Ideas on Using Asset Criticality Inference (ACI) Through Gathering and Processing of Asset Contextual Data Utilizing Analytical Models and Processing Rules

D. Grzetich

6/26/2013

The Problem We Face Today

Security services such as vulnerability management and security monitoring and response would benefit if context and asset criticality were used to rebalance or re-rank identified security issues. Today, we often rely on static data sets that define our assets and are obtained from asset management repositories or configuration management databases (CMDBs). The issue that arises from the use of this static data set is that criticality rankings are often not assigned or updated and do not take into account the actual usage of the asset and its role in revenue generation for the organization. Furthermore, as assets change in role or as they are added or removed from the network the static data in these repositories is often not updated. In order to be useful in measurement the asset information, criticality, usage, and available services need to be updated on a schedule of frequency that is not manageable through manual entry and periodic recertification of asset criticality.

By defining the asset criticality more accurately and through automated near real time processes we open up a wide range of uses. We can now apply the "asset criticality lens" to multiple problems to allow for better decision making and overall risk reduction. This new criticality data may be applied to any number of questions such as: what vulnerabilities in my environment pose the most risk, or in what order should vulnerabilities be remediated? What is our current risk posture for a specific business unit, process, or application? Are groups or types of assets posing more risk in our environment? If we receive alerts concerning a possible security issues with a set of assets, which should be investigated first?

The Problem Summarized:

- Asset criticality often does not exist or is assigned upon the asset's entry into a central tracking mechanism or CMDB
- The effort to manually determine and recertify asset criticality is often so great that manual processes fail or produce inaccurate data
- In order for asset criticality data to be useful we may need near real time views of the criticality that change in concert with the asset's usage
- Without accurate asset inventories and criticalities we cannot accurately represent overall risk or risk posture of an organization

A Potential Solution

Given the growing use a big data solutions and our ability to store, process, and query large data sets it is possible that using these data sets, through data analytics, to infer asset criticality may provide a solution. By leveraging the power of security data to infer the criticality of the asset based on its operating properties may allow the real time identification of asset criticality. This isn't an attempt to replace the asset management repositories that we have today, and will continue to have; it is an attempt to enhance this data with context that was previously unavailable information to better rank or rate the criticality of assets in a near real time fashion.

This potential solution of criticality inference also includes a mechanism for the organization to change asset criticality in near real time as the system use or functionality of the asset changes. For example, an asset which was previously not considered critical takes on a new role in an application stack which processes sensitive data. In the previous static data set model asset criticality would only be assigned and updated periodically if at all.

One could argue that some of the required data exists in the information technology environment today, albeit the data may be spread across several source systems and not centralized. While this may not pose an immediate problem it extends the solution time as interfaces to the various data sources may need to be built so that the data is meaningful to the security analysts. In the best case scenario we would define a central repository of the required data, import the data for analysis, and then represent the asset criticality as an output. It is also possible that the data required may not exist due to a lack of audit logs, infrequent vulnerability assessments, or lack of the necessary technology which represent only a few examples. Lack of the required data would also extend the solution time as well as lower the confidence levels of the inferred criticalities.

Potential Issues

Realizing a potential solution to solve our set of problems also comes with a set of potential issues that would need to be addressed. Since our solution requires data to feed the analytical models, and that data needs to be accessible to or by the solution, many of our issues are centered on data availability, access, and storage. In summary, the issues present are:

1. The security data does not exist or is not currently being collected
2. A central location to collect, store, and process the data may not exist or scale to enterprise levels
3. Interfaces and connectors to obtain the data required for the model or analysis will need to be built

Potential Data Sources

As our solution requires various data and/or data feeds to the analytical models and processing rules, the table below presents so potential data sources that could be included. While this is not a comprehensive list of data sources it presents a baseline set of data that is likely to be available in our environment currently. The table discuss each data source, a description of the source, and how it would potentially be used to build our analytical models or processing rules.

Data Source	Description	How The Data Could Be Used in ACI?
Asset Management (AM) Repositories	Static and periodically and manually updated inventory of IT assets.	Used as a baseline set of "known" assets in the environment. May require some normalization if we do not have one but many sources of truth for asset repositories.
Change Management Database (CMDB)	Configuration management database which contains information on the configuration elements of IT assets including ownership.	Used as a baseline of "known" assets in the environment in addition to mapping applications (potentially by assigned or ranked criticality) to their supporting infrastructure. This could extend the view of the critically of a single asset to its role in supporting a specific application. ACI could also be extended to support the automatic mapping of infrastructure and assets to applications.
Vulnerability Assessment (VA) Scan Data	Data about the services (ports/protocols) and vulnerabilities of IT assets that are scanned by the system. Also included in this data may be network address and physical (MAC) address information.	Used to feed information about currently reachable and/or existing assets in the environment. May be able to determine stale/orphaned assets in AM and/or CMDB based on non-existence in VA scan results. Enumerates and highlights the services, open ports, and applications that are installed and running on the target systems.
Network Mapping Data	Data obtained about the current logical layout of the network including IP address subnets and allocation.	A network map could be used in the processing rules to map identified assets by IP/subnet to a physical location. This data may be static (i.e. one-time assessment of logical network architecture) or updated regularly through network mapping (i.e. Solarwinds, Lumeta IP360, etc.). Network accessibility may also be a factor in determining the criticality of an asset in terms or risk (i.e. publically accessible versus secure subnet).
Active Fingerprinting Data	Data obtained about the operating system through active scanning of the IT asset. May be included in vulnerability scan data and port and service scan data or through additional tools.	This data could be a combination of asset data, such as installed operating system, patch levels, installed and running applications and services, etc. that is obtained from authenticated vulnerability assessment scans in addition to installed agents and applications (i.e. Tanium, HPCA/HPAM, SCCM agents, etc.).
Passive Fingerprinting Data	Data about the operating systems of IT assets through passive monitoring of network traffic.	Data related to observing network traffic to determine active assets and applications/services may indicate the presence of a previously unknown asset and/or application or service if missed by vulnerability assessment scans or does not have an host-based agent installed. This may be an overlap with network monitoring data or netflow as discussed below.
Dynamic Host Configuration Protocol (DHCP) Data	Data about the scopes and zones used by DHCP to dynamically allocate network addresses. Also included by be reservations or reserved subnets which are used for static assignment of IP addresses for static IT assets (i.e. network infrastructure, servers, other non-end user systems).	Data about asset, such as host name, location (based on network mapping data), and physical switch port (MAC address mapped to switch port mapped to physical jack locations). For ACI this may be another sources of asset/host name information for mobile/end-user/end-point type assets. There is a potential impact if the organization utilizes BYOD and co-mingles corporate-owned and user-owned assets on the same network segments or subnets.
Port and Service Scan Data	Data from active scans which enumerate open ports and services on IT assets. May be included with vulnerability scan data or obtain independently through additional tools.	Scan data, in addition to vulnerability assessment scans, that map an asset to listening ports/protocols/services to enumerate accessible services. Could be used to enhance and/or validate data received from vulnerability assessment and/or host-based agents.
Data Loss Prevention (DLP) Data	Data obtained from DLP tools which enumerate the type of data stored on the asset the sensitivity of the data (i.e. PII, PHI, PCI, etc.).	Data related to both the storage of "sensitive" information (at-rest) to map the type of data stored on a per asset basis which would increase/decrease the inferred criticality. Comprehensive discovery scans may indicate previously unknown critical assets

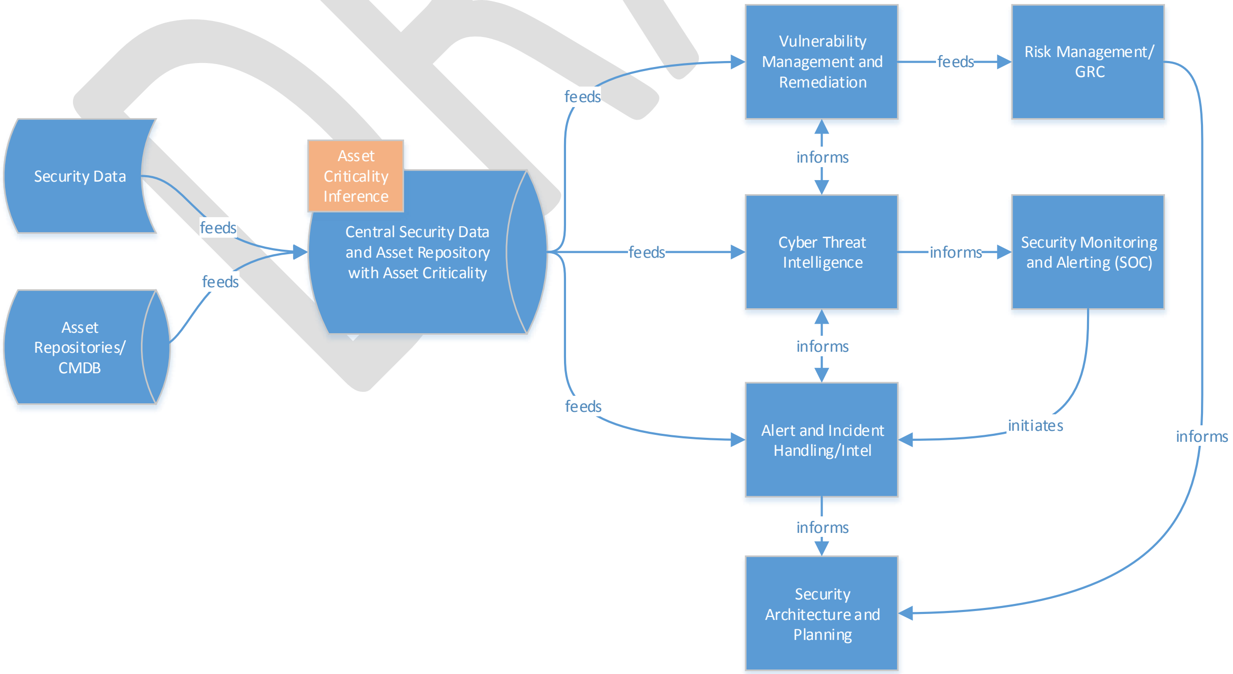
		and/or supporting infrastructure which could be used to update AM repositories or CMDB. Monitoring for data in motion may highlight the use of an asset in terms of sensitive or critical data (i.e. as asset is very, or not, active in sending and/or receiving sensitive information).
Software Agents on Assets/Asset Management Applications	Any software installed asset which can enumerate the type of system, running services, and installed software. Examples of agents include both asset management-focused agents (i.e. HPAM/HPCA/SCCM) as well as security-specific agents (i.e. Tanium, RSA ECAT, SilentRunner, etc.)	Data to enhance and/or validate the applications and services running on an asset (i.e. a server is running LDAP and DNS, or a web server is running Apache, FTP, and SSH). This could also include data about applications that are installed but not active (i.e. IIS is installed but disabled).
Netflow	Netflow data captures the interaction, at a network level, between various systems as they pass through a router capable of capturing Netflow/JFlow/Nflow data which shows the source/destination network address and source /destination port and protocol of the traffic.	Data on the number, type, frequency, volume, and size of transactions between that could be used to highlight the "activity level" of an asset. In terms of ACI, the types of transactions and frequency may be important values in the formula or processing rules that affect an asset's inferred criticality. For example, a highly trafficked asset (such as an Active Directory Domain Controller) that processes user authentications via LDAP would be increased in the ACI models.
Network Monitoring Data	Encompasses a large set of data on network activity through monitoring and capture of network traffic data. Data may be captured at the perimeter through firewalls, proxies, and full packet capture solutions, or internally by network capture tools.	Data similar to neflow but one level deeper in terms of granularity. Netflow would highlight traffic flows across routed boundaries, network monitoring data would highlight data flows, port/protocols, and volume/size across firewalls and perimeter devices as well as intra-subnet communications that would be missed by Netflow as the traffic may not cross a routed boundary (i.e. server-to-server same subnet communications, or an application logic and database pair on the same subnet/virtual or physical switch).
Security Information and Event Management (SIEM) Systems	Only shows alerts, but may be a supplementary source of data since we'll know the "type" of system including name and IP which can be used in asset address to name mappings?	Provides additional data on the asset names/IP mappings, active users of the implicated system or target system from alert data. SIEM may also be expanded to include elements of log management (i.e. utilizing Splunk prior to SIEM, or the log management features of modern SIEN systems).
Intrusion Prevention/Detection Systems (IPS/IDS)	Only going to show alerts, not necessarily the attributes of the asset. Linux attacks against Windows are of no value in this context...	Debatable source of information as most of the asset information would have been gathered from one or more sources above. However, in low maturity organizations it is not uncommon to see IDS/IPS deployed that may indicate the presence of an asset and/or network-accessible service. Would consider this a tertiary source of asset information at-best.
BCP/DR or BIA	More of a static source of information...may hold more data on the asset in terms of RPO/RTO which speaks to criticality, or can be used as a factor in the calculation.	Static source of information on assets and their defined criticality through the BCP or BIA processes. May be useful in the absence of AM or CMDB systems.
Application Access Logs	Shows application access by user account, frequency of access, type of access (general user account versus privileged account).	Data, on a per application basis, in conjunction with DLP information about the type of information stored/processed by the application that would indicate the scope of connecting, data transmitted, and or application usage that could affect the inferred asset criticality. For example, applications accessed by many users that contains sensitive information that is highly utilized would be considered a higher criticality asset.
Database Access Logs	Shows usage by number of users and type of interactions (frequency, amount, type of data, etc.).	Same as above, but adding database usage parameter on a per applications basis to infer criticality of database systems.
IAM/IDM/AGS/PUAM Data	User account context based on group membership, role, or access levels. Can be used to add context to application, database, and user account logins to determine the type of access being requested and the frequency of access.	Data that could highlight the level of access users have to applications and data, and in conjunction with the application logs (usage), DLP (sensitivity of the data), network data (location, running services, etc.) and asset information (ownership, location, role) would help infer or modify statically defined asset criticality ratings.
User Account Data	General user account data from Active Directory, PeopleSoft, etc. to add context to user transactions and access attempts.	Data on the context of the user account not managed or maintained in an AGS/IDM solution. For example, if AD is used to group super users and separate them from the general population, this data may not be included in a AGS solution that

		is focused on financial/SOX applications only).
Mobile Device Management (MDM) Systems	MDM systems may contain information about user or system to IP or MAC address mappings as well as posture assessment information from active scanning of asset upon registration or connection to the network.	May be useful to update and/or validate AM repositories and CMDB (if necessary) on which mobile assets utilizing corporate resources. Data that would be valuable for ACI may include information gleaned from posture assessments (i.e. 802.1x auth plus scanning, or applications such as MobileIron)

Open Questions ad End Notes

This is a work in progress and this paper is simply a draft of ideas on asset criticality inference . To that end I have a series of open or unanswered questions that need to be considered, these include:

- How can we provide a solution that takes in available data about assets and come out with a criticality ranking based on the attributes in the data (i.e. processing rules versus machine-learning models)?
- In the absence of required data for the rules or model (or machine-learning), how do we assign confidence score to the asset criticality?
- What is the right analytical solution to process this type of information (i.e. existing providers in niche-markets such as identity analytics, standard analytics platforms such as SAP HANA, Palintir, IKANOW, etc., or a newly designed platform)?
- How do we ingest this new criticality ranking in other services (i.e. alert handling, incident response, monitoring, etc.)? Or, make it available in a form that can be utilized by multiple processes (i.e. a service-integrated model as depicted below)?


Security Data
Asset Repositories/ CMDB
feeds
feeds
Asset Criticality Inference
Central Security Data and Asset Repository with Asset Criticality
feeds
feeds
feeds
Vulnerability Management and Remediation
feeds
Risk Management/ GRC
informs
Cyber Threat Intelligence
informs
Security Monitoring and Alerting (SOC)
informs
Alert and Incident Handling/Intel
initiates
informs
informs
Security Architecture and Planning