Models Breakout Session 1

What is that next big epiphany that will shape the future of modeling?

*What should the world of models look like in order to be able to feed this nirvana of understanding the cyber threat?*

Seeking research direction. What research directions are needed to find solutions?

Danko – Analyst may need to figure out who someone is. For example, if a military guy captured someone and then they empty their pockets and they find a name. *Be able to take a scrap of information and be able to build a resume about this person and then maybe be able to generate a model of what that person has been up to.* Maybe find if this person has been posting things on blogs or maybe this person goes into chatrooms. Use that information and generate a model. Is there a way to apply this type of problem to cyber?

Pamala – The system, in real time, will make real time decisions based upon the data.

*Be able to build systems in real time.*

2 aspects of data. 1) how to gather data. 2) how to represent data

*What is the next thing in depth beside just indexing data*

Fragment of information that doesn’t complete identify an individual. We want to be able to provide a direction based upon the fragment. *Reducing uncertainty* by being able to give actions to the analyst suggestion

Adversarial behavior:

* we really don’t know how to define an adversary
* Unless we have a fairly precise definition or a model we won’t be able to find them in volumes of data.

Adversary has ‘two’ components:

* goals
* Capabilities
* Computational power

Given a better definition of an adversary can we build a concrete method to identify and then focus the direction of the analytics to create the model.

*Cannot talk about an adversary without talking about specific technologies*

Victor – two ends of the spectrum of where the problems lies. Interesting problem “ripple in the force” kind of problem. Where there are massive collection of data and trying to discern what is useful in that data and what anomalies are present in that data. *Wondering if we should just invert the problem*. Not to think about the problem as deductible but think of it as inductively and look for the counter-positive of evidence and look for something interesting for us to do

Rita- need to identify what is non-adversarial and weed out the ‘noise’ to be able to weed out what is normal or abnormal

Problem of adversarial detection is what happens when there is an outlier that is not adversarial?

Searching the large data set for patterns may not be difficult but determining what patterns to look for may be difficult. *What determines bad behavior?*

*Find a way to retrain a model* after you feed back info that you have determined to be adversarial or normal.

*Not want static models to check again but a feedback system that allows analyst to interface*

* *Called “active learning” need active learning techniques for these models that the analyst is in the loop*

Pam - Not talking about one monolithic model. Discussing many models.

Parallelize information across multiple models and use humans to validate.

Sean – Heavy human in the loop need to be heavy humans on the loop. Not rely on humans so much. More AI centric model. Get *context* into the situation. “information qualifiers”. ***“The information about the information needed to make a decision.”***

Pam – are there new ways, what kind of research, how do we see that happening in the future vs. what is now.

Sean – weed out contextual information that is *not* relevant to decision making

Difference between meta data and meta information

Meta data – things that describe data. The contextual environment of the data.

Meta information –

Deborah – human data is messy, what people can legally write down, what they believe is ‘wrong’. Power grid – what does it take to hook up. Anticipatory data, find things to make proactive.

Meta object facility – primarily for static based modeling. Is being extended to include dynamic.

“Meta-meta-data” – Victor Harrison

Questions regarding models is the topic of conversation. Not the language of models (coding language).

How to come up with adversarial models.

How we operate, what’s important to us, how we use the network, model our *own* use of the network. There may be many models of adversarial models but where they merge with the models of *ourselves* is where is really matters so we can create a filter to automatically graph and model behavior.

David – an adversary is a person but the he or she user-name/data is presented differently on many different places. It is hard to build adversarial intent because

***How can you tell if you are dealing with the same entity in the context of no identity (how to fingerprint who is doing what)?*** – Pam

Chris – it is very hard to trust data without having the fingerprinting.

There is little success on how many users are on a machine based upon their behavior.

\*\*\*\*Fingerprinting:\*\*\*\*\*

***Resolving the behaviors of the data and attaching it to a users/person/entity and machine (and amount of users on a machine)***.

Attaching the data sent to a person

Attaching the behaviors of a person to the data

Attaching the behaviors and person(s) to a machine

Be able to build a history of an entities’ behavior based upon the attachments above.

The entity can be: human, malware, an individual host on a network changing IP address.

Extract the entity *purely* based on its behavior. The username ID’s, IP address, and routing are always changing.

Lowest layer get raw data dump and then get more abstract of data as you go up the hierarchy models of different scales. Built through statistical analysis or humans – So

Want to understand attacks via multiple levels of abstraction.

\*\*\*\*Hierarchal State Models\*\*\*\*

Adapt the research from the area of hierarchical state machines

Attacks are ‘only’ exploiting bugs.

The problem is we are stuck in the foreseeable future of the architecture on the internet that has no accountability.

The Internet architecture is “broken” – John

IP does not have security but people don’t care. The problems are built on an architecture of IP is broken.

***Research: give a coherent critique of the security problems of the current IP system***.

**This community can talk not just about how to adjust the existing architecture but what if/what could be fixed to change these attacks.**

* Have to get rid of IP
* Don’t want physical address of your laptop across the network
  + Your physical address is exposed to the world and a bad guy could easily hack it and pretend to be anywhere in the world

Current attacks are mostly at the software level. The malware attacks are at the browser level. The drive-by downloads are software.

Many attacks work against users that are not administrative of their machine

---- In their architecture you only have limited scope based upon authentication and there are some scopes without any authentication and those scopes are completely separate of other scopes requiring authentication.

The 7-layer OSI model wasn’t based on the attacker model but based on communication

**What is a model for a network that can keep an attacker out. What would be the modification to keep a network secure**.

***What is the model that enables us to use an attack free computing environment?***

Is this possible? Be aware that it is going to be compromised and be better suited to deal with the compromise. Design a system that built with resilience in mind and recoverable. Use a model to design this (need a simulation).

\*\*\*Mixed Initiative\*\*\*

Design models that allow humans to participate in the group and modify model in real time. Models that delegate interaction to the humans but has a risk assessment that will make a decision based upon a lack of human interaction in time and a certain risk value. *Must allow the system to take initiative and make decisions*. Do not completely eliminate the human. Human must decide if this is an attack or is not an attack. *How do you do it so that a human is not creating a bottleneck on data flow*; How to allow for mixed initiative systems that the decision can be delegated to the computer if it is not in made in time; even if the decision is wrong. Have a built in recovery mode.

The system will make decisions based upon whatever model it is following but will defer to a human via feedback or when it can’t make a decision. *Can present a question to a human but will feed back to the model*.

Simulation is a key component to probabilistic modeling. (Monte Carlo simulation)

\*\*\*Scope\*\*\*

In the current internet you do not have a ‘natural’ way to limit scope. So at the physical layer we have NATS and … at the application layer have limited scope.

So that if I subscribe to a scope, if it is the kind of scope that requires authentication, it requires very strong authentication. “If we are doing gay porn we might want to have a less strict scopes” - John. With current DNS we cannot incorporate such scopes. Only people in a scope get the packets from a scope. Scopes must be subscribed to.

Any security mechanism material we build must have a registration and a bureaucracy. Who will administer that scope?

Centralized or non-central authentication/authorized scope – how to implement the scope and authenticity

Model what we are trying to protect.

*What happens when you are your own adversary?*

* can we predict our next adversary

\*\*\*\*Model Combination and Integration\*\*\*\*

Many different researchers, many different models

Many different analysis’s with many different preferences

Which things do you abstract out of model?

Time speed of a model is so different.

“Little things become big things” – Deborah

Two different models, same level of abstraction, same language but when you try to put them together they break in some way. *What has to happen so that the models don’t break apart*? What would the solution look like and what sort of research needs to be done reach that solution.

How would you do that in an automated sense?

DoD in 90’s biggest problem is operator error.

*Need a metamodel of inconsistencies and how to fix it.*

* Simply understanding what you can fix and can’t fix

**Technology that will automatically merge models**.

* Automatic way of resolving meta model

**Fundamentally,** where are all the models coming from? How are you making them? Which models already exist in the cyber realm?

***\*\*\****What is a model\*\*\*

***What are we going to represent with the model?*** Models are a representation of *something*. What is that *something?*

How do we merge two models of different types so that we can make a bigger prediction?

*So what matters and how do you know?*

What are the different protocols that you are going to use?

How to represent non-artificial artifacts (behavior)?

*If we can’t describe it we will never be able to model it*.

Can we model the behaviors of a compromised system? How do you know you’ve been had? **Can you describe a non-compromised system**?

Define the arguments that define those behaviors and the anomalies are when the evidence is collected that are not consistent with the claim.

Understand what different groups/stakeholders want?

\*\*\*Measuring how badly your system is compromised\*\*\*

*What do I do when something is compromised? What is the recourse?* ***Feasible Remediation***

If you do have failures that are localized the goal is to: minimized reliance on similar systems. Identifying the compromise the not the attacker. What is the notional value of what we are trying to protect.

*How much do you value the system that is compromised?*

Need to first understand the characteristics of the person or stakeholder that has the value proposition that has what is normal or proper behavior.

The user is able to separate value from non-value.

*Be able to model the analyst as well as the enemy.*

*Be able to model the mission (what we want to protect).*

Modeling the basis’s of the analyst

Tell them what they need to know when they are making the wrong decision.

Modeling the behavior of: children, drunk, people that are making bad decisions because they are upset; while they are on the net.

Authorization of the user towards the machine depends not just on their ‘rights’ (admin rights) but their state (mental state, age, etc.).

Need a normalized behavior model.

*Value proposition problem: Value vs. Risk*

Really fast flexible code that we can verify – goal

Need to be able to ‘corner solution’. Lock the code down

*Click this button to see what would happen if you do this. And then a VM opens up to demonstrate what happens*. (‘Sandbox mode’)

* Virus downloads and runs your computer slow
* You see you credit card number get sent off

No market for this software

Market is for features, free music, free…

*Security is an “economic” problem.* – the market won’t pay for it

A safe mode would be too technichal for the average person using the computer. Any kind of sand box mode would require too much assumption of the knowledge of the user OR too much estimation from the machine. Unless the machine can interpret *intent*. – ET

*Use of dynamic context as a first class object for data and modeling interpretation*.