Quantitative Evaluation of Moving Target Technology

Paula J. Donovan, Jeffrey W. McLamb, Hamed Okhravi, James Riordan, Charles V. Wright**
Cyber Security and Information Sciences Division
MIT Lincoln Laboratory, Lexington, Massachusetts 02420
{*pjdonovan, mclamb, hamed.okhravi, james.riordan*}*@ll.mit.edu, cvwright@cs.pdx.edu*

***Abstract*—Robust, quantitative measurement of cyber technology is critically needed to measure the utility, impact and cost of cyber technologies. Our work addresses this need by developing metrics and experimental methodology for a particular type of technology, moving target technology. In this paper, we present an approach to quantitative evaluation, including methodology and metrics, results of analysis, simulation and experiments, and a series of lessons learned.**

I. Introduction

ROBUST, quantitative measurement of cyber technology is a critical need [1] that this work addresses by developing metrics and robust experimental methodology for measuring emerging cyber technologies. An additional related objective also addressed in this work is demonstrating the plausibility of developing such metrics and methodology.

We begin this paper by presenting a threefold approach to quantitative evaluation. Then, we present the the quantitative evaluation of two moving target technologies. Finally, we discuss the results and key lessons learned.

A. Quantitative Evaluation Methodology

The metrics designed and evaluated in this work follow the requirements for scientific risk assessment according to a threat derived methodology [2][3]. The following are adapted from these requirements:

1) The mathematical approaches used to model threats, the overall system, and defense models are clearly stated and agree to the extent possible with known threat, system, and defense characteristics. Assumptions, simplifications, limitations, and constraints are clearly defined to make sure they can be reviewed for accuracy and the results can be replicated.
2) The analysis focuses on determining the risk or probability of different outcomes for different threats and with different defenses.
3) The analysis and results are reliable, meaning that a different group of researchers making the same assumptions would obtain similar results, and they are valid, meaning that they estimate the true underlying risk.

*This work was supported by ASD(R&E) under US Air Force under Air Force contract FA8721-05- C-0002. The opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

**Worked performed while at MIT Lincoln Laboratory. Wright is now at Portland State University.

978-1-4799-1737-2/15/$31.00 ©2015 IEEE

We derive metrics from parameterized threat models. In each of the technologies evaluated, there is stated function to the technology which, implicitly, specifies a possible collection of threat models. The metrics mathematically formalize these threat models with parameters given by the capabilities of the adversary paired with the configurational parameters of the technology being tested. We considered these metrics measures of effectiveness (MoEs), as they address how well the technique defends against a particular adversary. Measures of performance (MoPs) (e.g. resource overheard associated with the implementation) and measures of safety (MoSs) (e.g. compatibility with the host or environment) are also important; however, they are not presented in this paper.

B. Threefold Approach to Quantitive Evaluation

Each effort in technology measurement has tracks of analysis, modeling and simulation (M&S), and experimentation. This threefold approach to quantitative evaluation offers the most benefit when executed in combination as the results and insights from each tend to inform and enrich the others.

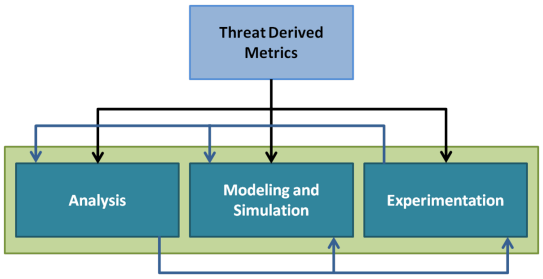

Fig. 1: *Analysis, M&S and Experimentation*

By analysis, we mean the decomposition of the behavior of a technology into to a collection of base components that can be mathematically described in a tractable fashion. Analysis tends to provide a deep understanding of a technology, and is also often the least expensive of the three modalities.

M&S remains generally inexpensive and is comparatively easy to explore variations on an experiment. It can capture emergent behaviors and phase changes and provide data and examples of an idealized model for the purposes of analysis. Simulation is often possible where direct analysis is not.

Experimentation provides the greatest fidelity but generally carries the greatest cost. Experimentation can vary degrees of

realism. Live experiments often capture environmental effects that none of the other modalities can. Experiments can provide realistic input values to simulation thereby significantly improving the fidelity of the simulation.

C. Evaluation Focus

The technological focus for this study is Moving Target (MT) technology with a goal of quantitatively measuring resiliency and agility in controlled, repeatable laboratory experiments. These evaluations are not intended to provide a comprehensive assessment of the security of these techniques, or to measure the protection provided against all adversaries. Instead, the focus is on identifying one or more specific aspects that each technique is intended to improve, then quantitatively measuring the technique's ability to realize that improvement with respect to a relevant, realistic threat model. To achieve the goals described, particular focus was dedicated to the following areas:

- **Metrics** that quantifiably evaluate the MoEs, MoPs, and MoSs of MT technique. Note, only MoEs are described in this paper.
- **Methods** that provide an accurate, repeatable means of creating and capturing the data necessary for calculating the defined metrics.
- **Threat Models** that are realistic and specifically address the attacks the technology is protecting against.

D. Overview of Moving Target Technology

MT technology is currently an area of emphasis for those responsible for protecting systems and networks, as it provides a less deterministic attack surface (limits adversary reconnaissance ability and probability of adversary success), a less static environment (reduces time window for attack) and a less homogeneous environment (complicates exploit development, limits possible damage, reduces the scale of exploits).

MT techniques have been proposed, and implemented in several domains. In this paper we specifically address dynamic network and dynamic platform techniques, as these were the types of prototypes available to us at time of this evaluation effort.

II. LPS Evaluation

Lightweight Portable Security (LPS) is a bootable operating system burned onto a read only-medium that provides resiliency by enabling rapid recovery of workstations to a known clean state. It achieves this by operating without persistent storage (such as hard drives or flash drives), thereby eliminating an adversary's ability to persist on a system. It allows a user to set up an untrusted computer into a secure configuration and connect to DoD (or other protected) resources over a Virtual Private Network (VPN).

A. Objective

The objective of LPS quantitative evaluation is to measure the degree to which LPS achieves resiliency thereby impacting the adversary's mission. Toward this end we defined three adversary mission categories as described in the next section.

B. Threat Model and Derived Metrics

The threat model for the experiments is a client-side attack, where an adversary in the network is able to plant malicious content on popular web pages on the Internet. The adversary uses the malicious content planted on a server to attack any client system that browses to those pages. The client system can be protected with LPS. This is a realistic threat model, as cyber criminals have previously placed malicious content in banner advertisements on major websites successfully.

The metrics are derived from the adversary missions:

- **Exfiltration** - The adversary's goal is to exfiltrate data from the network. We assume that the exfiltration bandwidth is proportional to the number of devices under adversary control so that the metric is simply the expectation of this value.
- **Command and Control (C2)** - The adversary's goal is to maintain a persistent C2 channel past a firewall with at least one compromised device on a network. Here the metrics are the percentage of time that the adversary has at least one compromised device together with the expected duration of periods when the adversary has no compromised devices.
- **Fractional Control** - The adversary's goal is to maintain some fraction of the devices on the network. Here the metric is the percentage of time that a percentage of devices under adversary control is above some threshold. This mission category might apply in settings where those devices using LPS participate in a voting or averaging scheme such as a sensor network.

As analysis shows that each of these metrics can be reduced to the the fraction of time that the adversary controls any given device, our results presented in this paper will focus on this one particular metric. This metric is parameterized by:

- **Reboot rate** - The rate at which virtual users reboot their systems. Rebooting can either be periodic or Poisson. We considered the cases where the users reboot their systems as a Poisson process with averages every 2 hours, every hour, every 30 minutes, and every 10 minutes.
- **Infected web page ratio** - Fraction of web pages that contain malicious content. Due to time constraints, this variable was held constant at 1 out of 40.
- **Web browsing rate** - Rate at which each virtual user browses to a new page. Due to time constraints, this variable was also held constant for the experiment, such that each virtual user browsed to a new page every 4 minutes.

C. Analysis and Simulation

We analyzed an LPS system and expressed the base metric (fraction of time that an adversary controls any given device) in closed form. We further expressed combination formulas to combine the values of the base metric of a collection of devices into values for each adversary mission. Then, a simulation was written that verified both the analytic formulations and real world testing.

978-1-4799-1737-2/15/$31.00 ©2015 IEEE

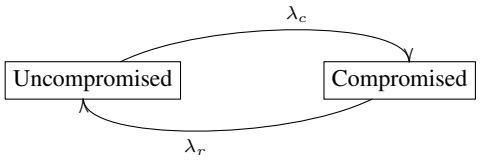

Fig. 2: *Combination of Poisson Arrivals of Compromising Events and Reboots*

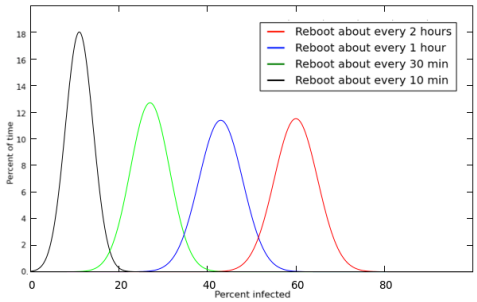

Fig. 3: *Predicted adversary's fractional control of the network for LPS experiment*

The full model, depicted in Figure 2, combines the arrival of compromising events with rate λ_c and rebooting (hence transition from compromised to uncompromised) with rate λ_r. Note that rebooting from an uncompromised state leads back to the uncompromised state so this transition is not explicitly included. The resultant probability of being compromised P_c at any given time is $\lambda_c \cdot (\lambda_c + \lambda_r)^{-1}$.

Sequential browsing where each page is visited for a uniform duration from 0 and 4 minutes, and where the infected page rate is $\frac{1}{40}$, is approximated by a Poisson arrival rate:

$$\lambda_c = \frac{1 \text{ arrival}}{2 \text{ minutes}} \cdot \frac{60 \text{ minutes}}{1 \text{ hour}} \cdot \frac{1 \text{ compromises}}{40 \text{ arrivals}} = 0.75.$$

The Poisson process with reboot rate having mean of 2 hours, 1 hours, 30 minutes, 10 minutes corresponds to to an reboot arrival rate λ_r equal to 0.5, 1, 2, 6 reboots per hour respectively.

These values of λ_c and λ_r result in the following probabilities:

λ_c	Δ_r	λ_r	P_c	$\neg P_c$	Samples
0.75	2 hours	0.5	0.40	0.6	11
0.75	1 hour	1	0.57	0.43	22
0.75	30 minutes	2	0.73	0.27	44
0.75	10 minutes	6	0.89	0.11	132

We set the number of devices to 55 (to match later experiments) and generate binomial distributions for the various P_c. The probability mass functions of these distributions are displayed in Figure 3.

D. Experiment Design

The environment for the LPS experiments consists of an enterprise network, an "Internet" network, and several small remote ("home", "airfield", or "Internet café") networks, all connected via the Internet. The enterprise network offers a VPN gateway for remote clients to connect to its internal network, and from the internal network to the Internet. Figure 4 shows the network environment for this scenario.

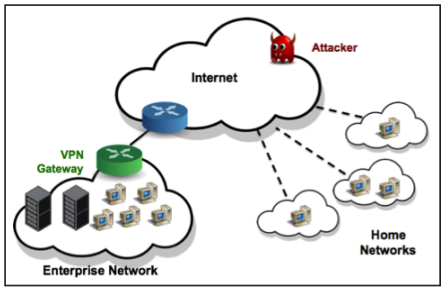

Fig. 4: *Network for LPS scenario*

The key tunable parameters for this experiment are the reboot rate, infected web page ratio and web browsing rate - all as described above. We let each scenario run once for 22 hours, which means that the different λ_c will have a differing number of samples. In particular, we expect the prediction for $\lambda_c = 6$ where there are 132 samples to be significantly more accurate than that for $\lambda_c = 0.5$ where there are only 11.

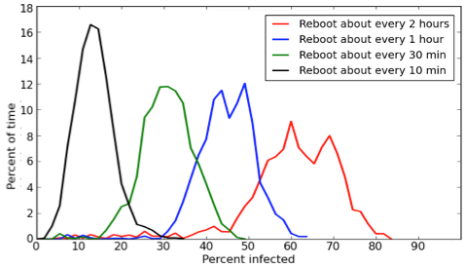

Fig. 5: *Adversary's fractional control of the network for LPS experiment*

E. Results

By increasing the rate of recovery by increasing the rate at which users reboot their machines, we observed a general decrease in the adversary's fractional control of the network. Figure 5 shows the fractional control of the network that the adversary was able to achieve under the different experimental configurations. Also, although the adversary was still able to quickly gain a small foothold in the network (less than 10%), the time required for him to first capture even 30% of the target devices increased dramatically, as depicted in Figure 6.

III. IP Hopping Gateway Evaluation

IP Hopping may be used to improve the agility of virtual private network (VPN) gateways by constantly varying their externally-visible network addresses. This way, even if there are vulnerabilities in the gateway's hardware or software or even if an adversary has used a lifecycle attack to plant a "back door" deep in its code, IP Hopping can still make it

978-1-4799-1737-2/15/$31.00 ©2015 IEEE

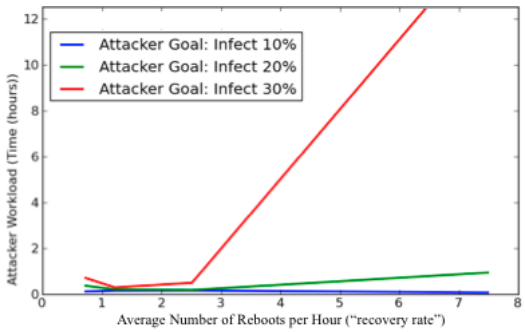

Fig. 6: *Adversary's time to success for LPS experiment*

very difficult for an external adversary to break in to the protected enclaves of the VPN. IP Hopping is generally termed a network address randomization technique.

A. Objective

The goal of the IP Hopping Gateway quantitative evaluation was to measure the relationship between the randomness of the Gateway's IP address and the cost to one realistic class of adversary.

B. Threat Model and Derived Metrics

In order to complete the experiments in a reasonable period of time, we focused on evaluating IP Hopping Gateway against one particular class of adversary. We made no attempt to comprehensively measure the security of the IP Hopping Gateway system against all adversaries or to measure the performance impact to the protected VPN or subnet.

The adversary considered is one that can remotely exploit vulnerabilities on the gateway. We assumed that this adversary is outside of the protected enclave, and does not have control of the router that connects his victim to the rest of the network.

As the only adversary mission under consideration is compromising the gateway, the metrics considered are the probability of adversary success and average time to successful attack.

C. Analysis and Simulation

The analysis, simulation and experiment are set up such that the IP Hopping Gateway chooses a new IP address at some regular interval from a fixed IP subnet ("IP Hop Space"), without immediate repeat. The adversary scans the IPs in his target's IP Hop Space at a constant rate, and on each scan attempt, he chooses his target IP address uniformly at random. The adversary can remotely exploit some vulnerability in the IP Hopping Gateway, but to do so he must successfully send several packets back and forth to the victim.

The key parameters considered are:

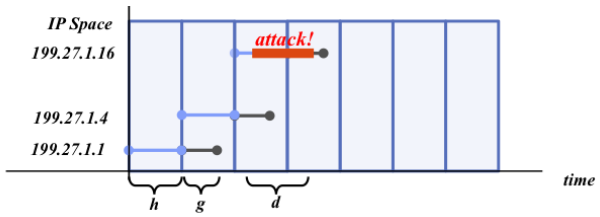

Fig. 7: *Time lapse of IP hopping, showing h, g, and d.*

n	number of addresses in IP Hop Space
a	number of *active* IP addresses
h	intra-hop time
g	grace period
d	interval required for attacker success
s	interval required to scan
t	duration of trial

We are assuming that:

- h, g, d are all significantly less than the duration of the trial t so that boundary effects are negligible.
- The adversary knows all of IPs in the IP Hop Space, as these IPs are utilized on an externally-visible interface.
- There is a grace period g where an old IP address will still be accepted (such functionality is expected to prevent poor TCP performance due to the protocol's exponential backoff).

In the cases where the grace period g is greater than the intra-hop time h, there will be multiple IPs active at a time. Therefore the maximum number of active IPs, a_{max}, is $\lceil (g+h)/h \rceil$ and the minimum number of active IPs, a_{min}, is always 1 less than a_{max}. An attack will begin during any one intra-hop interval, and therefore will begin either when there are a_{min} IPs active ("case 1") or when there are a_{max} IPs active ("case 2") as shown in Figure 7. The probability the attack starts during case 1 is $P_1 = a_{max} - (g+h)/h$, and the probability the attack starts during case 2 is $P_2 = 1\text{-}P_1$. Now we want to calculate the probability of successful attack in either of these cases. To do so, we must use the probability of successful attack given the adversary selects an active IP:

$$P_{s:IP_{active}} = max(min(\frac{Y-d}{X}, 0)),$$

where the attack starts before time X and must end before time Y for success, given that the attack must complete in time d. For case 1, if the attacker selects an active IP, each of the active IPs have an equal $1/a_{min}$ probability of being the one selected. Therefore, the probability of adversary success in case 1 given an active IP was selected is:

$$P_{s1} = \frac{1}{a_{min}} \sum_{r=0}^{a_{min}-1} (max(min(\frac{g-rh+hP_1-d}{hP_1}, 1), 0)).$$

Similarly derived, the probability of adversary success in case 2 given an active IP was chosen is

$$P_{s2} = \frac{1}{a_{max}} \sum_{r=0}^{a_{max}-1} (max(min(\frac{g-rh-d}{hP_2}, 1), 0)).$$

978-1-4799-1737-2/15/$31.00 ©2015 IEEE

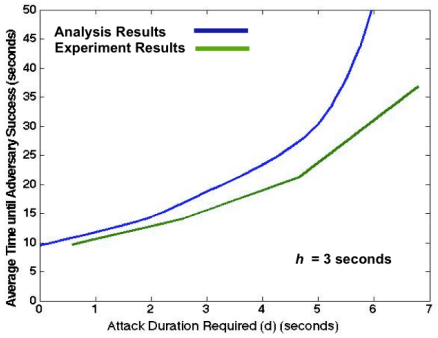

Fig. 8: *Probability of attack success for a range of attack duration requirements*

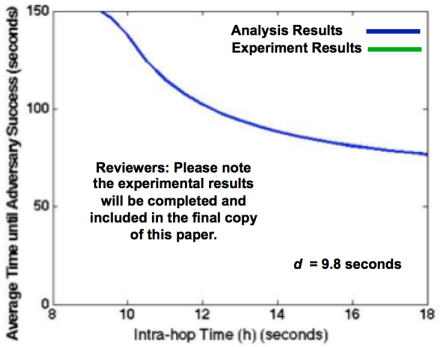

Fig. 9: *Average time to successful for a range of intra-hop times*

Next, we can calculate the probability of adversary success when selecting an IP out of all IPs (active and inactive) in the IP Hop Space using

$$P_s = \frac{(P_{s1} + P_{s2})\frac{g+h}{h}}{n}.$$

Finally, the expected number of trials is $E_T = \frac{1}{P_s}$.

D. Experiment Design

We developed software that acts as a surrogate for an IP-hopping architecture, and implement an experiment using all of the parameters list in the table above. The IP-hopping process loops continuously, and is able to effect the notion of IP-hopping with grace period by using Linux address aliasing with deterministic addressing. The attacker process also runs in a loop, making an attempt to telnet to the IP-hopping host every s seconds. A successful telnet login, a specified wait period and logout is considered a successful attack.

For the experiment, n is 252, g is 4.9 seconds, and s is 0.1 seconds. Other key parameters are given within the results presented.

E. Results

Overall, the experimental results agreed with our analytical predictions and the results of our simulations. In general, by increasing the randomness of its external IP address, the IP Hopping Gateway was able to cause large increases in the delay experienced by the adversary. The grace period, hop time, and necessary attack duration all play a significant role in the results. Figure 8 shows the average time until successful attack which increases non-linearly with increased attack duration requirements. Figure 9 shows the decrease in average time to adversary success as h increases, for d set to 9.8 seconds.

IV. CMC1 Lessons Learned

As stated early in this paper, an objective was to explore the possibility of developing experimental methodology and metrics that are applicable to a type of technology. For all evaluations, the general methodology applied was very similar, comprised of the following steps:

1) Determine the intended objectives of the technology under test on which to focus effectiveness testing.
2) Identify all viable threat models to align evaluation modalities and metrics.
3) Understand the implementation details of the technology well enough to measure the effectiveness.

Metrics were tailored to the specific threat models and thus were more disparate for each MT technology than the overall methodology. Although these metrics are MoEs, a broader set of metrics including the MoPs or MoSs have to be developed in the longer term to study the impact of a new technology. Moreover, the metrics should be generalized to study a family of technologies, such as entropy-based metrics that can be used to evaluate the MT dynamic runtime environment techniques. The study did give the authors the insight that the objective of developing experimental methods and metrics in the cyber domain for families of technology is feasible. More specific lessons learned derived from this study are described in the following subsections.

A. Lessons Learned in Quantitative Evaluation

- Experimentation with real systems is necessary to fully understand complex behaviors. Effects due to differences in implementations, environments, etc. cannot be captured in representations, and corner cases found in the real systems cannot be accounted for reliably.
- The complexity of experiment design creates a challenge of examining results for small subsets of variables at a time and thus must be a experimental design consideration. The ability to establishing which effects are caused by a particular variable is a key.
- Defining well-specified and realistic scenarios prior to evaluation is critical. For some of our early evaluations, a clear threat model had not yet been defined when the experiment was initially designed. As a result, data that was necessary for computing some of the MoEs was not collected and had to be generated later via simulation. Although we had a work around, it was preferable to get the necessary results during the experimentation phase.
- The combined use of analysis, simulation, and experimentation in quantitative evaluations offers significant benefit over isolated execution of these disparate means of

978-1-4799-1737-2/15/$31.00 ©2015 IEEE

understanding. For the IP Hopping Gateway, simulation and analysis quantified the dependency of the key agility metric on different parameters and experimentation validated the results.

B. Lessons Learned in Metric Development

- The theory of probability and stochastic processes provides a solid foundation for measuring and analyzing many other notions of resiliency and security. Examples of this are the calculations of recovery rate and adversary success rate described throughout this paper.
- Current cyber range capabilities are sufficient to capture many general measures of performance, such as changes in network throughput. However, any given cyber range environment may require augmentation with special tools and instrumentation to measure the necessary characteristics of each new system under test.
- Deriving metrics from threat models enables the creation of reasonable MoEs for new cyber defense technologies [2]. These threat models may be reasonably abstract so that the resulting metrics are widely applicable. Currently these threat models, and thus metrics, are limited to known threats to our networks and cyber assets [1].

C. Lessons Learned in Resiliency and Agility

- Rapid recovery and heterogeneity have the potential to substantially increase the resiliency of a system.
- Resiliency and agility that address realistic threat models can significantly limit the probability and extent of an attack.
- Real systems must provide resiliency and agility as well as low cost of use (albeit this MoP was not presented in this paper). Further research is needed within the cyber technology development community to better achieve these goals.

V. Conclusion

The study described here met its objective of establishing that development of metrics and experimental methodology for a family of cyber technologies is feasible. Quantitative evaluation of two MT technologies, LPS and IP Hopping Gateway, were completed and a series of lessons learned captured. The metrics defined were rather specific to technology and relevant threat model. However, we gained some insight on how more generalize metrics may be developed in time. The methodology used for each technology was similar and could feed into an overall, high-level methodology for cyber experimentation.

References

[1] Multiple: Defense Science Board Task Force. Resiliency Military Systems and the Advanced Cyber Threat. Technical report, Department of Defense, January 2013.

[2] R.P. Lippmann, J.F. Riordan, T.H. Yu, and K.K. Watson. Continuous Security Metrics for Prevalent Network Threats: Introduction and First Four Metrics. Technical report, MIT Lincoln Laboratory, May 2012.

[3] T. Aven. *Quantitative Risk Assessment, The Scientific Platform*. Cambridge University Press, 2011.

978-1-4799-1737-2/15/$31.00 ©2015 IEEE

Presenting Author Biography:

Paula J. Donovan
Assistant Group Leader
Cyber System Assessments Group

Paula Donovan joined MIT Lincoln Laboratory in 2000 and is an Assistant Group Leader the Cyber System Assessments Group. Ms. Donovan's research efforts currently focus on establishing metrics and experimental methods for assessing cyber security technology. She also provides leadership in cyber assessment and evaluation efforts. Ms. Donovan spent her first decade at the Laboratory in the Advanced Networks and Applications and Airborne Networks groups where her research activities included developing and analyzing communication data links for airborne networks, testing and evaluating a broad range of technologies, developing simulations and analyzing data of many network types, including all layers of the network stack on networks of all scales, developing algorithms and protocols for a variety of communication networks, such as wireless ad hoc networks, and identifying then deploying existing technologies that align with warfighter needs. Ms. Donovan received her MS in Applied Mathematics with a concentration in Computer Science from the University of Massachusetts at Amherst. She received her BA in Mathematics with a concentration in Pre-Medical from the College of the Holy Cross.

978-1-4799-1737-2/15/$31.00 ©2015 IEEE