Reviewer Comments

CIF21 DIBBs: DIBBs for Intelligence and Security Informatics Research Community

NSF Proposal #1443019

Context Statement

The Data Infrastructure Building Blocks (DIBBs) program solicitation (NSF Solicitation 14-530) encourages development of robust and shared data-centric cyberinfrastructure capabilities, to accelerate interdisciplinary and collaborative research in areas of inquiry stimulated by data. Solicitation NSF 14-530 was issued on January 8, 2014 and proposals were due on April 9, 2014.

The program received 52 proposals; one was returned without review. Each of the remaining 51 proposals was considered by a panel of experts. At least three reviewers provided independent reviews on each proposal prior to the panel meeting at which the proposal was to be considered. Reviewers evaluated proposals using two National Science Board (NSB)-approved merit review criteria (intellectual merit and broader impacts), and additional review criteria as specified in the DIBBs solicitation. Each proposal was discussed during the panel meeting, and the panel arrived at a recommendation on each proposal. In cases where there were conflicts of interest, the panelists with conflicts did not review the proposal, nor did they participate in the discussions on proposals where a conflict existed. Panel recommendations to NSF were: "Highly Responsive," "Responsive," or "Not Responsive."

Panel Summary

Panel Summary

Objective of the proposed work

Propose development of a research testbed and archive for the Intelligence and Security Informatics community. Will construct a portal interface for searching and downloading, and develop computational tools.

Panel Discussion

Intellectual Merit
The proposal is based upon a prior Dark Web CRI project which ended in 2012. Based on community use, three areas were identified for improvement:
1. Extend the size of the collection by including contributions from ISI research groups.
2. Support sharing of open source tools

3. Develop collection manager and testbed APIs for data ingestion and content management, and develop collaboration environment.
Will develop testbed builder and extend spider software to run on multiple virtual machines, index terms using Lucene, monitor the workload and status of the crawlers.

Will build testbed manager with relational database, 100 terabyte tape library for backup, and web service APIs.

Will develop portal for search, download, and analysis.

Does have good security control.

STRENGTHS
Will apply to Hacker Forum collection, Honeypot collection, and AZPhish web testbed. The PIs should check for overlap with CERT and other security assessment organizations.

WEAKNESSES
There is an issue with whether the prior Dark Web research will be sufficiently extended. Is this a continuation of effort or a new effort? The project has a very large scope, and there is concern whether all of the tasks can be completed. How much of the work has already been done?

A second concern is that the project is highly internalized to a specific use case. How can this be applied to a broader community? Need broader applicability of the technology, such as analysis of anomalous behavior on the web for data curation.

How will the project be coordinated across the participating communities (Arizona, Drexel, University of Utah)?

Challenges include appropriate interactions with the tape library for managing web pages (10 Kbyte sizes), and the management of links to original sources.

Broader Impacts

The project has a high probability of success. The investigators are leading the field and have the expertise. The subject area will be of greater importance over time, with analysis of security attacks becoming relevant to all data collections. There are adequate resources available for the research.

The project will train graduate students, and the material will be incorporated into multiple classes. The testbed will be used within Data Mining and Business Intelligence classes. Through the NSF SFS program, a testbed will be provided for Ph.D. and Master's students.

DIBBs Responsiveness

The proposal is responsive to the needs of the Intelligence and Security Informatics community. Data harvesting methods will be developed, and a data repository will be created that will be

available to the community. The research conducted on the testbed will be highly relevant to cybersecurity challenges. The project has a high probability of success.

Security and confidentiality of data are appropriately addressed. Integrity is assured through regular backups.

Panel Ranking: Responsive

Summary Statement: A resource will be created for the Intelligence and Security Informatics community that will be used in both research and training of students.

This summary was read by/to the panel, and the panel concurred that the summary accurately reflects the panel discussion.

PANEL RECOMMENDATION: Responsive

Review #1

Rating: Good/Fair

Summary

In the context of the five review elements, please
evaluate the strengths and weaknesses of the proposal with respect to intellectual merit.

The proposal intends to create a litany of components that would benefit the emerging ISI (intelligence and security informatics) community, including a testbed, analytical tools, and web portal. The proposal would build upon prior efforts, and has the support of leading figures in the ISI field.

The opening articulation of vision is rather long, and does go into details on the formation of the ISI virtual organization - not much of which is relevant to the overall description of the work itself.

The proposed budget is missing some details, in particular there is mention of the purchase of a server, but no vendor quotes. Seeing this information would have given the reviewer more confidence in the estimate.

The qualifications of the PIs, and the facilities to be used, are well described.

The reviewer does have concerns about the scope of work to be done in the time allotted. There are many facets to this project, and there is significant dependence on graduate student labor. Historically this form of development resource may not produce a professional product in the end, that could be scaled beyond the life of the project and to other use cases.

In the context of the five review elements, please

evaluate the strengths and weaknesses of the proposal with respect to broader impacts.

The broader impacts of this work are hard to realize. It is true that the proposal would address the needs of the ISI community, and do it well. It is hard to imagine how this would translate into other disciplines beyond the traditional role of computer sciences. There are mentions of benefits to social sciences, but it is hard to see this tie since there are no social scientists mentioned in the grant as being available to provide input or test final results. Beyond this - the narrow scope limits the good the funding would provide, and does not fully address the cross discipline nature of DIBBS.

There is a short passage on sustainability that was helpful to see, and this reviewer does acknowledge thinking about this is a positive move.

Please evaluate the strengths and
weaknesses of the proposal with respect to any additional solicitation-specific review criteria, if applicable

The data management plan is comprehensive and complete for the work that is being accomplished.

The proposal does not fully answer the needs of DIBBs in being cross discipline. This reviewer has no doubt the PIs and senior staff are qualified to speak to the needs of the computer and information sciences, but there is much doubt if and how the work could translate beyond those boundaries.

The letters of support are useful to see that there is buy in from the participating organizations.

Summary Statement

This reviewer found that the opening, with a mention of September 11th, was a little disturbing even if to motivate the need for cybersecurity (there are other ways to motivate this need, without invoking such dark events). It is not necessary to discuss such provocative topics when describing a proposal for CI software.

The proposal plans for much, and could benefit the emerging ISI communities if funded. This reviewer has doubts about the broader impacts of the work, and finds it hard to imagine this would be useful outside of the computer science/engineering disciplines, despite suggestions it may also benefit a narrow grouping of the social sciences.

Review #2

Rating: Good

Summary

In the context of the five review elements, please
evaluate the strengths and weaknesses of the proposal with respect to intellectual merit.

STRENGTHS
Well balanced team across different institutions.

WEAKNESSES
How will the coordination across different universities be done?

In the context of the five review elements, please
evaluate the strengths and weaknesses of the proposal with respect to broader impacts.

STRENGTHS
Detailed and good proposal on teaching and training of students and sharing of tools and results, and engaging other members of the scientific community.

WEAKNESSES
There are no postdocs? Why?

Please evaluate the strengths and
weaknesses of the proposal with respect to any additional solicitation-specific review criteria, if applicable

Providing DBBSs for intelligence and security informatics research.But not very concrete on how exactly they will improve on the current Dark Web and other platforms that they use as a starting point.

Summary Statement

Providing DBBSs for intelligence and security informatics research.

Review #3

Summary

In the context of the five review elements, please
evaluate the strengths and weaknesses of the proposal with respect to intellectual merit.

The proposal addresses the needs of a specific community, namely the Intelligence and Security Informatics community. The authors are experts in the field, and have very strong interactions with the security informatics community.

They will build upon a prior Dark Web CRI project which ended in 2012. Based on community use, three areas were identified for improvement:
1. Extend the size of the collection by including contributions from ISI research groups.
2. Support sharing of open source tools

3. Develop collection manager and testbed APIs for data ingestion and content management, and devbelop collaboration environment.
Will develop testbed builder and extend spider software to run on multiple virtual machines, index terms using Lucene, monitor the workload and status of the crawlers.

Will build testbed manager with relational database, 100 terabyte tape library for backup, and web service APIs.

Will develop portal for search, download, and analysis.

In the context of the five review elements, please
evaluate the strengths and weaknesses of the proposal with respect to broader impacts.

They will build three collections for use by the community: Hacker Forum collection, Honeypot collection, and AZPhish web testbed.

It is not obvious whether equivalent collections are already being assembled by organization such as CERT.

Please evaluate the strengths and
weaknesses of the proposal with respect to any additional solicitation-specific review criteria, if applicable

The project could profitably build upon existing data management systems supported by other NSF infrastructure projects. This includes digital libraries for managing collections.

The authors should consider managing the results of analyses as products supported by the repository.

Summary Statement

There is potential for major contributions to the Security Informatics field. The implementation could be stronger.