Devlan: Automated Acquisition of Digital Evidence from Large Networks

NIJ 2016-IJ-CX-K002
Final Report FR-001R1

Grier Forensics
8903 Greylock Road
Pikesville, MD 21208
DUNS:
EIN:
Jonathan Grier, Principal Investigator
Contact: 410-220-0962, jgrier@grierforensics.com

NIJ 2016-IJ-CX-K002
Project Period: 1 Jan 2017 – 30 Sep 2019
Report Period: 1 Jan 2017 – 30 Jun 2019
Submission Date: 30 Jul 2019
Report Term: Semi-annual
Final Report: Yes

1. ACCOMPLISHMENTS

1.1 Major goals of the project

Devlan's mission is to enable investigators to acquire digital evidence from enterprise scale networks

1.1.1 Problem

Investigators need to obtain digital evidence from computer systems. Current practices and tools are excellent at extracting items from a single laptop or hard drive.

Large-scale computer networks would also be a potential source of valuable digital evidence in a criminal justice investigation. However, law enforcement organizations, especially at the state and local level, typically lack the resources (both technology and staff) to acquire this valuable evidence.

Five key factors make acquisition of evidence from networks, especially large-scale networks (thousands of devices), challenging:

1. **Data Size and Scale**: A large-scale network can easily contain 1 PB (petabytes) of data (5000 devices x 200 GB/device = 1 PB). Collecting that much raw data is infeasible, overwhelming even the most powerful technology. Additionally, and more importantly, this type of collection would rarely be within the limits of a warrant.
2. **Distribution**: A large-scale network is a collection of physically distinct but interconnected devices. This fails to provide that single, concrete entity that is typically subjected to traditional forensics (seized and processed).
3. **Disruption**: Large-scale networks contain vast stores of different types of data. Furthermore, they are multipurpose. Traditional forensics allows a legal authority to seize and examine particular devices. However, authority to seize or disrupt an entire network is rarely granted. Indeed, the *NIJ Electronic Crime Scene Investigation Guide* warns of severe potential civil liability for even inadvertent disruption. Thus, evidence from networks must be acquired without causing disruption.
4. **Diversity**: Networks contain a bewildering array of different devices, operating systems, applications, and data types. No one tool can hope to encompass such a range of diversity.
5. **Dynamics**: Technology is innovating at a rapid pace, far beyond the speed that any tool-vendor can match. The recent RAND Corporation report describes this as a "whack-a-mole" problem trying to catch up with innovation.

1.1.2 Solution

Devlan searches for and collects evidence from large networks, spanning potentially thousands of devices, with minimal interruption of business operations.

Devlan addresses the problems of scale and diversity in large-scale networks by using the functions and computing power of the network under investigation. It leverages the libraries and processors of the network itself to locate and acquire forensic evidence. For example, Microsoft's Windows Search, using an open architecture of filters, continuously indexes and searches textual and metadata information from files, emails, and other sources of data.

Devlan addresses the problems of disruption through being a digital forensics system that is brought to the Network Under Investigation. Devlan is deployed, searches are executed, evidence acquired, and then Devlan is removed and the evidence taken back to the investigators' office.

While the bulk of the Hosts in the Network Under Investigation are typically Windows based, Devlan's architecture (and APIs) are completely operating system agnostic.

To support a variety of search techniques, Agents use an open, pluggable framework. This common paradigm allows both the search and acquire mechanisms of the Devlan Agent to be easily extended, by Grier Forensics or by third parties.

1.2 What was accomplished under these goals

1.2.1 Architecture, Workflow, and Software Implementation of Devlan

We defined, implemented, and tested, the architecture, workflow, and software implementation of Devlan, as described in detail below.

Architecture

Devlan's architecture consists of three distinct types of components, as shown in the diagram below:

- User *Consoles*, the sole point of investigator interaction with Devlan,
- A central, rack-mounted *Controller*,
- A network of *Agents* deployed on the hosts of the network under investigation.

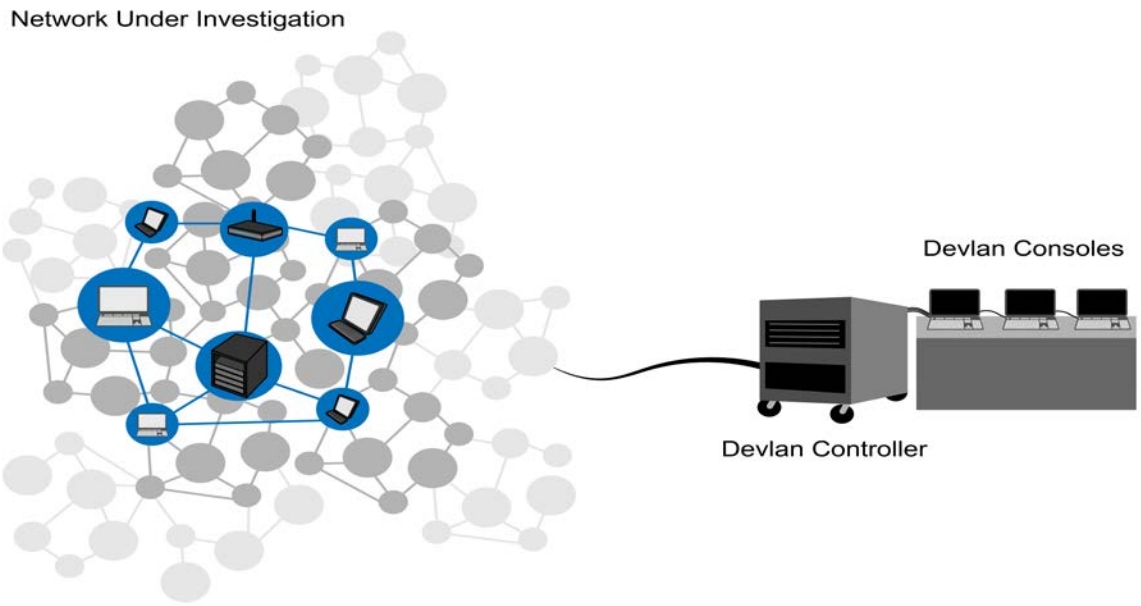


Figure 1 Devlan Physical Architecture

Note that the investigators' Consoles are not directly connected to the network under investigation; the Controller serves as a bridge between the Consoles and the deployed Agents.

Investigation Workflow

The primary Devlan functions are to search for evidence items and then to acquire the items selected by an investigator.

Search

A Devlan search is depicted in the following illustration.

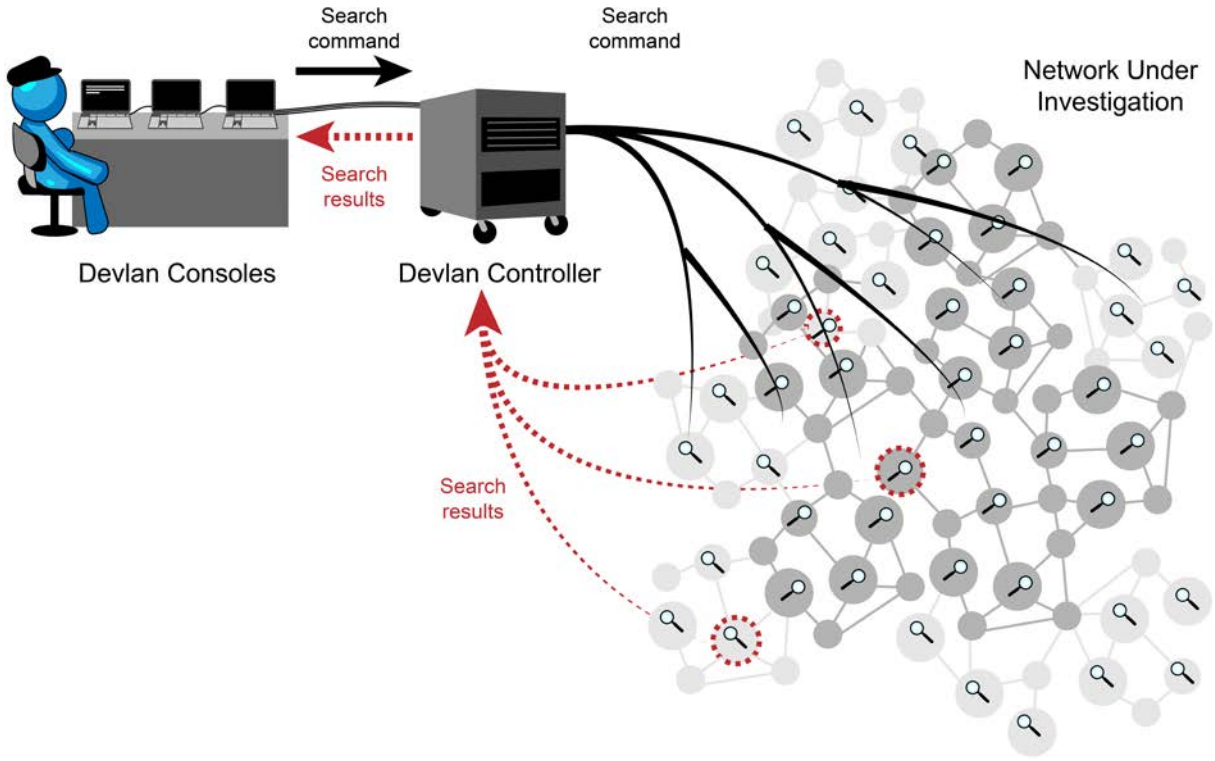


Figure 2 Devlan search workflow

The workflow of the search is:

1. An investigator initiates a search request from a Console. (An investigator may have multiple searches in progress, and there may be multiple investigators.)
2. The Console sends a command to the Controller.
3. The Controller relays the search request to each Agent.
4. Each Agent executes a search of its files, emails, etc. There may be multiple searches in progress on an Agent, and Agents execute searches concurrently with each other.
5. Each Agent sends search results in the form of URIs and metadata back to the Controller.
6. The Controller continuously aggregates the search results from different Agents. Note that each search has its own collection of aggregated results.
7. The investigator can examine the search results on his/her Console.

Acquire

Using Devlan to acquire evidence is depicted below:

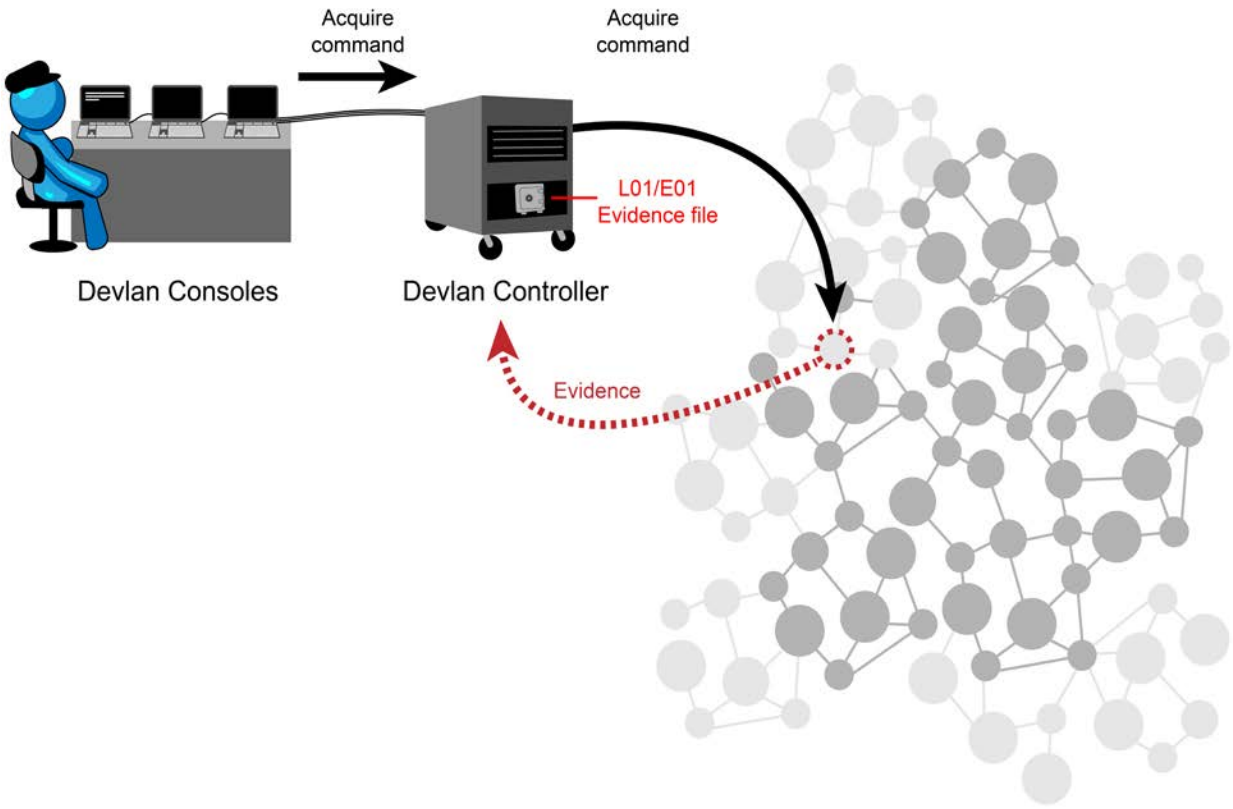


Figure 3 Devlan acquire workflow

The workflow for evidence acquisition is:

1. An investigator uses his/her Console to request that a selected evidence item be acquired. Multiple requests may be in progress.
2. The Console sends this request to the Controller.
3. The Controller sends a command to the specific Agent that advertised the evidence item.
4. The Agent uploads the item to the Controller. Multiple uploads may be in progress on an Agent and across Agents.
5. The Controller stores the evidence item in L01 or E01 forms of Evidence Compression Format.

Status of the upload can be shown on the Console (not shown above)

1.2.2 Model Network

We built a model network with the key features of a large network that Devlan may be used to investigate, and completed the engineering necessary for Devlan to successfully investigate it.

Specifically, in order to test, measure, and demonstrate Devlan, we implemented a model network, small enough to be operable within our lab resources yet still maintaining the key features of a large network that Devlan might be used to investigate. This network features multiple operating systems, platforms, and roles, all networked together in a single domain. The network is populated with data and documents from the Naval Postgraduate School's Digital Corpora Govdocs, a corpus designed specifically to foster scientific research into digital forensics, as well as data from the Enron investigation, which entered the public record in the Enron trials. The network also includes experimentation harnesses, allowing us to "snapshot" the network and rollback to the state before the experiment. We have successfully deployed Devlan to the network and used the network to test and refine Devlan.

1.2.3 Forensic Acquisition

We implemented a forensic acquisition mechanism via Devlan, as well as a storage and preservation mechanism. It is critical that Devlan be able to acquire large amounts of evidence from high volume networks without disrupting network usage. Consequently, we built an acquisition mechanism that can handle large volumes of data seamlessly; makes proper use of compression; offloads compression to the network edge, to reduce computational load; uses proper representational state transfer semantics; and asynchronously acquires data from across the network. Moreover, the mechanism stores data in a new format we specifically invented, based on the widely adopted ZIP standard, augmented for forensic purposes, including preservation of audit trails and other forensic metadata.

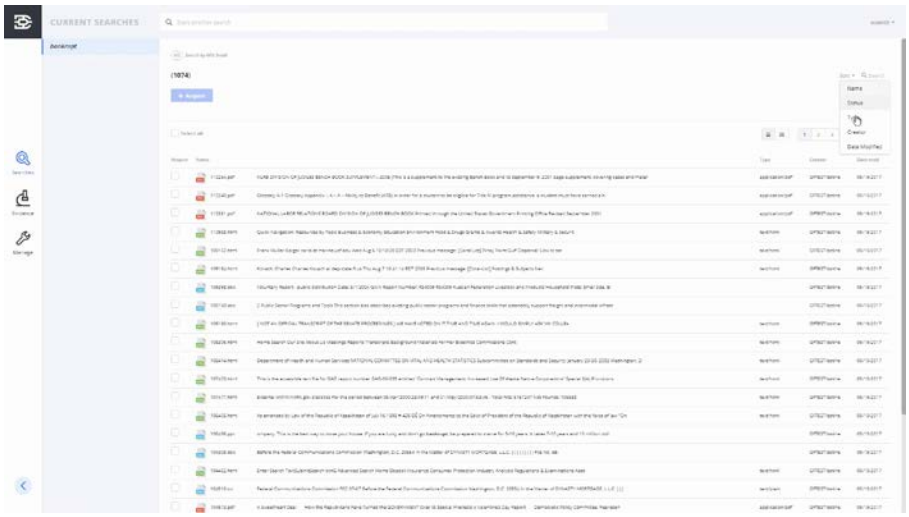
1.2.4 Integration & System Engineering

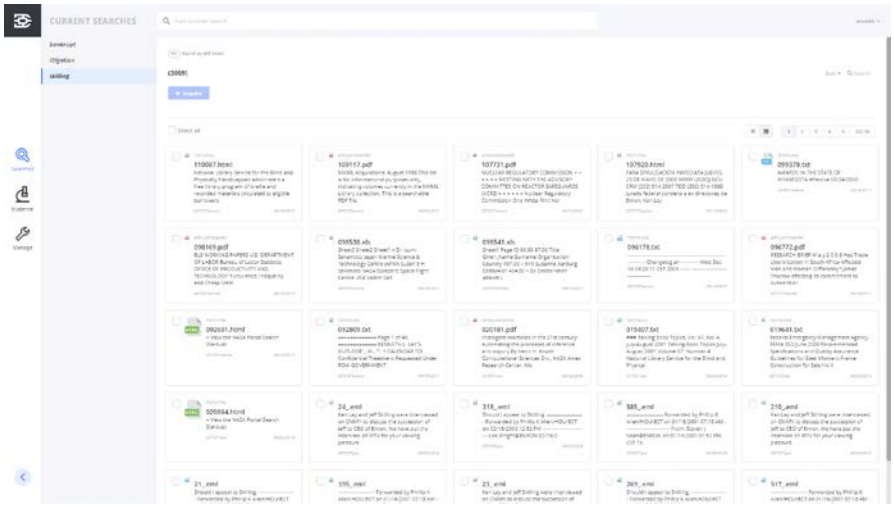
Having successfully implemented and tested the underlying components of Devlan, we integrated them into an end-to-end system, completed interfaces between them, addressed gaps and required modifications, tested and documented the system, culminating in end-to-end demonstration of Devlan on the model network.

1.2.5 User Interface

For law enforcement personnel to use Devlan to investigate large networks, it is critical that it be usable without advanced computer training. We designed, implemented, integrated, tested, and demonstrated the Devlan User Interface, based on a workflow designed to be useful and immediately familiar to investigators. The Devlan User Interface enables any investigator to use a graphical user interface (GUI), similar to an

Internet search engine, to forensically search for and acquire evidence from a large network. Screenshots of the Devlan User Interface are shown below.





1.2.6 Benchmarks & Measurements

To measure the performance, behavior, and accuracy of Devlan, we developed a set of benchmarks and measurements, as well as a measurement procedure, involving controlled testing and measurement of Devlan on the populated lab-scale test network. Furthermore, we created a model investigatory environment, populated with real-world data from the Enron investigation, which entered the public record in the Enron trials, as well as documents from the Naval Postgraduate School's Digital Corpora. In this environment, we tested Devlan as well as a conventional forensic tool (to serve as an experimental control), and measured accuracy and performance. Devlan achieved 95.1% precision, 95.8% recall, and 95.5% F1, at a speed 170x faster than the conventional forensic acquisition tool.

1.3 Opportunities for training and professional development

In conjunction with National Institute of Justice personnel, we presented *Big Forensics: Investigating Very Large Organizations* at the Techno Security and Digital Forensics conference, providing opportunities for State, local, Federal, and private forensic investigators to learn about the challenges in investigating large networks as well as how this work can be used to solve them. We also demonstrated Devlan to NIJ personnel and visiting researchers at our office.

1.4 Dissemination of results

We have submitted our work to the *Journal of Digital Investigation* and the *DFRWS Digital Forensics Research Conference* as a peer-reviewed scientific paper, currently under review. Notifications of acceptance are typically expected in April or May. Drafts of the paper have been submitted to NIJ.

The paper's abstract notes:

> We present a new approach to acquiring evidence from large organizations with thousands of networked devices. Given their scale and complexity, it is impossible to build a self-contained tool encompassing all the logic, processes, and power needed to satisfactorily acquire evidence from such organizations. However, we assert that there is no need to do so, as large networks already contain all the tools and resources needed for acquiring forensic evidence. Devlan leverages these resources, and the network itself, to provide an asynchronous, stateless, and pluggable means of search and acquisition of digital evidence in large-scale environments. Devlan acquired evidence from a small-scale testbed at 95.1% precision and 95.8% recall (F_1 = 95.5%) in just 4.6 seconds, and was 8.4% more accurate and 170x faster than a conventional forensic acquisition tool. Devlan presents a powerful and scalable approach to investigators, tackling the challenges involved in locating and acquiring evidence across large-scale networks.

Additionally, in conjunction with the National Institute of Justice, we presented Devlan at the Techno Security and Digital Forensics Conference.

1.5 Plans for next reporting period

N/A.

2. PRODUCTS

2.1 Software

This effort has resulted in the Devlan software system, as described in this document.

2.2 Model Network

The effort has resulted in a model network, complete with key aspects of a large enterprise network, allowing Devlan to be tested, demonstrated, and measured. This network has been populated with data and documents from the Digital Corpora Govdocs Corpus, published by the Naval Postgraduate School for scientific research into digital forensics, as well as data taken from the Enron investigation.

2.3 Technologies

The effort has resulted in development and demonstration of a network-wide, heterogeneous, asynchronous, federated search and acquisition, and shown how this technology can solve the key obstacles identified in Section 1.1.1 above.

The effort has resulted in development and demonstration of multiple system-specific network endpoints for forensic investigation, and shown how this technology can be used as part of the federated search and acquisition described above.

The effort has resulted in a plugin application programmer's interface (API), allowing direct integration of new technologies and network components into the forensic investigation.

3. IMPACT

In addition to previously reported demonstrations, we performed and documented during this period a system wide end-to-end demonstration, showing Devlan operate on the model network described above. In this demonstration, Devlan is deployed to a model network that investigators wish to investigate, and used to investigate it.

Devlan is posed to enable law enforcement to investigate very large networks, an ability they have to date lacked. Specifically, despite the fact that large-scale computer networks are often a potential source of valuable digital evidence in criminal justice investigations ranging from combating terrorism to economic crimes, law enforcement organizations, especially at the State and local level, typically lack the resources, technology, and staff needed to acquire this valuable evidence. Indeed, the *National Institute of Justice Electronic Crime Scene Investigation Guide* explicitly cautions law enforcement of such, noting "...Processing a crime scene where the computer systems are networked poses special problems... [which] can result in loss of evidence and potential severe civil liability.... When investigating criminal activity in a known business environment, the presence of a computer network should be planned for in advance, if possible, and appropriate expert assistance obtained.... The possibility of various operating systems and complex hardware configurations... make the processing of a network crime scene beyond the scope of this guide." Thus, law enforcement is, to date, often denied the evidence that these networks contain. As activities and communications increasingly move to digital networks, this gap, if, not resolved, may ultimately blind law enforcement to a large segment of criminal activity. Devlan stands posed to address this, restoring to law enforcement the ability to properly investigate digital networks of all size.

Crimes such as white-collar crimes and fraud, organized crime, terrorism, and public corruption, are becoming increasingly rooted in large-scale networks. By enabling law enforcement to effectively investigate these crimes, Devlan will assist law enforcement to keep up with the 21st century, and preserve public safety, criminal justice, and constitutional rights in the digital age.

Scientifically, Devlan has demonstrated technologies enabling rapid investigation of large, heterogenous, dynamic, fluid, evolving networks, a capability we feel will have a large impact on many problems in computer science and technology.

4. CHANGES/PROBLEMS

4.1 Changes in approach and reasons for change

During this reporting period, there were no substantial changes in approach.

4.2 Actual or anticipated problems or delays and actions or plans to resolve them

During this reporting period, no actual or anticipated problems or delays manifested.

4.3 Changes that have a significant impact on expenditures

During this reporting period, there were no changes that have a significant impact on expenditures.

4.4 Significant changes in use or care of human subjects, vertebrate animals, and/or biohazards

During this reporting period, there were no significant changes in use or care of human subjects, vertebrate animals, and/or biohazards.

4.5 Change of primary performance site location from that originally proposed

During this reporting period, there was no change of primary performance location.

5. PERFORMANCE MEASURES

Objective: Conduct research in science, technology, engineering, and/or mathematics having clear implications for criminal justice policy and practice in the United States.

	Performance Measure	Status
1.	Relevance to the needs of the field as measured by whether the project's substantive scope did not deviate from the funded proposal or any subsequent agency-approved modifications to the scope.	Project's scope was devoted exclusively towards research and development into *Devlan: Automated Acquisition of Digital Evidence from Large Networks*, as detailed in our funded proposal.

2.	Quality of the research as demonstrated by the scholarly products that result in whole or in part from work funded under the NIJ award (published, peer-reviewed, scientific journal articles, and/or (as appropriate for the funded project) law review journal articles, book chapter(s) or book(s) in the academic press, technological prototypes, patented inventions, or similar scientific products).	Project has produced a scientific journal article currently undergoing peer-review. Project has been presented, in conjunction with NIJ personnel, at a leading industry conference.
3.	Quality of management as measured by such factors as whether significant project milestones were achieved, reporting and other deadlines were met, and costs remained within approved limits.	Significant project milestones were met and demonstrated. Reporting deadlines have been met. Costs have remained within approved limits.
4.	Number of technologies fielded as a result (in whole or in part) of work funded under the NIJ award.	Project has produced *Devlan*, which has achieved an 170x acceleration over conventional forensic approaches, while maintaining 95%+ accuracy under, as measured under laboratory testing simulating its intended operational environment.