Army Science Planning and Strategy Meeting: The Fog of Cyber War

by Alexander Kott, Ananthram Swami, and Bruce J West

Approved for public release; distribution is unlimited.
NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer’s or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.
Army Science Planning and Strategy Meeting: The Fog of Cyber War

by Alexander Kott and Ananthram Swami

*Computational and Information Sciences Directorate, ARL*

Bruce J West

*Army Research Office, Durham, North Carolina*

Approved for public release; distribution is unlimited.
Army Science Planning and Strategy Meeting: The Fog of Cyber War

The Army Science Planning and Strategy Meeting on The Fog of Cyber War took place on January 7–8, 2016, at the US Army Research Laboratory’s Adelphi Laboratory Center. The meeting examined the theoretical foundations of the “fog of cyber war” concept for Army battlefield operations.

The workshop identified several key research questions associated with this concept:

1) What theoretically grounded models can help characterize the complex tradeoff inherent in radical dispersion of information among mobile tactical edge devices (and related diversification of channels and protocols), including tradeoffs of communications overhead, energy consumption, and security impacts?

2) What approaches, including semantic-based techniques, can help minimize the impact of dispersion on timely, secure and efficient regathering of information in a fashion that would support formation of situational awareness appropriate to the time, place, and mission of the user?

3) Could risk or other related metrics, along with new analytical methods that are in part game-theoretic, serve as a comprehensive framework for characterizing the “goodness” of cyber fog?

4) What formal models, theories, methods, and tools can be devised to execute and manage successful obfuscations of friendly information within cyber fog?

Type of Material

Information dispersion, data splitting, cyber warfare, cyber resilience, secret sharing, obfuscation, deception, fog computing, situational awareness

Distribution Statement

Approved for public release; distribution is unlimited.

Abstract

The Army Science Planning and Strategy Meeting on The Fog of Cyber War took place on January 7–8, 2016, at the US Army Research Laboratory’s Adelphi Laboratory Center. The meeting examined the theoretical foundations of the “fog of cyber war” concept for Army battlefield operations. The workshop identified several key research questions associated with this concept:

1) What theoretically grounded models can help characterize the complex tradeoff inherent in radical dispersion of information among mobile tactical edge devices (and related diversification of channels and protocols), including tradeoffs of communications overhead, energy consumption, and security impacts?

2) What approaches, including semantic-based techniques, can help minimize the impact of dispersion on timely, secure and efficient regathering of information in a fashion that would support formation of situational awareness appropriate to the time, place, and mission of the user?

3) Could risk or other related metrics, along with new analytical methods that are in part game-theoretic, serve as a comprehensive framework for characterizing the “goodness” of cyber fog?

4) What formal models, theories, methods, and tools can be devised to execute and manage successful obfuscations of friendly information within cyber fog?
Contents

1. Introduction .................................................. 1
2. Feasibility, Value, and Challenges of Dispersion ............ 4
3. Dispersion and Effective Regathering ........................ 5
4. Situational Awareness and Information Semantics .......... 5
5. Risk and Mission ........................................... 7
6. Deception and Obfuscation .................................. 8
7. Applicability of Formal Methods ............................ 10
8. Conclusions .................................................. 10
9. References .................................................... 12
10. Other References Suggested by the Workshop Participants 13
List of Symbols, Abbreviations, and Acronyms ................ 14
Distribution List .............................................. 15
List of Figures

Fig. 1   The Fog of Cyber War concept explored at the workshop seeks to enhance survivability of battlefield information by splitting and dispersing it among the multiplicity of edge devices.........................2
1. Introduction

The Army Science Planning and Strategy Meeting on The Fog of Cyber War took place on January 7–8, 2016, at the US Army Research Laboratory’s (ARL) Adelphi Laboratory Center and was organized by ARL. The focus of this meeting was to examine the theoretical foundations of the “fog of cyber war” concept for Army battlefield operations. Clausewitz’s fog of war spoke of uncertainty in information, at a time in history when information was synonymous with knowledge, a situation that no longer exists. More recently, the development of Internet technologies has led to cloud computing, which, depending upon the situation, some refer to as a fog rather than a cloud. These seemingly disparate notions of fog merge when one considers how cyber space is used now in conflict versus how it will be used in the future. One possibility that retains the security of friendly networks and information is to maximize the “fogginess” of the friendly information as it appears to the adversary. Networks at the tactical edge—and the tactical information they carry—must be resilient to cyber and electromagnetic operations by a capable adversary; even when partly compromised, they should remain opaque to the adversary and effective for friendly forces.

One concept for achieving such an opaqueness is radical fragmentation (splitting) of friendly data into a large number of fragments (cyber fog) and to continually maneuver those fragments across multiple devices at the battlefield edge (i.e., end user) networks (Fig. 1). Data splitting for security and scalability is practiced in many modern commercial data stores (Voldemort of LinkedIn, Dynamo of Amazon, etc.), but not for tactical, edge devices and networks. With growing interests in fog computing and fog networks (e.g., Chiang 2015), and maturing of edge-network distributed data stores (e.g., GaianDB, a product of ARL’s UK-US International Technology Alliance [Bent et al. 2016]), the Army should explore the first tactical use of data splitting. A focus of the meeting was to determine the consequences of such fragmentation at the tactical edge.
Fig. 1 The Fog of Cyber War concept explored at the workshop seeks to enhance survivability of battlefield information by splitting and dispersing it among the multiplicity of edge devices

While potentially offering a number of military-relevant benefits (e.g., resiliency to adversary electronic warfare, cyber and kinetic attacks, and intercept; agile maneuvering of data, rapid recovery, obfuscation and deception), concepts of this nature also present formidable challenges of complexity of network management, data management, and reassembly; demands on bandwidth, storage, and battery power; latency of reassembly; and impact of intermittent connectivity. The excessive amount of data also results in information uncertainty, which becomes crucial in the reaggregation of strategically fragmented data.

The goal of the meeting was to identify fundamental research issues that need to be addressed, which may enable future military-relevant capabilities. Participants were asked to identify gaps in scientific understanding and describe how to apply existing scientific understanding to establish bounds on performance. The meeting encouraged structured yet open and broad-ranging discussion and exploration of multiple perspectives on the issues.

The meeting’s topics included the following:

- Methods for and of underlying theoretical models for securing information by fragmenting, dispersing, and moving it in fragmented form across multiple tactical heterogeneous devices.
- Analysis, synthesis, and prediction of behaviors, structure evolution, and emergent phenomena in such highly dynamic systems of information and networked devices; phase transitions; controllability and system identification and state estimation; role and behavior of human elements in
such a system, including the dynamics of human comprehension, trust, and confidence in the system.

- Approaches to characterizing tradeoffs of potential benefits and added vulnerabilities (e.g., lower vulnerabilities to capture of devices, keys, and data in tactical environments; data exfiltration; loss of availability due to cyber electromagnetic activities [CEMA] effects; less need for long range communications; increase in replication without greater danger of data loss to adversary; obfuscation of friendly [EOB] and portrayal of deceptive EOB; complexity of information management; tradeoffs between computing power, storage, communications bandwidth, energy consumption and latency).

- Formal languages for representation, analysis, synthesis and provably correct construction of deceptions; techniques for effective, near-automated execution of deceptions against a near-peer adversary that eavesdrops and otherwise attacks the friendly information via CEMA; theory and methods of control for data flows that allow deceptive modifications of apparent communications patterns.

- Computational methods and underlying theory for analyzing and quantitatively managing risk to friendly information, particularly the survivability of information in terms of maintenance of confidentiality, integrity, availability; and characterizing uncertainty of such assessments.

- Approaches to continually assess information needs of Soldiers from context and mission information; potential enhancements, such as reduction in latency, by learning the user information demand model and using the model to modulate the degree of splitting, the distance of dispersions, and prepositioning of data.

- Approaches and supporting theory for fusion and (re)generation of needs-relevant information from highly fragmented and dispersed data; ensuring high-quality, fused information to friendly forces; maintenance of situational awareness (SA) for the Soldier in spite of extreme volume, dynamics, and dispersion of the information.

The following annotated and extended notes capture some of the discussions and findings of the meeting. A few disclaimers apply to this report. First, not every author of the report or participant of the workshop agrees with every (or any) opinion presented in the workshop's report. Second, the views presented in this report are those of the authors or of the workshop’s participants, and do not reflect positions of their employers.
2. Feasibility, Value, and Challenges of Dispersion

Research and practical successes of the database security community have already demonstrated, to a large extent, the feasibility and value of data dispersion, and to a lesser extent, frequent repositioning of data fragments. Research and successful products exist that use some forms of fragmenting, dispersing, and frequently repositioning data “shards”. For example, Mei et al. (2003) presents a distributed algorithm that uses replication and fragmentation schemes to allocate the files over multiple servers. The file confidentiality and integrity are preserved, even in the presence of a successful attack that compromises a subset of the file servers. Further exploration of the cyber fog concept would benefit from interactions and collaboration with the database security community.

However, dispersion of data is but one of many ways of increasing diversification—and thereby uncertainty to the adversary—within a communication system. Other examples include diversification of channels, protocols, and media. Software-defined networks (SDNs) are potentially effective mechanisms for increasing such a diversification. Diversification is helped by applying it over the large scale of the network of devices and channels, which suggests that there are benefits in dispersing information not only over friendly networks but also over civilian networks and even over the adversary network.

The workshop participants frequently mentioned Shamir’s Secret Sharing (Shamir 1979) scheme as either a metaphor or an actual component of a cyber fog approach. Roughly, a Shamir-like scheme of dispersion may enable sharing of information in such a way that even if the adversary captures a significant fraction of shards, any meaningful information could not be reconstructed from it. It was also speculated that such a scheme might help balance the bandwidth requirements over time (e.g., the bulk of shards could be distributed during the lull in communications demands, while only the final and critical shards—a few—would be sent over the network during the busy periods). Secret sharing, particularly when verification of reconstructed secrets is required, could be made computationally efficient (Subbiah and Blough 2005).

Granted, challenges of dispersion are formidable. First, there are the obvious challenges of developing, validating, and managing the complex mechanisms required to perform dispersion with desired effects. Increased diversification (e.g., dispersion of data) and the complex mechanisms required to manage the diversification also create new surfaces and venues for cyber attacks. For example, if SDN is used, a centralized SDN is a single point of failure; thus a more complicated, distributed SDN will be needed. (Indeed, one anticipates that use of

Approved for public release; distribution is unlimited.
SDN in a tactical coalition network will be distributed.) In particular, a cyber fog approach may increase a network’s vulnerability to availability attacks, even as it improves its resilience to confidentiality attacks. Therefore, a complex tradeoff between availability and confidentiality may need to be managed in real time depending on mission and circumstances of the friendly forces.

Consistency, too, is complicated to achieve (e.g., updates across the system) in this scheme, although local consistency may be easy enough. Increased diversification also makes it more difficult to ensure that friendly users obtain all the information they need; this could be mitigated by relying on the background knowledge that friendlies have and adversaries probably do not have. We subsequently discuss this point in detail.

3. Dispersion and Effective Regathering

Dispersed information will be eventually requested by users and will have to be regathered in a timely and efficient fashion and reconfigured into a useful form. This could be helped by intelligent dispersion—put shards where they are more likely to be accessible at the time when they are more likely to be needed by the users. One way to achieve improved regathering of information is to account for the semantics of information while splitting it into shards. This could help intelligent prepositioning of related shards. While doing so, care must be taken not to introduce some regularities into the dispersion scheme that would make it easier for the adversary to find and gather that information. In fact, this creates a tradeoff between data security and the anticipated availability of the data. For example, CYRUS (Chung et al. 2015) ensures user privacy and reliability by scattering files into smaller pieces across multiple clouds, so that no one cloud can read users’ data; an algorithm selects clouds from which to download user data to minimize latency.

To determine which data are more likely to be required by the user and when, it is important to have means of automatically determining relevance of information to the user. Cyber fog complicates determination of relevance: unlike in a conventional files system where file A is likely to be relevant to the same issue as file B in the same folder, co-location of 2 shards tells us nothing about their common relevance.

Data provided to the user must be not only relevant but also timely. Timing issues are also complex. The way a collection of information is dispersed (e.g., how small the shards are and how far they are dispersed) depends on when and how rapidly these bundles of information will be needed by the user and the overhead for distributing and gathering each shard. Not only are such real-time tradeoffs of security versus timeliness complex, the timeliness is even difficult to define (e.g., I
need message M by time T, or how much of message M do I need to have by time T and still derive sufficient value from M). The timeliness versus security tradeoff is dependent on the nature of the mission: if maximum security only needs to be maintained for a short period of time, it may be acceptable that an adversary has a higher chance of obtaining the information, after a given time interval. Researchers Bilbray et al. (2015) have considered how to geographically distribute fragments and replicas to minimize expected latency for retrieving data and how to optimize a utility function, which incorporates both aggregation latency and storage overhead. Notions of caching in a dynamic delay tolerant network have been explored under ARL’s Network Science Collaborative Technology Alliance (Zhuo et al. 2011; Gao et al. 2014).

Consideration of timeliness also depends on the intended or likely purposes of the data: whether the data are needed for real-time execution (in which case it needs to be dispersed in a way that allows rapid and reliable regathering), or intended for postoperation analysis, in which case it can be dispersed with less attention paid to the resources needed for regathering. Network structure, dynamics, and characteristics (e.g., the network’s dynamic profile of connectivity and the network diameter) also influence the optimal ways of dispersion. In some cases, timeliness can be improved by avoiding regathering, (e.g., using distributed analytics to obtain the desired answers without regathering the shards). The joint design of fogging/defogging must take into account the tempo of the network and of the information: the mobility and dynamics of the nodes—both the requesters as well as the nodes where the shards are stored, how soon the sharded information will become stale, and how soon stored information may be needed.

4. Situational Awareness and Information Semantics

Ultimately, SA is the goal of information, and even timely and relevant information delivery does not guarantee high-quality SA. For one, not all shards are equally valuable from the SA perspective. A shard could be used for creating multiple different pictures or drawing multiple conclusions, depending on how the shards are glued together for SA purposes—does it make that multipurpose shard more or less valuable? To a large extent, SA presents us with the problem of not merely regathering, but also discovering, information. Where and how do I find what I need to achieve the required SA? Which shards need to be collected to get the right SA? A centralized or distributed indexing scheme may be required to help this process; however, this too introduces security concerns. Novel methods of information fusion will be required to achieve adequate SA, especially when regathering is incomplete due to an adversary action or network failures.
Semantics are significant for successful SA formation. To estimate the level of significance, consider a game where the players are given a few letters and asked to guess a phrase to which the letters belong. As the players are given more and more of the appropriate letters, they eventually recognize the phrase. The lowest fraction of letters when recognition becomes possible is called the “phase transition threshold”. Note that the phase transition threshold is significantly lowered when the phrase is familiar to the players or when they know something about the phrase. Thus, background information or context matters. Knowledge of the semantics of the information and the semantic context of the information are highly influential on how correctly the information is understood by the recipients. Ideally, phase transition should occur rapidly for the friendlies (who possess background information) and less rapidly for the adversary (who presumably does not possess such background information). But such background or side information may be provided by an agent that may be friendly or adversarial.

To reiterate, semantics of information—including the dispersed data, the semantic context of the friendlies’ mission, and the background knowledge of the users—are critical for effective and accurate “defogging”. Semantic information theory seems highly relevant to challenges of cyber fog. Sheaf theory was mentioned as relevant in this context.

Mission context is particularly important because the success of the mission is the true measure of “goodness”. An adversary may need only very little information to disrupt a key element of the mission. Thus, understanding of the value of information is critical. The dispersion, the regathering, and SA formation processes must be designed and executed in a way that information has high value for the friendlies and low value for the adversary. This implies, inter alia, the need for a thorough knowledge (model) of the adversary’s intent and prior knowledge.

5. Risk and Mission

Risk could serve as a comprehensive framework for characterizing the goodness of cyber fog. It is recognized that cyber fog scheme could potentially increase risk in certain aspects, and decrease it in others. Because poorly understood and modeled phenomena like obfuscation and deception play important roles in cyber fog, new risk models are unquestionably needed.

Although it is tempting to formulate risk in cyber fog in terms of data (e.g., a fraction of data captured by the adversary), it would be misleading. Rather, risk should be analyzed in terms of impact to the mission. Consequences of failures of cyber fog should be best assessed in terms of its consequences to the mission objectives. This implies a need for an adequate model of a mission (including its
dependencies on network and computing assets)—a modeling problem that is known to be highly complex. Other complexities arise in seeking ways to measure (quantify) consequences to the mission. Some may be indirect and involve impact on the adversary (how much the adversary lost or invested, how long our deception story holds, etc.) Time plays a role. The same event can have very different consequences depending on its timing, and time decay of the importance of the information may be involved (loss of dated information could be less important than that of freshly obtained information). Additive properties of failures such as information losses are important too (e.g., if you know A and B, information obtained is of high value; if you know only A or B, little information is obtained). Uncertainty of failure increases risk: if I know I lost data A, I can decide to do something about it, but if I am uncertain, my effectiveness is impaired.

Understanding the risk to mission in an adversarial environment could clearly benefit from a game-theoretic treatment. Risk is highly dependent on the decisions and actions of the opponents, who are interdependent. The game here is far from classical. It deviates strongly from the traditional zero-sum game; conducted under partial information, bounded rationality, etc. In fact, even the mission itself (i.e., the goals of the game) can be subject to change if some supporting assets fail or are captured by the adversary. Further, this is a game involving deception.

### 6. Deception and Obfuscation

Both dispersion and obfuscation share the key idea to increase uncertainty (to the adversary) through increased diversity. Arguably, dispersion helps to perform obfuscation and possibly its stronger form that is deception. The workshop participants discussed possible differences between obfuscation and deception. One interpretation suggests that while obfuscation intends to present the adversary with information that leads to multiple seemingly equally possible interpretations, deception aims to present the adversary with information leading to a specific interpretation beneficial to the friends. Very little rigorous, quantitative research has been directed at either deception or obfuscation. In the following, we use the term deception implying both deception and obfuscation, unless only obfuscation is discussed.

Within the cyber fog concept, deception may take multiple forms. The dispersion and frequent repositioning of information by itself presents adversaries with uncertainty as to where they could find information relevant to their interests, and how to reconstruct it from the shards they captured. Examples of other types of deception include presentation to the adversary of false software and hardware vulnerabilities, thereby inducing the adversary to expend efforts and resources on
unsuccessful attacks. Diversity of channels helps deception (e.g., the deceiver could use one channel for real communications and another for deception). An SDN could be used to present the adversary with a misleading view of the network. Honeypots and honeynets deceive the adversary as well. This may include cyber-physical honeynets (e.g., a honeynet that looks to the adversary like a friendly tank).

Still, even with multiple ways to create deception, it is hard to create a believable deception. For example, creation of believable battle plans and other unstructured documents is very challenging. Also very challenging is the problem of creating believable network traffic that the adversary’s traffic analysis mechanism would perceive as a particular EOB of the friendlies. Other examples include placing false shards into the fog; designing believable feint attacks that effectively support a real attack; and generating complex multistep deceptions. Creating a believable deception is harder when the adversary observes the friendlies along multiple dimensions (e.g., physical movements, cyber activities) and when the friendlies are uncertain as to what the adversary can actually observe. Machine-learning techniques might be applicable to generating believable deceptions. Parenthetically, because in cyber fog the adversary is likely to spend more efforts discovering the desired information of the friendlies, the deceiver might benefit from observing these efforts and determining better ways to formulate the deception.

Counterdeception (discovery of a deception) is no less challenging. Much research is needed to determine fundamental limits on counterdeception, as well as actual techniques for performing counterdeception. Detection of deception might be assisted by the fact that a deception, a purposeful human creation, is likely to be far less complex and rich in details than real-world information. Lessons might be learned from work on code de-obfuscation, such as truth maintenance approaches. Machine learning might also be applicable to detection of anomalies indicative of a deception. However, sophisticated adversaries may specifically target machine-learning techniques to defeat them. If so, research is needed on limits and verification of how a particular classifier (machine learning) can be fooled by particular inputs.

As mentioned earlier, deception requires game-theoretic approaches. Examples of highly challenging and poorly studied issues are payoff function or metrics of goodness for a deception; modeling of deceivers’ behaviors; modeling of humans (and humans with computational tools) who are targets of deceptions. Considering that the battlefield of the future will be populated by many artificially intelligent (AI) systems, it is important to study how AI and human differ (or not) with respect to perceiving a deception.
7. Applicability of Formal Methods

Given the extreme challenges and complexities inherent in the world of cyber fog, designing tools and planning specific activities within such an environment may greatly benefit from formal methods. If successful, such formal methods would ensure the friendlies that their environment and plans are guaranteed to exhibit certain properties. Unfortunately, the current state of capabilities in formal methods presents a number of limitations. For example, formal methods suffer from lack of insights into formulating the right questions to ask (i.e., which property to verify). Formal methods developed for one domain do not transfer well to another domain (e.g., methods developed for verification of hardware do not transfer well to verification of software, and even less so to verification of deception plans). Some difficulties can be mitigated by designing structures that lend themselves to formal methods (e.g., some language primitives lend themselves to verification by formal methods; check points in software make it easier to verify formal methods). Perhaps it might be possible to create a cyber fog that would lend itself to formal methods.

Furthermore, it is unknown how well, if at all, formal methods apply to human factors, such as the role of cognitive factors in deception. It may be possible to prove formally the consistency of a deception story but it may not be possible with respect to the cognitive aspects of that deception. If formal proof of deception may not be possible for a human receiver, one might speculate that it might be possible for an AI system that is a receiver of a deception. A possible starting point in research exploring applicability of formal methods to deception could be a problem of proving that a deceiver is producing and delivering to the receiver a picture that the deceiver intended, and that meets the deceiver’s specification.

8. Conclusions

The key scientific questions that emerged during the workshop could be summarized as follows:

Research Question 1: What theoretically grounded models can help characterize the complex tradeoff inherent in radical dispersion of information among mobile tactical edge devices (and related diversification of channels and protocols), including tradeoffs of communications overhead, energy consumption, and security impacts? While there are a growing number of commercial products that use related methods of dispersion, they have not been attempted empirically or studied theoretically in the tactical environments characterized by physical mobility, uncertainty of connectivity, energy and bandwidth constraints.
Research Question 2: What approaches, including semantic-based techniques, can help minimize the impact of dispersion on timely, secure, and efficient re-gathering of information in a fashion that would support formation of SA appropriate to the time, place, and mission of the user? Semantics of information and the need for related theoretical advances, including the dispersed data, the semantic context of the friendlies’ mission, and the background knowledge of the users, are critical for effective and accurate re-gathering of information for SA.

Research Question 3: Could risk or other related metrics, along with new analytical methods that are in part game-theoretic, serve as a comprehensive framework for characterizing the goodness of cyber fog? The consequences of failures of cyber fog should be best assessed in terms of its consequences to the mission objectives. This implies a need for an adequate model of a mission and quantitative measures of consequences to the mission (e.g., the timing of information determines its importance to the mission).

Research Question 4: What formal models, theories, methods, and tool can be devised to execute and manage successful obfuscations of friendly information within cyber fog? Arguably, dispersion helps to perform obfuscation and possibly its stronger form—deception. However, very little rigorous, quantitative research has been directed at either deception or obfuscation.
9. References


Approved for public release; distribution is unlimited.
10. Other References Suggested by the Workshop Participants


### List of Symbols, Abbreviations, and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>artificial intelligence or artificially intelligent</td>
</tr>
<tr>
<td>ARL</td>
<td>US Army Research Laboratory</td>
</tr>
<tr>
<td>CEMA</td>
<td>cyber electromagnetic activities</td>
</tr>
<tr>
<td>EOB</td>
<td>Electronic Order of Battle</td>
</tr>
<tr>
<td>SA</td>
<td>situational awareness</td>
</tr>
<tr>
<td>SDN</td>
<td>software-defined network</td>
</tr>
</tbody>
</table>
INTENTIONALLY LEFT BLANK.