AFRL-RI-RS-TR-2021-017

INTERNET RISK ASSESSMENT AND MITIGATION (I-RAM)

PARSONS GOVERNMENT SERVICES, INC.

FEBRUARY 2021

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th ABW, Wright-Patterson AFB Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2021-017 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /
FRANCES A. ROSE
Work Unit Manager

/ S /
GREGORY J. HADYNSKI
Assistant Technical Advisor
Computing & Communications Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
FEBRUARY 2021	FINAL TECHNICAL REPORT	JAN 2018 – MAY 2020

4. TITLE AND SUBTITLE

INTERNET RISK ASSESSMENT AND MITIGATION (I-RAM)

5a. CONTRACT NUMBER
FA8750-18-C-0039

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
N/A

6. AUTHOR(S)

SURESH KRISHNASWAMY

5d. PROJECT NUMBER
DHS0

5e. TASK NUMBER
IR

5f. WORK UNIT NUMBER
AM

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Parsons Government Services Inc.
100 W Walnut St
Pasadena CA 91124-0001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/RITGB
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSOR/MONITOR'S REPORT NUMBER
AFRL-RI-RS-TR-2021-017

12. DISTRIBUTION AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited. PA# AFRL-2020-0565
Date Cleared: 25 Jan 2021

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The goal of the Parsons Internet Risk Assessment and Mitigation (I-RAM) project is to provide risk analysis, risk management and decision-making support needed by enterprise owners and operators across the federal government, critical infrastructure, and private sectors, to support an integrated, holistic understanding of the risk inherited through dependence on the Internet infrastructure for critical Internet services. In this report we document the work that was accomplished as part of the I-RAM project. We discuss the research that was performed and the results from our work. We also discuss the data and capabilities that were made available based on our research, the lessons learned and our plans for future evolution of the I-RAM capability.

15. SUBJECT TERMS

Internet Risk, Risk Assessment And Mitigation, Internet Analytics, Internet Infrastructure

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			**FRANCES A. ROSE**
U	U	U	UU	37	19b. TELEPHONE NUMBER (*Include area code*) **N/A**

Table of Contents

List of Figures

List of Tables

1 SUMMARY

Traditional enterprise risk analysis is usually performed in the context of some well-defined organizational boundary. In cases where an organization's core information system assets can be confined to some security perimeter such risk assessment methods are fully appropriate. However, there are two reasons why traditional risk assessment approaches are likely to be insufficient for modern enterprises. First, the enterprises of today, rather than being monolithic entities, are more likely to resemble an interconnection between a number of disparate entities, where certain pieces of the enterprise's information systems are external or managed by third-parties, where not all pieces of information that are necessary for determining enterprise risk are readily available. Second, the ability for enterprise systems to interconnect with other information system elements and for its users to reach such elements depends on certain core Internet protocols and services, which the enterprise, itself, does not control fully.

Enterprises often take their dependence on the Internet for granted. However, understanding the nature of its dependence on the Internet such that it can analyze and mitigate the risks that are most likely to produce harm to its organizational mission is, arguably, a critical element of the enterprise's modern-day risk management strategy that it can ill-afford to ignore. At the same time, there also exists a gap in the risk management landscape in terms of the set of available tools and methods for quantifying and managing the risk associated with an enterprise's dependence on the Internet.

The goal of the Parsons Internet Risk Assessment and Mitigation (I-RAM) project is to bridge this gap, by providing the risk analysis and decision-making support needed by enterprise decision makers and operators across the federal government, critical infrastructure, and private sectors, to support an integrated, holistic understanding of the risk inherited by the enterprise due to their dependence on the Internet infrastructure for critical Internet services. In this report we document the work that was accomplished as part of the I-RAM project. We discuss the research that was performed and the results from our work. We also discuss the data and capabilities that were made available based on our research, the lessons learned, and our plans for future evolution of the I-RAM capability.

2 INTRODUCTION

The past decade has witnessed a rapid transformation in the way that enterprises are provisioning their Information Technology (IT) infrastructure through the use of cloud-provisioned applications and services. To put the scale of this transformation into perspective, from recent figures [1], it is predicted that data center traffic over the next two years is likely to increase from 6 ZB to around 20 ZBs, a roughly three-fold increase from its current value, with most of this increase expected to be a result of cloud-provisioned applications and other Internet of Things (IoT)-like services that rely on some cloud-provisioned backend for providing an emerging set of enterprise IT capabilities. The use of cloud-based applications, many of which are often provisioned and managed by third-party providers, has resulted in enterprises now having dependencies and inter-dependencies that extend beyond their traditional network perimeters to span a diverse set of components, functions, organizations and geographies, all using the Internet as the common interconnection medium (see Figure 1).

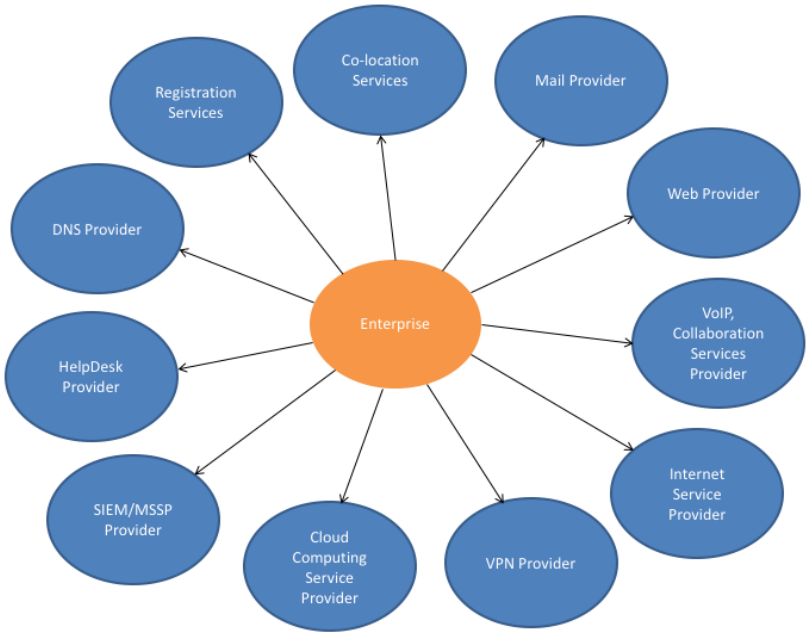

Figure 1. Dependencies outside the enterprise boundary

There are advantages for an organization to use third-party providers for the provisioning of their information system services. For example, an organization that needs to provide a highly available web service to a large and distributed set of users will often use Content Distribution Networks (CDNs) to provide such content with high availability guarantees. Similarly, Domain Name System (DNS) providers are generally able to provide a much more distributed and redundant infrastructure for the provisioning of Internet naming services than the organization would be able to provide on its own. Finally, certain third parties, such as mail service providers, may also provide some critical value as part of their service offering, such as malware and Uniform Resource Locator (URL) scrubbing, which makes delegating to such third-party providers appealing.

While the use of a cloud-based infrastructure has resulted in a reduction of operating costs and improvement in efficiency, the dependence on the Internet has also resulted in the enterprise's digital attack surface growing significantly. An enterprise's attack surface, today, not only includes the IT elements that are provisioned and managed in-house but also includes the multitude of external providers to which the enterprise interconnects, including DNS providers, Internet mail providers, web service providers, Managed Security Services Providers (MSSP), Security Information and Event Management (SIEM) providers, and providers of Voice over Internet Protocol (VoIP) service as well as collaboration services and many other services (see Figure 1).

There is growing awareness of the transformative forces that are shaping the enterprise's digital attack surface. Yet, most enterprises do not have a clear strategy to translate the risk associated with their dependence on the Internet into something actionable. There are two primary reasons for this: one, enterprise stakeholders do not have the means to translate the different dependencies and inter-dependencies on the Internet into a construct of risk (a conceptualization problem); and two, enterprise stakeholders are unable to quantify risk in a manner that enables them to make trade-offs between alternatives (a quantification problem). At the same time, the inability to form a strategy to evaluate the risk associated with Internet dependence is of particular concern, given that most of the users of modern enterprises, such as its remote workforce, partners, and vendors, often reside outside the enterprise network perimeter and rely on the shared Internet infrastructure to access critical enterprise information system services.

This shared Internet infrastructure that enables access to the different enterprise services, is composed of certain core protocols and services, whose proper functioning forms a prerequisite to enabling trustworthy inter-connectivity between different network assets in cyberspace. For example, the routing and delivery of an organization's data packets through the Internet relies on the proper functioning of multiple entities, including the set of Internet Service Providers (ISPs) that provide the Internet connectivity paths for the enterprise, the organizations that hold IP prefixes associated with the organization's various public facing Internet services, the Autonomous Systems (AS) that originate a route to the appropriate IP prefix, the various backup and hosting arrangements that can exist as part of business continuity and disaster recovery planning, and the set of peering relationships between different Autonomous Systems (AS) that lie in the network paths between multiple yet interdependent systems.

The potential ramifications of attacks that target dependencies within the Internet infrastructure can be very severe, with effects ranging from loss or denial of access to critical enterprise service, to more advanced forms of traffic redirection and interception. Similarly, attacks that target a particular trust point (such as Certificate Authorities) or some critical piece of the global DNS may compromise the legitimacy and authenticity of enterprise service endpoints, thereby hindering the enterprise's ability to meet its mission requirements. Risks may also manifest at the spatial level, when the enterprise mission and the topological structure between users and services intersect in a manner that increases the criticality of a particular geographic region.

The goal of the I-RAM project is to provide a quantitative basis for assessing the risk associated with an enterprise's dependence on the Internet in order to enable enterprises to take an active role in managing such risk. The project accomplishes this goal by providing the methodology, a set of metrics and an analytics capability that enables enterprise stakeholders to obtain a multi-dimensional view of the inter-dependencies between their Information Systems, in order to help them assess and mitigate their risk exposure to adverse events that target such inter-dependencies. By doing so, it provides the decision analytics to support an integrated, holistic understanding of the risk environment in order to enable effective interventions in the form of investment and other activities, to prevent, protect, mitigate, and recover from cyber disruptions and harm.

The I-RAM project lays the groundwork for enterprises to assess their Internet Exposure Risk (IER). We define the IER for an enterprise as the probable frequency and magnitude of loss associated with the inability to fulfill enterprise functional objectives through attacks on the Internet infrastructure. IER is also a reflection of an enterprise's procedural maturity in that it characterizes the degree to which the enterprise has measures in place to identify, protect, detect, respond and recover from attacks or conditions that prevent it and its users from accomplishing certain objectives when such objectives are dependent on the Internet infrastructure.

The ability to assess IER through I-RAM is structured around three broad components, as illustrated in Figure 2.

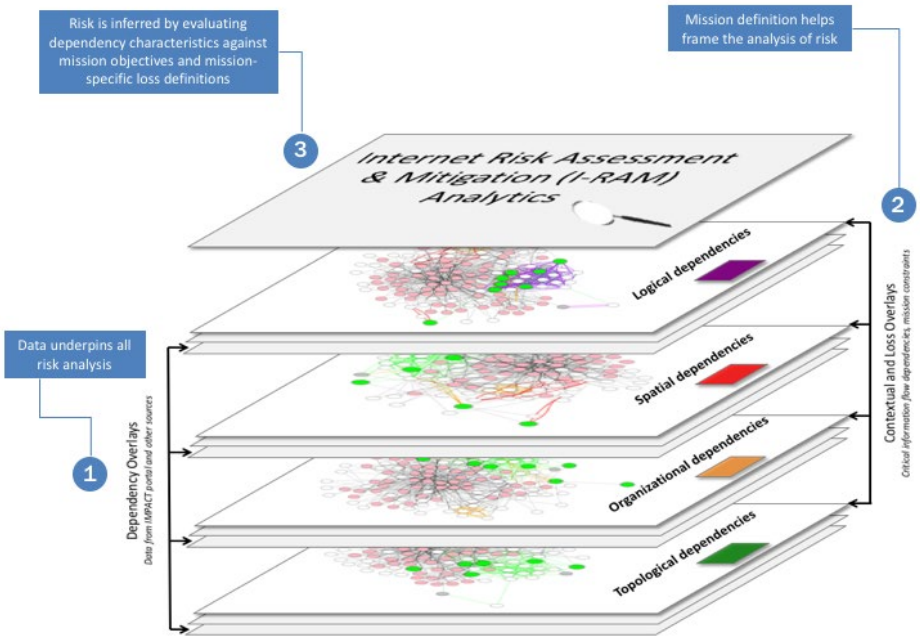

Figure 2. I-RAM concept diagram

The first component (labeled "1" in the figure) is a corpus of data that enables one to infer a set of dependencies that exist between different information system elements, where the dependencies themselves may manifest along multiple dimensions. The second (labeled "2" in the figure) is a set of capabilities that help scope the assessment of risk through an examination of different dependency overlays within a particular evaluation context. The third component (labeled "3" in the figure) is a set of measures and metrics that provides a way to characterize risk through the dependency patterns that exist between the different dependency elements.

In order to infer dependency relationships, the I-RAM capability uses data from existing sources of curated data, including the Information Marketplace for Policy and Analysis of Cyber-risk and Trust (IMPACT) platform[1]. In order to support the use of the I-RAM capabilities by a diverse set of end-users, the capabilities developed through I-RAM project, in the form of new data and analytics, are also postured to integrate with the IMPACT Platform, as illustrated in Figure 3. In this way, the I-RAM project fulfills its dual-role as an IMPACT Data Provider (DP) as well as that of an IMPACT Data Analytics Service Provider (DASP).

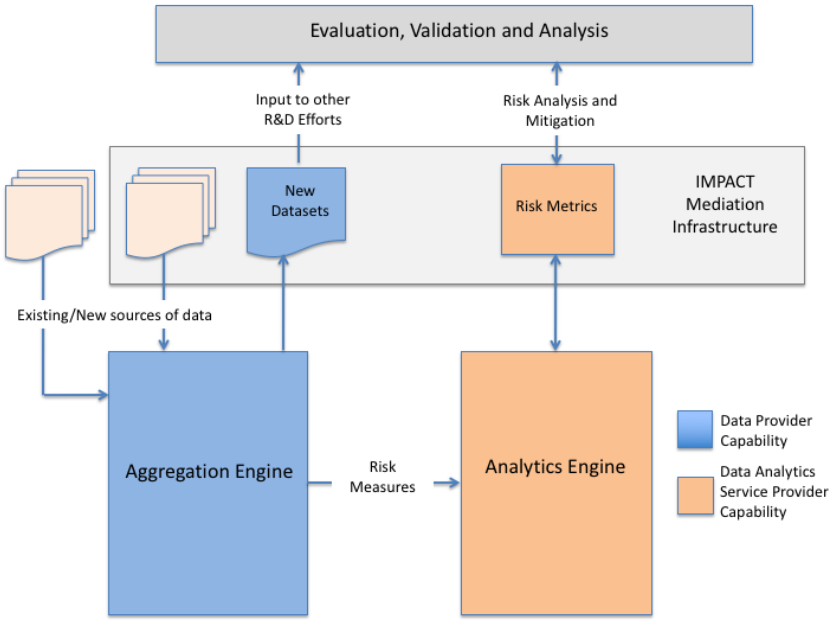

Figure 3. I-RAM capability integration with the IMPACT platform

[1] See https://impactcybertrust.org. The IMPACT platform enables empirical data and information sharing between and among the global cyber security research and development (R&D) community in academia, industry and the government, by making available a centralized brokering and distributed provisioning capability between the providers, hosts and researchers.

The primary outputs from I-RAM that integrate with the IMPACT platform are new datasets and a set of analytics capabilities associated with evaluating risk within a particular context. The core capabilities within I-RAM that generate the above outputs are an "Aggregation Engine" that synthesizes existing sources of curated data and generates measures of dependency, and an "Analytics Engine" that uses data from the Aggregation Engine in order to provide the necessary metrics and analytics that are used for the assessment of IER.

This document is organized as follows. In Section 3 we describe our research methodology. This includes the research work that we performed in order to develop the concepts for performing risk assessment, and the measures and metrics required to support the analysis of risk in the context of an enterprise's dependence on the Internet. In Section 4, we describe the results from our effort, including the capability, research and data outputs that were created as part of this effort. Finally, in Sections 5 we provide our conclusions, including the lessons learned and ideas for future development of the I-RAM capability.

3 METHODS, ASSUMPTIONS AND PROCEDURES

The overall methodology for developing the framework for risk analysis within I-RAM was comprised of the following steps:

- Performing a survey of the available data in order to identify potential sources of dependency data along multiple dependency dimensions.
- Developing an ontology describing the relationship between threats, vulnerabilities, loss and impact in the context of Internet Exposure Risk.
- Developing the unit of risk analysis for I-RAM, namely an I-RAM "mission".
- Developing various measures and metrics associated with measuring risk.
- Developing a threat event library for quantifying exposure to hazards that may be relevant within the context of an I-RAM mission.
- Developing the methodology for performing Internet Exposure Risk assessments.

We describe each of these steps in greater detail in the sub-sections that follow.

3.1 Survey of available data and inference of dependency dimensions

Our process for developing a framework for IER assessment began with a survey of existing data that were available from IMPACT and other sources (see Figure 4). We organized the data by providers of data and the different types of data that were available. An analysis of the different data types revealed that there were essentially four dimensions of dependency data, and a number of other sources of data that provided reputational or loss proximity information associated with various threat events.

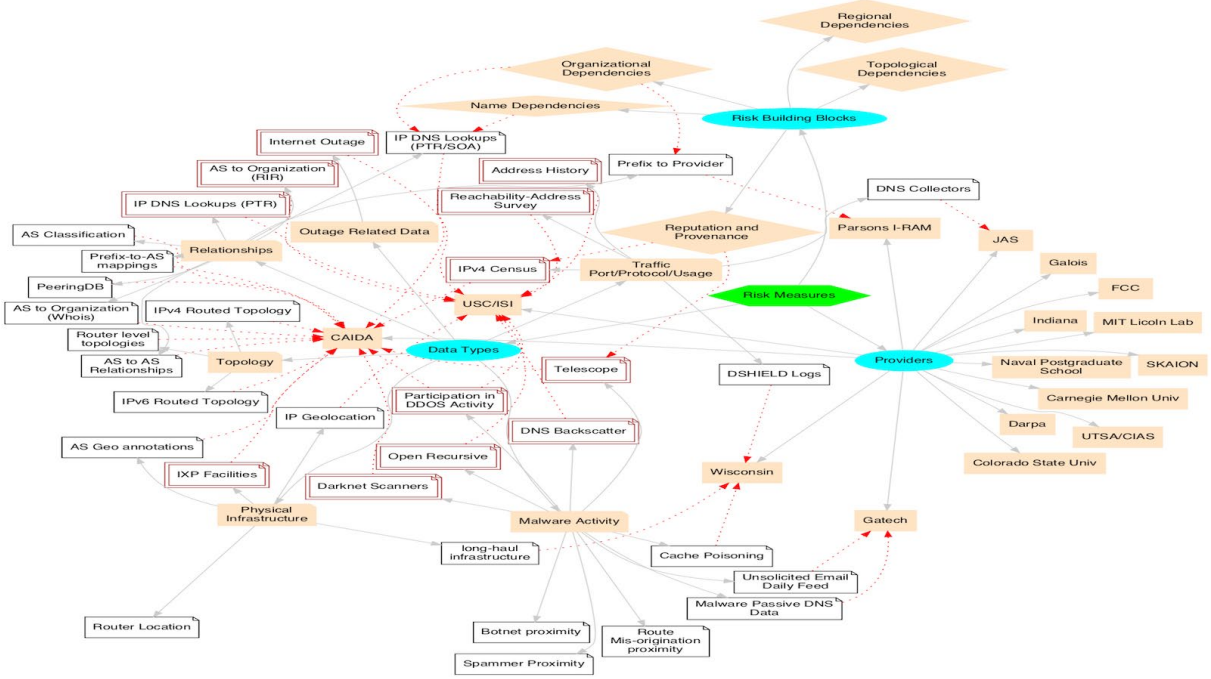

Figure 4. Survey of existing Internet-related data

The four dimensions of dependency data serve as lenses through which the vulnerability induced through dependence on the Internet can be analyzed. Viewed in the aggregate, these dimensions also represent the interdependency mosaic through which failures or loss events cascade across enterprise components even though they are physically dispersed.

A description of the four dependency dimensions and some of the sources of data that we used in order to infer the relationships between entities along those dimensions are described below.

3.1.1 Topological dimension.

Dependencies within the topological dimension include those that pertain to interconnecting elements within the Internet. These connections may be physical, such as the infrastructure links that establish physical connection between network entities; peering-based, which represent the links between two routing entities or Autonomous Systems (AS); or route-based, which relates to the specific path that traffic from a specific source destined for a particular destination will traverse, on the basis of the business relationships established between the different network providers.

We observed through our survey that a number of data sources that enable one to model the topological dependencies were available, including physical network infrastructure maps such as those generated by the Internet Atlas project[2], AS-link information collected by the Center for Applied Internet Data Analysis (CAIDA)[3], and route information collected by the RouteViews project[4]. Of these, we used the AS-link data from CAIDA and the route-level information from RouteViews in order to represent topological dependency relationships within I-RAM.

3.1.2 Geo-spatial dimension.

Each AS within the Internet is comprised of one or more routers where each router has some notional representation of spatial geolocation[5]. Thus, the geospatial plane forms another dimension where enterprise network resources may lie at the proximity of one another, even if they are topologically dispersed.

Determining the geolocation for a router is hampered by a number of factors, including the inherent difficulty associated with inferring such data, and the existence of certain edge cases, such as multiple instances for the same AS. However, many sources of geolocation data for ASes exist. The data provided as part of the CAIDA Internet Topology Data Kit (ITDK)[6] dataset, for example, uses a traceroute and heuristics-based approach [2] to assign routers and geolocation information to ASes. Since we were already using the CAIDA AS-link data to infer the topology-level interconnections within I-RAM, we also used the CAIDA ITDK data in order to obtain geo-spatial information associated with those ASes.

---

[2] http://internetatlas.org
[3] https://www.caida.org/data/active/ipv4_routed_topology_aslinks_dataset.xml
[4] http://www.routeviews.org/routeviews/
[5] While the use of virtual routers may render this statement somewhat incorrect, widespread use of virtualized router functions is still not in common use today.
[6] https://www.caida.org/data/internet-topology-data-kit/

3.1.3 Organizational dimension.

Consideration of organizational affiliation is important because it enables one to analyze data provenance and supply-chain issues in the determination of risk.

For a given network topology, organizational affiliation can be derived in multiple ways. In the basic case, the organizational affiliation refers to the organization handle associated with the AS that originates one or more Internet Protocol (IP) prefixes of interest. A separate organizational affiliation may be derived in cases where the prefix-holder and the AS that announces a path to that prefix are different entities. Finally, the organization handles associated with identifiers used within the routing system may themselves be part of a larger network of organizational inter-dependencies. For example, different organization handles could be assigned to different points of contact within the same enterprise, or two organizations could become part of a larger conglomerate as a consequence of organizational mergers and acquisitions. In such cases, organizational affiliations represent linkages with different organizational clusters rather than single organizational identifiers.

Organizational relationships can be usually inferred from the Regional Internet Registry (RIR) Whois databases. Additionally, tools that help derive organizational clusters based on the relationships between resources within the RIR databases also exist[7]. For I-RAM, the organizational dependencies were inferred from the CAIDA Inferred AS-to-Organization dataset, which "uses Whois information available from Regional and National Internet Registries to infer a mapping from AS numbers to the organizational entities that operate them" [8].

3.1.4 Logical dimension.

The fourth dependency dimension relates to those relationships that link enterprise resources by names or IP addresses, protocol dependencies, or through business-level relationships that do not reveal themselves directly within data sent over the Internet.

Name dependencies may exist in the form of the reverse DNS information associated with enterprise network prefixes, external names referencing enterprise resources, the specific choice of internal and external application end-points, the use of names registered within a particular domain, and the use of private namespaces or namespaces that are subject to name-space collision. IP-level dependencies, on the other hand, may exist when different enterprise service names or application endpoints map to the same IP address or address range, or when prefixes are aggregated in route announcements.

Protocol-level dependencies refer to those situations where an event associated with one protocol can also trigger a separate event in another protocol. For example, if the reverse DNS zone for a certain prefix is served by a resource whose address is in that same prefix, then a routing problem for the prefix could affect reverse DNS resolution for the prefix as well. The dependencies between DNS and NTP is another example of such protocol-level dependencies

[7] See the tools produced by the Secure Routing Project at https://www.securerouting.net
[8] https://www.caida.org/data/as-organizations/

where the ability to check the source authenticity of signed DNS data is dependent on an accurate source of time.

Finally, business-level dependencies refer to the set of services that an organization may acquire from external providers, including DNS, mail, website, VoIP, virtual private networks, training portals and other enterprise support services. They also include the multilateral agreements that may exist, such as when two ASes exchange routes at a particular Internet Exchange Point, and instances where an organization co-locates some portion of its own infrastructure within a different provider facility.

While there is no singular data source that enables one to model logical dependencies between enterprise resources, certain dependencies can be inferred through available datasets. These include reverse namespace lookup data and other "reputation" based tools that attempt to spider through the world-wide web looking for name relationships that establish linkage to a given enterprise. In most cases, however, identifying logical dependencies for an enterprise requires approaches that are tailored to the enterprise, such as by performing content analysis over their internal and external web pages, data from application firewalls, or must be made available by the enterprise directly. For I-RAM we used data from a University of Southern California-Information Sciences Institute (USC/ISI) survey[9] of reverse names in order to infer the dependencies between IP addresses used by an organization and their corresponding names.

3.2 Developing an Ontology for Internet Exposure Risk

A significant piece of our research was focused of developing the ontology for Internet Exposure Risk. We leveraged a number of existing models and frameworks for conceptualizing and quantifying risk. These included the Factor Analysis for Information Risk (FAIR) ontology [3], NIST Special Publication 800-39 [4], NIST Special Publication 800-30 [5], the NIST Criticality Analysis Process Model (NISTIR 8179), and the STRIDE model [6], among others.

The Internet Exposure Risk ontology describes the manner in which the different dependency relationships extracted through available sources of data can be combined through known risk constructs so as to convey some notion of risk. It also describes the relationship between vulnerabilities, threats, loss and risk events in the context of one or more I-RAM "missions", where the I-RAM missions are defined by existing enterprise Governance, Risk and Compliance (GRC) functions. In addition, it also identifies certain enterprise objectives, that if violated, could produce loss events for an enterprise.

An illustration of the relationships explored within IER ontology is provided in Figure 5. The initiating actions for IER are business decisions that are made in relation to the enterprise GRC functions, which result in the creation of various dependencies along the different dependency dimensions. The dependencies that exist within and across the dependency dimensions, in turn, produce vulnerabilities that refer to either structural points of stress within a single evaluation context (critical dependence), or points of stress that span multiple evaluation contexts (cross dependence). Actions initiated by one or more threat agents that target critical dependencies

[9] https://www.impactcybertrust.org/dataset_view?idDataset=702

could produce loss events that hamper certain mission objectives. Finally, the nature and criticality of affected mission objectives shape the mission loss magnitude and the mission loss frequency, with the combination of impact and probability shaping Internet Exposure Risk.

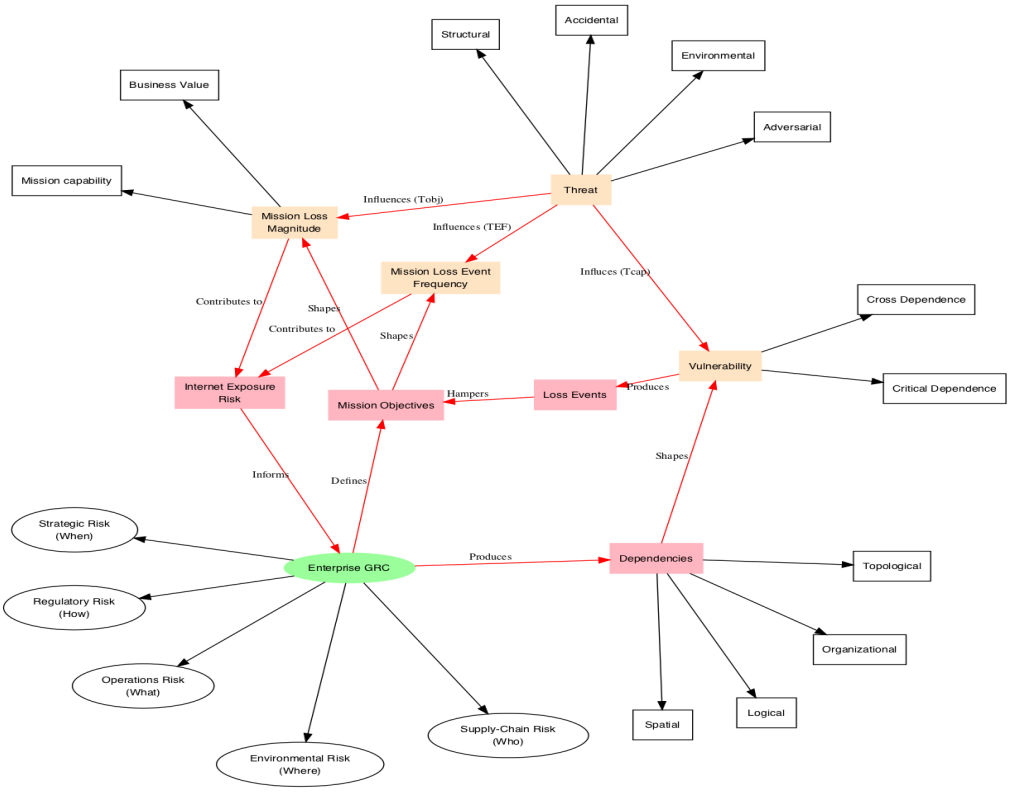

Figure 5. Ontology for Internet Exposure Risk

The threat communities that were considered as part of developing the I-RAM ontology included the basic types defined in NIST 800-30, as well as a more comprehensive collection of types defined in other frameworks such as FAIR and the Intel Threat Agent Library [7]. In general, threat agents can interact independently with vulnerabilities (1) through the threat's capability (Tcap), (2) the loss frequency, through the Threat Event Frequency (TEF) and, (3) the loss magnitude, through the threat's objective (Tobj). For the purposes of the I-RAM ontology, the main interactions that were considered were in relation to the TEF[10].

There are two ways in which the magnitude of loss for a mission could be described, one in terms of business value and the other in terms of mission capability. A detailed description of loss types in terms of business value, and a way to perform a full quantitative-based analysis for such loss events is described in the FAIR methodology for risk assessment. However, in certain cases, quantifying the loss magnitude in terms of lost business value is not the best gauge of impact[11]. In this case, the severity of loss may be best described through the use of a qualitative

[10] The terms Tobj, Tcap and TEF are derived from the FAIR ontology for risk analysis.
[11] This is particularly true in the case of the Homeland Security Enterprise, where the business value is essentially a function of successful mission execution, and where losses are often not quantifiable in terms of a dollar amount.

scale. In I-RAM we used a mixed approach to analyzing risk, by using quantitative measures of vulnerability and loss exposure, and a qualitative scale for loss magnitude. This approach was used, primarily, to focus on those use-cases where the qualitative scale for loss was believed to be more appropriate. However, extensions that enable a fully quantified approach to analyzing risk are envisioned for the future.

The distinctiveness of the I-RAM ontology lies in its placing emphasis on the mission as the unit of risk analysis. This notion of mission was borrowed from NIST 800-39, which identifies the mission as the tier that focuses on the enterprise architecture and embedded information security architecture. The mission, in the context of I-RAM, encompasses the critical endpoints, user communities, and functional objectives being satisfied by the mission. Since risk for a given mission can be analyzed by those who are closest to it, a more granular definition of loss types and loss impact is possible. Specifically, the overlay of different enterprise missions supports a risk accumulation strategy, where it is possible to consider interdependencies and loss events that have the potential to cascade across interconnected enterprise information systems.

3.3 Scoping IER assessment

Once the overall concept and framework for I-RAM were developed, the next phase of the research delved into the details associated with formalizing the specifications for the different I-RAM capability elements. These included defining the parameters of an I-RAM mission, the definition of different mission functional objectives, and the loss types to be considered. These are further described below.

3.3.1 Defining the I-RAM mission.

An I-RAM mission scopes the analysis of risk to a specific contextual environment. We defined a number of parameters that were useful to scope the I-RAM mission definition, including the following:

Vantage points: these refer to the vantage points associated with various topology-related dependency data, including the different route collectors associated with Route Views data, and the sensors from the ITDK dataset discussed earlier. However, these could also refer to vantage points that are associated with other dependency-related datasets, including private sources of data available to the organization, such as local route collectors or looking glasses.

Sites: These relate to attributes that convey some notion of organizational presence or relevance in a particular geographic area. This may include, for example, the site of a field office, location of key customers or users, and the locations where key enterprise operations are performed.

AS interconnection points: These relate to enterprise ASes that interface with the larger Internet. Such ASes may be owned and operated by the enterprise directly or may be ASes associated with the upstream provider that announces the enterprise prefixes on its behalf.

Prefixes: In some cases, it may be more useful to define the enterprise scope in relation to the prefixes that are used, rather than the ASes that announce such prefixes. The prefix parameter provides for this.

Services: This parameter refers to names that are associated with the locations of critical Internet services used within the context of the mission. Additional parameters that help qualify the nature of dependence include the type of service (such as DNS, mail etc.) and identification of whether the service represents a third-party hosting arrangement or not. The advantage of including the hosting type is that it enables one to separate the analysis for two very diverse categories of service providers in terms of the additional data that are required for analysis.

Application Programing Interface (API) endpoints: This parameter refers to a second type of name dependency, specifically in relation to application endpoints that may be used or provisioned in the context of the mission. As in the case of services these endpoints may include additional parameters to qualify the nature of dependence, including the use of hosting arrangements. It is useful to keep the definition of API endpoints separate from service endpoints since the usage context of these types of endpoints are often very different.

Uniform Resource Locators (URLs): URLs are yet another way in which an enterprise mission may interface with the larger Internet. They however encompass more than just the name contained within the URL in that they also include additional logical dependencies that can be inferred from an examination of the web content associated with those URLs.

Names: The mission scope could also be defined in relation to other names that are provided by and relied upon by the mission, which do not fall under the earlier mentioned categories of services, APIs and URLs. Name associations, in general, represent logical dependencies that enable one to take into account threat events that target names in the consideration of risk. Thus, the names parameter enables one to extend the mission scope to cover a range of dependency types including trust points and DNS and Internet resource registration points.

Constraints: A final parameter used in the specification of a mission is the list of constraints associated with the mission, in the form of functional objectives to be met, the source and sink end-points relevant to the construction of the dependency overlays, and the loss types associated with the failure to fulfill a particular functional objective.

3.3.2 Defining functional objectives for a mission.

The constraints for a mission are specified in relation to the fulfilment of certain functional objectives. Since our effort was primarily focused on assessing risk that extended outside the organization's traditional perimeter, we focused on those functional objectives that were most aligned with Internet dependence. Five functional objectives were defined as part of the I-RAM effort, as described below. In each case the vulnerability was defined in relation to the degree to which the given functional objective was likely to be fulfilled.

Diversity: provides a characterization of the structural chokepoints associated with the flow of data between the given set of source endpoints and the given set of destination endpoints. The

lower the diversity the greater the potential for a loss associated with one of the choke points to compromise mission objectives.

Specificity: provides a characterization of the variance in the paths between a given set of source endpoints and a given set of destination endpoints. The lower the specificity, the greater the number of alternative paths, and therefore greater the potential for route manipulations to go unnoticed.

Capacity: provides a characterization of the degree of connectedness along the path from a given set of source endpoints to a given set of destination endpoints. The lower the capacity, the greater the possibility that a single, or group of, nodes in the dependency network can be overwhelmed to compromise reachability between a given set of sources and sinks.

Exclusion: provides a characterization of the concentration of node attributes of a particular type among nodes that match a given criteria. The lower the exclusion, the greater the potential for a given set of environmental factors to produce a negative influence on mission-relevant nodes.

Independence: provides a characterization of the degree of overlap between two missions. The lower the independence, the greater the potential for loss events in one mission to affect functional objectives associated with another mission.

3.3.3 Specifying loss types for a mission.

We surveyed a number of different sources in order to develop a list of loss types for the mission. These included the FAIR ontology, various questionnaire-based approaches for gauging impact and existing literature on loss types covered by the cyber-insurance industry. The list of loss types that was developed through this survey is provided in Table 1.

Table 1. Loss Types

Label	Enterprise Loss Type	Description
L_OP_BI	Operational	Business Interruption
L_OP_DS	Operational	Loss of Data and Software
L_OP_PH	Operational	Physical asset damage
L_OP_IN	Operational	Incapacitation (Ransom)
L_SC_CBI	Supply Chain	Contingent Business Interruption
L_SC_CP	Supply Chain	Compromised products
L_ST_IP	Strategic	Intellectual Property Theft
L_ST_IR	Strategic	Investigation, Incident Response
L_ST_RE	Strategic	Reputational
L_RE_PR	Regulatory	Privacy Breach
L_RE_FP	Regulatory	Fines and Penalties
L_EN_CO	Environmental	Collateral
L_EN_HU	Environmental	Human (e.g Communication and media)
L_EN_ED	Environmental	Environmental Damage

3.4 IER-related measures and metrics

Once the parameters for the mission scope were defined, the next step in our research methodology was to develop risk metrics for each functional objective being analyzed. The process for developing these metrics was comprised of two phases. In the first phase we gathered relevant data in order to serve as our risk baseline. In the second phase we studied various network statistical properties and analyzed the dependency patterns associated with the risk baseline to derive a working set of metrics that could be used to measure our parameter of interest. We describe these phases below.

In order to develop our risk baseline, we needed a representative set of nodes that had a high likelihood of being used as a dependency endpoint within missions, and that were diverse enough to provide a range of network structures that could be analyzed for their structural characteristics. In order to meet these requirements, we decided upon the use of "dominant providers" of certain critical Internet services as our representative set of nodes. The process used for generating this dataset was as follows

1. We identified the IP addresses associated with the DNS, mail, and web-related records of the top 5000 URLs[12].

2. We identified the AS that originated these IP addresses[13], in order to generate a mapping between URLs and their provider ASNs.

3. We identified the top one-hundred providers that had the most common occurrences.

4. For each provider that was identified, we generated the network overlay representations and computed various statistical properties associated with these graphs. The different statistical properties that were examined included graph-level attributes such as node and edge counts, density, assortativity and transitivity, and also a range of different network centrality measures at the node level including the eigenvalue centrality, clustering coefficient, and current flow betweenness centrality[8].

The I-RAM risk metrics were defined as transformations over certain statistical properties that had semantic alignment with the different functional objectives being considered. The measures and transformations also needed to be such that the metrics computed over the risk baseline resulted in a distribution of values that supported the analysis of divergence from the mean. The metrics that were developed through our analysis are summarized in Table 2, while the distribution of values for these metrics are summarized in Figure 6.

It can be observed that the distributions of the metrics identified are not strictly Gaussian; however, the above metrics form a useful starting point in trying to estimate the degree to which metrics derived in the context of a mission deviate from the baseline. We defined the deviation

[12] http://s3-us-west-1.amazonaws.com/umbrella-static/index.html
[13] http://archive.routeviews.org/dnszones/

from the mean, measured in terms of the z-score to comprise the vulnerability exposure in the context of risk evaluation.

Table 2. Risk Metrics

Functional Objective	Measure	Metric (as transformation of measure)	Node Group
Diversity	current flow betweenness centrality	log (averaged measure)	cut sets between sources and sinks
Specificity	clustering coefficient	log (inverse (averaged measure))	set of shortest paths between sources and sinks
Capacity	eigenvalue centrality	log (inverse (averaged measure))	set of shortest paths between sources and sinks
Exclusivity	current flow betweenness centrality	normalized (averaged (measure))	nodes matching a given criteria
Independence	current flow betweenness centrality	edge overlap between nodes that have non-zero value for measure	intersection of nodes across the given set of missions

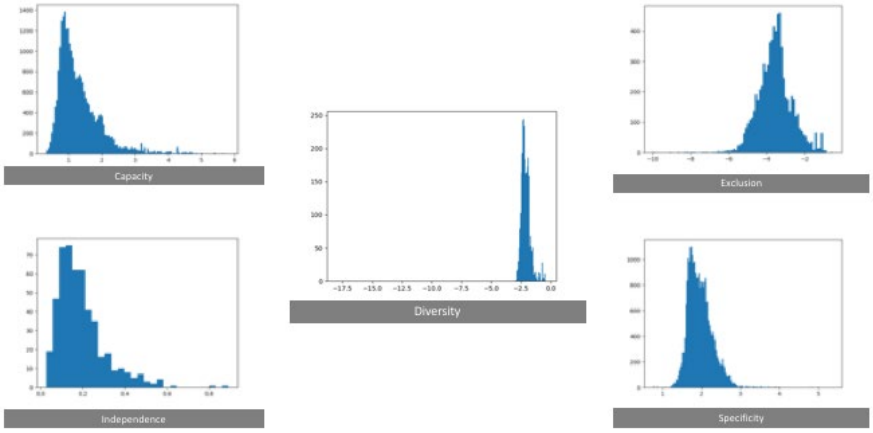

Figure 6. Risk metrics baseline.

3.5 Constructing the threat event library

The measures and metrics discussed in the previous section provide a way to express whether a particular subset of nodes within a mission context are exposed to a condition that results in a failure to accomplish certain mission functional objectives. In addition to this vulnerability exposure, the nodes' proximity to various threat events may also exacerbate the probability of loss. In order to incorporate this element of risk into our analysis, we also developed the conceptual foundations for a threat event library as part of our research.

The design for our threat-event library was based on the examination of two types of threat event data. The first was the set of geospatial datapoints associated with the San Andreas Faultline (see Figure 7). The Faultline represents a natural hazard that has the potential to disrupt the operation of numerous ASes that are in its vicinity. At the same time, the level of exposure for a single AS is also a function of the number of router instances that announce routes from this AS. Thus, the greater the proportion of instances that are in the proximity of the Faultline, the greater the exposure to the hazard associated with proximity to the Faultline.

Figure 7. Geospatial threat-events

The second dataset that we examined was the dataset made available by CAIDA in relation to their analysis of serial hijackers [9]. In the context of I-RAM, proximity to serial hijackers represents a hazard in that there is an increased potential for malicious route advertisements to affect the organization's missions, either through use of such routes directly, or through collateral damage suffered on account of filtering mechanisms that consider organizational network identifiers to be "harmful by association".

While the targets of each threat-event listed above are ASes threat-events could, in reality, target any of the dependency dimensions associated with enterprise missions. We therefore constructed a threat event schema that defined various "risk atoms" to correspond with the various dependency dimensions. The benefit of treating threat-events through the risk atom construct is that it simplifies the consumption of such threat feeds and makes the addition of new threat events and risk atom types more seamless.

3.6 Developing the IER assessment methodology

The final step in our research methodology consisted of developing and formalizing of the idea of risk assessment in the context of Internet dependence. The broad principles associated with mitigating risk through a formal risk management process are discussed in NIST 800-30, which

points toward an iterative process combining the elements of risk framing, risk assessment, risk response and risk monitoring. A similar process was used to model the risk assessment methodology for IER, where the different elements were defined in terms of the constructs developed as part of the IER ontology (See Figure 8).

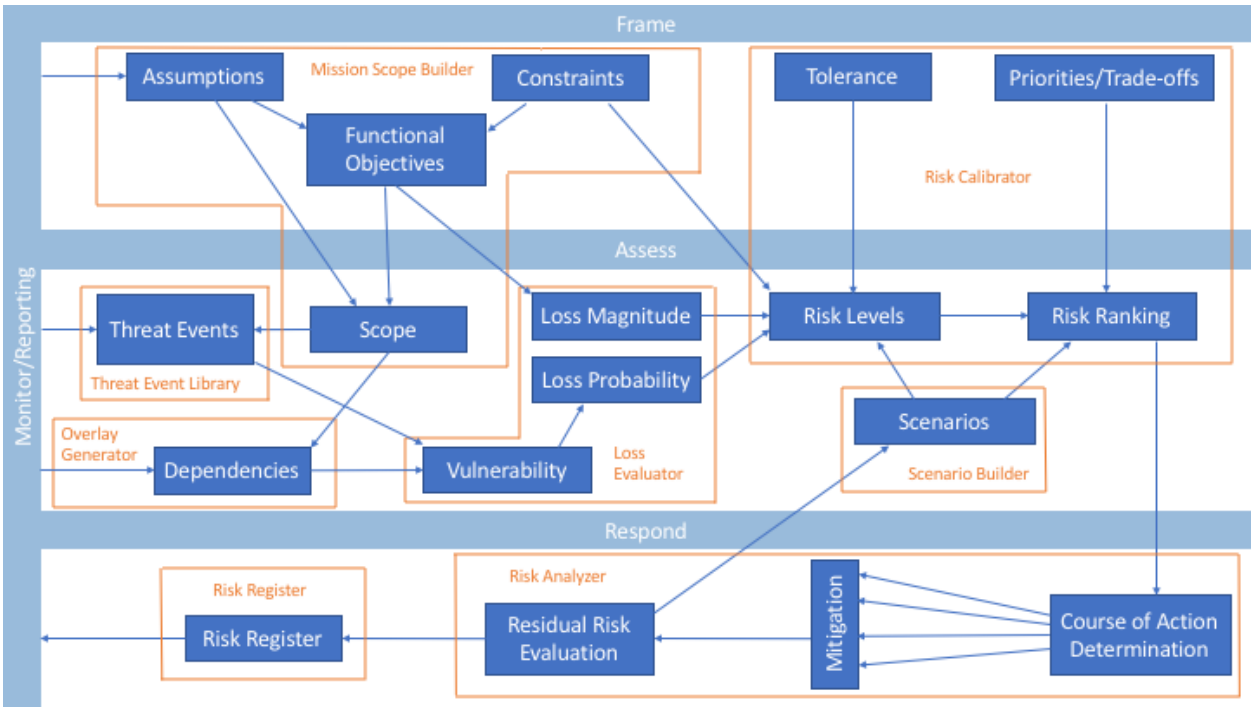

Figure 8. IER Assessment Methodology

The IER assessment methodology builds upon the main elements described in NIST 800-30, but also contains additional components for building a mission scope and evaluating losses in the context of a mission. They are further discussed below.

The "risk framing" element is comprised of two different components, the Mission Scope Builder and the Risk Calibrator. The Mission Scope Builder defines the mission scope though the consideration of assumptions, constraints and functional objectives, and through the selection of threat events of interest. The Risk Calibrator incorporates the organization's views on risk appetite and risk tolerance, in order to construct the different risk levels, and the organizational strategy used in defining a particular risk ranking structure. For the current scope of work, we primarily focused on the Mission Scope Builder, recognizing that development of the risk calibrator would require greater interaction with the stakeholders within an operational setting[14].

The "assess" element combines ideas developed as part of our earlier work on developing the dependency overlays, the ontology for risk, the mission construct, the various measures and metrics associated with vulnerability exposure, and the threat event library. The assessment scope defines the endpoints used in the construction of the dependency graph, and consequently

[14] While many of the capabilities envisioned as part of the IER assessment methodology were out of scope for the funded effort, our overall vision of the set of capabilities was designed to serve as a template for the creation of a more advanced set I-RAM capabilities in the future.

defines the set of threat events that are to be considered either directly or indirectly. The Overlay Generator builds the dependency overlays for the provided mission scope, while the Threat Event Library manages access to threat event data.

The Loss Evaluator computes risk levels through loss magnitude and probability. The loss magnitude is specified directly through the mission scope definition. However, the process used to arrive at a loss probability value from threat events and vulnerability exposure, and the process used to derive a risk level from the composition of the loss probability and magnitude values require some additional explanation.

Our initial attempt for composing a loss probability from threat events and vulnerability exposure values explored the use of Bayesian Networks. Bayesian Networks are a means to include conditional probabilities into a model, while avoiding the unscalable explosion of data that would be required to represent all possible conditional probabilities. Unfortunately, even with the scaling improvement of using Bayesian Networks, there are still a large number of measurements of ground probabilities when dealing with graphs that have the scale of the Internet. Some current work has suggested using Bayesian Network techniques to extend the FAIR model, by improving the FAIR loss event frequency determination to provide a numerical result and to provide for ordering threats in the same threat group. It is possible that we might be able to use such techniques with the data we hold, as part of future work.

As an alternative to using Bayesian Networks we used a vector-based approach in order to compose hazard exposure values from multiple threat events into a single hazard representation. In this approach, each threat event type is defined as a "feature" within a vector representation, with feature values being the normalized hazard exposure values associated with the different threat event types considered for a particular risk event within the mission. The magnitude of this vector is further normalized within the context of the mission to provide a probability of loss. The product of this value with the vulnerability exposure level yields an index for measuring the joint loss frequency. The vulnerability exposure values are also translated to an index along a qualitative scale, with values that are more than two standard deviations from the risk baseline classified as "high", values between one and two standard deviations from the risk baseline classified as "moderate", and other all other values classified as "low".

We used results from existing NIST publications, including NIST 800-30 [5] and NIST 8286 [10] in order to make the determination of risk level based on probability and impact values. Our final mapping between the risk level and the corresponding probability and impact values is tabulated in Table 3.

Table 3. Risk level mapping

Probability/Frequency (computed)	Impact/Consequence (from mission definition)	Risk Level
High	High	High
High	Moderate	Moderate
High	Low	Low
Low	High	Low
Moderate	High	Moderate
Moderate	Moderate	Low

The IER risk assessment methodology also includes elements related to risk response and monitoring. For the current scope of work, the functions are provided by a Risk Analyzer and some elements of a Risk Register. However, the risk mitigation process for IER, as part of a more advanced capability, is expected to be iterative, using a combination of scenario development and course of action selection.

In terms of mitigation measures we considered a number of possibilities for how different risk events could be mitigated in practice. These included changing business processes to minimize exposure to risk through inherited dependencies; changing the precise nature of dependencies, for example by reducing certain types of dependence, using different providers of services or using different geographical areas for the location of critical services or users; performing greater monitoring of dependencies and threat-events that try to subvert them; enforcing SLAs; adding greater redundancy; and asserting ownership over organizational resources such as names and route objects through the use of DNS Security (DNSSEC) [11] and the Resource Public Key Infrastructure (RPKI) [12]. In order to support the use of many of these mitigation measures, however, a more advanced scenario-based analysis component is necessary, which will be explored as part of future work.

4 RESULTS AND DISCUSSION

This section discusses the results from the research work performed as part of the I-RAM project, including the infrastructure that was provisioned and the capabilities that were developed.

The simplified logical architecture for I-RAM is depicted in Figure 9 below. It consists of four components corresponding to the ingestion, aggregation engine, analytics engine, and user interface functions. The different architectural components are provisioned over a set of virtual machines with network containerization techniques being used to simplify the deployment of individual application components. Kernel-based Virtual Machine (KVM)[15] is used as the virtual machine hypervisor, while Podman[16] is used as the containerization framework, since these are the de-facto choices for the Operating System environments used within I-RAM. Network segmentation measures are also used in order to keep the different networked components isolated.

The implementation of each component is further discussed below.

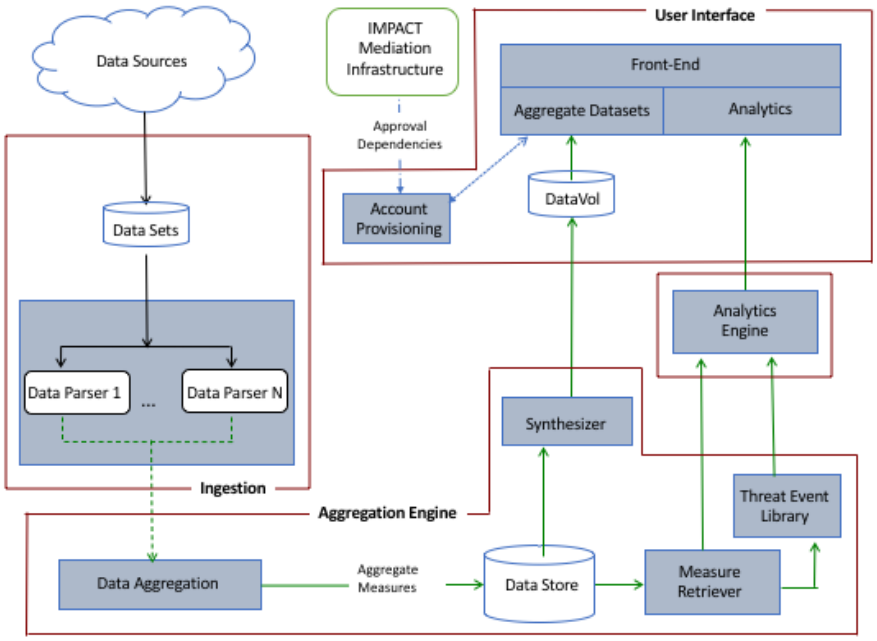

Figure 9. I-RAM architecture

4.1 Ingestion

The ingestion component is responsible for transforming various pieces of raw data into a form that can be stored within the data aggregation component. There are two types of data that are handled by the data ingestion component, namely dependency data and threat event data.

[15] https://www.linux-kvm.org/page/Main_Page
[16] http://podman.io

Dependency data are obtained from IMPACT and other sources and are used to infer dependencies along the topology, organizational, logical and geospatial dimensions. Since the format for each dataset is different, parsers have been developed for each type of data in order to transform them from their raw form to the Javascript Object Notation (JSON) structure[17] suitable for storage within the aggregation engine.

Certain types of data require additional processing prior to being converted to a JSON structure for storage. For example, in order to support the analysis of logical dependencies associated with the use of prefixes, we process the reverse name lookup information from the USC/ISI dataset[18], which provides a mapping between IP addresses and their reverse name, and create an aggregated set of names for every /24 grouping (that is, groupings of 256 contiguous addresses. In this way we are able to associate the name dependencies gleaned from the above dataset with individual prefixes, which support the use case where missions are scoped based on the prefixes in use.

The second type of data that we process through the data ingestion function are threat-event data. The data transformations that are performed here are in keeping with the requirements of the schema that we have defined for our threat-event library. The San Andreas Faultline data, which was the first dataset used within our threat-event library, has been transformed into a set of 'asn' risk atoms where the exposure to the threat-event for each AS is computed based on the number of router instances that are in the proximity to the San Andreas Faultline. The CAIDA serial hijacker dataset, similarly, has been transformed so that in addition to merely conveying the presence or absence of an AS in the serial hijacker dataset, a more continuous range of exposure values have been derived based on the topological proximity of any AS seen in the routing table data to one or more of such serial hijackers. The proximity values have also been encoded as risk atoms and have been stored within our data store as threat-feed data.

4.2 Aggregation Engine

The aggregation engine component provides a consistent way to store data, aggregated and transformed by the ingestion component, to a backend data store. The aggregation engine also provides the ability to retrieve such data through a well-defined API. We use Elasticsearch[19] as our backend store for its ability to scale with growing amount of data. We also investigated the use of a graph database, Dgraph[20], in order to store dependency information, but rely primarily on the Elasticsearch capability for the current effort, since it supports our immediate needs for storing dependency as well as threat-event information very effectively.

The Data Aggregation function within the aggregation engine is implemented using Kafka[21], a widely used distributed streaming platform. The advantage of using Kafka here is that it streamlines the flow of various data between the ingestion component and the backend store.

[17] https://www.json.org/json-en.html
[18] https://www.impactcybertrust.org/dataset_view?idDataset=702
[19] https://www.elastic.co
[20] https://dgraph.io
[21] https://kafka.apache.org

Furthermore, the use of this platform also enables us to ingest, in the future, other third-party sources of dependency and threat-feed data that are distributed over Kafka.

Access to aggregated data is provided through a Measure Retriever module. This module provides the basic interface to the data store for other components such as the Analytics Engine. In addition, the Threat Event Library, which is responsible for the construction of normalized hazard frequencies for the different risk events also retrieves data from the Measure Retriever module.

The primary outputs provided by I-RAM in its role as an IMPACT Data Provider (DP) are aggregate datasets derived from the dependency graphs associated with dominant providers of Internet services. The aggregated datasets containing dependency-related measures are constructed and packaged by the Synthesizer module. The Synthesizer operates in an offline fashion where the necessary measures are constructed through a process that is run on a separate container. The dependency measures so computed are exported, packaged and then copied to a data volume so that it can be shared with external researchers via the IMPACT mediation platform. The dataset provided as part of this capability is shown in Figure 10 below.

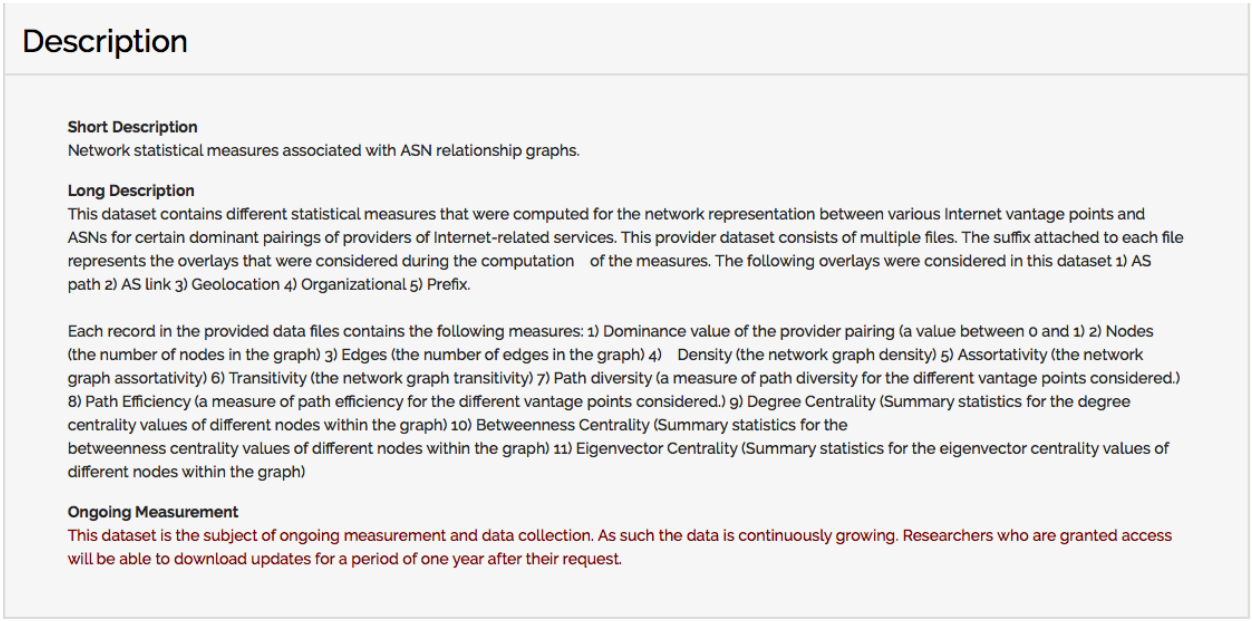

Figure 10. I-RAM data shared on IMPACT

4.3 Analytics Engine

The Analytics Engine builds overlay representations for a given mission context and evaluates risk in terms of the intersection between the different dependency dimensions and mission constraints.

The Analytics Engine is comprised of four sub-components, as illustrated in Figure 11. These are the Mission Scope Builder, the Overlay Generator, The Loss Evaluator and the Risk Analyzer.

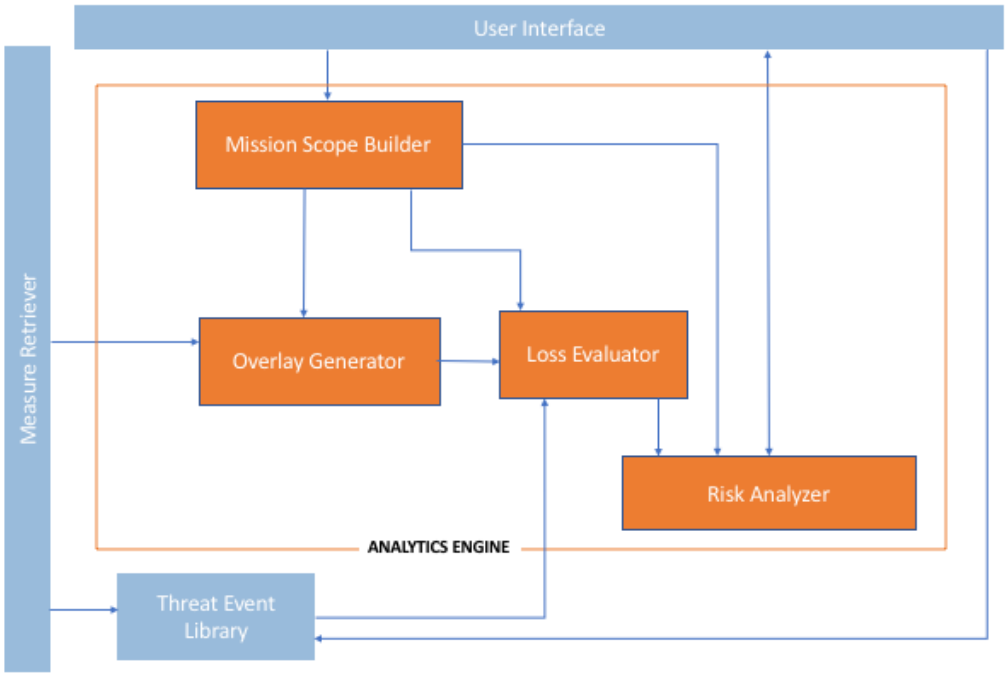

Figure 11. Analytics Engine sub-components

The primary purpose of the Mission Scope Builder is to create a consolidated representation of the mission based on provided mission parameters. Based on the mission definition, the Overlay Generator is responsible for constructing a representation of the aggregate set of dependencies associated with the different dimensions, using mission-relevant vantage points and enterprise network identifiers as sources and sinks. The outputs from the Overlay Generator and the Mission Scope Builder, in conjunction with threat event data obtained from the threat event library, serve as inputs to the Loss Evaluator, which is responsible for identifying the dependencies that are critical to the fulfillment of a given functional objective and the probability of their impairment. Finally, the Risk Analyzer, which is responsible for computing the risk associated with the loss of critical dependencies, serves as the mechanism through which external users can perform queries against the Analytics Engine and obtain risk levels for events through the assessment of loss probability and impact.

4.4 User Interface

The fourth component in the set of I-RAM capabilities is the user interface. This component has evolved over a number of iterations. The initial version consisted of a set of command-line tools that were capable of producing overlay representations for different dependency networks, in addition to highlighting those nodes that were most central based on their computed centrality values (see Figure 12 for sample output produced by these tools).

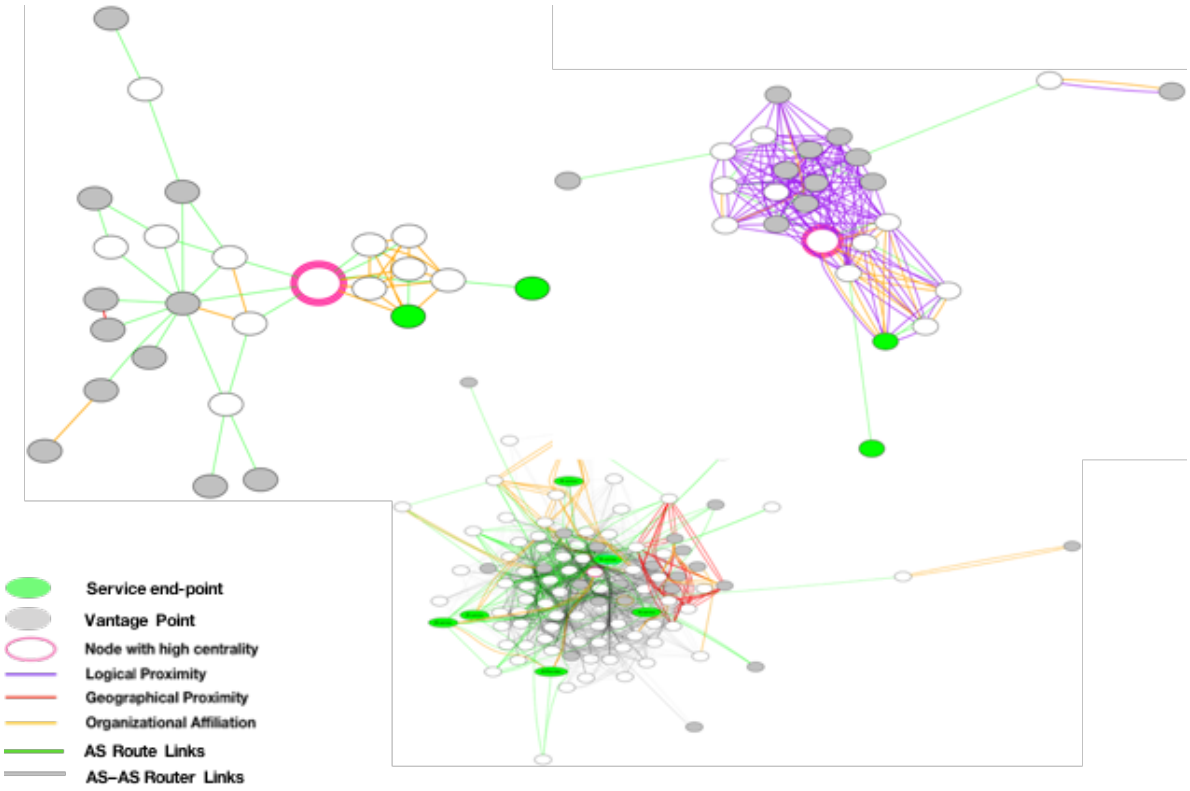

Figure 12. Overlay representation generated by command-line tools

After the analytics capabilities for I-RAM were developed, the user interface was enhanced significantly to support the evaluation of a more extensive set of risk metrics through a web API. This programmatic interface serves as the primary capability provided by the I-RAM project in its role as an IMPACT Data Analytics Service Provider (DASP). The integration of this tool with the IMPACT portal is shown in Figure 13.

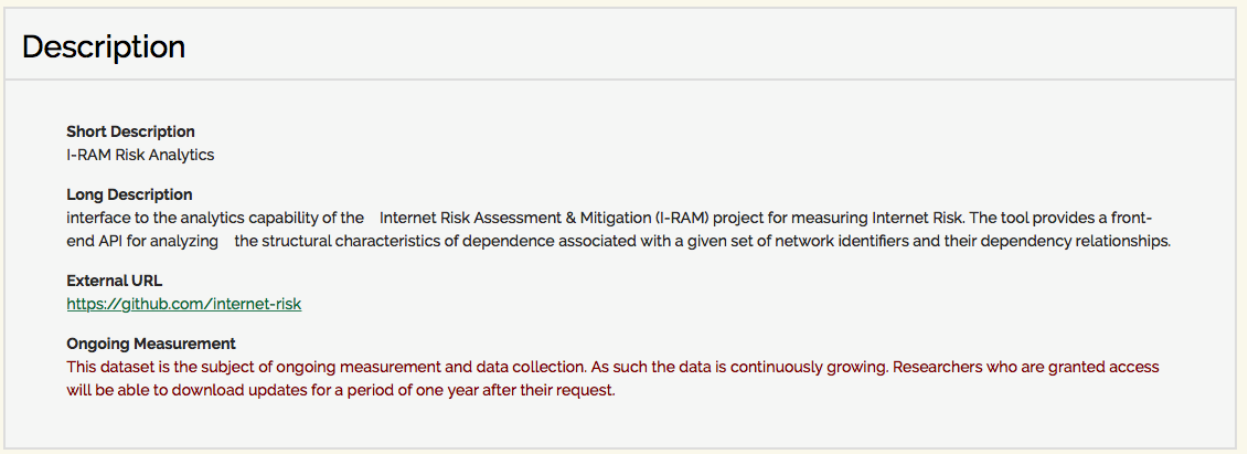

Figure 13. I-RAM Risk Analytics tool on the IMPACT portal

While the programmatic access to the analytics capability, through the API, provides a flexible mode of querying the I-RAM system for risk analytics, additional user interface support was added to provide a more visual representation of risk. A screenshot of this interface is provided in Figure 14. The visual interface to the I-RAM capability includes a number of pieces of information that are useful towards understanding the characteristics of Internet Exposure Risk. It supports the ability to add mission configuration files for analysis and the selection of multiple mission configurations when performing a composite analysis of risk across multiple missions. It also provides a graphical representation of the dependency overlays, where links can be shown either as a bi-partite affiliation network or as a network projection that only shows links between nodes of the same type.

Risk events that are computed through the assessment of dependencies against specified constraints are listed and ordered based on the level of risk inferred from vulnerability exposure values and threat-event proximity values. For each risk event that is selected, the dependency graph is updated to highlight only those nodes that are pertinent to it. The user interface also contains a risk profile plot that shows the manner in which risk events compare against each other and the risk events that are obvious outliers. This enables mission stakeholders to focus on a subset of risks whose mitigation is likely to produce the greatest return on investment.

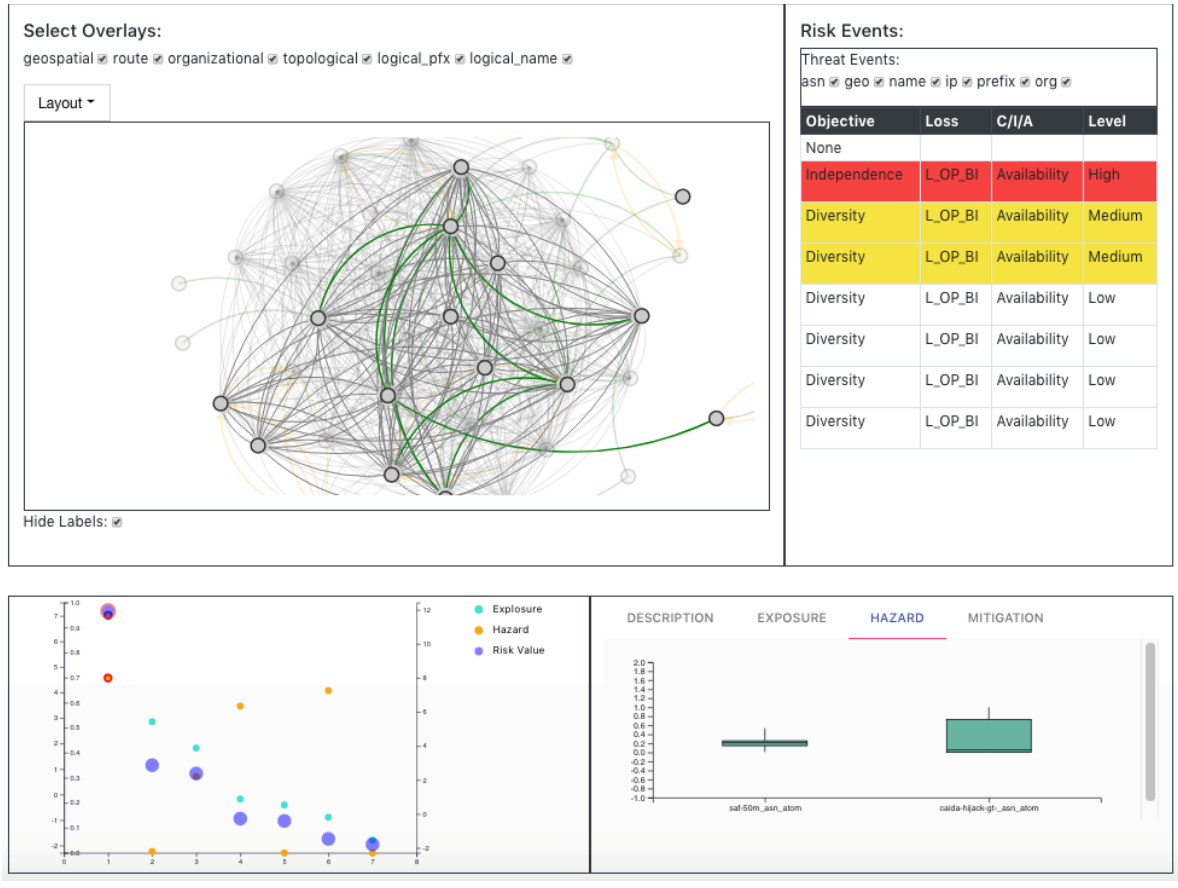

Figure 14. Visual user interface to I-RAM

5 CONCLUSIONS

The primary contribution of the I-RAM project is addressing an existing crucial gap in the risk assessment landscape, by increasing visibility of certain vulnerabilities that are likely to remain unknown without this research. While the idea of Internet Exposure Risk at the commencement of this project was largely abstract, the work accomplished through our effort has provided a concrete representation of that abstract idea and has thus made a key contribution in the area of enterprise risk management. In this section, we provide a number of our important conclusions on the value of the research performed, lessons learned, and examples of future uses of the research that was performed as part of the I-RAM project.

5.1 Value of work performed

The I-RAM project has made several contributions to the area of enterprise risk assessment and mitigation, in the context of managing risk associated with Internet dependence. Some specific contributions include the following.

- We have provided a methodology and an implementation of that methodology that enables enterprise stakeholders to extend their enterprises' risk scope to cover endpoints outside the enterprise boundary, in a way that aligns with existing Enterprise Risk Management best practices.
- We have developed a number of algorithms, scripts and components that serve as building blocks for the quantification and assessment of Internet Exposure Risk.
- We have also provided data to support the independent validation of research performed. The main use of the data is currently in relation to the I-RAM analytics capability for computing and analyzing Internet Exposure Risk. While initial use of this data outside I-RAM expected to be sparse and primarily exploratory in nature, the use and value of this data is likely to grow as more researchers begin to explore this important research space.

Our work has also provided several research contributions. Some of these include the following:

- Through our research, we have made significant inroads into trying to conceptualize and quantify Internet Exposure Risk.
- Through our analysis of different dependency related data, we have developed a useful set of metrics that enables one to characterize mission's alignment with certain specified objectives.
- We have also developed a threat-event framework that provides an extensible way for enterprise stakeholders to incorporate a diverse set of threat events in the computation of mission-level risk.

In addition to the specific contributions made by our research, some broader contributions associated with the use of the I-RAM methodology include the following:

- The I-RAM work provides value to datasets currently being published within the IMPACT portal beyond their primary collection intent.
- The analysis of risk at the mission level supports federation of the risk assessment process, and the aggregation of risk registers across a community of enterprises. Consequently, this risk accumulation strategy has the potential to serve as a bridge between procedural and outcome-based risk management.

- The approach also enables better return on investment of security investments by prioritizing targeted courses of action based on mission impact.
- Finally, the idea of assessing risk through a mission's functional objectives is synergistic with the National Critical Functions construct and the assessment of risk through a "functional" lens[22]. Thus, certain results from the I-RAM project may provide some useful insights on the ways in which an overlay of risk events can be analyzed at the macro or meso-level.

5.2 Lessons Learned

Our research covered many domains of knowledge, many technologies and many bodies of literature. Consequently, there were many insights that were gained but also many lessons that were learned. Some of these lessons are listed below.

- Full quantification of risk, as advocated by certain risk assessment approaches, may be possible in many situations, but it does not remove the need to also provide qualitative outputs of risk levels when presenting a risk scenario to certain types of stakeholders.
- Certain types of data (such as dependency data) are more suited for storage within a graph database, while certain other types of data (such as threat feeds) are more suited for storage within a data store such as Elasticsearch. In practice, a useful backend for risk analysis may have to combine the use of multiple types of databases in order to support needs for a scalable and responsive dashboard for Internet Exposure Risk.
- While a number of different sources of data exist, the aggregation of the various datasets in order to produce enriched representations of those data are not always feasible due to data rights and terms of use associated with that data. Furthermore, for a risk assessment system that relies on third-party data whose use is subject to terms and duration of use limitations, the risk assessment engine must support the ability to expunge selected pieces of data, and re-compute risk inferences based on only those sources of data to which it has access.
- The notion of risk is very context specific. Likewise, the choice of risk baselines may also have to be different based on the specific context of analysis, so that any inferences of risk levels made are not skewed.
- Threat events that target Internet dependency nodes are not necessarily independent. However, the size and density of the network associated with Internet dependencies is considerably large for the efficient application of Bayesian Network and related analysis techniques.

5.3 Future Work

Our work on the I-RAM project has opened up new research possibilities, a number of near-term capability evolution streams, and a number of longer-term strategic evolution possibilities. Some ideas for the near-term evolution of I-RAM include the following:

- Application of I-RAM risk analysis capability to specific mission scenarios and to drive active engagement with enterprise stakeholders, who can help provide guidance associated with risk tolerance and risk strategy.

[22] https://www.cisa.gov/national-critical-functions

- Considering additional sources of dependency data including secondary and tertiary level of dependence through the study of affiliations across various entity types and dependency levels.
- Infrastructure updates to support the use of I-RAM within an operational setting that contained large volumes and types of dependency data, and to support more scalable data ingest pipelines for dependency data and threat feeds.
- Enhancements to support more flexible representations of threat-events, loss, risk, and mitigation. Some threat events, for example, may have greater salience when observed from a particular vantage-point. Similarly, extensions that enable a fully quantified approach for analyzing risk are also likely to be useful in certain situations. In addition, the current risk mitigation alternatives provided by I-RAM are intentionally broad, in order to avoid the use of prescriptive measures based on the limited sources of data. Future evolution of the I-RAM project would need to take into account scenario-based approaches to assist in the selection of mitigation measures, including the use of advanced recommendation capabilities.

Some potential areas of research include the following:

- Refinement and extension of risk metrics based on operational use, validation through real-world examples, and as new sources of data become available.
- Performing a longitudinal analysis of risk metrics and their evolution over time, in order to support the creation of better risk baselines.
- Collaboration with other researchers who are exploring related ideas of dependence and risk at the macro-level.
- Development of additional risk mitigation metrics through the study of the resilience characteristics associated with various structural motifs, and the analysis of risk reduction through selection of alternative paths or providers, the presence of integrity mechanisms (e.g. DNSSEC, RPKI, etc), the use of alternative address identifiers (e.g. IPv6).
- Development and analysis of metrics associated with the provisioning characteristics of large providers of Internet services, as a way to understand the provider level supply-chain risk.
- Integration with third-party enterprise GRC capabilities, where feasible.

Finally, some of the longer-term, strategic goals for I-RAM include the following:

- Extensions to support risk assessments in other related contexts. While I-RAM was designed primarily for the analysis of Internet dependencies and assessment of Internet Exposure Risk, the framework is generic enough to be used in any scenario where dependencies can span multiple dimensions and risk are an emergent property of such dependency networks. The dependency endpoints themselves could consist of more flexible representations of entities, in order to make inferences about risk through dependencies within and across organizational boundaries.
- Development of sector-specific strategies, including Cyber Threat Intelligence (CTI) sharing across communities of enterprises with similar Internet Exposure Risk.
- Assessing risk associated with technological inter-dependence, such as in the cases of mobile 5G networks, zero trust architectures[13], and smart cities[14].

6 REFERENCES

[1] "2019 Cybersecurity Almanac: 100 Facts, Figures, Predictions and Statistics," *Cybercrime Magazine*, Jan. 11, 2019. https://cybersecurityventures.com/cybersecurity-almanac-2019/ (accessed May 18, 2020).

[2] B. Huffaker, A. Dhamdhere, M. Fomenkov, and K. Claffy, "Toward Topology Dualism: Improving the Accuracy of AS Annotations for Routers," *CAIDA*, 2010. https://www.caida.org/publications/papers/2010/as_assignment/index.xml (accessed May 19, 2020).

[3] F. Institute, "Measuring and Managing Information Risk: A FAIR Approach." https://www.fairinstitute.org/fair-book (accessed May 19, 2020).

[4] J. T. F. T. Initiative, "Managing Information Security Risk: Organization, Mission, and Information System View," National Institute of Standards and Technology, NIST Special Publication (SP) 800-39, Mar. 2011. doi: https://doi.org/10.6028/NIST.SP.800-39.

[5] J. T. F. T. Initiative, "Guide for Conducting Risk Assessments," National Institute of Standards and Technology, NIST Special Publication (SP) 800-30 Rev. 1, Sep. 2012. doi: https://doi.org/10.6028/NIST.SP.800-30r1.

[6] "Threat Modeling: Designing for Security." https://threatmodelingbook.com (accessed May 19, 2020).

[7] T. Casey, "Threat Agent Library Helps Identify Information Security Risks," 2007.

[8] M. Newman, *Networks*. Oxford university press, 2018.

[9] C. Testart, P. Richter, A. King, A. Dainotti, and D. Clark, "Profiling BGP Serial Hijackers: Capturing Persistent Misbehavior in the Global Routing Table," *CAIDA*, 2019. https://www.caida.org/publications/papers/2019/profiling_bgp_serial_hijackers/index.xml (accessed May 22, 2020).

[10] K. Stine, S. Quinn, G. Witte, K. Scarfone, and R. Gardner, "Integrating Cybersecurity and Enterprise Risk Management (ERM)," National Institute of Standards and Technology, NIST Internal or Interagency Report (NISTIR) 8286 (Draft), Mar. 2020. doi: https://doi.org/10.6028/NIST.IR.8286-draft.

[11] M. Larson, D. Massey, S. Rose, R. Arends, and R. Austein, "DNS Security Introduction and Requirements." https://tools.ietf.org/html/rfc4033 (accessed May 22, 2020).

[12] M. Lepinski and S. Kent, "An Infrastructure to Support Secure Internet Routing." https://tools.ietf.org/html/rfc6480 (accessed May 22, 2020).

[13] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero Trust Architecture (2nd Draft)," National Institute of Standards and Technology, NIST Special Publication (SP) 800-207 (Draft), Feb. 2020. doi: https://doi.org/10.6028/NIST.SP.800-207-draft2.

[14] "A Risk Management Approach to Smart City Cybersecurity and Privacy," Cybersecurity and Privacy Advisory Committee (CPAC) Public Working Group, Jul. 2019.

[15] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," National Institute of Standards and Technology, NIST Special Publication (SP) 800-145, Sep. 2011. doi: https://doi.org/10.6028/NIST.SP.800-145.

[16] J. Voas, "Networks of 'Things,'" National Institute of Standards and Technology, NIST Special Publication (SP) 800-183, Jul. 2016. doi: https://doi.org/10.6028/NIST.SP.800-183.

7 LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

API	Application Programing Interface
AS(N)	Autonomous System (Number)
CAIDA	Center for Applied Internet Data Analysis
CDN	Content Distribution Network
CTI	Cyber Threat Intelligence
DASP	Data Analytics Service Provider
DNS	Domain Name System
DNSSEC	DNS Security
DP	Data Provider
ERM	Enterprise Risk Management
GRC	Governance, Risk and Compliance
IER	Internet Exposure Risk
IMPACT	Information Marketplace for Policy and Analysis of Cyber-risk and Trust
IP/IPv6	Internet Protocol/Internet Protocol version 6
ISP	Internet Service Provider
IT	Information Technology
I-RAM	Internet Risk Assessment & Mitigation
IER	Internet Exposure Risk
IoT	Internet of Things
JSON	Javascript Object Notation
MSSP	Managed Security Services Provider
NIST	National Institute of Standards and Technology
NTP	Network Time Protocol
R&D	Research & Development
RIR	Regional Internet Registry
RPKI	Resource Public Key Infrastructure
SIEM	Security Information and Event Management
STRIDE	Spoofing, Tampering, Repudiation, Information disclosure, Denial of Service, Elevation of Privilege Authorization
Tcap	Threat capability
TEF	Threat Event Frequency
Tobj	Threat Objective
TCP	Transport Control Protocol
URL	Uniform Resource Locator
VoIP	Voice over IP
ZB	Zeta Byte (roughly 10^21 bytes)

8 GLOSSARY

Autonomous System (AS)

An identifier under the control of a single organization, that represents an administrative entity within the Internet routing infrastructure. The Autonomous System Number (ASN) is used to advertises reachability to a particular set of IP Prefixes within the routing infrastructure.

Cloud-provisioned applications

The National Institute of Standards and Technology (NIST) defines cloud computing as " a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." [15]

Data Provider (DP) and Data Analytics Service Provider (DASP)

These terms refer to the different roles that could be served by providers of different capabilities within the IMPACT platform. The I-RAM project provides capabilities that fulfill its dual-role as a DP and a DASP.

Enterprise Risk Management (ERM)

Refers to the practices associated with evaluating risk within an enterprise and taking the necessary actions to reduce the probability or magnitude of loss.

Internet of Things (IoT)

Interconnected systems, tethered to the Internet, that involve computation, sensing, communication, and actuation [16].

Internet Exposure Risk (IER)

The type of enterprise risk being conceptualized and assessed through the I-RAM effort. IER relates to the risk faced by an enterprise through its dependence on the Internet.

Internet Protocol (IP) Prefix

Network identifiers that cover a range of IP addresses, which serve as the one of the core units (the AS is the other) for advertising reachability to end-points within the Internet routing infrastructure.

Tobj, Tcap, and TEF

Attributes of threats derived from the FAIR ontology for risk analysis.

TCP/IP

The suite of addressing and transport protocols that enable communication between two endpoints on the Internet.

Whois

public information related to Internet resource registration data made available by the Internet registries that are responsible for the allocation of such resources