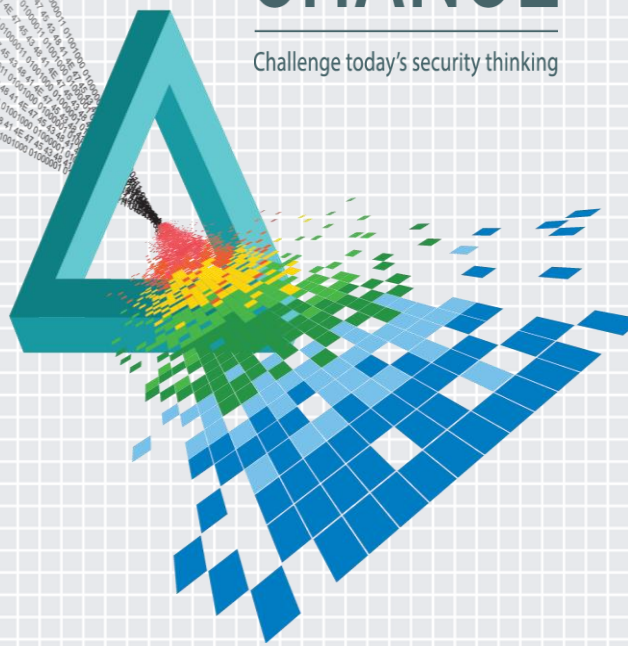
RSAConference2015

San Francisco | April 20-24 | Moscone Center

CHANGE

Challenge today's security thinking



SESSION ID: DSP-R02

Please DON'T Share My Data: Imparting Sensitivity Markings on Shared Data

Patrick Cain

Resident Research Fellow
APWG

The Dilemma

- I want to share some cyber-crime data with you
  - E.g., one of your systems is compromised, I have your password, etc.
  - I trust you will not disclose my data improperly
  - [Actually, I don't trust you at all]
- Instead, I will share data (maybe anonymously) with a data clearinghouse so they can share it with you
  - How do I say how I want the data to be further shared (a marking)?
    - (And maybe NOT to you) ☺
- I have 18,623 pieces of data to share, today
  - How do I make this a script?
  - And will you understand the data and markings when you get them?

The APWG

- Started In 2004 as Anti-Phishing Working Group
- Non-profit US corporation
- ~2100 members, 25 researcher groups
  - National Bodies, CERTs, LEA == free
  - Extreme International Composition
  - (Really) Big Company ←→ Sole Proprietor
- Goal: solve problems, share experiences and data
  - We host meetings, generate metrics, and share data
- Be vendor, country, and * agnostic

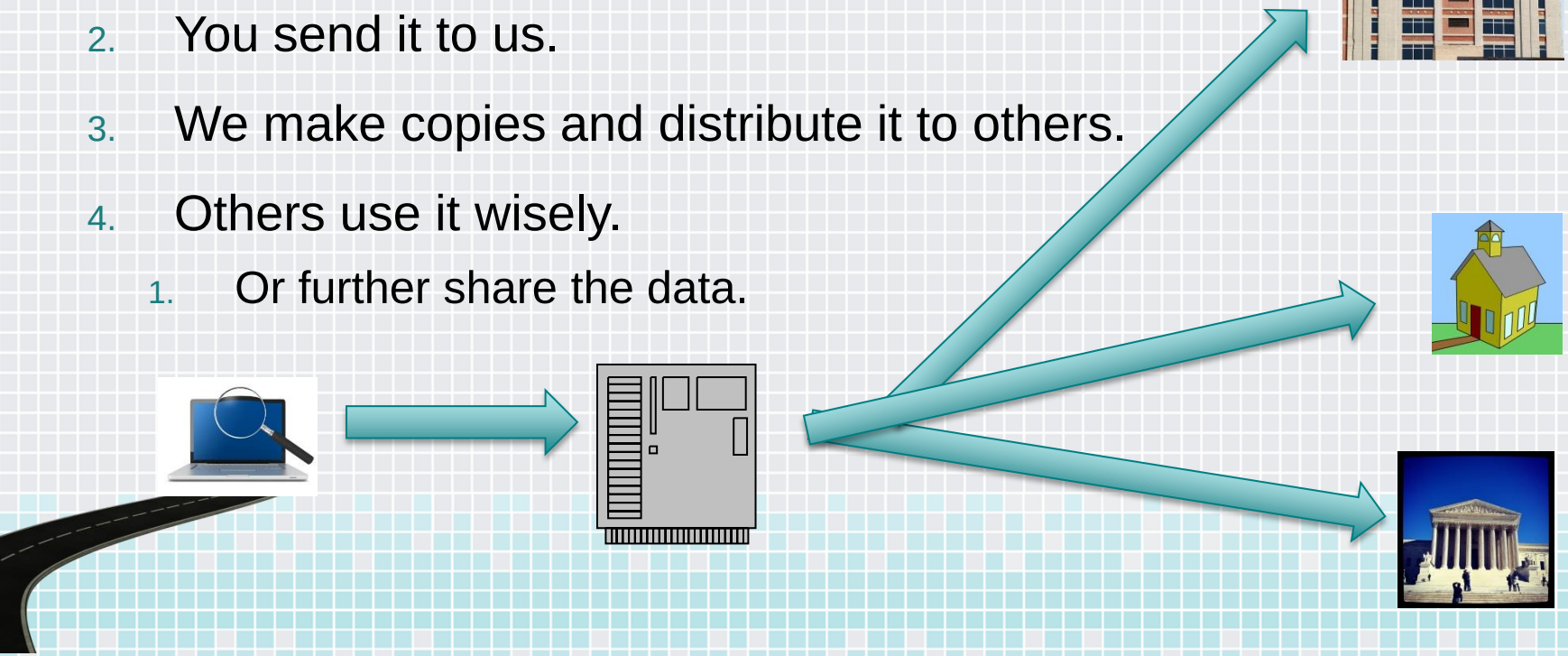
APWG Committed to Wiping Out Internet Scams and Fraud

Overview

- One of the APWG functions is as a data clearinghouse
  - We process 20,000 – 230,000 data items per day
  - It's very important that humans don't touch (slow down) the data
- Our big UBL is time-sensitive but not content-sensitive
- Some new lists ARE content-sensitive and less time sensitive
  - Think infected system or attacker identification, or future intel
- How do we note the sensitivity of that data in a way that makes the submitter, the recipient, and us comfortable?
  - Not all data is appropriate for all parties
  - How do I tell the collector NOT to spam my data to all parties?

An Example Clearinghouse

1. You detect badness.
2. You send it to us.
3. We make copies and distribute it to others.
4. Others use it wisely.
    1. Or further share the data.



Our Challenge

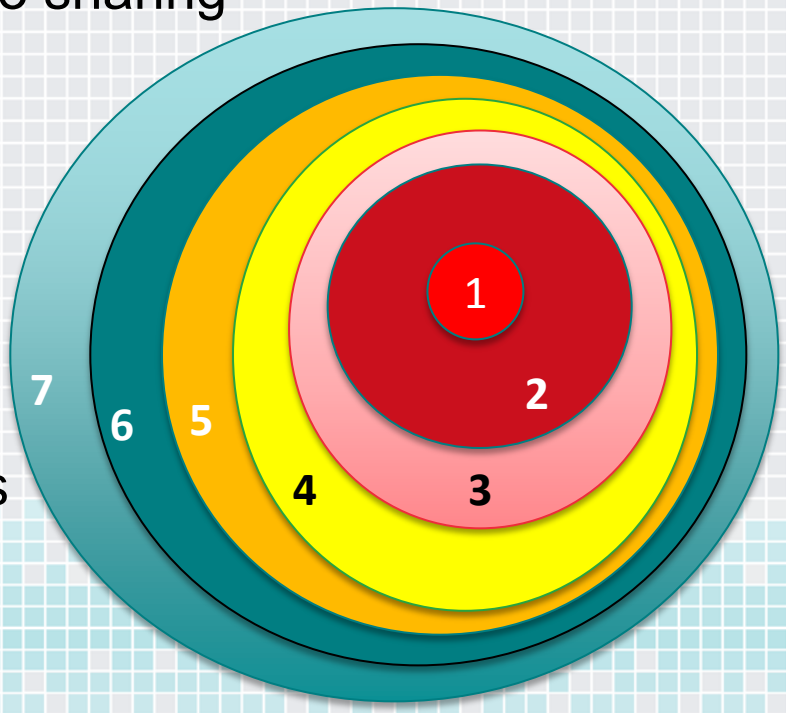
- We started with “How do we mark the data to guide its use and further sharing? TLP? The standard 4 categories?
- Then we backtracked to:
  - What is our data collection and sharing model?
  - Would any of the developed data marking schemes work for us?

Our Model

- Our users are international
- Everybody signs an agreement if they want data
- A submittor sends data – properly marked – to us
- We do minimal processing
- Recipients either get it or come and get it
- Recipients can do something with the data
  1. Process it themselves and destroy it
  2. Pass the data onto others
  3. Send us additional details about the data

Who to Share the Data With?

- OUR MODEL is a series of concentric ovals:
- 1 - No one, aka ‘recipient only’ or ‘no sharing’
- 2 - Coworkers in security groups
- 3 - To incorporate into products
- 4 - Share with affected users
- 5 - Share within the company
- 6 - Forward to other security groups
- 7 - Share with the public



Alignment to Regulatory or Legal Acts

- Most data sharing cabals are informal
- Could data markings be used for compliance?
  - Pink-marked data only shares with pharms
  - Blue-marked data only shared with banks
- Green-marked data can only be shared with certain countries?

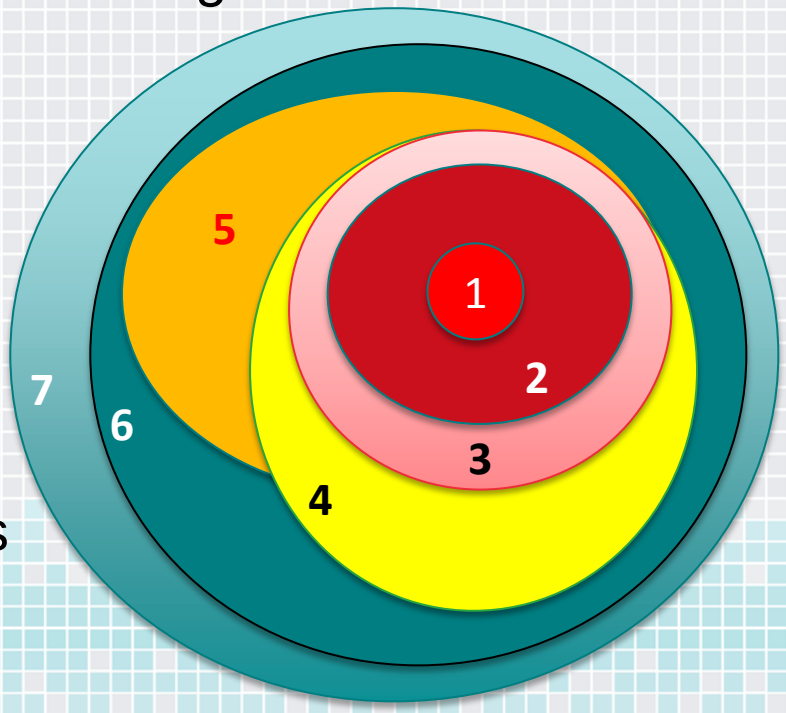
- We think defining sharing groups is easier than making conditional sharing decisions
  - How does one really hide the fact that sharing is restrictive?
  - What's the impact of sharing data with an improper party?

A Model with Disjoint Communities?

- The MODEL could be modified to a series of concentric ovals:
- 1 - No one, aka ‘recipient only’ or ‘no sharing’
- 2 - Coworkers in security groups
- 3 - To incorporate into products
- 4 - Share with bank users *OR*
- 5 - Share with pharm users

This gets harder in the outer circles:

- 6 - Forward to other security groups
  - Which data?
- We’re not doing this right now.



Example Document Marking Schemes

- IETF IODEF
  - 4 levels – Public, Private, Need-to-Know, Default
  - No real guidance on how to use it (as expected; IODEF is a format)
- Government or Intelligence Circles
  - Restricted, Confidential, Secret, Top Secret, Compartments, Caveats
    - Confusion amongst military recipients? APWG secret vs MOD secret?
- TLP – Traffic Light Protocol (the human marking scheme)
  - Red – recipient; Amber – limited ; Green – community; White – public
  - To meet our model we would redefine the “already excepted” terms
- REN-ISAC
  - Marks are grouped by community, TLP-ish

A New Marking Scheme?

- Many marking schema are
  - "if you can see it"…not "what can you do with it"…
  - Many do not distinguish "the details are for you; summarize to others"
- The MOD vs APWG marking issue is important; confusion is bad
- TLPv1 doesn't work for us
  - The general four colours seem to be targeted at human interpretation
  - It's hard to get our groups into four blobs where the scripts can decisively check the markings
- TLPv2 (whose v2? There are many divergent versions)
  - Amber still has confusion
- We need VERY clear definitions on what you can do with the data

We're Trying a New Marking Scheme

- The data marking is really an integer from 0 -> 99
  - Bigger integers mean bigger concentric circles
  - An (optional) phrase follows the mark to aid interpretation
    - E.g.: 1 – Community
- Each circle is divided into "Summary Info" and "Details"
- We added Extra 'caveats'
  - NA - No attribution
  - AI - Active Investigation, aka no touchee (NT)
  - HI – Historical
- Each marking definition has an id, e.g., APWG, apwg-2, bob
  - We don't give 'clearances'; we define communities

Marker	Description
0 - Recipient only	Not to be shared at all (Probably never used)
1 - Community	Recipient(s) should NOT share details of this data outside of community
11 - Internal-summary	Recipient(s) may share summary data with their internal groups
13 - Internal-details	Recipient(s) may share detail with their internal groups
21 – Impacted Party	Details may be shared with the target entity
23 – In Products	Details can be embedded within products
31 - Trusted-summary	Summary data may be shared with other trusted security types
33 - Trusted-details	Details may be shared with other trusted security types
81 - Public-summary	Summary data may be publicly
99 - No restrictions	Data has no sharing restrictions

Details of the Marking Scheme

- 0 – recipient, could be used to send data to us for aggregation, but not redistribution. This could also be administrative data
- The tags are arithmetically increasing
  - 21 is a bigger circle than 11
  - Comparison in the scripts are quite easy.
- The attached phrases can be different for internationalization
  - 0 – restricted === 0 – restringido === 0 - 受限
- There is currently no “but not these guys…”
- Up to this point, I didn’t say XML or CSV once!

Using the Marks – XML/STIX

```
<STIX_Header>
    …
    <Handling>
        <marking:Marking>
            <marking:Marking_Structure_marking_model_ref="apwg1">
                <apwgMarkings:tag>99 - No Restrictions</apwgMarkings:tag>
                <apwgMarkings:caveat>NA - No Attribution</apwgMarkings:caveat>
            </marking:Marking_Structure>
        </marking:Marking>
    </Handling>
```

Using the Marks - CSV

- The community, tag, and caveats are encoded as community/tag/caveats - followed by a comma
- ,”apwg/11 – Internal Summary/NA - no attribution”,
- Easy to parse by scripts ;)
- Communities could also agree to shortcuts
  - ,apwg/11/NA,

Initial Results

- We're currently trialing the marking. It's going 'okay'
- We've been asked to identify the TLP R/Y/G equivalents ☹
- Some people are hesitant to pick a mark
  - We've had to define a 'default' for each data type
    - As we gather more sensitive data this will change
  - The interface allows remarking
    - Original mark was improper
    - New data in the data record changes its sensitivity

Some Hard Operational Issues

- What happens if data is marked wrong?
  - We currently trust the submitter*
    - * - we may remark your data based on your history
  - Some sharing groups may prohibit certain marks
    - E.g., the “group identification” group may not use “public summary”
- Integration with others’ tools moves slowly
- The community still needs guidance/documentation
  - How to pick marks
  - What the marks mean to *you*

Tasks for *You*

- If your data sharing model looks like ours, feel free to use this
  - PLEASE let us know how it works
- If you are a data-marking-kind-of-person
  - Tell us what's broken or isn't working
    - There is a write-up on apwg.org
    - There is an XML schema and some examples
- If you send us data, try to mark it

- This is an evolution; expect changes.

Thank You

Patrick Cain

pcain@apwg.org