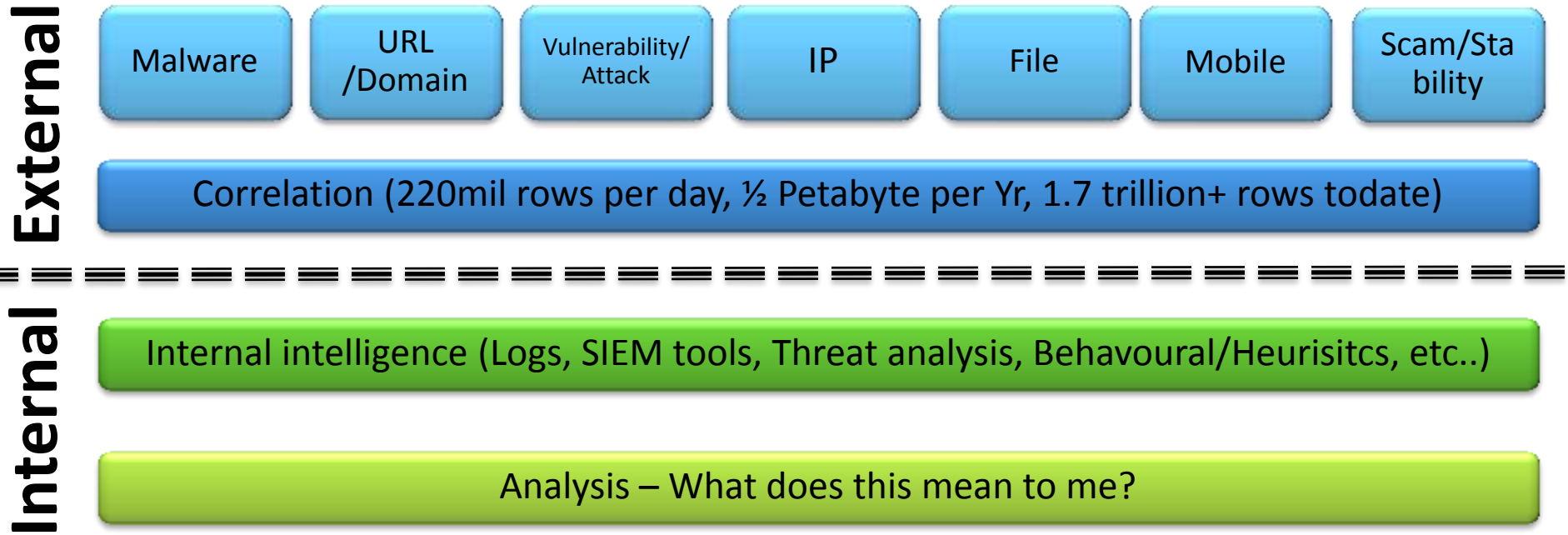
Enabling Cyber in a world of big data

1. Cyber Intelligence (what may and is happening)
   - to anyone or just to me?
2. What does this actually mean to me?
3. What should I do
   - Prevent
   - Respond

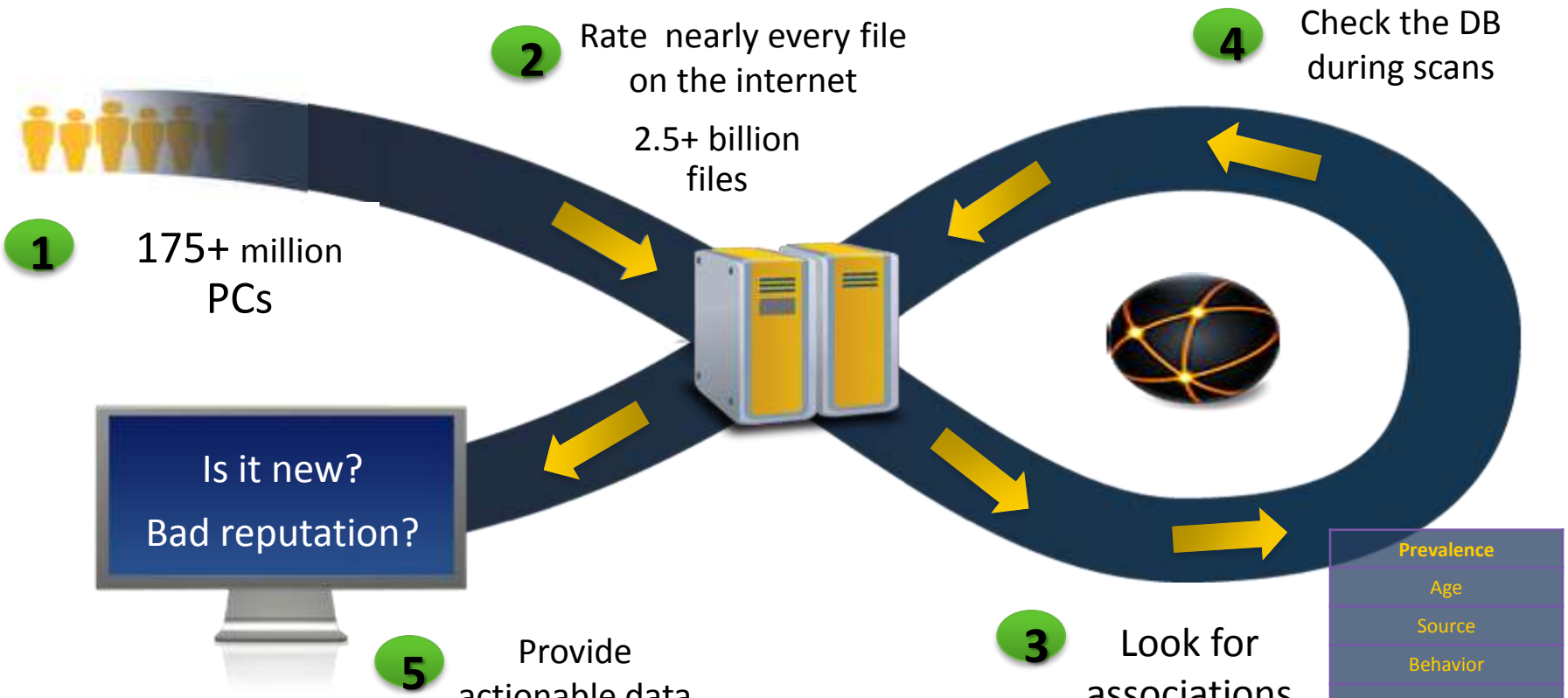


- Can I enhance my security with better Cyber Intel
  - Looking for better Intel/early warning systems
  - Specific contextual intelligence
  - Briefings on new specific threats and the mitigations
- Correlation and awareness
  - between our industry data and the your environment; help in prioritizing, defining risk and response
- Response to incident
  - Contextual data
  - Geneology
  - how to prioritize and action events and incidents

New Intelligence requirements – Big data!



Security Industry example – File reputation

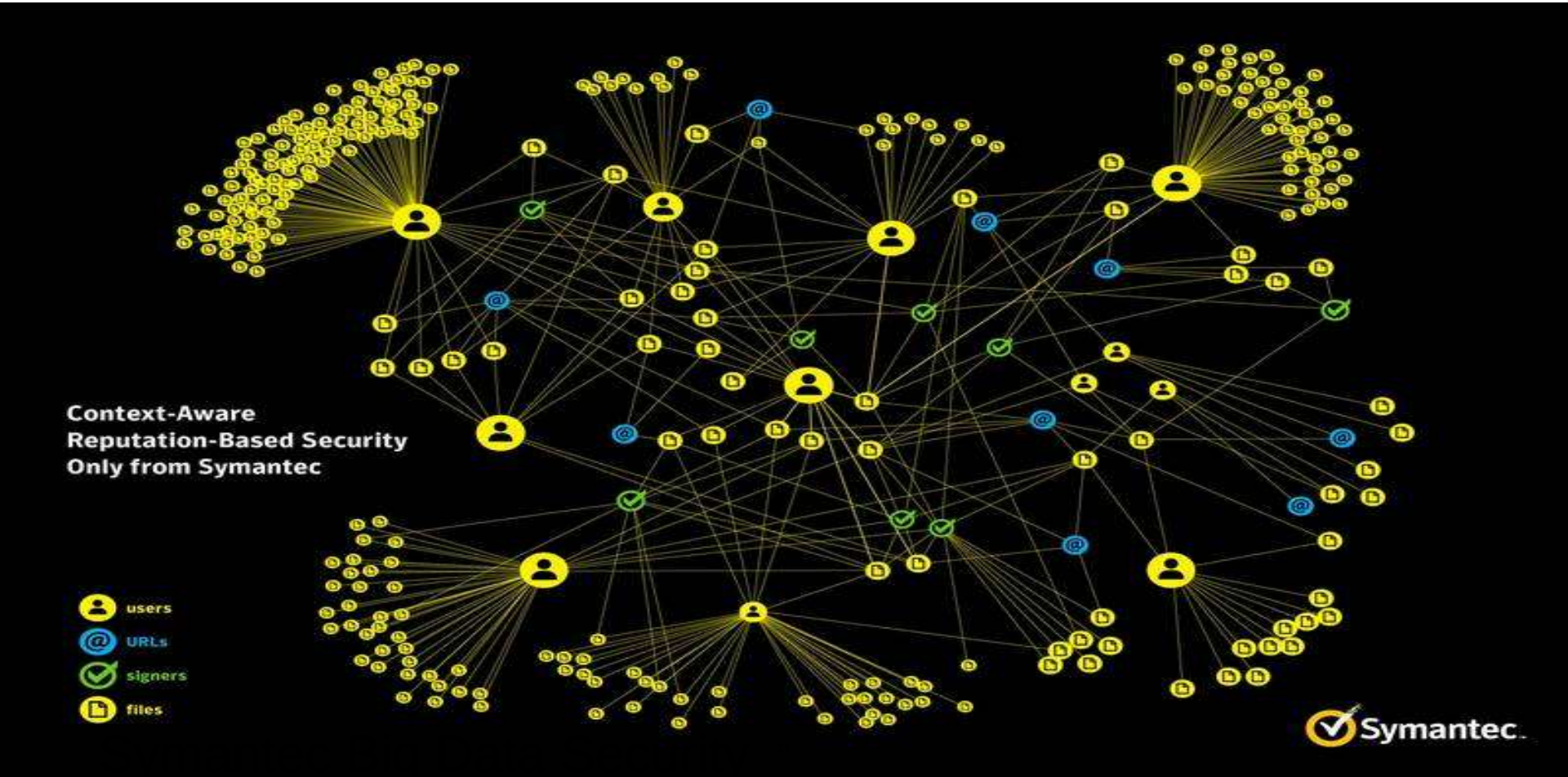


IP reputation – being able to mine the data (genealogy)

	address	consecutive_listings	listing_ratio	reputation	name	hostility	confidence
2	67.63.55.3	90	90	10	malware	4	5
3	92.242.132.15	90	90	10	malware	4	5
4	184.154.48.82	90	90	10	attack	2	2
5	184.154.48.82	90	90	10	bot		4
6	69.43.161.170	90	90	10	malware	5	5
7	69.178.145.238	90	90	10	attack	5	4
8	137.118.219.139	90	90	10	attack	3	5
9	27.255.64.111	90	90	10	malware	4	5
10	70.248.29.2	90	90	10	attack	4	5
11	72.3.199.7	90	90	10	malware	4	5
12	83.133.119.154	90	90	10	malware	4	5
13	83.133.119.155	90	90	10	malware	4	5
14	83.133.124.196	90	90	10	malware	4	5
15	41.223.119.129	90	90	10	spam	5	4
16	74.113.233.56	90	90	10	malware	4	5
17	74.113.233.58	90	90	10	malware	3	5
18	173.236.21.106	90	90	10	attack	2	2
19	173.236.21.106	90	90	10	bot		4
20	188.95.52.163	90	90	10	malware	4	5
21	204.232.137.207	90	90	10	malware	4	5
22	207.223.0.140	90	90	10	malware	4	5
23	208.73.210.29	90	90	10	attack	2	5
24	208.73.210.29	90	90	10	malware	4	5
25	208.73.210.29	90	90	10	spam	5	4
26	209.8.45.28	90	90	10	attack	3	5
27	63.223.106.16	90	90	10	malware	4	5
28	63.223.106.17	90	90	10	malware	4	5
29	64.74.223.47	90	90	10	malware	4	5

Smart File Reputation – All about the correlation

Over 100 Billion Associations



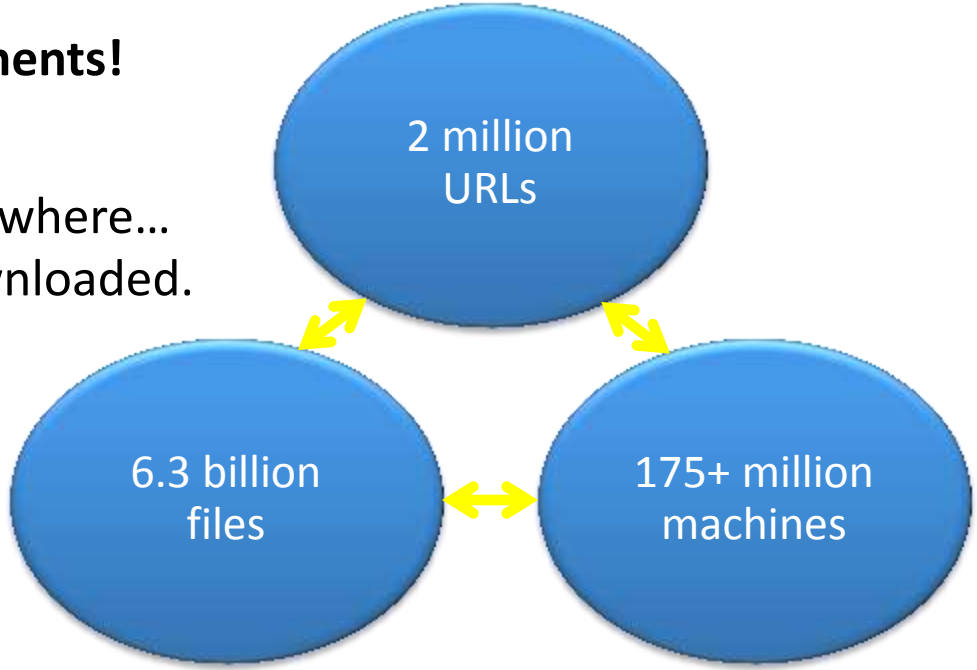
Big Data – scalability & speed

Understand your time frame requirements!

E.g: File reputation:

Based on who downloads - what from where…
- Not just how many times a file is downloaded.

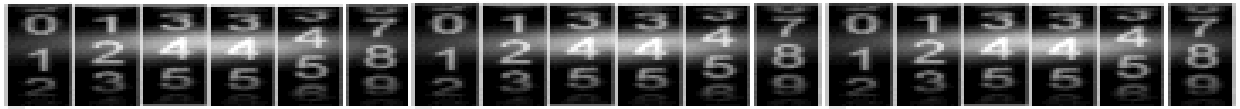
When one machine downloads a file,
all reputations must be re-calculated!



That’s over 100 billion associations

that must be refreshed every few hours from 2Billion+ queries!

It’s all about the information



How do you apply the data?

How to leverage cyber intelligence

Security products

- Summary feed
- Query-able data
- Contextual/Meta Data

Security industry data

Managed Security Services (MSS)

Logs, SIEMs
Threat Analysis

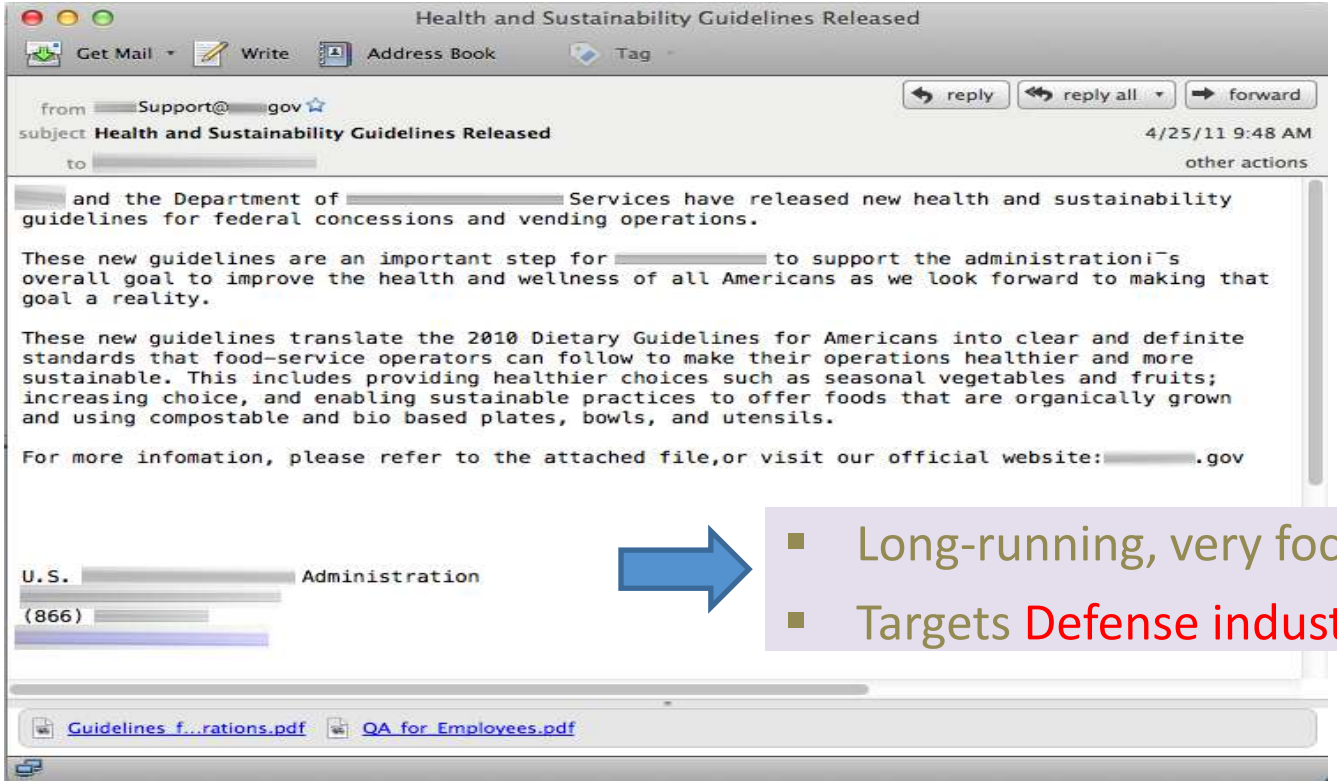
Some key questions for you to contemplate…

1. What intelligence do you gather today?
2. What intelligence should you be gathering?
   - How long do you keep it?
3. What external data do you receive?
4. How are you going to correlate the data?
   - So you have the expertise & resources?
5. How are you prioritizing the data?
6. What is an acceptable timescale to analyze?

Why do you need better intelligence?

- Increase detection of malware
- Identify/prevent targeted persistent attacks
- Understand the threat genealogy
  - Who what where and why
- Post attack forensic analysis
- Prioritize your responses?

Correlated cyber intel output: Sykipot (2011)



More Info:

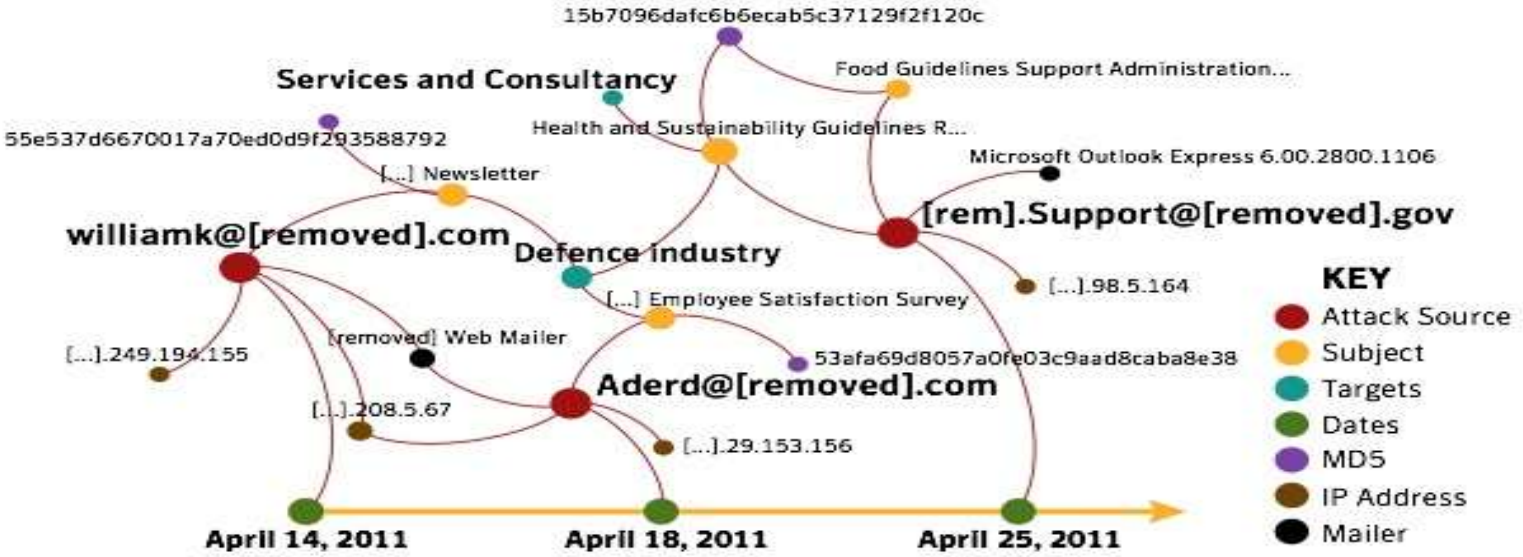
Detailed review in: *The Sykipot Attacks*, Symantec Connect Blog

http://www.symantec.com/connect/blogs/sykipot-attacks

- Long-running, very focused campaigns
- Targets Defense industries, Governments, etc

Understand the anatomy and geneology

3 attackers – 52 emails sent on 3 dates
Targeting 30 mailboxes of 2 Defense industries



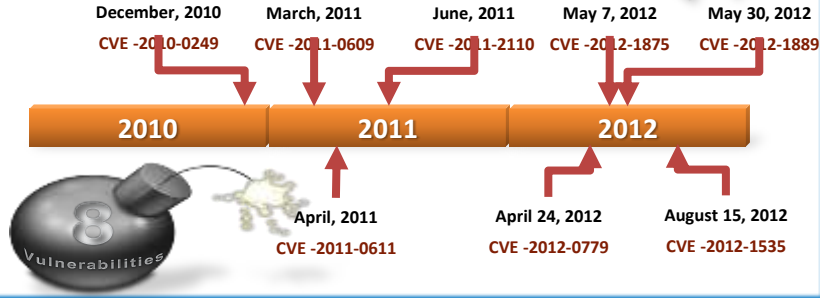
Understand the techniques/objectives

Commands received by Sykipot

ipconfig /all
netstat –ano
net start
net group "domain admins" /domain
tasklist /v
dir c:*.url /s
dir c:*.pdf /s
dir c:*.doc /s
net localgroup administrators
type c:\boot.ini
systeminfo

Elderwood project – connecting the Dots!

- Hydraq/Aurora + many more
- A platform is a technical term for the integrated building blocks used by software developers
- There is evidence that the Elderwood group have a platform that they use to build the components of their attacks
  - Consistent document file used across multiple campaigns
  - Common SWF file used to trigger exploit code
  - Automated registration of domain names
  - Automated intelligence gathering on targets
  - Tool to automate generation of web-based email accounts



How to leverage the data?

The Current Paradigms Have It Backwards….

Visualization & Risk modeling evolution

Summary - What do I differently tomorrow?

1. Do you know if you have been breached?

   a. We need to better understand our own environments (own big data)
      - Infrastructure, users, information

   b. Record everything or focus on high impact areas?
      - Profile to baseline normal

   c. 3rd parties/supply chain can be the weak link

2. Accept that you have or will be breached, what do you do?
   - Too much information, how do you spot the wood for the trees?
   - Do you have the forensics?

3. What is an acceptable breach?

Summary - What do I differently tomorrow?

3. What external feeds are you using today?

   a. Do you monitor social channels for chatter of a breach?

   b. What Intelligence feeds do you receive/use?

      - Are you gathering attack or attack attribute data?
      - Reputational data?

4. Do you have the skills and the tools?

   a. Big analyst skills requirement!

   b. What is the right big data tool for you?

   c. How do you correlate to business impact?

Summary - What do I differently tomorrow?

5. Be clear on your objectives for big data

   a. Detection (During or after)?

   b. Forensics?

   c. Better business response?

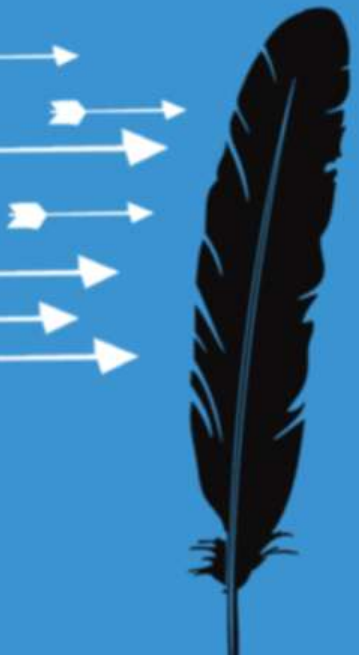
   What is the appropriate timescales for these?

6. Do I have the resource to do this internally or should I outsource?

Summary

Cyber Intel key to future success as targeted attacks evolve, requires correlation of your intelligence data and external data

1. Recognize there is more to threat mitigation than real-time defense
2. Do you correlate your existing internal security data (State/Incident)?
3. Are you leveraging the security industry big data that already exists today?
   - In Security products or your Incident Response teams?
4. How do/will you correlate suspicious activities into something meaningful?
5. How do/will you visualize/prioritize your response activities?
6. Do you have the budget/resource/capability to achieve this or should you outsource the solution?



Thank you

Greg Day
Security CTO EMEA
Symantec
Twitter: GregDaySecurity