Four Valuable Data Sources for Network Security Analytics

Table of Contents

SEI WEBINAR SERIES | Keeping you informed of the latest solutions 3

Carnegie Mellon University 3

Copyright 2017 Carnegie Mellon University 4

Four Valuable Data Sources for Network Security Analytics 4

Overview 6

Data and Process Description 7

Polling Question 1 8

Domain Resolution Data 9

Network Device Inventory/Configuration Data 11

Network Flow Data 13

Network Intrusion Detection/Prevention Alerts 14

Network Flow Data 15

Network Intrusion Detection/Prevention Alerts 16

Process 17

Explore 18

Model 19

Test 20

Analyze 22

Refine 23

Polling Question 2 25

Analytic Examples 27

Example: Co-located Generated Domains 28

Example: Assessing Patch Efficiency 33

Example: Quantifying Vulnerability Exposure 36

Understanding and Improving Security 39

Understanding and Improving 41

Contact Information 47

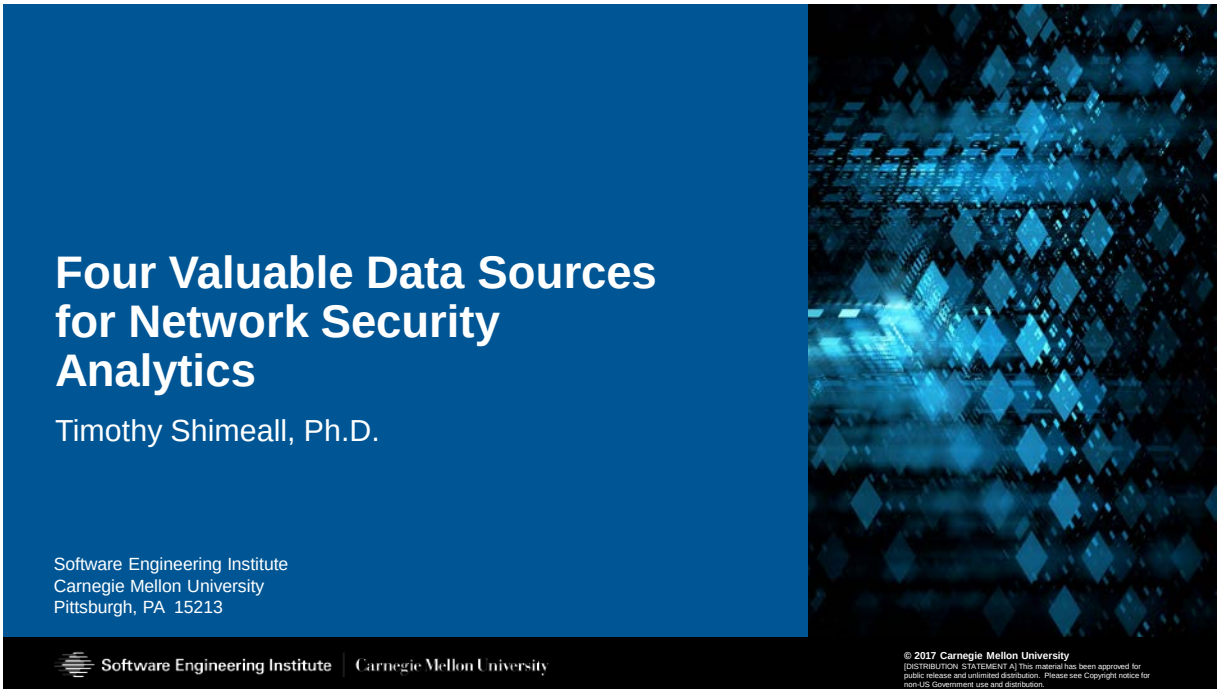
SEI WEBINAR SERIES | Keeping you informed of the latest solutions 48

SEI WEBINAR SERIES | Keeping you informed of the latest solutions

Carnegie Mellon University

Carnegie Mellon University

Four Valuable Data Sources for Network Security Analytics



**004 Presenter: And hello from the campus of Carnegie Mellon University in Pittsburgh, Pennsylvania. We welcome you to

Virtual SEI. Virtual SEI is our new streaming platform, where you can watch live events or access on-demand videos discussing our cybersecurity and software engineering research and best practices. Our presentation today is Four Valuable Data Sources for Network Security Analytics by Dr. Tim Shimeall. And depending on your location, we wish you a good morning, a good afternoon, or a good evening. My name is Shane McGraw, your moderator for the presentation. And I'd like to thank you for attending.

We want to make today as interactive as possible. So, we will address questions throughout the presentation and again at the end of the talk. And you can submit those questions to our event staff at any time by using the Q and A or chat tabs on the page interface. Now, we will also ask a few polling questions throughout the presentation. They will only appear as a slide in your video window on the screen. We ask that you type your response into the chat tab as we go along.

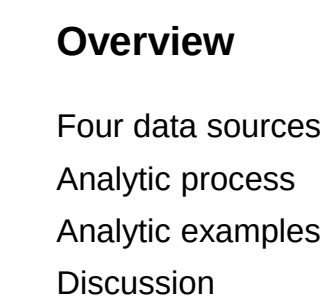
Lastly, a link to a PDF copy of today's slides are in the chat area now. And we ask that you fill out the survey upon exiting today's event. That link will be added to the chat, as well. And we appreciate your feedback. For those of you using Twitter, be sure to follow @SEINews and use the hashtag seiwebinar.

Now, I'd like to introduce our presenter for today. Dr. Timothy

Shimeall received a PhD in information and computer science from the University of California. Since 1999, Tim has served at the SEI, currently with the CERT Situational Awareness group. He is responsible for the development and analysis methods in cybersecurity at and above the enterprise level. Tim, welcome, all yours.

Presenter: Thank you. And today, we're talking about four valuable data sources. These are certainly not all of the data that could be available for network analytics. But we're going to be focusing on ones that are available, particularly applicable to network behavior at and above the enterprise level. So, we're looking at these four data sources as examples.

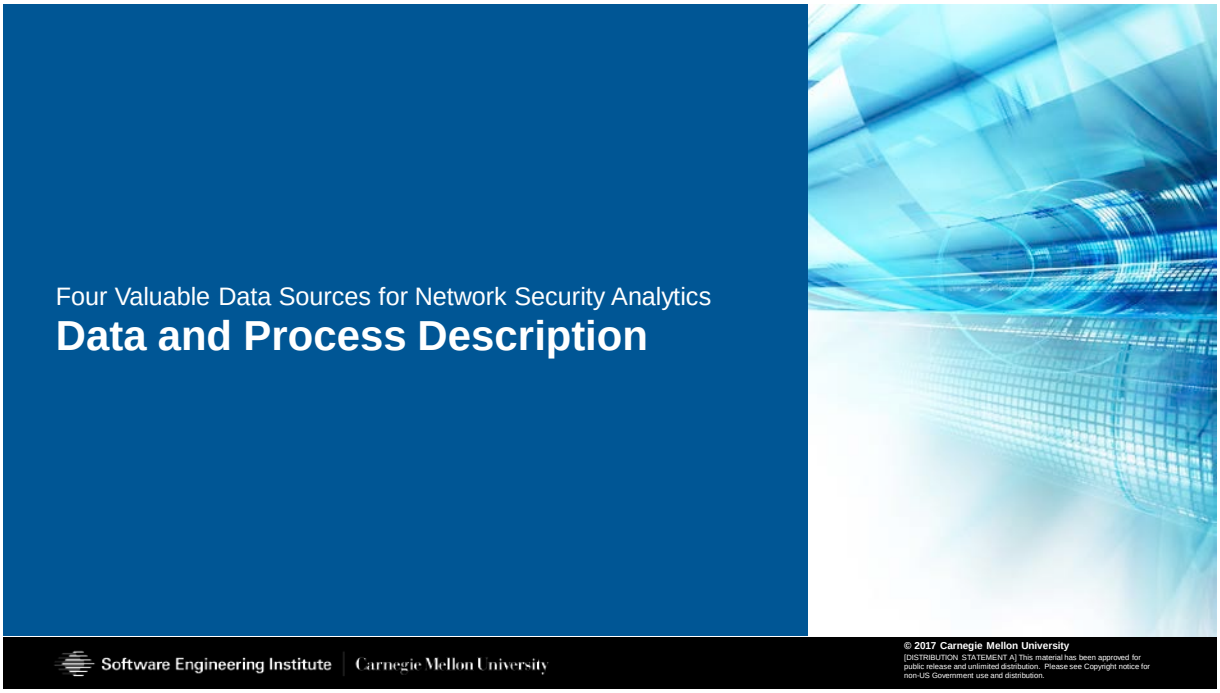
Overview



**005 The talk today is going to

look at what are these four data sources that we're presenting as examples. What's an analysis process for taking the data and moving it into information that could be used to base decisions on, to detect illegal behavior, or to understand the ramifications of putting controls in place. We'll be presenting some analytic examples. And we'll finish off with a discussion on how to use this information to improve cybersecurity within organizations.

Data and Process Description

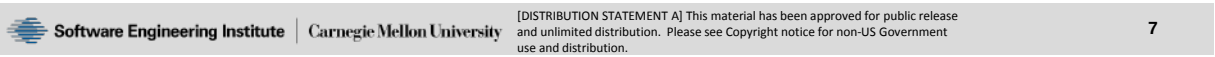


**006 We're going to start off with a data and process description talking about specifically what are these data sources that we're moving with. But before we do that--

Polling Question 1

What information sources do your organization use to inform network security?

a. Mostly intrusion detection/prevention alerts
b. Mostly network packet monitoring
c. Mostly network flow collection (or traffic traces)
d. Mostly third-party reports (Threat intelligence)
e. Mostly vulnerability scanning
f. Mostly host-based logs
g. A balanced mix of sources



**007 Here's a brief polling question.

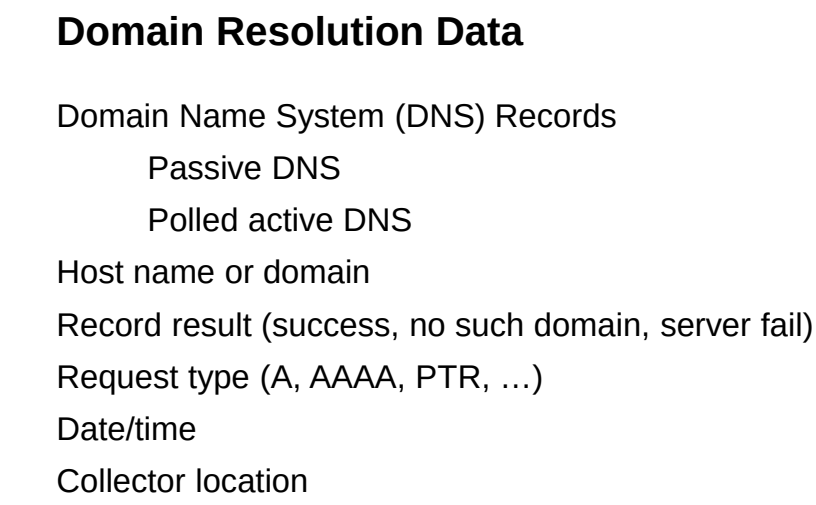
Presenter: So, that first question we wanted to ask is, "What information sources do your organization use to inform network security?" So, we ask that you just type that into your chat. Once again, the question will appear on your screen here in a second. That question is, "What information sources does your organization use to inform network security?" And you'll see A through G. If you just type in the letter of your response, I will tally that way. So, we'll give you about twenty or thirty seconds to see what the options are and respond. And just to remind everybody, today's event is being archived. Anybody can watch the recording as early as this evening just by coming back to Virtual SEI. You can access that archive. A

reminder, if you're just joining us now, a copy of today's presentation slides are available at the link within the chat window. They'll take you back to the SEI site where you can download.

And I'll start looking at some of the responses. So, we've got a number of Gs, Tim, which is-- that's the overwhelming mix so far is a balanced mix of sources. That seems to be the overwhelming favorite.

Presenter: All right.

Domain Resolution Data



**008 So, in considering those kinds of things, let's talk about the four that we're ready to talk about. But we recognize, you guys may-- you folks may be using a-- data sources that go beyond this, or you may not be using all of these data sources.

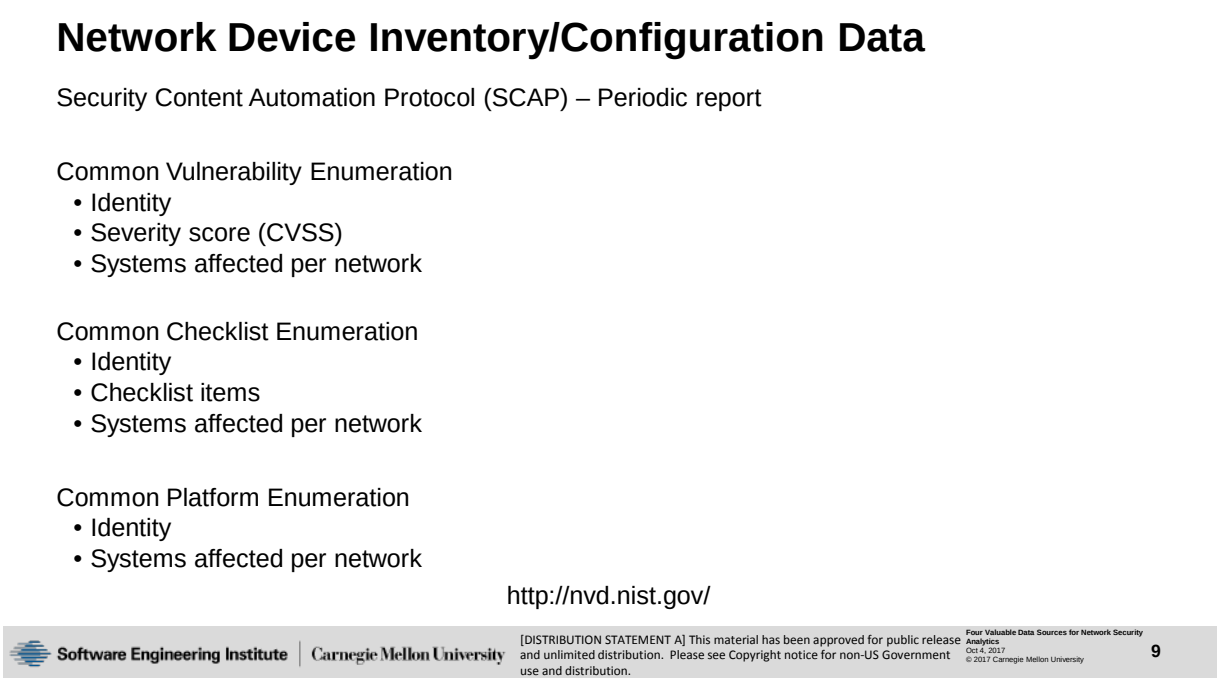
The first one I want to go into is domain resolution data. This is conventional data presented across your network in order to link services to hosts and those kinds of things. And the basic content you're looking for at this point is some understanding of what hosts you're reaching out to and where you're accessing those hosts in terms of network address. In some cases, you're looking for more information such as through what resolver or through what mail exchanger.

But there's two basic ways that organizations collect this at an enterprise level. The most common one we see is passive DNS collection, where essentially there are network traffic monitors. And they are searching for either outgoing requests, or more commonly, incoming responses, and looking at the results coming in those incoming responses, and archiving those for later analysis and processing, summarization. This includes recording what host name or domain is being queried for or is coming in place, what's the result if there's a success, what address is associated with the hostname or domain. If it's a failure, is the failure because there is no such domain or because the server that would handle the domain failed to respond? We also typically archive the request type. That is is it a simple name to address request, or is it an IPv6 named address request, or is it the inverse request where you have and address, and you're looking for the name, or any of the other

variety of DNS requests and responses that can come in place.

And then along with that, in order to understand the flow of behavior, you're going for date and time that's in place. But again, what you're pulling out of all of this information is some understanding of what endpoints you're talking to and where those endpoints are being addressed to in terms of Internet address space.

Network Device Inventory/Configuration Data



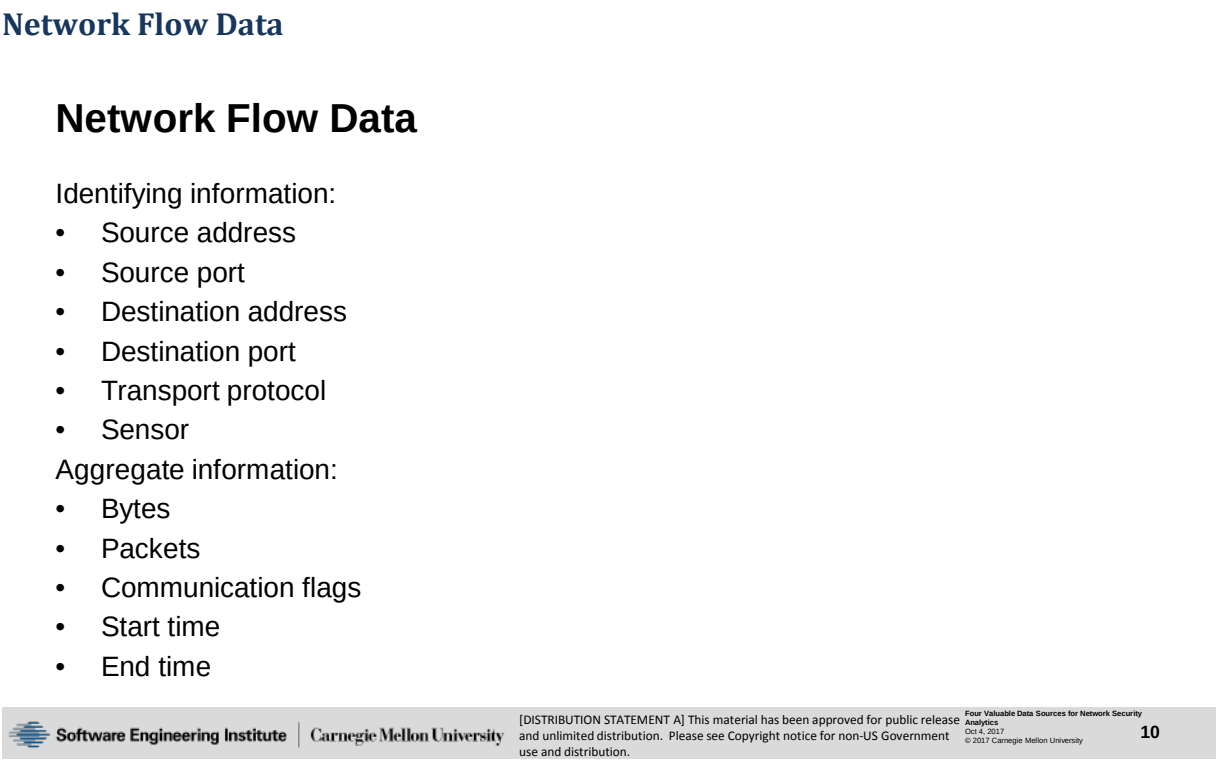
**009 In terms of your own local systems, it's frequently useful to archive as accurate a network inventory as you can. There are variety of approaches for doing this in terms of traffic recognition protocols and so forth. But one protocol for collecting this information developed originally by MITRE and now promulgated through

NIST is the Security Content Automation Protocol. And the basic--this pulls information off of network management systems or security scanners that are associated with your network and archives that information in summary format through a periodic report.

The particular information that typically is part of an SCAP report would be common vulnerability enumeration information, where the identity of particular vulnerabilities and their severity is recorded along with how many systems on the network appear to be vulnerable to this particular vulnerability, security checklist enumeration where there is some specific checklist for making sure key security controls are being enabled, and what items are included within that checklist, and, again, how many systems that checklist has been applied against or have been identified as being in compliance to.

Finally, common platform enumeration specifies what software and hardware configurations are present and how many of those are present across the network. And the goal here is to try and understand not just what's a snapshot of the state of our network at one point, but also, over time, which direction is the network behavior shifting. Applied at an enterprise scale, this is sometimes a challenge to keep track of. But it's useful in terms of making security decisions, and understanding what controls could be applied where.

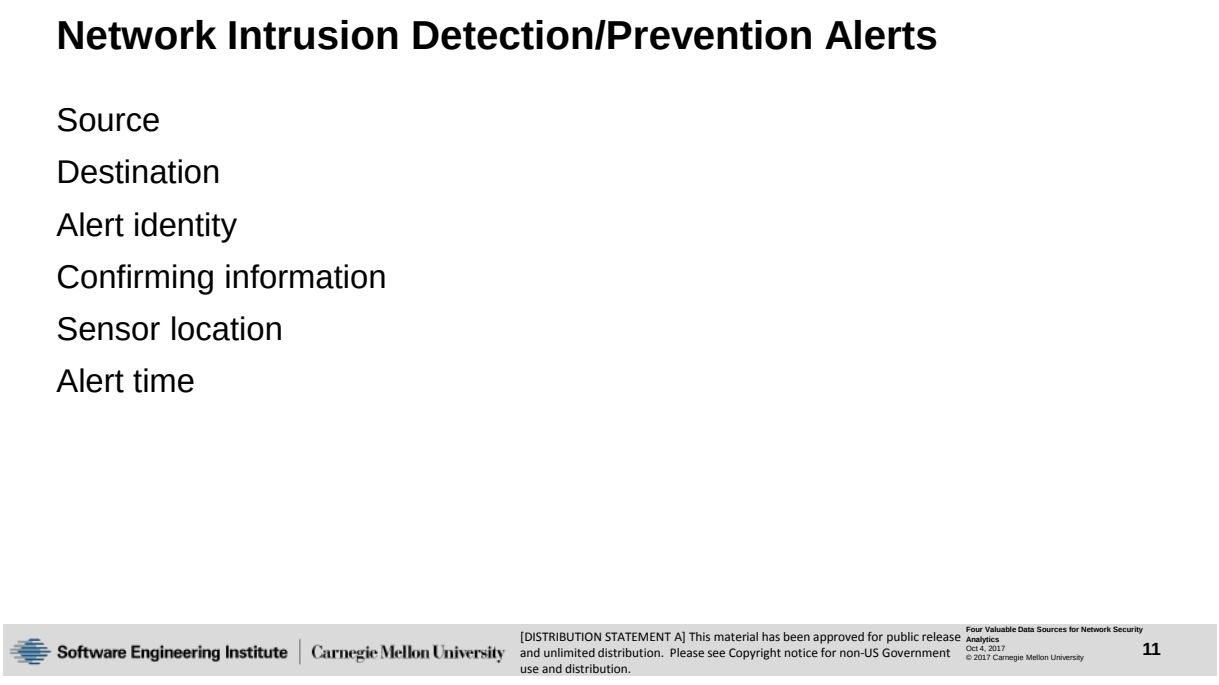
Network Flow Data



**010 Generally, you want to also keep track of the network traffic. And while this can be done at the packet level, or in packet traces, which basically snap the header content and leave off the packet bodies, we find if more useful to look at the network flow level. The network flow is aggregated packet header information. It produces much more concise summaries. And across large enterprises, that conciseness allows you to have very practical network traffic collection. This includes things like okay, what systems are talking to which, client server or source destination, depending on the kind of flow data collection that is going on, across which traffic protocols, at what time, through which sensors are being collected.

And the a long-- then for each connection, how many bytes, how many packets, what communication flags, possibly what applications are detected as of running through packet communication? All of that information being archived together to try and pull together a clear understanding of what traffic is moving across your network, as well as the previous thing, what end points are you communicating with and what configurations you're dealing with on your network.

Network Intrusion Detection/Prevention Alerts

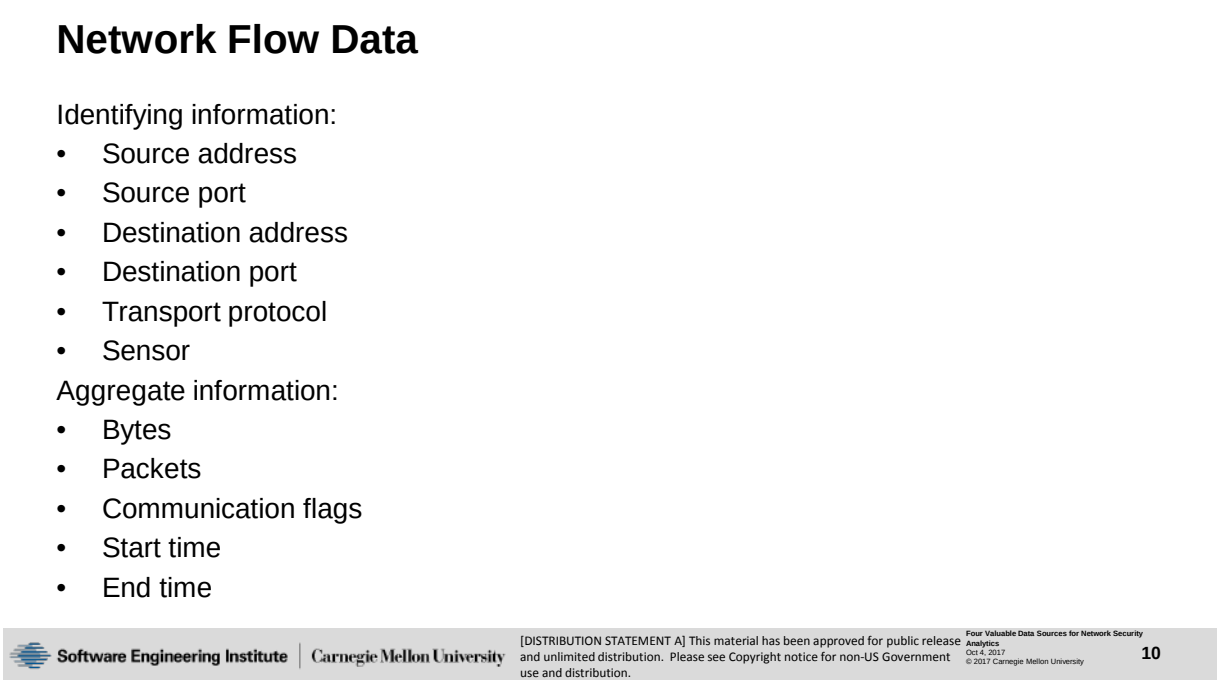


**011 Presenter: So, Tim, we had a relevant question to this actually from Tom asking, "What are some of the ways that intruders evade being tracked when they're in your network?"

Presenter: Well, that's a very timely question for this particular slide.

Often, many organizations us their intrusion detection and prevention systems as very key information sources for triggering security protections. And the difficulty is the intruders know this. And they will frequently tune their attacks to evade common IDS signatures or to deal with those kinds of things which is why having--

Network Flow Data



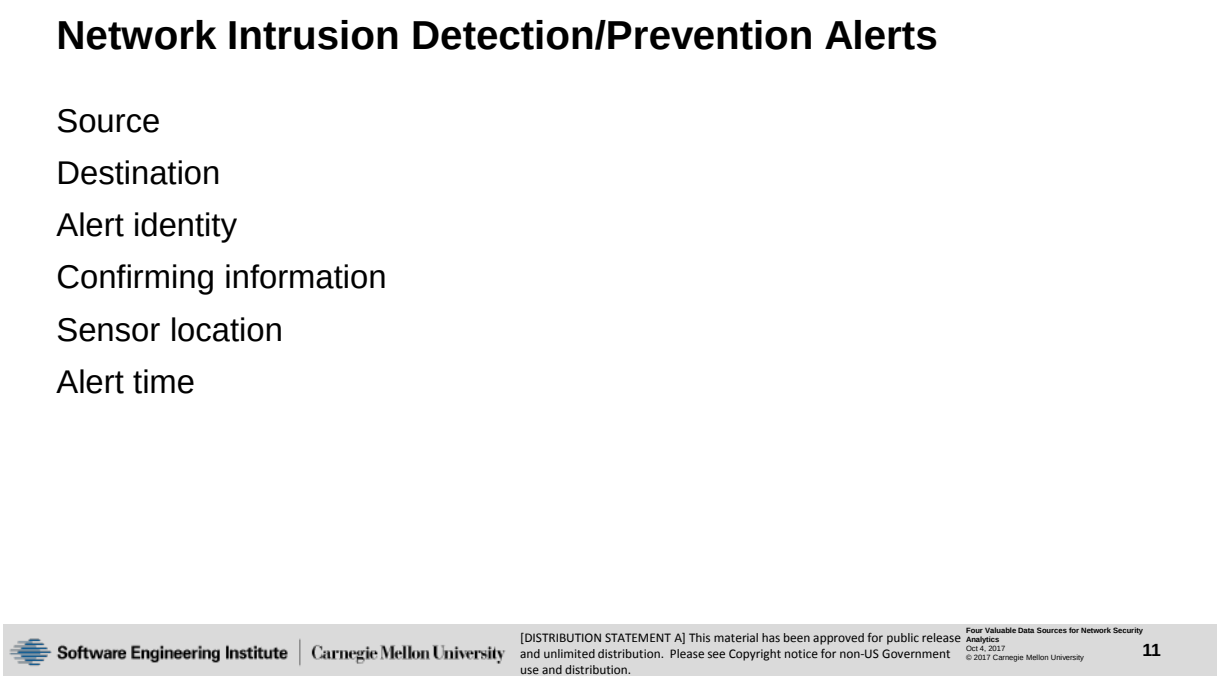
**010 Something like a network flow data solution where you can capture all of the traffic moving across can be useful in spotting attempts to evade or spotting unusual traffic patterns that might indicate such evasion is in place. The other thing that intrudes typically do is they bury their attacks among the common communication.

So, the attempt is to pass under the radar either by pacing your attack

very slowly or doing those kinds of things. And the attackers frequently need to deal with things at that level. Or the defenders frequently need to deal with attempts to evade at that level.

Presenter: Great, thank you.

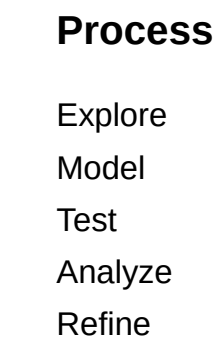
Network Intrusion Detection/Prevention Alerts



**011 Presenter: Moving back to the intrusion detection systems, and saying that attackers sometimes evade them, doesn't mean they're--certainly does not mean that they're useless. They are often very important confirming information to be applied. And again, you've got source and destination information. You may have a specific alert that is being triggered. You may have some packet content to give context to that alert or other confirming information that's there. And certainly, also

knowing where it was detected and at what time it was detected are very key factors for determining how you can blend together information sources. If you'll notice, virtually all of these information sources have some degree of time location. And that's because network behavior is inherently dynamic.

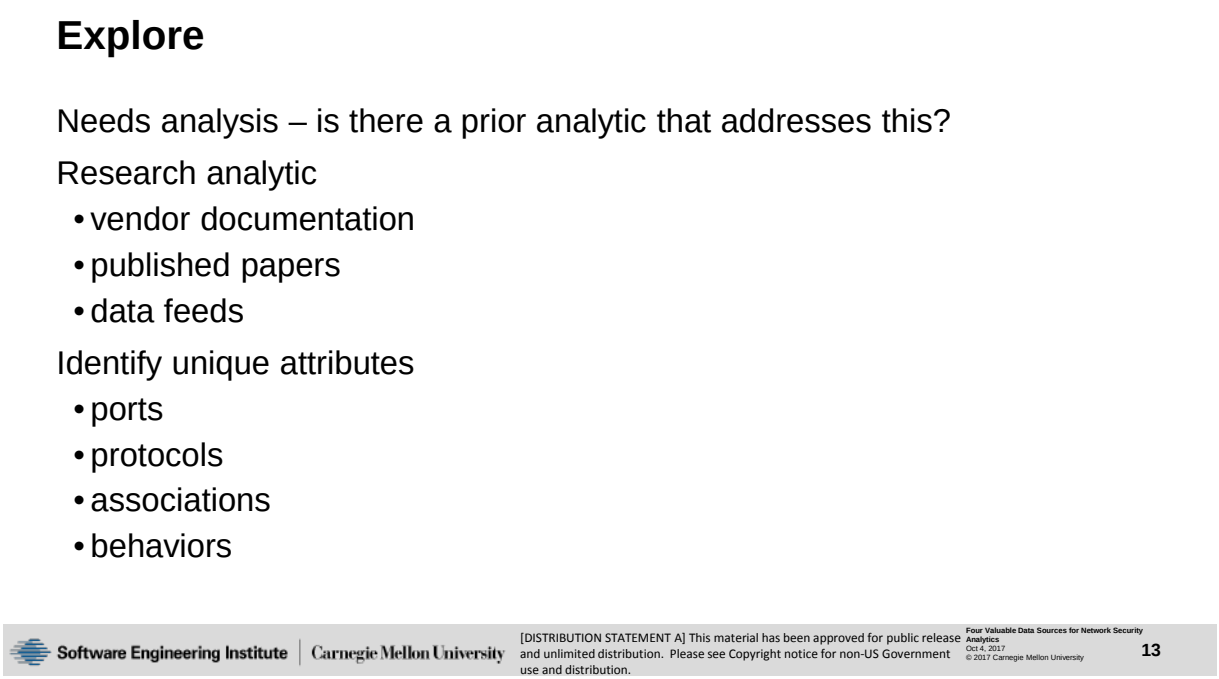
Process



**012 So, let's move from okay, what are these four data sources to more how we're going to use them. And talking about how we're going to use them, we're going to recap a process that was introduced at a webinar last year. And that is a five-step model of exploring the attribute or the behavior that is wanting to be detected, modeling it both in terms of a theoretical model effectiveness and also an executable model to allow you to actually run the behavior

against it, then applying it in a test environment either against capture of real data or against manufactured data. Then you analyze the test results, and you refine the analytic, repeating test, analyze, refine as often as you can in order to get good detection characteristics.

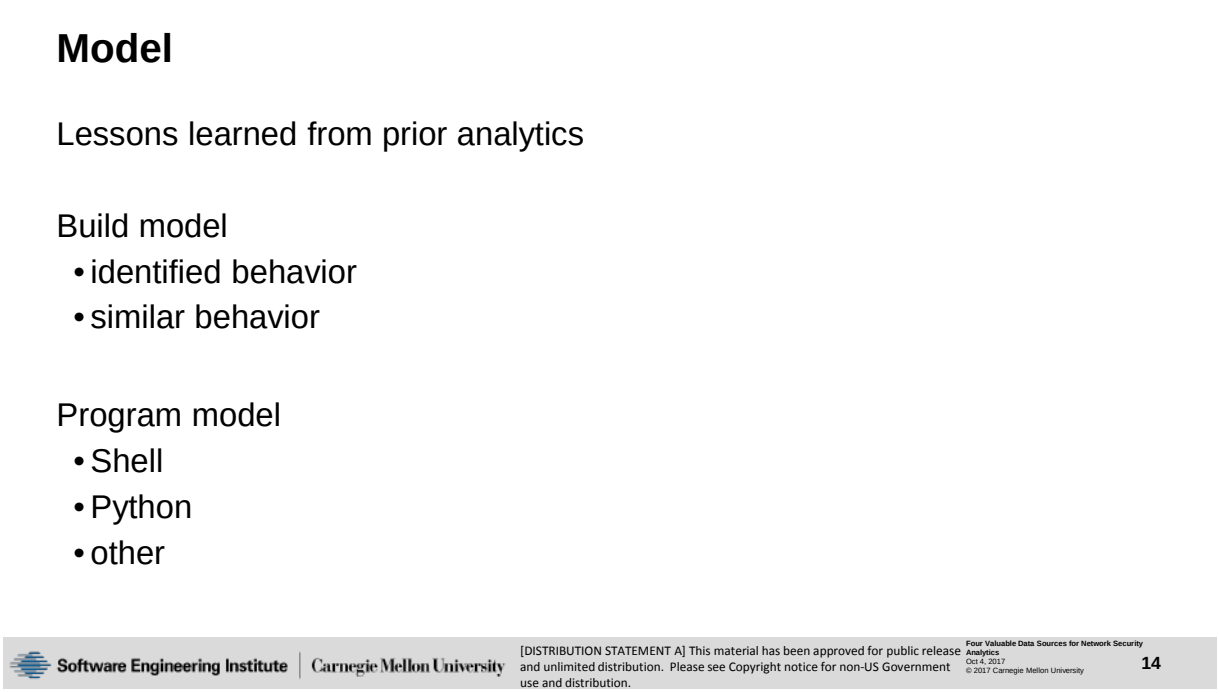
Explore



**013 So, the explore step, this is where you do needs analysis. What's going on? What do we need to detect? What-- is it illegal behavior? Is it changes in configuration or manipulation? Is it simply understanding what the normal behavior or configuration of the network is, so we can start spotting abnormalities? And then the question becomes okay, how is this already being done. Is there some prior analytic, and if so, what's weak about this analytic? Prior analytic doesn't

necessarily mean just an analytic you have in-house. It can be an analytic which is developed elsewhere. So, you may have some research looking at security vendors, looking at published papers, looking at threat feeds and other data feeds that are there, and identifying what attributes of all of this information you should focus in as your trying to deal with--develop your analytic.

Model

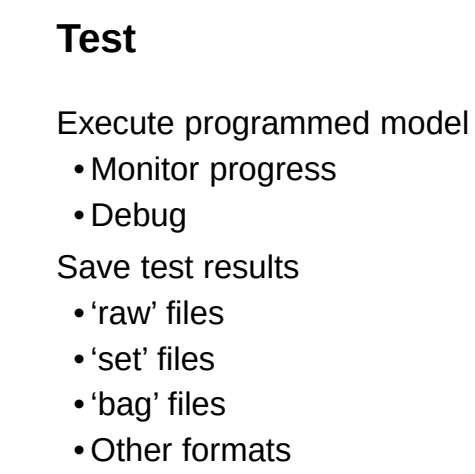


**014 Once we've got those prior analytics, we build off the lessons learned of them and start building a model. We want to model both what are the behaviors or attributes that we really want to measure here, and what are similar behaviors or similar attributes that either might be confirming factors or confusing factors. If they're confusing factors, frequently, we want to eliminate

them. If they're confirming factors, we want to correlate them with what is detected as a quality check.

And then we're looking at building an executable model, program it. Most commonly, you want some type of interpretive language here, frequently Shell or Python. Sometimes you'll use other approaches, particularly compiled languages if the volume is simply too large. Occasionally, you'll see some analytics, which are actually enabled by hardware. But that's unusual because the pace of change of network attack and defense argues against doing these kinds of things at the hardware level.

Test



**015 Once we have the analytical-- the programmed model, now we want to try it out. And we're trying it out typically against-- at least initially,

against captured data or manufactured data to try and see how well it behaves, how it handles edge cases. We'll monitor the processing that is inherent in the analytic and looking for places where it goes wrong, in which case, we debug it where it is slow or excessively computing bound. In which case we'll search for more efficient solutions.

And as we go through the test, frequently, we're not interested in a Boolean pass/fail. We're interested in understanding the overall flow with the behavior of the analytic. And that means saving around what test results you can, either raw test results, or summarized test results either as sets, or as multi-sets, or sets with counts, which would be bags. Sometimes, you're looking at archiving traffic that is identified for more detailed examination of what's going on.

Analyze

Review test results.

Reduce false positives.

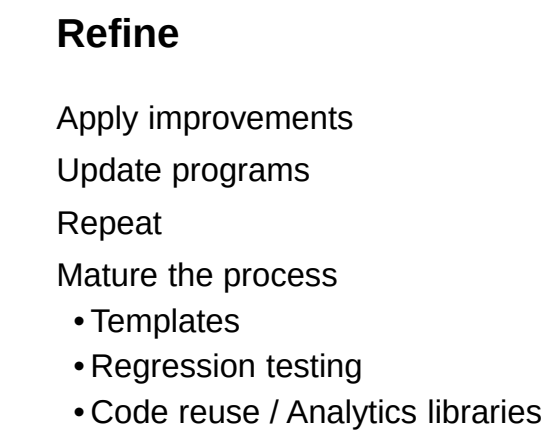
Reduce false negatives.

Identify improvements.

**016 And then as we have these results, the next step is to review them. You want to very carefully identify, particularly in your test data, set what are false positives and what are false negatives. False positives, something's being alerted that should not be alerted on. It's being identified as behavior that it's not. And that's often a problem for security teams because they-- it wastes their time. They spend time chasing down something that isn't real. They also want to reduce false negatives. False negatives indicate blind spots in the execution process. And that is also a problem. Missing things that you need to detect means your organization is left more vulnerable than it should be. And the other feature here is, again, where we can improve performance, where we can improve data summary to make the

results more actionable. We want to go ahead and analyze the results to see where we can do that.

Refine



**017 And once we've identified the places where we feel like they need change, then there's the refine step. The refine step you examine the behaviors. You identify possible courses of action to improve them, evaluate them to come up to the ones that are selected, update the programs. And then you go through a repetition process where you're testing, analyzing, refining, testing, analyzing, refining, until you've got the rates at about where you want it to be. I'm seeing results. The results are real. The results are presented in a manageable fashion.

As we go through the refine step, this is the phase at which process

improvement can also take place. Process improvement is a very classic SEI thing. Here we're looking at process improvement in terms of can we build reusable templates to really understand common situations and common summaries that are in place. Can we reuse test sets both to understand what behaviors are being done, and so we can more easily compare the results that we get from different analytics? And also, the classic code reuse, can we build up libraries of mini-analytics or small common analytic steps and enable our analytics to be dealt with much quicker and much more efficiently and generally of higher quality through that means?

As we go through this five-step model, we're building analytics. In this particular presentation, we're going to be looking at several analytics. And we're going to be discussing how we go through those five steps to combine the data in ways that give enterprises more information and more useful combined information on which to make security decisions.

Polling Question 2

What is a difficult step for your organization in developing security analytics?

a. Getting dependable data
b. Handling large data volumes
c. Turning data into behavior observations
d. Prioritizing significance on behavior observations
e. Matching behaviors with threats
f. Automating the process with the tools available
g. Communicating efficiently with management

**018 Presenter: Okay this leads us to our second polling question, which you'll see on your screen in a moment. And that question we'd like to know is, "What is a difficult step for your organization in developing your security analytics? What is a difficult step for your organization in developing security analytics?" And again, we're just looking for the letter reply there. And then we'll tally that. While we give you about thirty seconds to vote, a question from Kim came in, Tim, asking, "Which parts of this model can be the most difficult for organizations when building analytics?"

Presenter: Often the most difficult part is the initial explore step. It's more difficult because typically, the analyst will assume oh, I understand the problem and not spend enough

time searching for solutions and clearly establishing the need. The fact that there is a needs analysis that needs to take place, often an abbreviated one, but a needs analysis to take place is often a big challenge for analysts to understand and for managers to guide the analysts in developing their analytics.

Presenter: Okay. A couple of responses so far, we've got two As, a G, an F. So, again, with a diverse audience, we're going to get some different questions here. Let's work on another question form Joseph asking, "Which are difficulties that organizations need to handle in analysis?"

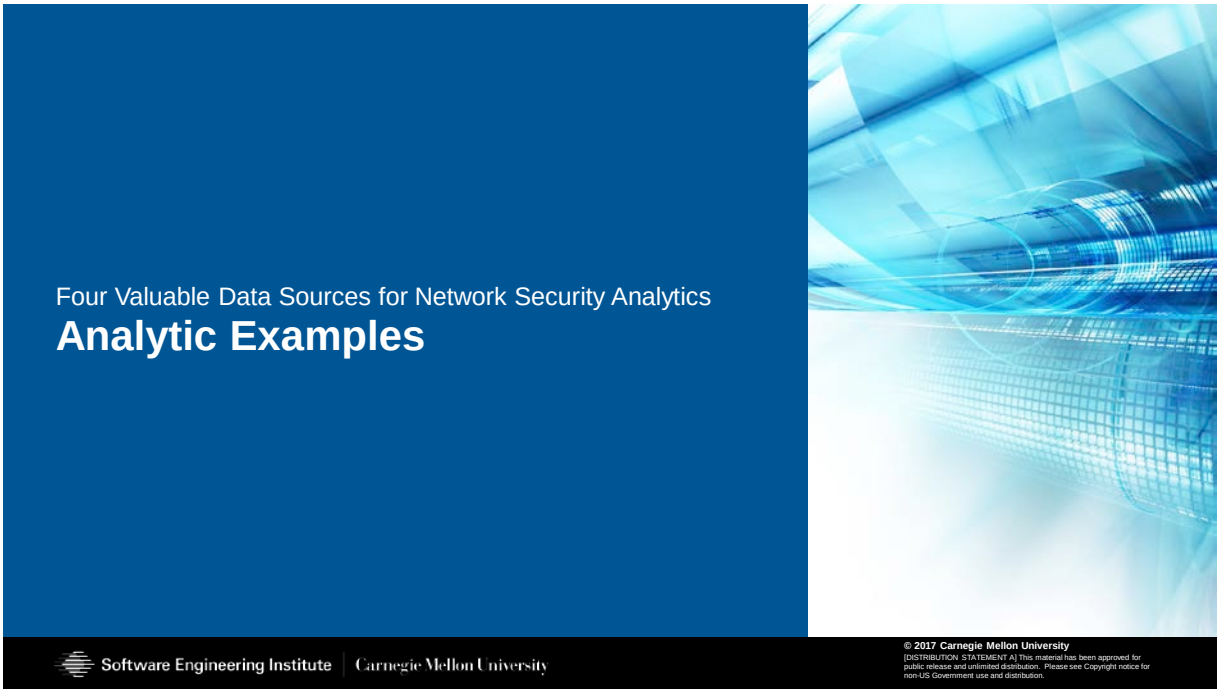
Presenter: Often the most difficult one is A, getting dependable data. Being able to understand where do I need to put sensors in place in order to get a good snapshot and a good ongoing trend. If I place it too close to an endpoint, I may not get the summary view that I want. If I place it outside security zones, often I will find that I'm abstracting away the very behaviors I want to see.

Presenter: Okay another couple responses, a couple more Gs, a couple-- a C. So, we're going to turn it back to you.

Presenter: So, it sounds like what I'm seeing from the polling question is some organizations have the pain point of how do I get the dependable data so I can deal with. Others are looking at automating and

communicating and really seeing the pain point there. And often communicating efficiently with management basically involves giving enough context, putting-- not just giving them numbers or a quick assessment, but giving them some background in which we can move things forward effectively.

Analytic Examples



**019 So, let's walk through some of these examples, and I'll try and hit some of the ways that they have been used to communicate effectively with management.

Example: Co-located Generated Domains

Example: Co-located Generated Domains

Explore: Scan DNS queries for computer-generated domains (several algorithms), couple with network flow data looking for propagation attempts, and intrusion detection alerts for compromise attempts.

Model: Identify timeframe from DNS queries. Identify sources from IDS alerts or from scanning / service probe detection through network flow.

Test: Apply at several scattered points throughout day (early workday, morning peak, noon, afternoon peak, end-of-workday, late evening). Watch for recurring sites and ongoing activity

Analyze: Correlate against third party reporting; remove contracted or internal security scanning

Refine: Revise model to improve throughput and to whitelist sources

**020 So, this initial example is looking at coupling DNS information with network flow information and intrusion detection information. The goal here is you want to be able to find computer generated domains often because a lot of malware will use that-- those computer-generated domains for command and control, or for coordination among network threats. What we're often looking-- what we're also looking for is attempts to spot domains that we may not detect initially but could be applied. So, if there's a common location that is used in several different domains, we might look at other domains which resolve to the same address and pull things forward. But we only really want to do that if we have some assurance that this is a computer-generated domain that's really associated with

behavior that we could-- that could indicate malware.

So, the initial explorer concept was okay, let's scan for DNS queries looking at domains that don't look like they were constructed by humans, fleabag.com or SEI.cmu.edu, but rather things which look to be computer generated, XX, YY, PG3, or some attempt at taking pronounceable syllables and slamming them together, or other sorts of algorithms that are typically used. Then once we've identified the resolution of such domains, now we can start looking and saying okay, at the timeframe which those resolutions were taking place, what was the network doing. Where-- was there propagation going across there? Were the intrusion detection alerts being found at that point? Can we couple the generation of these domains with other indications that there's been compromises on our networks and be able to identify them? So, the model then becomes okay, from the DNS queries, we can establish a timeframe, often a fairly imprecise timeframe, but a timeframe of what's going on. So, we'll want to look both a little before and a little after that timeframe. And then look at the IDS alerts or models with-- from the information flow that detects scanning or service probes and couple those sources with the timeframe to try and identify okay, these sets of queries were likely computer-generated and potentially malicious.

Great, good concept, let's see how well it works. We took this information. We applied it at several scattered points throughout a day, looking at an early workday snapshot, a morning peak, a noon, an afternoon peak, and end of workday, and a late evening. The reason for selecting several of those scattered points is two-fold. One, intruders attack at different parts of your work cycle. They will often attack-- attempt to hide in your traffic during peak times. So, looking at morning peak and afternoon peak, those are often very key attempts that intruders have in evading behavior. Looking at off peak times, the early workday, the late workday, the noon, the late evening, one, the intruders may not be working on your work schedule. They're working on their own. So, it may be decoupled from your timeframe. The other aspect is they may be trying to hit when they think you're not watching, when you're out to lunch, literally, or when you are off during the evening. So, you want to watch for the scheduled things. You want to look at to see okay, what's being detected when, are there recurring sites, is there recurring activity, is there activity that seems to be not coupled with the workday or coupled odd with the workday. And then we pull out of that co-located domains, identifying those kinds of things.

For analyzing this, well computer-generated domains and misbehaving domain servers are something you can get from a threat feed. So, there

may be third party reporting that you can pull in place to bring things-- to help refine that information. On the other hand, you may find that the co-generated domains were part of internal security scanning, part of contracted network security services, part of ephemeral communication services and would not be specifically maligned. So, you look for ways to say okay, these are things that we can confirm and are likely to be associated. These are things which are probably distractions and we want to minimize them.

And then we apply the refinement model to improve the throughput, where we do know of specific sources, which would be authorized co-located domains. We might whitelist those off and exclude them from further analysis.

So, here's an example of that five-step process working through combining several data sources, generating results that are more clear than what you could get from a single data source on its own. Even if you're just looking at just the DNS query information, you could spot that they were querying for computer-generated domains. But it's difficult to make an assessment that it's a malicious domain without other data sources to give you information on what's going on on the network.

Presenter: And tracking all this stuff, how many people does it take to do this stuff? Are these a large

staff of people monitoring this stuff? Is this tools that are being done?

Presenter: Often, it's not that large a staff. Often, one, to develop a good analytic of this sort there are enough- this is enough information out there where you can have a very modest staff, one to three individuals, that are developing these analytics, and then applying them across your network data within your enterprise. Interpreting the results of the domain will depend on how rapidly you find them. Initially, it may be quite large, and you're farming it out to the teams that are associated with different parts of your organization. In other cases, once you've matured this process and eliminated some of the immediately responding threats, there you're looking at a much more modest workload and one that your enterprise security team might easily be able to continue.

Presenter: Great.

Example: Assessing Patch Efficiency

Example: Assessing Patch Efficiency

Explore: Patch efficiency – mitigations are applied for significant (serious and exploitable) vulnerabilities prior to exploitation. Couple detected responses to scanning against reported vulnerability patching and detected changes in behavior.

Model: Use inventory data to identify decreasing vulnerability, then query network flow data for responses to identified scans on previously-vulnerable services. Apply more in-depth flow analysis to profile and contrast service behavior before and after patching.

Test: Apply for very common services (web, email, DNS) and less common (database, file transfer) services.

Analyze: Compare changes in behavior for patched services against those for non-patched services.

Refine: Revise model to distinguish significant vs. coincidental changes.

Campbell, G. *"MEASURES and METRICS In CORPORATE SECURITY." Security Executive Council Publication Series.* January 2008.

**021 Presenter: Another issue that organizations face, there are many, many patches which are being issued, both for network services and for common applications. And the question is how well are we keeping up with our patches. How much are we patching responsibly versus. how much are we patching proactively and really protecting our organization? And that really rings up the idea of patch efficiency. This is not one that you can find from a single data source, much like identifying co-located malicious domains. Instead, you need to look for the patch-- the mitigations applied to significant vulnerabilities and couple that with some indication of both how often are exploits being attempted and if there have been any successful exploits around the time interval when we're deploying patches.

Well, the patching information is often pulled from your network inventory information. You can assess this information from the changing rates at which vulnerabilities are applied to systems. And you can look at those kinds of things. The severity frequently is also available in your network inventory systems through the CVSS data as we apply it. The-- finding scanning, finding changes in behavior, that's where you typically will be interested in bringing on network flow or network intrusion detection information or dedicated scan detection information to bring in place. So, we know we've got data sources that are potential for here. The model, then, would be okay, let's use the inventory data to identify vulnerabilities that are decreasing. So, that classifies where the patching is taking place and what's going on in that. Then we want to look at a time window around the patch report, the report of decreasing vulnerability, to determine had there been attempts at exploiting it before we started patching, have there been attempts during the patching, or have-- and of those attempts, how many are likely to have been successful, how many are likely to be somewhat problematic.

You can then apply more in-depth analytics and enable more in-depth analytics to really focus in on what's the behavior for the particular effected network services and how does that behavior change, both before and after patching. So, are they revealing changes that indicate

the patch was successful and blocked things that were in place?

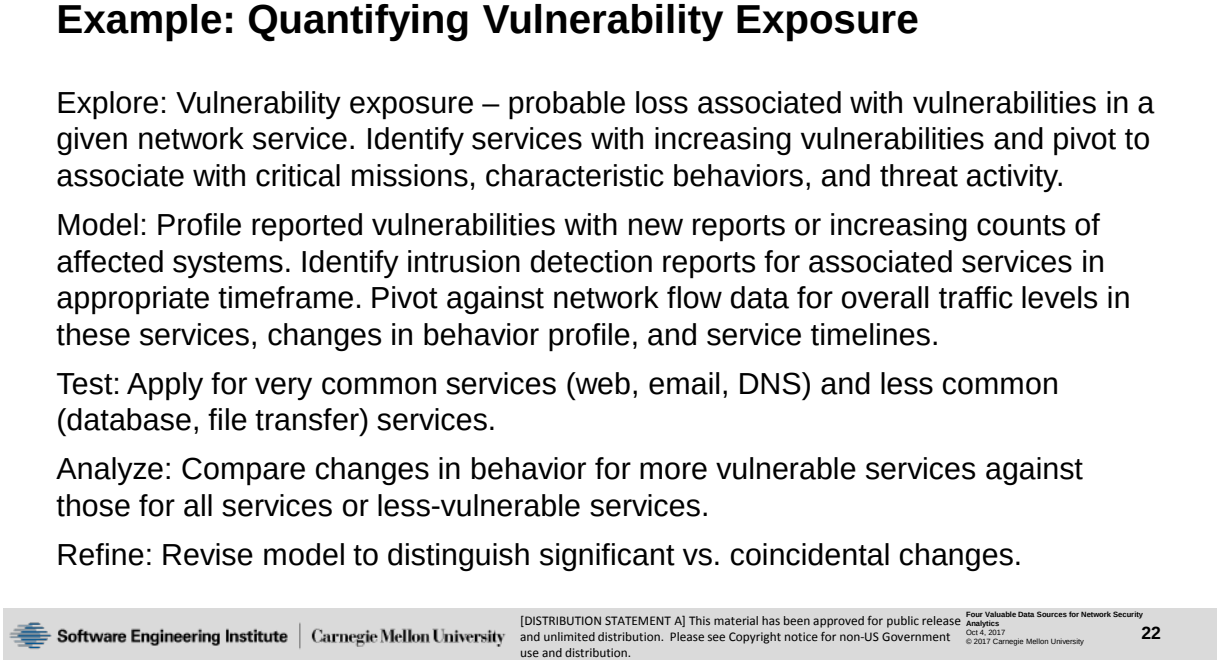
Great, now we want to test it. So, let's test it against both very common services, and the big three are very typically web, email, and domain name-- DNS. Sometimes, if you're doing a lot of remote terminal access, you want to look at VPN behavior or SSH behavior as well. You may also have some less common services. You may have critical database services or critical file transfer services that you need to see how well are applied. Which services you're going to test against depends on what the configuration of services for your particular network actually is.

And then in analysis, we want to compare-- we want to examine the changes in behavior in the patch systems, and again, matching patch systems by looking at the vulnerabilities, identifying the services, pairing up the services against the servers. And as we're going through that, look for changes in behavior that indicate the patch itself is changing behavior versus the patch is really effective in preventing vulnerability. And then we want to revise the model to specifically focus in on those that are significant changes with respect to fixing a vulnerability versus those that are coincidental changes with applying any change to the system and move things forward there.

This particular analytic actually came from a third party source. And the

source is there cited on the screen. It's a fairly interesting work on doing models and metrics across-- in computer security across the enterprise.

Example: Quantifying Vulnerability Exposure



**022 Third example, and again, it's the explore, model, test, analyze, and refine model. The question here is what's our level of vulnerability exposure. Where previously, we looked at seeing where are we being effective and how effective are we being, in this case, we're looking at where are we being ineffective. Where are we seeing increasing vulnerability? Where are we seeing increasing exposure? And often, the kinds of decisions you want to reach here are what kinds of configuration changes do we need to make. Do we need to reduce the number of servers fielding a particularly vulnerable

service? Do we want to change the vendor that we're doing to get this? Is this something we want to outsource because it's too much vulnerability for our organization to put in place?

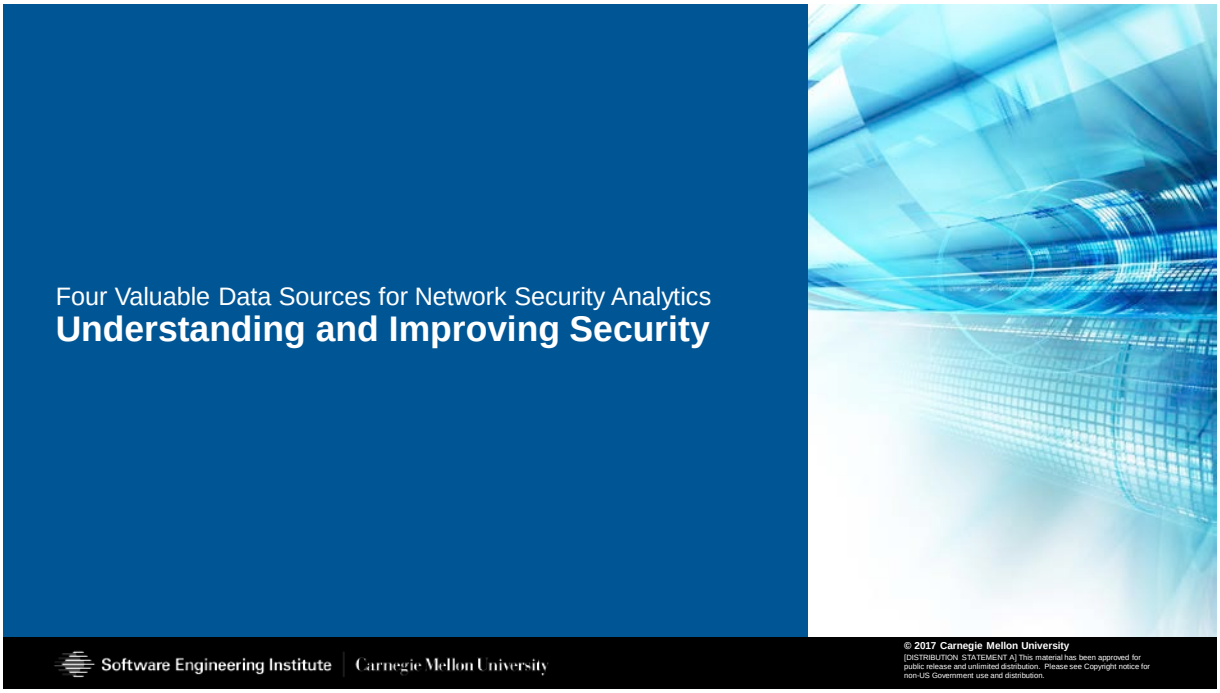
But to get a good quantitative number affiliated with this, you often need to, again, mix data sources. So, from the network inventory data, we're looking at what sort of network services are involved. We couple that with what we know about what's critical to our enterprise to get a feel for what's the deg-- probable loss associated with those services. We identify services that have the increasing vulnerability. Then we pivot to look at characteristic behaviors. We look at threat activity that's been associated with those services and exploits to get some understanding. This all helps us to more clearly understand the need-- the specific need for this particular analytic.

Then the model is okay, let's profile the vulnerabilities with new reports or increasing counts of reports from the network inventory data and then identify intrusion detection reports and automated-- associated services in the appropriate timeframe, similar to the last analytic that we did. But then look at overall traffic levels, overall changes in behavior using network traffic data, or network flow data that may give us a clue as to which are more critical and which are more in high volume in our enterprise.

Then in testing, again, we want some mix of very common versus less common. And in analyzing, again, we want to look at what's giving a clue that our vulnerability's getting worse as opposed to the last one which says we're improving, and what's giving us a feel for what might be more effective services. If we're applying several different vendors for a given network service, several different implementations, which ones seem to be behaving better than others? Or what's the overall risk depending-- in comparison to our dependency on this. And is this something we might want to outsource?

Finally, we look at revising that model. And again, you're looking at distinguishing what factors may clearly address the questions that remain versus what are coincidental changes that we simply want to wash out.

Understanding and Improving Security



**023 So, those were our three basic examples that are in place. The concept here is that by applying the analytic development model, combining a group of data, a given data source-- each given data source is, in fact, a part of the solution. But none of them are a complete solution. So, what we want out of this is some understanding of what our net enterprise lies and what we can do to improve security because I don't know of any enterprise that's out there that says okay, we're secure enough. That's often a very, very dangerous statement. Organizations that think they are pretty secure are altogether proven disastrously wrong.

Presenter: And I heard you mention outsourcing it, though, a couple of times.

Presenter: Right.

Presenter: What's the risk associated with that? I mean there has to be something.

Presenter: The big risk is you-- frequently, in outsourcing, you lose visibility. Because you've made the security of this critical service somebody else's problem, it's also part of somebody else's data stream. So, as you're outsourcing, you really need to be able to understand what is-- what do we need-- information do we need to get in order to be able to continue to track the risks associated with this service.

Of course, the other aspect of it is how reliable is our outsourcing agent. Are they on top of the security fixes? Are they--

Presenter: Right, how do you know that there of the-- is there a--

Presenter: And frequently, that's just something that you develop over time. And you look at what's provided for in the contract. And you do your due diligence in contract follow-up to try and improve that.

Understanding and Improving

Understanding:

- Data overload
- Observer bias
- Incomplete observation

Improving:

- Response to change
- Associating threat and risk
- Focusing on what can be improved

**024 So, in trying to understand our network security situation, whether it's outsourced or in-sourced, one of the things to recognize is that, frequently, your security teams are drowning in data. A busy network, a single busy network host will generate gigabytes of data easily. Networks, as a whole, easily generate terabytes of data. And much of this data frequently is pounding through so fast that the security team simply have no time to look at it. This is where analytics are key. Have in place tracking and summarizing analytics if nothing else that allow you to look at several different data sources and cross-correlate them. Just doing time series and overall spark lines or trend lines for the behavior would at least give you some capability of rapidly looking at the data sources and saying oh,

there's something going on there, or we had a data gap, or we had an unusual spike. Let's focus in on that particular aspect really, really quickly.

The other thing to look at is all of us, as observers, have our own biases. Certain analysts are really key to tracking normal and expected services. And they really want to provide expected services in an effective way. Other ones are looking particularly for threat and wanting to push threat. Neither one of those examples are wrong, but you need to recognize that the bias is there and what's going on. Finally, we need to recognize that all of our observations of network behavior are inherently going to be incomplete. Computer networks are very complex and very diverse and very rapidly changing. And it's-- we have to track with the data that we can get and summarize at an enterprise level. And that tends to be often a little less complete than we'd really like. There's no magic bullet that says this is good, this is bad. And the attackers don't want us to have that magic bullet.

So, as we're looking at these for improving network security, then we take a look at analytics such as these and the development process and combining data sources like this and saying okay, what changes are really needed in our enterprise at this point, where are we doing well, what do we want to continue doing, what used to be effective but is no longer effective, what do we need to rethink, what do we need to redevelop, what are blind

spots that we didn't realize that we had, and how do you respond to those changes. Frequently, you want to associate risk with threat saying because something is possible, it's going to be damaging to us. In practice, that's not always a good use of resources. You can over secure. And so, what you want to really focus in on is what can I measure that gives me a clear understanding of the threats that are active against those features of my network that I depend upon for my business. And focusing on those threats and tracking data with respect to those particular services is often what's key. And often it reduces the data overload that you will see by focusing in on the key enterprises and attributes.

And finally, you want to focus on what you can improve. We are not free to completely change our network behavior. Particularly as security professionals, we often aren't. So, you want to pick your, if you will, pick your threats wisely, or pick your fights wisely, and move forward with what-- where improvements can be made, planning things through in an iterative and justified fashion. And again, tracking these from observation may help you to focus-- to prioritize and to focus on the key improvements that would give you the biggest bang for your buck.

Presenter: So, you hear about some of these companies and security breaches in the news almost daily.

Presenter: Yes.

Presenter: And of companies not being aware the someone's on their network for months. Where are they going wrong? Do you have a guess? I mean is that something you could speak about? What are they not doing here?

Presenter: I suspect it is a combination of data overload and the observer biases. They assume behavior is ordin-- is expected and benign when, in fact, it's malicious. And while there are indicators that, after the fact, might have been really clear as to what goes on, during the incident, there's so much data coming at them that they have difficulty understanding. That's where your tracking and trending becomes so key to be able to clearly break out hey, what is unusual or unusual today. And is that unusual behavior, in fact, threatening? Or is it within the realm of what we see as day-to-day variation?

Presenter: All right.

Presenter: But in terms of what at specific organization A went on, I'm not free to lay it out, and often cases, I don't know.

Presenter: Which leads us to Flocon and why people should attend. So, Tim's a frequent speaker at our Flocon conference. And we wanted to invite everybody to attend Flocon 2018 this year. And this year's theme is using data to defend. The conference will take place in Tucson, Arizona January 8th through 11th. And

in general, Flocon explores largescale next-generation data analytics in support of security operations. So, everyone's attended. Anybody that's interested in data driven security, we invite to attend. All webinar attendees that have registered for this will get a discount of ten percent. We will send out a discount code via email tomorrow announcing the archive is available and how-- for more information on the conference.

Presenter: One of the things that's interesting about Flocon, having been to several-- been to all of them, in fact, one, it tends to be a very accepting bunch. So, if there are-- while the program is currently fixed, if there are topics that you want to discuss, frequently there are times during breaks and with speakers where they welcome the discussion. About a third of the attendees at every Flocon tend to be first-time attendees. And that tends to keep it a very dynamic and very applicable one. It's also a ver-- not an academic conference. This is a conference in which there are some academic depth going on, but there's also people with real current business problems. And there are also people that work for government agencies and vendors present. We have a very active vendor room with actually vendors sending engineers not salespeople to talk about things. So, this is-- it's a very involving environment. And I would welcome anybody to use your discount code and come onto Flocon. And Tucson in January should be pretty nice.

Presenter: Not too bad, right? A little bit warmer than East Coast.

Presenter: At least warmer than this part.

Presenter: Yeah. All right, so we're going to get the Q and A now so folks, use that chat tab or the Q and A tab. Feel free to type anything in there. We have one in the queue, so we're going to ask this one. If we don't get any more after that, we'll let everyone go home a little bit early here today. Amir wrote earlier asking, "What are scripting languages preferred to do analytics? Or preferred to do analytics?"

Presenter: Well, frequently, because we are iterating several times on the analytic to get-- to isolate the behaviors that we're interested in, it's not really get it right, but to look at ways to improve it, you want something that has a very quick development cycle. And scripting languages really let you do that.

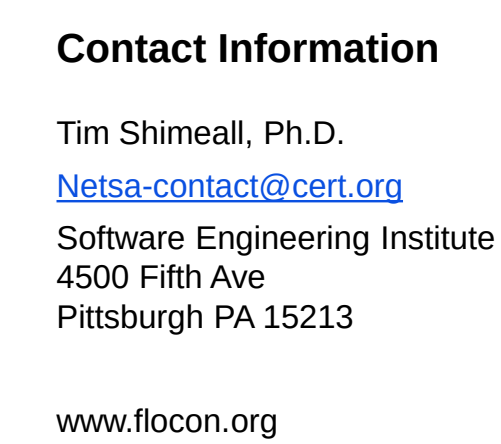
Presenter: Okay. Let me see what else we've got. Looks like the queue is empty. So, we'll wait about another thirty seconds or so. Someone just chimed in, "I like Python."

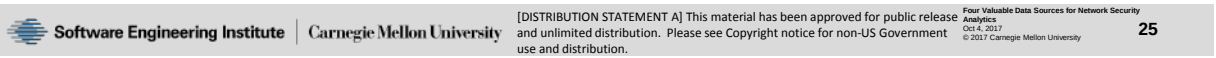
Presenter: Good. I do, too.

Presenter: And we'll wait there. While waiting there, just a reminder, I did add a survey link into the chat area. So, upon exiting today's event, we do ask that you fill out a survey, provide your feedback, because that's

greatly appreciated. Also, the slides are available.

Contact Information





**025 At the top of the chat window you'll see a link to the SEI website that will take you to the slides for today.

Presenter: Finally, if you do have questions coming out of this webinar that you would specifically be interested in exploring with me, the contact information is on the last slide. There's a contact address that doesn't go just specifically to me, but, in fact, goes to the situational awareness group. It's called netsa-contact@cert.org. And the URL at the bottom, I've left off the HTTP part, but the URL at the bottom will lead you to more information on Flocon.

Presenter: Great. Tim, thank you. Thank you for an excellent presentation today.

Presenter: No problem.

Presenter: Very well done. Thank you for everyone attending. We appreciate your time. We hope to see you at Flocon 2018. And we'll send out an email tomorrow with that relevant information and for upcoming webinars, as well. Thanks, everyone. Have a great day.

Presenter: Have a good day.

SEI WEBINAR SERIES | Keeping you informed of the latest solutions