Bilateral Analysis of Information Sharing Efforts: Determining the Expected Effectiveness of Information Sharing Efforts

David Mann
The MITRE Corporation
202 Burlington Rd.
Bedford, MA 01730
+1 (781) 271-4719
damann@mitre.org

Stuart S. Shapiro
The MITRE Corporation
202 Burlington Rd.
Bedford, MA 01730
+1 (781) 271-4676
sshapiro@mitre.org

Deb Bodeau
The MITRE Corporation
202 Burlington Rd.
Bedford, MA 01730
+1 (781) 271-8436
dbodeau@mitre.org

ABSTRACT
Cyber security and threat information sharing efforts involve a variety of groups of practicioners and stakeholders. This paper presents a methodology for analyzing information sharing efforts, to determine whether and how well the efforts will succeed. An effort to share information between two groups is represented by a directed graph. Each edge is characterized in terms of the detail of shared information products and the diversity of the work practices of the two groups. The result is mapped onto a Diversity/Detail tradespace, in which successful and unsuccessful efforts can be situated. The framework can be applied to sharing of information security and cyber threat information across different groups, to aid in determining what types of information products could usefully be shared.

Categories and Subject Descriptors
K.6.5 Security and Protection. D.4.6 Security and Protection

General Terms
Security, Standardization

Keywords
Information sharing; cyber security; cyber threat information

1. INTRODUCTION
Cyber security and threat information sharing efforts involve a variety of groups of practicioners and stakeholders. Some efforts are highly successful. Others founder, due to a variety of political, operational, economic, or technical (POET) factors similar to those that affect systems engineering efforts [1] [2].

This paper presents the Bilateral Analysis of Information Sharing Efforts (BLAISE) methodology to determine whether and how well the efforts will succeed. We define an information sharing effort to be the combination of a regularly published information product, the groups that produce and consume the information product and the established work place practices of those groups. An effort is successful if it demonstrates an established ability for both the producer and consumer to collaborate with each other relative to the information product. An effort, or an information product designed for use in an information sharing effort, is considered feasible if BLAISE predicts that the effort will be successful.

We developed BLAISE as an extension of our work in developing cybersecurity information sharing systems and standards. Cybersecurity is an interesting domain because, among other reasons, it is a relatively young discipline. This means that many of its practices and concepts are still evolving. System designers in cybersecurity cannot simply rely on applying automation to existing and well-established collaborative business practices in the way that system designers in other domains have often done. By necessity, we have been forced to look for ways to predict the success or failure of proposed information sharing solutions to support cybersecurity collaborations.

In doing so, BLAISE amounts to taking step back and inferring rather than implying the nature of the problem to be addressed. Rather than assuming, for example, that an information sharing problem is by definition one of data interoperability, BLAISE assists in determining the form of a feasible solution, which may or may not be grounded in data interoperability. (Conversely, BLAISE indicates infeasible solutions as well.) Data interoperability becomes, therefore, one possible analytical conclusion rather than a starting assumption.

For the past six years, we have engaged in cross-disciplinary research with sociologists who specialize in the sociology of language, collaboration and workplace practices. Two key insights have emerged. First, knowledge structures are not universally comprehensible across group boundaries. Second, the degree to which groups can mutually recognize and use knowledge structures is related to the degree to which they share workplace practices. The goal of BLAISE is to frame these sociological insights in a manner that can be understood and acted upon by decision makers and system designers.

We emphasize that BLAISE is designed to shed light on the question of the feasibility of two groups being able to collaborate vis-à-vis an information product. While this is a critical success factor, we recognize that the design and deployment of information systems that span group boundaries can fail to achieve their goals for other reasons. We recommend using BLAISE as one of several analytical tools.

2. BLAISE Overview
The central theoretical insight on which BLAISE is founded is that mutual comprehensibility of shared information products can only be analyzed and understood in the context of a specific collaboration between the two groups who are attempting to share

WISCS'14, November 3, 2014, Scottsdale, Arizona, USA.

http://dx.doi.org/10.1145/2663876.2663880

information with each other. This runs counter to the commonly held assumption that knowledge structures can or should be universally comprehensible.

BLAISE's consideration of an information sharing situation has three major phases. The first involves the development of a "collaboration graph" in which each node represents a different group and a directed edge is added between two nodes if and only if the first group regularly publishes an information product or set of products for consumption and use by the second. One of the primary purposes of this graph is to explicitly identify the groups, information products and collaborations under consideration. This phase is very similar to the construction of flow-chart or concept of operations diagrams.

The second and third phases are applied to each directed edge. In the second phase, the published information product is considered in terms of the detail it encodes and the groups are considered in terms of their work practice diversity. The detail of the information product and the diversity of the groups' work practices are analyzed by way of a set of defined factors, the goal of which is to better understand the potential for shared understanding of the information product. This phase of analysis squarely moves the focus to understanding the POET issues of information sharing from a socio-technical systems perspective.

The third phase of the analysis is to plot the results on a Diversity/Detail tradespace [3]. The key insight of the Diversity/Detail tradespace is that as work practice diversity increases, the level of detail in the shared information products must go down. The purpose of this phase is to discern if the current (or proposed) information sharing scheme is feasible and, if not, to help identify strategies to achieve feasibility. Typically this means either seeking ways to reduce diversity in the groups or reduce detail in the information product. In some cases where the diversity among the groups is very high, it is entirely infeasible for the two groups to directly share information and a new translating mediator capability is needed.

2.1 Building the Collaboration Graph

The purpose of developing a collaboration graph is to help explicitly identify the groups and information products involved. This phase is similar to the familiar approaches of developing a flow chart or concept of operations diagram.

Our understanding of groups is similar to the idea of a community of practice or a professional group. A community of practice is defined as "a group of people who share a craft and/or a profession" [4]. Alternatively, they can be defined as "groups of people informally bound together by shared expertise and passion for a joint enterprise" [5]. Professional groups can be defined as "exclusive occupational groups applying somewhat abstract knowledge to particular cases" [6].

To this we add that a professional group (or community of practice) can be modeled as a node on a collaboration graph if and only if it either publishes an information product to be consumed by another group or it consumes a published information product. We define an information product to be artifacts that are "published and distributed as authorized by a formal organization with the express purpose of being consumed as input by another analytical process or information system." [7] The graph is completed by the addition of directed edges from A to B if and only if A publishes an information product that B consumes.

Figure 1. Directed Edge of a Collaboration Graph

There is a level of analytical choice involved in determining when groups should be joined or split in the model. The idea of a producer or consumer node is distinct from the sociological reality that node represents. When faced with this confusion, it can help to start with the information product itself and then let the group definitions be shaped by considering who produces that information product and who consumes it. We emphasize that it is common for an information product to have multiple consumers, so this approach does not entirely eliminate the analytical choice of when to split or merge groups.

There are two recurring cases that warrant mentioning. First, when the same profession or community of practice exists within two different organizations, it is generally helpful to model this by two different nodes. For example, the doctors in Hospital A might share medical records with doctors in Hospital B.

Second, when members of a group save information products in an archiving system for future use by other members of the same group, we consider this to be a form of publication. In this case, we model this with a directed loop edge from A to itself. For example, it is common for any given doctor in Hospital A to edit and save changes to a patient's medical records for the use and consumption by another doctor in the same hospital. The act of submitting the change in the record is a form of publication.

2.2 The Factors of Detail and Diversity

For any edge of interest in the collaboration graph, the relationship between the producer, consumer and the shared published information product are analyzed in two ways: the level of detail encoded in the information product and the amount of diversity within each and between both groups. Understanding each of these dimensions involves considering several factors. We briefly describe these factors in the following sections and provide more detailed descriptions and scoring guidance in the appendices.

It is crucial to bear in mind that the analysis is driven by the structure of the graph, not the groups the graph represents or the set of nodes in the graph. We do not aim to analyze information sharing combinatorially for an arbitrary set of groups, nor do we aim to analyze information sharing combinatorially for an arbitrary set of nodes. Rather, we focus on explicitly defined edges and, even then, only those edges that warrant attention based on the context of the situation.

Each of the factors are scored on a scale from 1 (less detail or diversity) to 4 (more detail or diversity). The choice of a 4 point scale is somewhat arbitrary and chosen primarily for the sake of simplicity. In the appendices, we give detailed definitions associated with the score, which may be modified as needed. Fractional scores may also be useful in some circumstances. The individual diversity (detail) factor scores are averaged to produce a single overall diversity (detail) score. In cases where one or

more of the factors is particularly significant, a weighted average may be used. The process of scoring the detail and diversity factors is typically done through a series of interviews with primary stakeholders and group members and through the review of documentation that describes the published information products and the work practices of each group. The definitions of the scores given in the appendices can be used as the basis for interview questions with the caveat that some factors may need to be translated to be more effectively applied to a given situation. We also caution against over analysis of the numerical aspects of the quantified scores; quantitative manipulation beyond the simplest is more likely to produce false precision than additional insight. While we present the scoring numerically, BLAISE is a qualitative tool, not a quantitative tool.

2.2.1 The Factors of Detail

The technical aspects of the exchanged information can be characterized according to the fundamental building blocks of computing: syntax (how the information is rendered), semantics (the conceptual organization and accessibility of the information) and computation (how the information is processed). We briefly summarize the factors of detail here and provide more detailed descriptions of the scoring in the appendices.

Syntactic Rendering – The form in which the group's published information products are composed can cover a broad range: Text/Prose (1), Digital Documents (2), Digital/Mark-Up Language (3), Digital/Binary (4).

Conceptual Structure[1] – Successive published documents may each conform to a common, standardized conceptual (semantic) format. The conceptual (semantic) formats can range across: No Consistent Structure (1), Simply-Structured (2), Structured (3), Complex Structure (4).

Semantic Exclusivity[2] –The information captured in exchanged documents can vary in terms of the expected accessibility and comprehensibility of the terms used within the document. The terms and concepts codified within a group's published documents or artifacts can range in audience accessibility across: Outsider (1), Trainee (2), Established Practitioner (3), Expert (4).

Publication Modality – The modality of the initiation and completion of the information exchange can range across: Human/Ad hoc (1), Human/Routine (2), Automated with Human Review (3), Automated (4).

2.2.2 The Factors of Diversity

The factors of diversity are sub-divided into two groups: those related to the work practices associated with the professional domain of the group (professional ambiguity, internal diversity, comparative diversity) and those related to the socio-organizational factors that bear on the group's commitment to the cooperative work (cooperative resistance, process novelty).

Professional Ambiguity[3] – This factor measures the amount of ambiguity in the group's professional work. The type of work a group performs can range across: Routine (1), Largely Routine with Some Esoteric Elements (2), Largely Esoteric with Some Reduction to Practice (3), Esoteric (4).

Internal Diversity – This factor measures the degree to which the work practices within each group are the same or different. Most directly, it is a measure of the degree to which members of a group are fit to perform the work of other members of the group. The similarity of the member's work practices can range across: Same (1), Closely Related (2), Somewhat Related (3), Unrelated (4).

Comparative Diversity – This factor measures the degree to which the work practices of the groups are the same or different. This dimension is defined comparatively relative to the other group. The similarity of the group's work practices can range across: Same (1), Closely Related (2), Somewhat Related (3), Unrelated (4).

Cooperative Resistance – This factor measures the degree to which a group supports or resists the collaboration with the other group. The group's willingness to cooperate in the information exchange can range across: Central to Mission (1), Supporting (2), Peripheral/Opportunistic (3), Unrelated/Hostile (4).

Process Novelty – This factor measures the degree to which the information exchange and the processes that support it are new, or conversely, the degree to which they have been codified in the group's work practices. The extent to which the information exchange is novel can range across: Mature (1), Recent but Established (2), New (3), Desired (4).

2.3 Diversity/Detail Tradespace Analysis

The Diversity/Detail tradespace [3] asserts that there is a basic trade-off between the amount of detail that is encoded in an information product and the amount of diversity that can exist among and between the groups that successfully utilize that information product to collaborate with each other. If there is little diversity, high amounts of detail can be successfully used. Conversely, if there is a lot of diversity, groups will only be able to successfully collaborate with information products with less detail, or with highly diverse groups, not at all. The tradespace is bounded by a frontier along which maximally feasible information products tend to be located. Information products that are inside of this frontier (to the lower left) represent a sub-maximal combination of diversity and detail. The diversity of the group may support more detail (moving up) or the level of detail may be useable by a more diverse group or groups (moving right). Information products that are infeasible are located outside of the frontier (to the upper right), an area we refer to as the "zone of infeasibility".

As we move along the frontier the potential maximally feasible information products fall on a spectrum that ranges from those that support computer automation (upper left) towards those that support human-human communication (lower right) [3]. For the purpose of BLAISE, we identify four major groupings of solutions:

[1] This factor closely follows the "semantic spectrum" as defined by Leo Obrst [8].

[2] This factor is based on observations by Bowker and Star on the relationship between group membership and terms [9].

[3] This factor closely follows the characterization of professional work by Andrew Abbott [6].

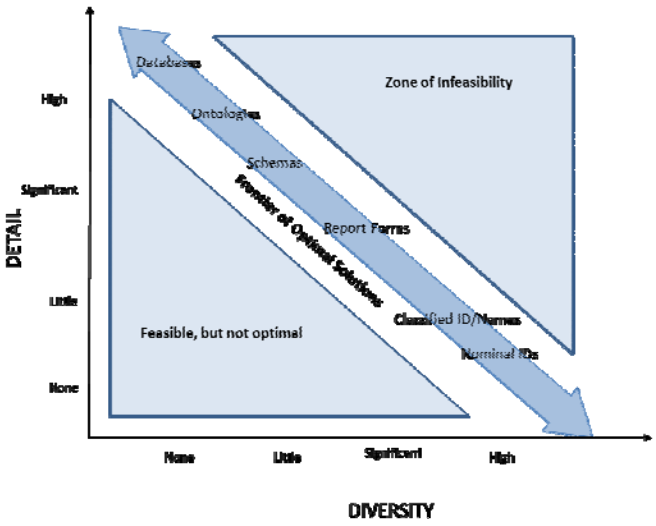

Figure 2. The Diversity/Detail Tradespace

Automation – Databases, formal schemas, implemented domain ontologies.

Reports & Mark-Up Languages – Non-digital structured reporting formats (e.g. police accident report form) and structured data mark-up languages (e.g., XML).

Human Process Aids – Naming schemes, indexing systems, controlled vocabularies, tagging schemes.

Mediated Translation – Situations requiring the intervention of a human analytic translating capability (e.g., medical billing).

2.3.1 Plotting on the Tradespace

As we consider how to translate the individual factor scores into cumulative diversity and detail scores to plot onto the Diversity/Detail tradespace, we must re-emphasize that despite the use of numeric scores, BLAISE remains at its heart qualitative, not quantitative. Our proposed scoring and plotting approaches are offered as aids for qualitatively understanding the socio-technical tensions that relate to the use of the information product in question. While our application of BLAISE to several information sharing efforts has produced scores that are consistent with the asserted frontier, a much larger set of cases studies would need to be performed to begin to place BLAISE on a more quantitative footing. With this important caveat noted, our approach is to take a straight average of the individual factor scores to produce the detail and diversity scores. There may be situations in which there is reason to believe that a particular factor has a more dominant role, which would justify the use of a weighted average or some other approach.

The scores for maximally feasible information products are: (1, 4), (2, 3) and (3, 2), where the scores are in the form (diversity score, detail score). A score of (4, 1) corresponds to a situation where no information product will sustain the collaboration and the use of a translator capability is necessary. We can now associate the scores with the four groupings of maximally feasible information products.

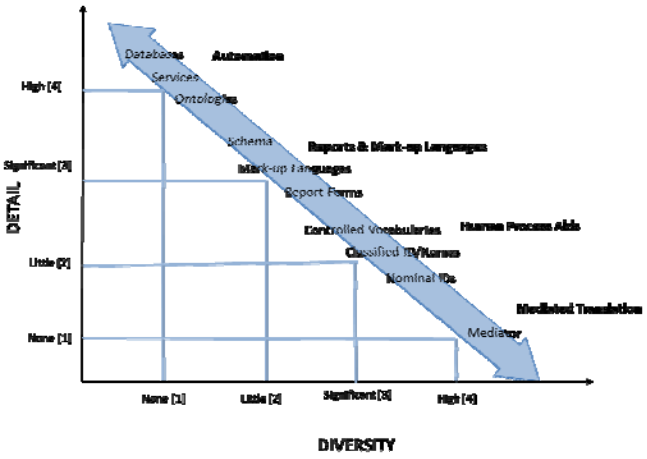

Figure 3. The Four Major Groups of Feasible Solutions

(1,4) Automation – These solutions involve highly automated technical systems that include databases, formal schemas and/or implemented domain ontologies. The producer and consumer groups are in the same profession or in very closely related professions. Much of the work of members of both groups is highly routinized.

(2,3) Reports & Mark-Up Languages – These solutions involve systems that use highly structured and regularly formed information products that contain significant blocks of free text (or its equivalents such as video or audio capture). These may be non-digital structured reporting formats (e.g., police accident report form) and or digital documents captured in non-executable mark-up languages (e.g., XML). The producers and consumers of such information products tend to perform mostly routine work with significant amounts of non-routine work associated with processing anomalous events which are often described in text fields. The producers and consumers work in closely related fields but typically have significant differences in background and training.

(3,2) Human Process Aids – These solutions tend to rely on information products that are not regularly structured but are augmented by the use of standardized elements such as naming schemes, indexing systems, controlled vocabularies and tagging schemes. The information products tend to be captured in digital document form (e.g., Microsoft Word, Adode PDF). The producer and consumer groups tend to work in related fields but have either significant differences in training both within and between the groups or one of the other of the groups may regularly need to analyze anomalous events using non-routine processes. Solutions of this kind are often referred to as boundary objects [7] or as "loose couplers" [11].

(4,1) Mediated Translation – These solutions do not rely on an information product that is directly shared between the producer and consumer. Instead, they rely on a 3rd professional analytical group who can take information products from the original producer and translate information and findings in a manner such that the original consumer can successfully consume them. Just as the first class of solutions is fully automated, this class of solutions is fully reliant on the intervention of a human-analytic translating capability (e.g., medical billing).

2.3.2 Infeasible Solutions

Scores that fall above and to the right of the frontier tend to be associated with collaborations that fail. Two primary strategies are available to attempt to adjust the system so that the information sharing can become feasible: lower the amount of detail in the published information product or lower the amount of diversity within or among the publisher and consumer groups.

For system designers, adjusting the level of detail in the published information product may be the most natural approach. This is particularly true if the producer and consumer simply must collaborate directly, leaving no room to adjust the diversity of those involved.

It is also sometimes possible to move towards feasibility by reducing the amount of work practice diversity of the groups involved. Steps may be taken to reduce the amount of analytical ambiguity in the work, the internal diversity of the producer and consumer groups, the comparative diversity between the groups, the cooperative resistance of one or both of the groups and/or the process novelty involved in the production and consumption of the information product. This might be accomplished in several ways.

First, work place practices could be made more stable and routine. This may help reduce the analytical ambiguity, internal and/or comparative diversity and the process novelty. This may be successful for work practices that are still in the process of maturing and stabilizing. In the cyber domain, both malware and vulnerability analysis have become more routine than they were ten years ago. However, that history can't be rushed and it is reasonable to expect that threat analysis, as a discipline, will be unstable and non-routine for a significant period of time. Further, some disciplines will fundamentally resist complete routinization. Medical diagnosis and criminal investigation offer two obvious examples. Moreover, process routinization can result in a reduction of scope for the work if it is pursued while implicitly ignoring anomalous events that can't be handled with routine processes.

Second, there are cases in which diversity of the groups can be lowered by redefining the groups so as to exclude sub-groups that may be the cause of the diversity. This approach is an explicit reduction in scope in terms of the utility of the information product. Either some sub-groups will be excluded from publishing the information product or some sub-groups will be excluded from consuming and using the information product.

Third, it may be possible to reduce the diversity by adding more economic resource support and/or applying more political pressure to one or both of the groups. Additional resources or political pressure may help facilitate greater work practice alignment. This might be particularly important in newly emerging collaborations in which collaboration paths have yet to be defined and solidified. However, this approach may not work if one or both of the parties are actively resisting the burden of extra work that distracts them from their perceived "real work" [10].

3. BLAISE in Use

We have applied BLAISE to six information sharing efforts that MITRE has been involved with. They highlight how BLAISE scoring can be used to recognize feasible solutions, regain feasibility through corrective actions or consider the introduction of a translating capability.

3.1 Feasible Solutions

We begin by considering two efforts that are feasible. The first is a cybersecurity continuous monitoring effort that seeks to integrate data from a large number of subordinate business units. This effort is a typical attempt to automate the exchange of highly structured data through the use of data exchange standards. The organization desires to collect and correlate cybersecurity assessment data from a large number of subordinate business units. These business units operate with some degree of autonomy and each has their own, well established cybersecurity assessment processes and supporting tools in place. However, the cybersecurity assessment groups at each business group have a data integration role that translates their local assessment data into normalized forms as specified by the central continuous monitoring effort.[4]

The collaboration between the data integrators within the business units and the central continuous monitoring capability is facilitated by a set of data normalization standards. A BLAISE analysis of this information exchange reveals that a key feature of these standards is their inclusion of text fields which allow integrators to attach non-normalized information to their submitted reports. This design decision gives integrators a way to gracefully report anomalies and exceptions that can't reasonably be expressed in delimited attributes; a common feature of successful report formats. The combined diversity score for the exchange is a 2 and the combined detail score is a 3, placing the information products squarely in our "reports and mark-up languages" solution set. While the system is still in development, BLAISE predicts that it is feasible because they have made a commitment to limit diversity to enable that detail.

The second feasible effort is MITRE's Common Vulnerabilities and Exposures list (CVE)[5], which assigns nominal identifiers to software vulnerabilities. The BLAISE Diversity/Detail scores for CVE (during the time of its initial phase) is (2.7, 2), which is consistent with the (3, 2) Human Process Aids category of solutions. The detail score of 2 is consistent with low detail boundary object solutions such as controlled vocabularies and tagging schemes. The diversity score of 2.7 is relatively high, which is consistent particularly with the diversity of the cybersecurity vendors who participated in CVE. Each sub-group thought of vulnerabilities in very different ways and attempted to do different kinds of jobs related to vulnerabilities. Today, CVE IDs are placed in non-standardized documents allowing humans to "pivot" to other information sources as needed.

3.2 Regaining Feasibility

We next turn our attention to two information exchange efforts that were initially infeasible but that took corrective actions to regain feasibility. The first regained feasibility by reducing diversity. The second did so by reducing detail.

The Common Platform Enumeration (CPE) effort was launched in 2007 with the goal of providing a standardized way to refer to IT

[4] The local business unit integrators are an example of a successful use of a mediating translating capability. We discuss other examples of mediated translation in section 3.3.

[5] See: http://cve.mitre.org.

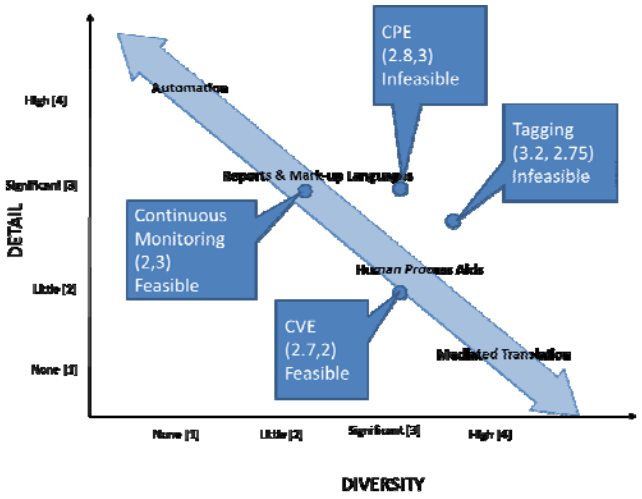

Figure 4: Feasible and Infeasible efforts

products and platforms.[6] It did this through the use of a hierarchical naming structure. However, discussions on the CPE e-mail list through 2008, along with group discussions at the MITRE Developer Days forum held in April 2008 revealed that the CPE community was strongly divided on the best way to structure the naming hierarchy. During the summer and fall of 2009, MITRE conducted a series of interviews with several prominent participants in the CPE community and in February 2009, MITRE released the results of the interviews and asserted that the four primary sub-groups within the CPE community had strong needs for irreconcilably different hierarchies.[7] Application of BLAISE to the 2009 state of CPE finds it infeasible with a diversity score of 2.8 and a detail score of 3. In the fall of 2009, the CPE effort officially decided to only support a single use-case instead of four. After reducing the scope of the effort (and the resulting diversity), CPE became feasible with respect to the shared use of the naming scheme.

The second effort regained feasibility by moving down in the tradespace, to less detail. The effort involved a large repository of data elements collected by researchers. To facilitate further research based on the shared data, the owners of the repository established data tagging requirements which included, among other things, the use of an ontology to describe the physical environment from which the data were collected. This ontology contained nearly 1400 different classes and the data capture using the ontology was not widely adopted. The BLAISE analysis produced a diversity/detail score of (3.2, 2.75), which is infeasible. In 2010, MITRE began to advocate that the complex environment data tagging ontology be replaced with a simpler scheme based on less than 20 classes. Initial reports after the introduction of the simplified tagging scheme were positive, suggesting a successful "move down" in the tradespace to a more feasible, lower detail solution. However, other stakeholders have remained committed to having the data tagged with the more detailed tagging ontology. This conflict underscores that moves to reclaim information sharing feasibility generally require a reduction in either the detail captured or the diversity of those involved, either of which may be actively resisted.

3.3 Mediated Translation

We conclude by considering two different infeasible efforts and some considerations regarding the introduction of a mediated translation capability.

The first effort involved what can be characterized as a business operations group and their supporting IT operations group. When IT failures occurred that negatively impacted business objectives, the business operations group demanded regular reports from the IT group. These reports were not effective and produced an infeasible diversity/detail score of (3.3, 2). To resolve the problem, the organization created a "Cyber Impact" team to act as a mediator between the groups. This group was staffed with senior former members of both the IT and business operations groups. Communication between the IT group and Cyber Impact team was feasible and communication between the Cyber Impact team and the business operations group was feasible. In this way, the Cyber Impact team facilitated effective mediated information flow between the IT group and business operations group in a manner that is analogous to the way that medical billing facilitates information flow between doctors and insurance claims processors.

The second effort involved another cybersecurity continuous monitoring program that wanted to augment the risk management data in its continuous monitoring system with information from threat intelligence data feeds as published by several commercial providers. The premise of this effort was that changes in the threat environment should prompt changes in their risk assessments of the various security posture controls that are being monitored. The BLAISE score for the effort was (2.8, 3), which implies it is not feasible and it's difficult to suggest a good path for rendering it so. It's not clear how the level of detail of the intelligence reports could be reduced nor is it clear how the level of diversity could be meaningfully reduced as threat analysis and posture management are distinctly different from each other.

These two efforts should be considered jointly regarding the option of introducing a mediating translator capability. The first organization that successfully introduced the Cyber Impact group did so by marshalling a significant amount of internal political will. They created a new sub-organization specifically tasked with the translation task, which required significant resources, the development of operational procedures for the new organization and a multi-year commitment to evolve their Cyber Impact capability. In stark contrast, the organization attempting to integrate threat information into their continuous monitoring capability tried, unsuccessfully to pass the translation cost to the threat analysis vendors. But the vendors had no economic incentive since few customers were asking for similar services. Mediated translation may be a way to effectively close the gap in some situations, but it's a costly option.

4. Conclusion

Bilateral Analysis of Information Sharing Efforts is premised on the sociological insight that shared understanding and utilization of information is based on the degree to which groups share common work practices. But professional groups are not static; they evolve and change. Over time, some professions disappear

[6] See: http://nvd.nist.gov/cpe.cfm

[7] See: http://making-security-measurable.1364806.n2.nabble.com/CPE-Technical-Use-Case-Analysis-td2323628.html

entirely. Others emerge and are formed as advances in knowledge and technology make new work both possible and necessary [1].

As professional groups change, the knowledge structures that are associated with them change, as do their collaborative relationships with other professional groups (which are also evolving) [1]. In fields like information technology and cybersecurity where the pace of technological change is rapid, the changes in the landscape of professional groups and their work practices can be staggeringly fast. This change poses a significant challenge to designers of systems and standards for the purpose of sharing cybersecurity information that are akin to playing chess on a moving chess board.

We believe that careful, sociologically informed knowledge structuring can help achieve best possible collaboration in the near term while simultaneously laying the foundation for deeper information sharing in the future and avoiding the costs associated with trying to do too much, too soon. Indeed, as cybersecurity matures as a discipline and as sub-disciplines emerge and resolve their jurisdictional and collaborative boundaries with other sub-disciplines, we are right to expect that cybersecurity work practices will stabilize and align. And as practices align, it is possible for more informational detail to be shared [1].

This type of evolutionary progress from low detail to high detail and automation is borne out by the example of MITRE's CVE list. Fifteen years ago, all that could be shared regarding vulnerabilities was a non-descriptive ID such as CVE IDs of the form CVE-1999-0135. Ten years ago, the industry had the ability to share assessment check logic with standards like the Open Vulnerability and Assessment Language. Today, more fully automated continuous monitoring capabilities are being built.

We assert that effective system design for cybersecurity information sharing requires that designers be able to gain actionable insights into the social factors that constrain shared use of knowledge structures that cross group boundaries. The designers' goal should be to correctly gauge the amount of shareable detail given the amount of diversity among the participants. BLAISE is a tool that can be reasonably applied to any cybersecurity information effort as aid in achieving this balance.

5. References

[1] J. Kruse, S. Landsman, P. Smyton, A. Dziewulski, H. Hawley and M. King, "The POET Approach: A collaborative means for enhancing C2 systems engineering," in *Proceedings of the International Command and Control Research and Technology Symposium*, Fairfax, VA, 2012.

[2] J. Kruse, S. Landsman, P. Smyton, S. Chin, A. Cooper, A. Dziewulski, H. Hawley and M. King, "POET: Integrating Political, Operational, Economic, and Technical Factors into Systems Engineering (PR Case No. 11-1825)," 2011. [Online]. Available: http://www.mitre.org/work/areas/research/2011iebriefings/20MSR058-CB.pdf.

[3] D. Mann, J. Brooks and J. DeRosa, "The Relationship between Human and Machine Oriented Standards and the Impact to Enterprise Systems Engineering," The MITRE Corporation, Bedford, MA, 2011.

[4] J. Lave and E. Wenger, Situated Learning: Legitimate Peripheral Participation, Cambridge: Cambridge University Press, 1991.

[5] E. Wegner, Communities of Practice: Learning, Meaning, and Identity, Cambridge: Cambridge University Press, 1998.

[6] A. Abbott, The System of Professions: An Essay on the Division of Expert Labor, Chicago, IL: University Of Chicago Press, 1988.

[7] D. Mann and J. Brooks, "Information Standards and Their Use: Implications for Design Patterns," The MITRE Corporation, Bedford, MA, 2010.

[8] L. Obrst, W. Ceusters and T. Janssen, "Ontologies, Semantic Technologies, and Intelligence: Looking Toward the Future," in *Ontologies and Semantic Technologies for Intelligence*, Fairfax, VA, IOS Press, Inc., 2010, pp. 213-224.

[9] G. C. Bowker and S. L. Star, Sorting Things Out: Classification and Its Consequences, Cambridge, MA: MIT Press, 1999.

[10] H. Garfinkel, "Studies in Ethnomethodology," Englewood Cliffs, NJ: Prentice Hall, 1967.

[11] R. W. Miller and D. G. Winkowski, "Loose Couplers as a Information Design Strategy," in *Military Communications Conference (MILCOM)*, 2007.

[12] A. W. Rawls and D. Mann, ""The Thing is... What is Our 'What'?": An Ethnographic Study of a Design Team's Discussion of "Object" Clarity as a Problem in Designing an Information System to Facilitate System Interoperability," The MITRE Corporation, Bedford, MA, 2010.

6. Appendix: The Factors of Detail

The technical aspects of the exchanged information can be characterized according to the fundamental building blocks of computing: syntax (how the information is rendered), semantics (the conceptual organization and accessibility of the information) and computation (how the information is processed).

Syntactic Rendering – The form in which the group's published information products are written can cover a broad range:

- Text/Prose (1) – The documents are recorded in a physical, non-digital format, such as print or hand-written documents. Alternatively, the documents are non-machine parseable scans of print document (e.g. XPS, jpeg) or audio or video transcriptions.
- Digital Documents (2) – The documents are recorded and transported in a digital document format such as Microsoft Word, Rich Text Format, Adobe PDF or Open Document Format.
- Digital/Mark Up Language (3) –The documents are recorded and passed digitally in a structured mark-up language such as XML with a corresponding schema or ontology, in a well-defined digital text format such as comma separated or a standardized spreadsheet format.
- Digital/Binary (4) –The information is digitally stored and transmitted in executable, non-human readable binary form that can be directly imported and processed by information processing systems such as a database system.

Conceptual Structure – Successive published documents may each conform to a common, standardized conceptual (semantic) format. The conceptual (semantic) formats can range across:

- No Consistent Structure (1) – Each publication is different in terms of its conceptual structure. Examples include: system documentation manuals, white papers and business correspondences.
- Simply-Structured (2) – The exchanged information conforms to a simple, consistent format. The relationships between entities form lists or simple hierarchies. Each entity or concept is defined by a mixture of delimited fields and text fields. Some of the text fields are tightly constrained (e.g., Name, Street) and others permit free text for some prescribed purpose. Examples include: bank check processing, driver's licenses, medical prescriptions and inter-office memos.
- Structured (3) – The exchanged information conforms to a consistent conceptual format. The relationships between the entities or concepts form complex hierarchical structures (e.g., taxonomies, classification schemes, dictionaries). Entities or concepts are defined by a mixture of delimited fields and text fields. Some of the text fields are tightly constrained (e.g. Name, Street) and others permit free text for some prescribed purpose (e.g., Describe the Accident, Other Surgeries). Examples include: standardized reports (e.g., traffic accident reports, medical history reports, birth certificates) and classified indexes (e.g., Library of Congress, VINs).
- Complex Structure (4) – The exchanged information conforms to a consistent and highly complex conceptual format (e.g. strongly defined entity/relationship model, schema or domain ontology). The relationships between the entities or concepts are complex, forming an arbitrary conceptual graph. Entities or concepts are entirely, or are almost entirely strictly delimited attributes. Text fields that exist are highly constrained (e.g., Name, Street). Database transactions are an example of this type.

Semantic Exclusivity – Terms are considered as material artifacts with social implications. Terms exist as physical inscriptions within books and formal standards used by groups and become physically (or digitally) rendered in the documents published by the group. Thus, the information captured in exchanged documents can vary in terms of the expected accessibility and comprehensibility of the terms used within the document. For example, the health history reports completed by patients in a doctor's office are comprehensible to most people in the general public while, in contrast, the medical notes compiled by the doctor are only comprehensible to trained doctors and nurses. The background necessary to understand the terms and concepts codified within a group's published documents or artifacts can range across:

- Outsider (1) – Most terms and concepts are comprehensible to most people outside of the group.
- Trainee (2) – Contains many terms and concepts that are only comprehensible to trainees in the group and many people outside of the group.
- Established Practitioner (3) – Contains many terms and concepts that are only comprehensible to established practitioners in the group and relatively few people outside of the group.
- Expert (4) – Contains many terms and concepts that are only comprehensible to experts in the group.

Publication Modality – The modality of the initiation and completion of the information exchange can range across:

- Human/Ad hoc (1) – The initiation of the information exchange is irregular and initiated by humans on an "as needed" basis. Human-to-human discussion to clarify and finalize the exchange is expected.
- Human/Routine (2) – The exchange process is formalized and well-structured but is initiated by a human judgment or decision. Human-to-human discussion to clarify and finalize the exchange is not exceptional.
- Automated with Human Review (3) – The initiation of information exchanges are automated but, are reviewed prior to being executed. Ongoing human-to-human negotiation to clarify an information exchange is rare and exceptional.
- Automated (4) – The initiation, processing and completions of information exchange transactions are automated. No human-to-human negotiation is needed to complete the transaction.

7. Appendix: The Factors of Diversity

The sociological characteristics are sub-divided into two groups: those related to the work practices associated with the professional domain of the group (professional type and comparative fitness) and those related to the socio-organizational factors that bear on the group's commitment to the cooperative work (commitment and establishment).

Professional Ambiguity – In *"The System of Professions"*, Andrew Abbott describes a range of professional work in terms of the analytical judgments that are performed, ranging from routine to esoteric. Routinized work can be entirely described in terms of established procedures. Esoteric judgments handle anomalous cases that can't be handled by written rules. For example, medical billing is largely routine. Medical diagnosis is a mixture of established routine practices and esoteric judgments. Some work in dynamic new, emerging fields with no established practices may be mostly or entirely esoteric. Cybersecurity threat and malware analysis are such examples. Esoteric work tends to be more unpredictable in terms of outcome. For instance, different doctors might diagnose the same patient differently, hence the practice of getting a "second opinion".

The type of work a group performs can range across:

- Routine (1) – The work is highly regulated in terms of established procedures. The outcomes are predictable and relatively few anomalous cases are handled. Published outputs vary little depending on the individual worker performing the work or when the work is performed. In cybersecurity, the running of anti-virus agents on end systems, which includes the

regular download of new virus signature definitions is an example of this kind of work.

- Largely Routine with Some Esoteric Elements (2) – The work has a significant amount of established procedures. A mixture of routine and anomalous cases are handled. The outcomes for routine cases are fairly predictable. For routine cases, published outputs vary little depending on the individual worker performing the work. In cybersecurity, system patch installation and prioritization decisions, is an example of this kind of work.
- Largely Esoteric with Some Reduction to Practice (3) – The work has few established procedures. As new cases are handled, some attempt is made to find similar prior cases and to use those as a model. Published outputs vary dramatically depending on the individual worker performing the work, but strong similarities among outputs by the same worker can be found to increase over time. In cybersecurity, malware and vulnerability analysis are examples of this kind.
- Esoteric (4) – The work has little to no established procedures. The outcomes are unpredictable and most cases handled are new and anomalous. Published outputs vary dramatically depending on the individual worker performing the work and on when the work is done. In cybersecurity, threat analysis is an example of this kind.

Note that the type of work can change over time. Thus, retrospective analyses of information exchange efforts may need to account for shifts in practice.

Internal Diversity – This dimension measures the degree to which the work practices within the group are the same or different. Most directly, it is a measure of the extent to which members of the group are fit to perform the work of other members of the group. More generally, it is a measure of the work practice diversity within the group, without consideration of the other group involved in the information exchange.

For example, if the group is defined as "the medical billing agents at Liberty Hospital", we would expect that there is practically no work practice diversity and that any member of the group could perform the work of any other with no significant retraining. If the group is defined as "medical billing agents in the US", we would expect less mutual work fitness and greater work practice diversity. If the group was defined as "cybersecurity analyst", we would expect almost no mutual work fitness and very high work practice diversity.

The similarity of group members' work practices can range across:

- Same (1) – Members of the group could perform the work of other members of the group with minimal amounts of orientation. They typically have comparable education, training and work experience.
- Closely Related (2) – Members of the group could perform the work of other members of the group with significant but reasonably achievable retraining. Members of the group typically have similarities and overlaps in their education, training and work experience, but also have significant amounts of training, education and experience that are different.
- Somewhat Related (3) – Performing the work of other members of the group would require lengthy retraining and education and would carry a significant risk of failure. The training, education and work experience of a typical member is significantly different from that of other members of the group.
- Unrelated (4) – It is not expected that members of the group could perform the work of other members of the group, even with reasonably substantial retraining and education (e.g., significant career change). The training, education and work experience of a typical member of the group are very different from that of other members of the group.

Comparative Diversity – This dimension measures the degree to which the work practices of the groups are the same or different. For example, it is common for groups that perform the same kind of work in different organizations to be able to do each other's work, as in the case of emergency room doctors at two different hospitals. Conversely, cybersecurity threat analysts and corporate chief information officers cannot do each other's work.

In some cases, it may be possible for members of one group to perform the work of the other group, but not vice versa. For example, most threat and incident analysts are qualified to act as Intrusion Detection System (IDS) analysts, but not all IDS analysts are qualified to be threat or incident analysts, which are typically more senior positions. For this reason, it is important to consider this factor bilaterally.

This dimension is defined comparatively relative to the other group. The similarity of the groups' work practices can range across:

- Same (1) – Members of the group could perform the work of the other group with minimal amounts of orientation. They typically have education, training and work experience comparable to that of members of the other group.
- Closely Related (2) – Members of the group could perform the work of the other group with significant but reasonably achievable retraining. Members of the two groups typically have similarities and overlaps in their education, training and work experience, but also have significant differences in training, education and experience.
- Somewhat Related (3) – Performing the work of the other group would require lengthy retraining and education for members of this group and would carry a significant risk of failure. The typical training, education and work experience of members of this group are significantly different from that of members of the other group.
- Unrelated (4) – It is not expected that members of the group could perform the work of the other group, even with reasonably substantial retraining and education (e.g., significant career change). The typical training, education and work experience of members of this group are very different from that of members of the other group.

Cooperative Resistance – "It takes two to Tango", as the adage says. The same is true for cooperative work. We consider groups that are both highly committed to the cooperative work to be unified in their work practices, since they both share enacted practices that make the cooperative work a reality. Conversely, we consider groups that resist the cooperative work to be more diverse in their work practices, since they don't share work practices necessary to realize the cooperation.

We consider the degree to which a group is committed to cooperative work as enacted. That is, we do not consider their stated or self-described commitment. Nor do we consider the degree to which the group's parent organization (if there is one) or legal or contractual commitments compel the group to participate in or support the cooperative work. Instead, we focus solely on the degree to which the group participates in the cooperative work and fulfills its obligations to the other party (and possibly to its parent organization and legal authorities).

In considering the group's resistance or commitment to the cooperative work, we also consider the relationship to groups that may be a direct competitor of the other group. For example, when we consider the cooperative work between a car salesman and a car buyer, the fact that there are other car dealerships who are competing for the car buyer's business is a potentially important factor in understanding the car buyer's commitment to the cooperative work with the salesman. In this case, we would say that there are potentially disruptive competitors to the dealer, because the buyer's contact with the competitors may result in the buyer deciding to purchase a car from one of the competitors. Not all interactions with potential competitors is potentially disruptive to the cooperation.

Conversely, if a police department has an exclusive contract to buy all of its police cars from a particular dealer and, most importantly, if the police department demonstrably adheres to that contract, the fact that there are potential competitors to the dealership has no bearing on the police department's commitment to the cooperative work.

The group resistance to the information exchange can range across:

- Central to Mission (1) – The information exchange is critical to the group's ability to achieve its mission objectives. The group may engage in information exchanges with organizations that compete with the other group, but they are infrequent and do not disrupt the cooperative work.
- Supporting (2) – The information exchange is important to the group's ability to achieve mission objectives or to perform supporting functions effectively. The group may engage in information exchanges with organizations that compete with the other group, but they are unlikely to disrupt the cooperative work.
- Peripheral/Opportunistic (3) – The group may be hopeful about realizing benefit from the information exchange, but currently it provides limited value to the group's ability to achieve mission objectives, either directly or indirectly by facilitating supporting functions. The group engages in information exchanges with organizations that compete with the other group, and there is significant risk that they will disrupt the cooperative work.
- Unrelated/Hostile (4) – The information exchange provides no value, or is perceived to be a threat to the group's ability to achieve mission objectives, directly or indirectly. The group perceives significant threat from the information exchange or has a strong preference to engage in a collaboration with a competitor to the other group.

Process Novelty – This factor measures the degree to which the information exchange and the processes that support it are new, or conversely, the degree to which they have been codified in the group's work practices. Novelty is related to but different from resistance. Resistance is a measure of the degree to which the group does or does not do the work of executing the information exchange. In contrast, novelty is a measure of the degree to which the group has done prior work developing a persistent operational infrastructure to support the execution of the information exchange, including but not limited to: the production of documentation, the allocation of budget and staff, and the degree to which the information exchange is codified in contracts, laws and regulations

As is the case with resistance, we consider groups to be more unified in their work practices if the information exchange is well established in both groups' environment and more diverse if it is more novel. The extent to which the information exchange is novel can range across:

- Mature (1) – The information exchange has been successfully performed for a significantly long time and is very well-established. The information exchange is routinely funded and staffed. The collaborative work is described or documented in maintained written practices. The group actively participates in the information exchange to satisfy a binding or compulsory mandate (e.g. under law, regulation, or contractual agreement) with known sanctions for non-compliance and regular enforcement.
- Recent but Established (2) – The information exchange has been successfully performed in the past and is established. The group has just enough funding and staffing to exchange information effectively, but ongoing funding and staffing cannot be assumed. The collaborative work is not well documented in written practices. The information exchange is referenced in non-binding best practices.
- New (3) – The information exchange is new to the group and has only been established in the recent past. The current funding and staffing have only been allocated recently and ongoing funding and staffing have not been established. There are no stable, written practices. The collaboration is non-compulsory. There is no sanction for non-compliance.
- Desired (4) – The information exchange has not yet been established, but the group is attempting to establish it. Operational resources have not yet been allocated. Repeated procedures have not yet been created, nor documented.