Executive Summary

The goal of the Workshop on GENI and Security was to engage the security community in GENI's design and prototyping, to ensure that security issues are properly considered during its development. The specific issues of interest were:

1. What classes of experiments should GENI support, and what capabilities will GENI require in order to support them?
2. How can GENI itself be secured and protected from attack? Moreover, how can those networks and cyberphysical mechanisms connected to GENI be protected from attacks originating from GENI, or malfunctioning GENI experiments?

An additional goal of the workshop was to encourage the security community to respond to a solicitation for GENI analysis and prototyping proposals released in mid-December by the GENI Projects Office.

Key Points

All participants in the workshop felt that GENI must foster a culture of scientific experimentation from the very beginning. To do this:

1. GENI must provide capabilities to enable a science of security that involves the experimental validation of security-related hypotheses that could not be validated in current testbed settings.
2. The construction of formal security experiments with hypotheses, controls, and well-articulated measurements will require substantial care and review to assure reproducibility and scientific and statistical validity.
3. GENI must provide the capabilities to enable experimenters to capture *all* the data needed to enable others to reproduce the experiment.
4. The deployment of GENI will require the development of mechanisms to reconcile conflicting requirements, constraints, and customs in different parts of the network.
5. The operation of GENI will require careful planning to enable communication among the federated organizations to handle (security and other) problems. The GENI infrastructure should support security testing, to ensure that security breaches can be handled quickly and effectively.

The participants were enthusiastic about the need for security in GENI, and had a myriad of ideas about the subject. We anticipate that several responses to the solicitation of proposals will focus on security, thus achieving the additional goal of the workshop.

Elaboration of Requirements and Issues

The nature of the GENI network itself raised security problems. As GENI is not controlled centrally, but is composed of autonomous federated networks, different organizations (indeed, different *types* of organizations—academic, governmental, and commercial) must provide resources and access for GENI to succeed. This raises technological, policy,

procedural, and legal issues.

This narrative highlights some of those issues supporting the key points, above.

Resource management. The question of resource management raises several security issues. First, who has the right to use resources? This requires identification and credentialing of the entities involved, and the ability to track delegation of rights. GENI will require cross-federation agreements and mechanisms to enable such management. The enforcement mechanisms must be able to reconcile disparate organizational practices and researcher identity management systems, and translate capabilities between the federated constituents. In addition, the ability to account for actions—to tie actions to the entities that take them—is normally considered a critical aspect of resource management. Will GENI researchers be held accountable for disruptive experiments? Interestingly, the participants split on this, a substantial number holding that too strict accountability might violate privacy. This raises a key issue that is explored below (see "privacy").

GENI provides virtual networks running on a large number of systems, most of which use virtualization to support the virtual network. (For future reference, a virtual network is called a *slice*, and that part of a virtual network supported by a single system is called a *sliver*.) Managing and securing virtualization to support the virtual networks and machines, and managing and securing the slices, is a question of resource management, and one critical to the success of GENI.

Another key issue in GENI is isolation: how to prevent an experiment in one slice (or a set of slices) from interfering with experiments in other slices. If two slices share the same CPU on a particular system, do the two experiments interfere? Managing resources to both mitigate and make visible such interference is critical—and depends on an equally critical issue, the definition of "interference". The issue of covert channels is an old one, and still a vexing one; thus questions of interference are likely to involve shades of gray, rather than binary black/white clarity. Furthermore, the degree of possible isolation may vary substantially across the heterogeneous technologies embedded within GENI. The meta-issue of how the environment and very nature of GENI affects experiments run on GENI must be understood in order to determine whether the results of the experiment will hold in other environments.

Logging, Recording, and Capturing Events. The participants expected that GENI would enable experimental validation of security-related hypotheses on large-scale networks. A key aspect of experimental science is *reproducibility*, not only by the experimenters, but also by others in order to verify the claimed results. This basic requirement implies that GENI must have specific capabilities.

GENI must be able to *record* events that occur during the experiment. This means it must support various types and levels of logging, at the level of "bits on the wire". The ability to capture packets, for example using a program like *tcpdump*, is not sufficient because we expect GENI to be used to test new protocols, including those not based on IP (and therefore that conventional network analysis programs will not record). But the ability to measure and record everything, including background traffic and timings, leads to privacy issues (see "privacy"). The multi-national federation of networks forming GENI exacerbates this conflict.

Second, GENI must be able to *replicate* the environment of an experiment so the experiment can be repeated under the same condition as the original experiment. An experimenter should be able to take the data recorded for an experiment and from that recreate the relevant parts of the background traffic, the slice on which the experiment was run (including all components—internal slivers, end points, etc.), and any other parts of the environment. Then the experiment can be rerun and the results can be validated. As in other experimental fields, perfect replication may be impossible in many scenarios, which raises the important scientific question of the degree of replication and repeatability required for experimental validation of security-related hypotheses.

Privacy. Because GENI is a federated testbed, the definition of "privacy" will vary among the federated networks. In particular, the federation is planned to include organizations in Europe and Japan, where privacy laws are very different than those in the United States. This has several consequences.

First is the impact on what can be recorded. Synthesized data (especially synthetic background traffic) should not be a problem anywhere, but such data is often not realistic. For example, intrusion detection systems often use the synthesized 1999 IDS Challenge data set to demonstrate their effectiveness; in the research community, any such results are considered suspect. Various proposals for recording and replaying real network traffic would avoid this problem, but raise many others, both technical and legal. In the context of privacy, one is whether the traffic *can* be used, or whether it must be anonymized and if so, to what degree.

Two approaches for this were discussed. The first is simply to record data elsewhere, anonymize it, and construct a framework for seeding it with attacks should the experimenter decide to do so. Then one could replay this data for experimental purposes. The second approach is to encourage ordinary users to use GENI, in effect making GENI a network that the public (or segments of the public, such as students or academic institutions) could use. Both raise issues of privacy, but the approach for handling privacy is different. The first can be anonymized before it is used; the second would have to be anonymized on the fly or recorded and subsequently anonymized. Further, the transformed data would have to be shown to have the same characteristics (specifically, those that can affect the experiment) as those of the untransformed data. Finally, whether "perfect" anonymization is in fact possible is an open question; often private data can be reconstructed from anonymized data when the attacker has access to external information.

There was some discussion of requiring the users of GENI to consent to monitoring, but this was felt infeasible unless the set of users is tightly controlled. We could solve this problem by limiting the measurement and recording to those parts of the data relevant to network analysis and that did not violate privacy rights; but this raises other issues, such as the reuse of the data for other experiments.

To put this problem of balancing as starkly as possible: under what conditions can we decide whether an experiment is doing something that violates the rules of usage without compromising the privacy of the experiment? Indeed, who is the "we" that decides this? And how are disputes handled (see "architecture and infrastructure" below).

A further question of privacy arises in regard to visibility of the measurements themselves.

Are an experiment's measured results visible only to the researcher(s) running the experiment? Or must they be made open and transparent to all researchers? Since some researchers may wish to preserve their own privacy (e.g. until they publish), there may be good social reasons to keep measurements private at least for some time.

Thus, the entire process of data collection, and controlling the data once collected, is key not only to the success of GENI as an experimental testbed but also to the acceptability of GENI under the law, regulations, and policies of its constituent networks.

Architecture and Infrastructure. Considerable discussion of the infrastructure for GENI revolved around the human and policy aspects, as opposed to the purely technical. As security is primarily a human endeavor, this was not surprising. Several interesting questions emerged.

First, *what is security*? An early discussion brought this out. Consider an experimenter who is designing a new protocol with attribution of its packets as its goal—that is, every packet can be traced back to its host of origin. This enables one to deduce, for example, origins of distributed denial of service attacks—generally considered a good thing.[1] However, if a dissident is emailing anonymous messages to the press identifying corruption, the anonymity may be that dissident's only protection from trial and imprisonment; here, attribution would be considered a bad thing.[2] So, is attribution a security requirement? The best answer is that it depends upon policy—and the exact policy will undoubtedly vary among the various federated networks (especially among those in different countries, and therefore subject to different laws).

These considerations suggest that using automated mechanisms to monitor and enforce security is problematic. Several specific mechanisms were discussed in the workshop. One issue is whether these mechanisms could provide a high enough probability of detection at a sufficiently low probability of false alarm. More broadly, it is by no means obvious how these mechanisms can be aligned with the deep policy issues discussed above.

Second, what (security) support services must the organizations with networks federated with GENI supply? To a large degree, this question is poorly understood because, so far, very few federated systems have crossed national and international boundaries. The participants in the workshop knew of no direct experience with such systems within the field of computer science research. There may, however, be lessons that are directly relevant from "big science" federations of recent years, such as the Grid endeavors and large-scale physics experiments (for example, the Large Hadron Collider).

Further, our experience with the Internet is disquieting. As an example, consider incident response. Different incident response groups have tried electronic means for communicating among themselves; these usually are not effective enough. The most effective communication mechanism is personal contact, either because you know your

[1] But not always. In war time, if a country were to use a DDoS attack to hinder its adversary, that country would want the DDoS attack to appear to come from an ally of the adversary to sow discord among the country's enemies. There, attribution is exactly what the country does *not* want.

[2] Except by the government trying to identify and catch the dissident.

counterpart personally, or you can reach your counterpart with the aid of others who know you both. Whom do you call when one part of GENI is malfunctioning and blocking access to your resources? Further, if there is a dispute, how will it be arbitrated? While legal recourse is available, in the United States at least, this often takes a very long time and is expensive. International litigation is probably even more expensive and time-consuming, even in those cases where it is feasible. An arbitration mechanism would work better.

The infrastructure ideally would supply timely response to questions, and take action when problems are reported. This essentially requires that someone be available at all times. If the model for GENI is to federate academic, government, and commercial institutions' networks, many of the constituent networks will not be able to provide that level of support—for example, academic computer science experimental networks run by faculty and students. Decreeing that a certain minimum standard must be met in order to federate with GENI was felt to be impractical, based on past experience with other types of voluntary federations. Invariably, some constituent fails to meet the minimum standards; but unless the failure is egregious, it is in practice difficult to expel a volunteer, and much more so if the volunteer is supplying needed or valuable resources. In general, social and peer pressure work better to encourage conformance to a minimal standard than do consequences that are costly for the federation.

This then brings up the question of how the federation works. Each constituent brings resources into the federation. Who decides how to assign these resources, and to whom? This affects availability, a key security service. For example, a policy may require that disruptive or misbehaving experiments have their priority, and hence their access to resources, reduced.[3] If this is a centralized decision, then the central controller must have control of all experiments—many in the workshop believed that this was highly unlikely in a federated network or system, and felt it antithetical to the nature of GENI. (GENI Spiral 1 does by contrast posit exactly such a centralized control system, located within the clearinghouse.) A distributed decision must take into account local policies as well as global policies, and there must be a mechanism for reconciling differences.

This takes us back to the requirements—what support services must GENI provide? It is not clear that a single set of services would meet with universal acceptance because of the tension between privacy and accountability, as discussed above. Thus, ultimately GENI's stakeholders must set its requirements. The workshop identified three major types of stakeholders: those who provide the resources (the constituents), *those who provide the data* (for example, sample background traffic or measurements), *and those who will use the resources* (the experimenters, managers, and other users). There was some discussion as to whether the experimental subjects also represented a set of stakeholders that needed to be represented beyond a human experiment review board (IRB). Additional stakeholders may include governments, regulatory bodies, and other political, legal, and social entities.

Ultimately, the owners of resources must manage their resources, because few will voluntarily give full control of their resources to the distributed system (see "resource

[3] This also raises the question of what "disruptive or misbehaving" experiments are. See the discussions about defining privacy and security, above.

management" above). Some aspects will probably be done locally. Others would require a common clearinghouse. For example, if an experiment needs access to a SCADA testbed connected to GENI, the experimenter can query the clearinghouse asking where she can access a SCADA testbed connected to GENI and having specific properties. The clearinghouse can then suggest other constituents whose SCADA testbeds meet the stated requirements, and the experimenter can then schedule time on them with either the local controller or (better) using the clearinghouse.

Workshop participants also discussed the nature of experiments. Some larger, long-term experiments will take on a provenance of their own, creating a meta-structure within GENI. Participants also raised the issue of ownership of the experiment, and of how to handle the transfer of intellectual property regarding the experiment should it be transferred into production mode, for example as a new security service for which organizations would pay.

Participants pointed out that the GENI must be easy for the constituents to manage. As GENI is a federation of networks, the goal is to get institutions to allow GENI to use their resources. Without funds from GENI, this requires volunteers. Experience shows that if volunteers must spend great amounts of time, effort, and other resources to do their tasks, they quickly become "former volunteers". For GENI, this would be disastrous. Further, the principle of psychological acceptability says that if management is not easy, configuration and other errors will occur, possibly disrupting experiments, and the GENI testbed itself. Therefore, considerable care must be given to making joining the GENI testbed, and maintaining membership in it, inexpensive in both efforts and funds.

It was also noted that GENI should enable an experimenter to specify and acquire specific classes of resources. For example, an experimenter should be able to acquire a computer to use as a router, rather than being forced to use a slice of the computer as a router.

Experiments. GENI must provide capabilities to enable a science of security that involves the experimental validation of security-related hypotheses that could not be validated in current testbed settings. The participants viewed GENI as a vehicle for instilling a culture of scientific instrumentation and experimentation into the security research community. With the availability of such a testbed there would be no excuse for failing to experimentally verify security claims that cannot be verified using other means. Further, several participants pointed out that GENI could be used as a teaching tool for how to carry out scientific experimentation in computer science and, especially, computer security. Given the need for emphasis on rigorous scientific testing in computer science curricula, this use may be the most important for the future of computer science and computer security.

Basic scientific and experimental capabilities include mechanisms to collect information and make measurements. This raises privacy issues, as discussed above. As more people want to use GENI, teaching them how to implement experiments correctly, and analyze the results, becomes critically important. The participants all felt that GENI should provide a set of detailed experiments that users could modify to learn how to do experiments on GENI—even simple experiments would be very helpful. Recipes or cookbooks for constructing and running experiments will prove invaluable, too. A supportive experimental community willing to share its knowledge and tools, combined with a GENI

help desk for experimenters, is an important asset

Creating a methodology for experimentation involving security, especially for experimentation on GENI, is important. This methodology should address topic such as the validation of the experiments themselves, validation of the data used by the experiment, and how GENI itself affects experimental results due to instrumentation effects, communication delays, and other attributes not present in the environment being experimented about.

Considerable discussion focused on the type of experiments users of GENI might want to run. Throughout the discussion, the focus was on experiments that are infeasible on current systems and testbeds because they are too small or not isolated; and infeasible on the Internet, again because it is not isolated.

The two experiments with the most immediate impact are the validation of models for distributed denial of service attacks and defenses on a large scale; and for the development of new architectures to inhibit botnets. Validation in this context requires the deployment and running of both types of attacks, because often experimental results show that the models we have developed do not match the reality of what happens in the network, and thus must be tuned. Worse, some phenomena may well be chaotic and so effects cannot be predicted, only described once they occur. Without experimentation, we will not know how good our models are, and whether they can be used to predict results on systems, especially those other than GENI (such as the Internet).

Both these experiments would disrupt the use of the Internet if tried on that. Other examples are cascading failure (where end or infrastructure systems begin failing), or simulations of changing large distributed networks with properties different than that of the Internet (such as the power grid). Thus, more generally, any experiment that would disrupt the Internet if run on the Internet would be appropriate for GENI. Further, GENI has a programmable infrastructure, so the routers and other infrastructure systems can be reprogrammed from other nodes (unlike the current Internet). This allows the edges (end nodes) and the core (infrastructure) to collaborate, for example on security defenses or measurements; this is not possible in the current Internet, in general.

Three other types of experiments were discussed. GENI offers the opportunity to evaluate the security of deployed solutions on a large-scale distributed network and/or system. For example, one can use GENI as the testbed for a large distributed system or application, and then analyze it to determine whether it *really* is secure, robust, and scalable. One can also use GENI to test (or simulate) very high cost, but low probability, events for complex scenarios and novel threats.[4] The third type was an exercise like CyberStorm to develop plans and procedures to deal with threats against large distributed networks and systems.

Concepts that start as an experiment may develop constituencies of users who depend on the implementation. Thus, the experiment may evolve into a service. As noted above, this raises the problems of handling intellectual property, and transitioning the experiment to a

[4] Some participants referred to GENI supporting an "Underwriters Laboratory" for security technology. This raises many issues such as quality control, requirements testing, and such that the workshop did not pursue.

production service. GENI needs to express the rules governing solutions to these problems in its environment, and develop mechanisms to support and promote this growth.

As more people want to use GENI, teaching them how to implement experiments correctly, and analyze the results, becomes critically important. The participants all felt that GENI should provide a set of detailed experiments that users could modify to learn how to do experiments on GENI—even simple experiments would be very helpful. Recipes or cookbooks for constructing and running experiments will prove invaluable, too.

Finally, several participants pointed out that GENI could be used as a teaching tool for how to carry out scientific experimentation in computer science and, especially, computer security. Given the need for emphasis on rigorous scientific testing in computer science curricula, this use may be the most important for the future of computer science and computer security.

GENI Itself. The workshop also discussed protecting GENI, and ensuring experiments stayed on GENI. The phrasing here is critical. It is not possible to prevent attacks on GENI, and undoubtedly some will succeed. The issue then is how to minimize the effects that those attacks have on GENI, and on the experiments being run; and how to ensure the experimenters are notified of the attack so they can take that into account when analyzing their results.

A key issue is legal liability. For example, suppose a malware experiment in GENI goes awry because GENI's mechanism for isolating the slice fails. What are the legal ramifications, especially when the network crosses international borders? How do we ensure that the GENI constituents can communicate among themselves to deal with terminating the worm, and repairing the damage to GENI and to others, effectively? As another example, suppose an attacker compromises a system belonging to GENI and implants a botnet on GENI. This not only compromises GENI, but it also renders many experiments (for example, those relating to network throughput) suspect.

One approach that many (especially the practitioners) thought would help would be to use "red teams" to compromise GENI to test the ability of the GENI organization, and the federated organizations making up GENI, to respond to attacks. The goal would be for the red team to disrupt some aspects of the GENI testbed (preferably those not being used for experimentation) and see how long it took to detect and restore those parts. This tests not only the technical protections but also the procedures, the availability, and the readiness of the constituents and the managers to act quickly.

Finally, the participants noted that GENI itself is an experiment: a federation and testbed of this complexity has not yet been created. Therefore, we should consider having social scientists study GENI itself and how users, organizations, and others interact with GENI and with one another. The goal here is to improve the usefulness and usability of GENI to make it as effective a testbed as possible.