keepgrabbing.py: additions, and documents.

Camille Akmut

January 13, 2020

Abstract

Important program continues to be reviewed from all perspectives.
Additions

- Python’s `str()` function takes an object and converts it to a string...

```python
>>> str(1)
'1'
```

In this example, the integer 1 becomes the string ‘1’ (or “1”).

```python
>>> type(1)
<class 'int'>
>>> type('1')
<class 'str'>
```

(Side note:

In Python, the single quotes ‘ ‘ do not signify a character as opposed to a string; for which double quotes “ “ would be used - in Haskell.

In that former language, they are the same (type of objects, they have the same type):

```python
>>> '1' == "1"
True
>>> '10' # '10' is not a character (but '1' is for example)
'10'
>>> type('10')
<class 'str'>
>>> type("10")
<class 'str'>
```

- Haskell’s version of Python’s `split()` is `words` (from the `Data.List` module):

```haskell
Prelude Data.List> words "article1 article2"
["article1","article2"]
```
Here I've re-written the proxy of the lambda function that I showed, in Haskell:

```haskell
Prelude> line = (\x -> x^2)
Prelude> line 3
6
Prelude> blocks = [3,2,1]
Prelude> map line blocks
[6,4,2]
```

(mapping is the functional programming equivalent of looping, as found in languages with imperative elements / Python)

or, closer to Aaron's application:

```haskell
Prelude> line = (\x -> ["curl"] ++ [x])
Prelude> line "--proxy"
["curl","--proxy"]
```

Now you may be asking yourself: how do we go from this list of strings, to something that could be passed to Curl/Wget or the Shell?

We do this:

```haskell
Prelude> unwords ["curl","--proxy"]
"curl --proxy"
```

(Haskell has a function just for this, unwords.

If it did not we would have needed to write our own, most likely a fold.

Here:

```haskell
Prelude> foldr (\x y -> x ++ y) [] ["curl","--proxy"]
"curl --proxy"
```

Can also be simplified to foldr (++), which turns out is the definition of `concat`. This is because of point-free style I believe.

We would have used `head` to extract the string if a list had been given back to us as intermediary result, but such is not the case.
e.g. head ["curl —proxy"] == “curl —proxy”)

- More context:

Timeline of events

<table>
<thead>
<tr>
<th>Dates</th>
<th>Actions</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sep. 25-26th 2010</td>
<td>&gt;450k .pdfs downloaded</td>
<td>IP address changed once, Class C network blocked</td>
</tr>
<tr>
<td>Oct. 9th 2010</td>
<td>8k .pdfs (only)</td>
<td>entire MIT network blocked</td>
</tr>
<tr>
<td>Late Nov. 2011 - Early Jan. 2011</td>
<td>? (lots, presumably)</td>
<td>new IP address, slowed below-monitoring downloading.</td>
</tr>
</tbody>
</table>

“Overall, more than 450,000 articles spanning 560 journals were downloaded between 5:00 p.m. Saturday and 4:00 a.m. Sunday. (…)

JSTOR engineers temporarily blocked further downloads directed to the MIT … IP address issuing the requests. But the downloading continued from a different IP address. (…)

The next day, Sunday, September 26, 2010, JSTOR shut off access for the entire range of addresses (Class C network) containing the two addresses. (…) The first IP address was 18.55.6.215, and the second address was 18.55.6.216. The blocked range consisted of all IP addresses beginning with 18.55.6 (…)

The network DHCP server maintains a log, called a DHCP log, which records the IP address assigned to a MAC address, as part of the DHCP process. At MIT, an IP address will often identify the building where network device is located.“

“Two weeks later, on Saturday, October 9, 2010, during the Columbus Day weekend, a second, similar incident occurred: a visitor downloaded more JSTOR articles, using a slightly modified MAC address. (…) JSTOR’s response was to shut down service, at approximately 11:15 p.m. on October 9, 2010, to all MIT’s IP addresses, that is, the entire Class A network, doing so quickly enough that only about 8,000 articles were downloaded during this incident.“

“December 26, 2010, JSTOR again noticed excessive downloading from MIT, originating from a new IP address. Significantly, this most recent downloading had been going on for some time, beginning in late November, but JSTOR did not realize this fact until much later. The manner of accessing downloads had been
slowed and altered in such a way that JSTOR's monitoring systems did not identify that the robotic harvesting had resumed."

(Abelson et al. 2013, 'Report to the President')

- "Cookie: TENACIOUS=" + str(random.random())[3:]

```
>>> "Cookie: TENACIOUS=" + str(random.random())[3:]
'Cookie: TENACIOUS=2113961635582544'
>>> "Cookie: TENACIOUS=" + str(random.random())[3:]
'Cookie: TENACIOUS=6451074056572803'
```

"It deleted its JSTOR cookie after each download, disconnected, and then re-accessed JSTOR, resulting in a new cookie being placed on the machine each time this occurred and making it appear that this was a new machine for each access." (Ibid.)

(I'm not entirely sure that Abelson's understanding of the program here is entirely correct.

It's not so much that the cookie was deleted, but a new one generated with every new curl execution. See above.)

...
Annex

- Document 1: Abelson et al., Report to the President of MIT.
PART I: EVENTS LEADING TO THE ARREST

I.A Downloading of JSTOR Articles
I.B Discovery of the Laptop
I.C Events of January 6, 2011: The Arrest
I.D Events of January 6, 2011: Seizure of the Laptop
I.E Access to the MIT Network
   I.E.1 Connecting to the MIT network
   I.E.2 JSTOR and eControl

Part I: EVENTS LEADING TO THE ARREST

I.A Downloading of JSTOR Articles

The history of the events leading to the arrest of Aaron Swartz in January 2011 begins the previous fall with the JSTOR (Journal Storage) digital library, a service that licenses scholarly journals to numerous academic and research organizations, including MIT.¹

On the evening of Saturday, September 25, 2010, JSTOR engineers noticed an extremely large number of requests for downloads originating from MIT. Overall, more than 450,000 articles spanning 560 journals were downloaded between 5:00 p.m. Saturday and 4:00 a.m. Sunday.² The volume of data transferred was enough to overload the affected JSTOR server. In response, JSTOR engineers temporarily blocked further downloads directed to the MIT Internet Protocol (IP) address issuing the requests.³ But the downloading continued from a different IP address.

---

¹ See Appendix 6 for more information on JSTOR and the MIT Libraries.
² Ordinarily, when someone requests a download from JSTOR, the system pops up a window that refers to JSTOR’s terms of use, and requires the user to click to confirm before the download can proceed. This not only informs the user of JSTOR’s terms, but also limits the rate at which automated downloads can be requested. In this case, however, the download script included a flag (acceptTC=true) that bypassed the acceptance step.
³ In an effort to warn the person causing these downloads to stop, JSTOR’s engineers caused a web page to be presented to the computer engaged in the downloading, reading:
   Access Suspended—We noticed content downloading activity from your IP address (18.55.6.215) that appears to be in excess of what is allowed under our Terms & Conditions of Use. Please review our terms for more information about allowable uses. If you have additional questions or need other information, please contact JSTOR Support.
The next day, Sunday, September 26, 2010, JSTOR shut off access for the entire range of addresses (Class C network) containing the two addresses. It sent an email to the MIT Libraries (“the Libraries”) notifying them of this fact, and explaining that JSTOR “rarely takes this level of response to abusive activity, but felt it necessary to maintain the stability of the Web site for other institutions and users.” JSTOR further noted in that email that the manner in which the download requests came into its system “clearly indicates robotic harvesting of PDFs [articles] which violates our Terms & Conditions of Use.” Reflecting its contractual agreement with JSTOR, the Libraries began to work with JSTOR in an effort to determine the source of the downloading and stop it from continuing.

JSTOR sought MIT’s assistance to prevent the incident’s recurrence. The Libraries responded: “We’re investigating this case and, because the origin of the activity was a guest visiting MIT, we believe that it will not recur. We hope that you will be able to restore the Class C range that has been suspended on this information.” In response, JSTOR turned all of MIT’s IP addresses back on, and decided to watch.

To provide some context for this event, we note that the Libraries handled 65 excessive use incidents during the 2010–2011 academic year. Typically, when an excessive use case is reported that is determined to originate from within MIT’s network, the Libraries report this to either the MIT Information Services and Technology (IS&T) network security team or MIT’s “Stopit” group, which deals with inappropriate behavior that occurs electronically. The Stopit group’s general response is to send the offender a warning email message. This is almost always all that is needed to get people’s attention and have them stop whatever it was they were doing that caused the problem. In this case, however, the computer was registered to a visitor, and the registrant used an anonymous

---

4 The first IP address was 18.55.6.215, and the second address was 18.55.6.216. The blocked range consisted of all IP addresses beginning with 18.55.6 followed by a number from 0 through 255. Such a range (four numbers, with the first three fixed, and the last one taking on any value from 0 to 255) is called a Class C network.

IP (Internet Protocol) addresses specify the network port where the device is attached. They are typically automatically assigned, when the device is attached, by a network service called a DHCP (Dynamic Host Configuration Protocol) server, although people with sufficient computer skills can change the IP address (as was done here). In addition to the IP address, each device has a MAC (Media Access Control) address, which uniquely identifies the device’s hardware network interface itself (essentially, the device itself, as opposed to where it is attached to the network). MAC addresses are typically assigned by the device manufacturer, although these, too, can be changed under program control. The network DHCP server maintains a log, called a DHCP log, which records the IP address assigned to a MAC address, as part of the DHCP process. At MIT, an IP address will often identify the building where network device is located.

5 Email sent Sunday, September 26, 2010, at 12:31 p.m.

6 Most cases of excessive downloading are due to misappropriation of MIT credentials, in which case the true MIT user is asked to change his/her password.
email address that could not be contacted.\footnote{As described in section I.E.1, guests may register to use the MIT network by supplying a name and an email address as contact information, and they obtain a registration that is valid for a limited time period.) IS&T therefore disabled the registration of the MAC address used by the offending computer, expecting that this would be a sufficient deterrent to further activity.}

Two weeks later, on Saturday, October 9, 2010, during the Columbus Day weekend, a second, similar incident occurred: a visitor downloaded more JSTOR articles, using a slightly modified MAC address.\footnote{This time, the requests and downloads stimulated a cascade of failures that brought down multiple JSTOR servers. Half the servers in one data center failed, and JSTOR engineers feared that the entire service might go down worldwide. Moreover, the requests seemed to be coming from thousands of machines.}

JSTOR’s response was to shut down service, at approximately 11:15 p.m. on October 9, 2010, to all MIT’s IP addresses, that is, the entire Class A network,\footnote{JSTOR’s response was to shut down service, at approximately 11:15 p.m. on October 9, 2010, to all MIT’s IP addresses, that is, the entire Class A network, doing so quickly enough that only about 8,000 articles were downloaded during this incident. JSTOR observed that the downloaded articles were not limited to a specific discipline, but were sequential across JSTOR’s entire database. To JSTOR, this indicated “a concerted effort is being made to download the entirety of the JSTOR archive.” JSTOR notified MIT Libraries of its findings by email, and spoke directly with personnel at the Libraries about its concerns.} doing so quickly enough that only about 8,000 articles were downloaded during this incident. JSTOR observed that the downloaded articles were not limited to a specific discipline, but were sequential across JSTOR’s entire database. To JSTOR, this indicated “a concerted effort is being made to download the entirety of the JSTOR archive.” JSTOR notified MIT Libraries of its findings by email, and spoke directly with personnel at the Libraries about its concerns.

\footnote{\textsuperscript{7}The addresses supplied with the registrations were generated by Mailinator (\url{http://www.mailinator.com}), a service that creates on-the-fly temporary email addresses.}

\footnote{\textsuperscript{8}(See footnote 4 for an explanation of MAC addresses.) The initial MAC address was 00:23:5a:73:5f:fb, registered on September 24, 2010, to a “person” named “Gary Host.” The second (slightly modified) MAC address was 00:23:5a:73:5f:fc, registered on October 2, to “Gary Host.” These were the same machine, as evidenced by the similarity of MAC addresses and the fact that they presented the same DHCP client ID, “ghost-laptop,” to the DHCP server. The MAC address provided is consistent with this being an Acer laptop. (See Appendix 7 for information on DHCP.) In addition, there was a second machine registered, on October 8, 2010, to “Grace Host” with the MAC address 00:17:f2:2c:b0:74. This machine had the client ID “ghost-macbook,” and the MAC address is consistent with this being an Apple MacBook. There was also a third registration, on October 22, 2010, with user name “Grace Host” and client ID “ghost-laptop,” with MAC address 00:4c:e5:a0:c7:55, and again a registration on November 28, 2010, with MAC address 00:4c:e5:a0:c7:56. These registrations are for the Acer laptop that was discovered in the closet. (See section I.B below.) The MAC addresses used here are invalid: they do not correspond to any manufacturer ID. This is apparently the same “ghost-laptop” machine as was registered on September 24. One can conjecture that the owner, having seen the first MAC address blocked, and the second, slightly modified address also blocked, set the machine to a radically different MAC address.}

\footnote{\textsuperscript{9}In actuality, there was only one machine. It deleted its JSTOR cookie after each download, disconnected, and then re-accessed JSTOR, resulting in a new cookie being placed on the machine each time this occurred and making it appear that this was a new machine for each access.}

\footnote{\textsuperscript{10}That is, the range of IP addresses starting with 18 and followed by three numbers (i.e., 18.x.x.x).}
Both MIT and JSTOR were anxious for service to MIT to be restored. JSTOR believed that it could monitor the system and stop further incidents because it had been able to notice and stop the latest one after only 8,000 articles had been downloaded. Based on this, JSTOR agreed to restore service to MIT and did so on Tuesday, October 12, 2010, after three days of the entire campus being blocked. At the same time, IS&T blocked access from the individual MAC address most recently associated with the downloading. Meanwhile, IS&T staff were able to determine that the downloading activity had originated from the Dorrance Building (Building 16), an academic building in the central campus.\footnote{The September 25, 2010, downloading was also from a wired connection in Building 16. IS&T did not pinpoint the exact location in either September or October 2010.}

Also on October 12, the Director of the MIT Libraries reported to MIT’s Academic Council\footnote{MIT’s Academic Council consists of the Institute’s senior leadership plus the elected Chair of the Faculty. It is chaired by the President and meets weekly during the academic year to confer on matters of Institute policy.} that a cyber-attack of the JSTOR database had caused a weekend shutdown of JSTOR to the entire campus.

Following the suspension of service to the entire MIT campus during October 9 through 12, JSTOR decided to monitor closely for additional downloading activity and be prepared to suspend access as necessary.

On the evening of December 26, 2010, JSTOR again noticed excessive downloading from MIT, originating from a new IP address. Significantly, this most recent downloading had been going on for some time, beginning in late November, but JSTOR did not realize this fact until much later. The manner of accessing downloads had been slowed and altered in such a way that JSTOR’s monitoring systems did not identify that the robotic harvesting had resumed.\footnote{There were over 4.3 million downloads during the period from late November through Swartz’s eventual arrest on January 6.} This time, JSTOR noted that the downloading activity originated from the same Class C network that IS&T had identified earlier as being in Building 16.\footnote{Beginning December 26, 2010, JSTOR took several actions to stop or at least impede the downloading. First, it blocked an entire Class C range of addresses for Building 16. JSTOR also kept open the one IP address that it observed the machine using, 18.55.6.240, and moved access for this address to a server separate from the rest of its network. Through this server, JSTOR responded to the machine’s requests by downloading strings of zeros and meaningless articles, in the expectation that: (1) the machine would not notice and would continue to download worthless material; and (2) this would slow any remaining downloading that might take place. However, the machine had also been registered with an IP address of 18.55.7.240, that is, with a 7.xxx instead of 6.xxx and entirely outside of the Class C range of IP address that JSTOR had blocked. The downloading continued unimpeded, without JSTOR realizing it.} JSTOR promptly notified the MIT Libraries about the new incident, by email, on the same evening, identifying Building 16 as the apparent location of the IP address. In this email, JSTOR also made the following request to MIT: “We are
requesting that every effort be made to identify the individuals responsible and to ensure that the content taken in this incident and those previously mentioned is secured and deleted.” An email sent the following day from JSTOR to the Libraries re-emphasized the urgency of the situation: “Once again, we are seeing extreme unauthorized activity from MIT. We really need to find out who is doing this; it is malicious and intentional and as best we can tell is coming from inside of MIT.”

Employees of the MIT Libraries had been furloughed for the winter holidays, and thus they did not see the messages JSTOR sent on December 26 and 27, 2010, until Monday, January 3, 2011. On the morning of January 4, the Libraries informed JSTOR that MIT did not expect to be able to identify the individual involved in these incidents based on the information available at that point.

I.B Discovery of the Laptop

At approximately 3:00 a.m. on January 4, 2011, IS&T staff sent an email to its network engineers requesting that they trace the exact location in Building 16 of the computer using the IP address. An IS&T network engineer began to search when he arrived for work in the morning, going to Building 16 and checking the basement closet containing the building’s network switches. Around 8:00 a.m., he entered the closet and saw a cable connected to a network switch and leading to a cardboard box on the floor. He lifted the box and saw a laptop computer. He telephoned an IS&T network manager, who quickly joined him at the closet.

\[15\] The MIT Libraries were closed for business from 6:00 p.m. Wednesday, December 22, 2010, through Sunday, January 2, 2011. The majority of these days were regular or special Institute holidays. However, December 27, 28, and 29, 2010 were furlough days, during which time the staff were required by the Libraries’ administration to take involuntary, unpaid leave as a cost-saving measure to meet budget reductions. Furloughed staff were explicitly prohibited from working during furlough days. Two technical staff members were asked to defer their furlough days to another time, so that basic technical support to the Libraries’ networked resources could be provided. Their directives included that (a) access to licensed resources was to be kept available, both on and off campus, to the degree possible; and (b) any outages caused by problems at MIT’s end were to be resolved promptly. In the unlikely event that a major aggregator or database should cut off service to MIT during this time, one of the two staff members was responsible for making best efforts to work with the vendor to resolve the outage. This staff member was copied on JSTOR’s email of December 26, 2010, to the Libraries but did not respond. The Libraries wrote to JSTOR on January 3, explaining that people had been on furlough and had not seen the JSTOR’s prior messages. This January 3 message was the Libraries’ first response to JSTOR’s December 26 and 27 emails.

\[16\] The January 4 email also suggested to JSTOR that it block the entire 18.55.xxx.xxx Class B network, since the downloading was coming from two different Class C networks. (See footnote 14.) At the time the email was sent, the Libraries did not know that that laptop had been discovered a few hours before.

\[17\] The closet had two doors connected in the middle by a common lock, with both doors swinging outward when opened. He used his key to enter the basement closet; however, he does not remember whether the doors were actually locked. According to this network engineer, even when locked, the closet could be opened by pulling on both doors simultaneously because the locking mechanism had been damaged.
Over the next hour, the two engineers contacted IS&T management and the IS&T Security Team. The Security Team also consulted MIT’s Office of the General Counsel (OGC). At 9:45 a.m. IS&T management notified the MIT Police that a laptop connected to a network switch had been found in an electrical closet in Building 16. Two additional IS&T staff members arrived. Minutes later, uniformed MIT Police officers arrived in Building 16 and were posted in the basement hallway. A network engineer then used an MIT laptop to connect to the network switch, in order to monitor the traffic (packet stream) to and from the suspect laptop. Through this monitoring, the Security Team observed the downloading of data.

Another member of the MIT Police arrived, accompanied by a photographer. They took photographs of the closet, including the box, the laptop, and a hard drive sitting under the laptop. The MIT Police decided that the situation required expertise in computer crime and forensics, which they did not have. They therefore telephoned the Cambridge Police Department detective who is their normal contact for assistance with computer-related crime activity.

The Cambridge detective they contacted was a member of the New England Electronic Crimes Task Force. When he received the call for assistance from the MIT Police, the detective was working at the Task Force field office in a federal building in Boston, together with other law enforcement officers whose agencies participate in the Task Force. He responded to the call, accompanied by two other Task Force members: a special agent of the U.S. Secret Service; and a detective from the Boston Police Department. They arrived at the Building 16 closet around 11:00 a.m.

We note that no one from MIT called the Secret Service. The MIT Police contacted the Cambridge detective by calling him on his individual cell phone. The special agent became involved because he accompanied the Cambridge detective. As a Task Force member, the detective would sometimes respond to calls alone, and sometimes respond in

---

18 This monitoring of the switch was accomplished by one of the engineers plugging a cable (connected to his MIT laptop) into a port on the switch. At no time did he or anyone else from MIT in any way connect to the suspect laptop itself or to its cable. With very small exceptions, the only communications observed were those to and from the suspect laptop. (See Appendix 7.) The packet stream was preserved on the MIT laptop and later made available to the Secret Service special agent who became involved in the investigation.

19 The MIT Police typically make calls to the Cambridge Police for assistance in computer-related matters about six times a year. In none of these incidents were federal agents part of the response.

20 The New England Electronic Crimes Task Force (<http://www.secretservice.gov/ectf_newengland.shtml>) is a Boston-based alliance organized by the Secret Service with participants from federal, state, and local law enforcement, as well as private industry and academia, to investigate electronic crimes, including computer system intrusion.

21 All federal agents who are authorized to carry firearms in the course of their normal duties are referred to as “special agents.” Agents of the Secret Service, as well as most federal law enforcement agents, have the title of “special agent.”
the company of other members of the Task Force. The MIT Police were aware that other members of the Task Force might accompany the detective, and that Task Force members included Secret Service agents.

When they arrived, the suspect laptop was still downloading data. Also, during the monitoring, the MIT network engineers had observed that the laptop was being queried from several sources, including on one occasion an IP address located in China. This information was communicated to the law enforcement officials. The fact that someone or some entity in China could be involved initially raised concerns that this might have been part of an international matter.\(^\text{22}\)

The special agent attached a USB device to the suspect laptop in an attempt to copy the hard drive, but this attempt was unsuccessful. Crime scene investigators from the Cambridge Police Department arrived and took fingerprints from the laptop and hard drive.\(^\text{23}\) The agencies and personnel worked together in a cooperative fashion, with no law enforcement group taking orders from any other.

The law enforcement group decided to leave the laptop and hard drive in place to see if the person who had set it up would return. Because it was not feasible to continuously post MIT Police officers in the basement corridor for an indefinite period of time, and doing so would reveal the surveillance of the closet, the decision was made to install a video camera in the closet that could be monitored from elsewhere within MIT. IS&T installed the camera at the request of the MIT Police. At around 3:00 p.m., the basement closet was restored to the way it was found, with the exception of the camera having been installed. IS&T engineers relocated the MIT laptop they were using to another room and reconnected it to the MIT network to continue monitoring the network traffic to and from the suspect laptop, and everyone left the closet area.

Half an hour later, an individual was seen on the video camera entering the basement closet. He changed the hard drive attached to the laptop, and put the old one into a backpack. Some of the law enforcement officers went to the closet to try to apprehend him, but he had left before they could arrive. No one recognized the person in the video. Still photos showing the suspect were taken from the video and provided to the MIT Police.

During the morning’s activities in the basement closet, the special agent had asked for whatever electronic records MIT might have on the matter. As it is IS&T’s protocol to obtain approval from MIT’s Office of the General Counsel (OGC) before releasing

\(^{22}\) Ultimately, MIT concluded that the communication from the IP address located in China was a—not unusual—“pinging” attempt by someone or some entity in China to determine what computer systems at MIT were available and accessible, and unrelated to the activity of this laptop.

\(^{23}\) MIT Police regularly rely on Cambridge Police for latent fingerprint collection.
information or materials to outside law enforcement agencies, IS&T contacted the OGC, which responded that it was appropriate to comply with the agent’s request in view of the fact that law enforcement was conducting an investigation into what was potentially ongoing criminal activity of unknown scope, and it did not appear to OGC that such information would disclose personally identifiable information.\textsuperscript{24}

IS&T turned over the following information to the Secret Service, at its request, on the afternoon of January 4, 2011:

1. Network flow data, which is made up of logs showing which IP addresses communicated with which IP addresses; when the communication took place; and how much data was transferred.

2. DHCP (Dynamic Host Configuration Protocol) logs, which are records of requests from computer clients for the DHCP service to assign IP addresses. These records contain MAC addresses, IP addresses, and when clients acknowledge the receipt of addresses.

3. RADIUS (Remote Authentication Dial In User Service) logs, which record requests to use various network services.

In addition, the following was made available to the Secret Service, at its request, and was provided to the special agent on January 25, 2011:

4. The packet stream captured by the MIT network engineer using his laptop, as described above; this consisted of copies of the JSTOR downloads and associated control information—some 87 gigabytes in all.

These categories of items, Nos. 1 through 4, were provided by MIT to the Secret Service without a subpoena having been issued to MIT. Thereafter MIT provided additional documents to the Office of the U.S. Attorney in response to grand jury subpoenas.\textsuperscript{25} A more detailed description of these items is available in Appendix 7. (Appendix 10 addresses legal issues concerning production of the records.)

\textsuperscript{24} At this time, IS&T knew about the JSTOR downloading, but they also were concerned that the laptop might be performing other actions.

\textsuperscript{25} One such item—consisting of six records from the network registration database showing registration for Gary Host and Grace Host in September and October 2010—was provided by MIT in September 2011, under the impression that it was doing so in response to a grand jury subpoena. However, at the time those records were produced, this subpoena (issued on January 27, 2011) was apparently no longer valid.
I.C  Events of January 6, 2011: The Arrest

At about 12:30 p.m. on Tuesday, January 6, someone entered the closet, as was recorded by the video camera. As he entered, he covered his face with a bicycle helmet, removing it after he entered and the doors had closed. The individual removed the laptop and hard drive and then left the closet. When the laptop was disconnected, the port status of the switch changed, and a monitoring script sent an email to the phone of one of the IS&T engineers, who was not on campus at the time. The IS&T engineer notified the MIT Police and other network engineers, but no one was able to reach Building 16 in time to stop or intercept the person who had entered the closet. Later the same afternoon, the suspect computer’s MAC address reappeared in MIT’s network logs, showing that it was connected, first in MIT Building 4, then subsequently in the Stratton Student Center (Building W20), in the offices of MIT’s Student Information Processing Board (SIPB—MIT’s student computing group).

At approximately 2:00 p.m. an MIT Police officer was driving to the Stata garage after his shift in an unmarked police cruiser. He was familiar with the investigation and had been informed by radio that the laptop had been removed from the basement closet. He had seen the January 4 video of the suspect, as well as stills made from the video, and he had a still with him in his cruiser. On Vassar Street, near Massachusetts Avenue, he saw a cyclist pass him heading in the opposite direction. Based upon the stills and video, and given the backpack and clothes the cyclist was wearing, the officer observed that the cyclist matched the description of the suspect from the basement closet. He made a U-turn to follow the cyclist, who turned onto Massachusetts Avenue and proceeded north towards Harvard Square. When the officer reached the cyclist and pulled alongside, he rechecked the still photos that he had in his car and concluded that the cyclist was in fact the person in the photos. He immediately called his department for backup. A second

26 A few minutes before this person entered the closet, two MIT employees were seen on camera, standing in the opening of the closet doorway and then leaving. It is possible that the person utilizing the laptop saw these persons leaving the closet and, for that reason, decided to obscure his face while he was in the hallway, still walking to the closet, to avoid being identified by them. It is also possible that his observation of these employees is why he decided to move the laptop out of the closet and to another location. We note that this is speculation on the part of the Review Panel.

27 See the maps in the front matter.

28 When the laptop was disconnected from the network and removed from the closet, the IP address that it had been using was released by the network. The DHCP server network reassigned this address to another user, in another part of the building. Since the IP address was being monitored by IS&T, this initially led to some confusion as to where the suspect initially went, and where he initially reconnected his laptop.
MIT Police officer, accompanied by the special agent, responded by car from the MIT Police station.\textsuperscript{29}

When the cyclist reached the north side of Central Square, the officer who was following him decided to pull ahead of him and stop him to ascertain his identity.\textsuperscript{30} While exiting his car, the officer held his credentials so that they could be seen and motioned for the bicyclist to stop. The bicyclist complied. The officer explained that he was an MIT Police officer and wanted to speak with him. The cyclist first said that he didn’t speak with strangers. The officer again displayed his badge, as well as his photo ID. The cyclist then said that MIT Police were not “real cops” and refused to talk to the officer. At that point the cyclist dropped the bicycle to the ground and started running back toward Central Square, on Massachusetts Avenue. The officer chased him briefly, but the individual was outrunning him, and the officer returned to his car, made a U-turn, and followed, maintaining visual contact. The suspect slowed to a walk, and the officer, still in his car, watched and followed him.

The first MIT Police officer radioed the second and told him where the suspect was located. Once near the suspect, both MIT Police officers and the special agent left their vehicles and chased the suspect around parked cars. They apprehended and handcuffed him.\textsuperscript{31}

At this time, the officers still did not know the suspect’s identity. One of the officers called the Cambridge Police, who arrived and took the suspect to the Cambridge Police Department for booking. There, he was identified as Aaron Swartz. Aaron Swartz refused to talk to the police. He made a phone call to his friend Quinn Norton, who arranged for another friend to rush to the police station with bail money. Shortly thereafter an attorney from the firm of Good & Cormier arrived at the Cambridge Police Department, completed the paperwork for Aaron Swartz’s bail, and departed with him.

\textbf{I.D \hspace{1em} Events of January 6, 2011: Seizure of the Laptop}

Later on January 6, 2011, after Aaron Swartz was apprehended, members of IS&T, accompanied by MIT Police and the special agent, went to the SIPB offices in the Stratton Student Center to look for the laptop. Together, they found the laptop with an external hard drive, plugged into a network jack. The special agent examined the laptop and the participants decided that there was no feasible way to collect evidence from the

\textsuperscript{29} When the backup call was received, the special agent was reviewing the relevant video at the MIT Police station. He joined the MIT backup officer.

\textsuperscript{30} The MIT Police are deputized under the Middlesex County Police Department.

\textsuperscript{31} Aaron Swartz was arrested in connection with an alleged violation of state law, not federal law. The special agent participated in the arrest.
laptop while it was operating. It was disconnected and turned off. An MIT detective took the laptop and the external hard drive as evidence. On February 3, 2011, custody of the laptop and hard drive was transferred from the MIT Police to the Cambridge Police.

I.E  Access to the MIT Network

As it is relevant for the post-arrest narrative, we briefly describe MIT’s procedures for network access and review how Aaron Swartz obtained access to the MIT network and to JSTOR.

I.E.1 Connecting to the MIT network

MIT community members who want to use the wired network register by presenting an MIT user name and password that were issued to them when they began employment, or first registered as students, or were given a formal appointment at MIT, and they obtain permanent registrations for their computers. All other individuals—“guests”—who want to connect to MIT’s wired network supply a personal name and an email address as contact information, and they obtain a registration that is valid for up to 14 days a year of cumulative use, as explained on IS&T’s information page:32

IS&T offers short-term network service to campus guests. Guests are allowed up to fourteen days of network service when they register on the wired MIT network (MITnet) . . . .

For wired connections, plug the Ethernet cable into the computer and to an MITnet network drop . . . .

The machine needs to be configured for DHCP (obtaining an IP address automatically) . . . .

Once the equipment is ready to connect, open a web browser and point it to any web page. A page will appear, prompting to select your registration option. After selecting Visitor registration, the returned page will display the MITnet Rules of Use, followed by a screen requesting the visitor's contact information, number of days of connectivity, and the event for which they are on campus.

Visitors can register between one and five (consecutive) days at a time, up to fourteen days per year.

More precisely, the requirement to register is triggered when the computer asks MIT’s DHCP server to issue it an IP address. It is also possible to configure a computer to use a self-assigned “static IP address,” in which case there will be no registration request (provided that the static IP address is in an appropriate range, and does not conflict with an address that has already been assigned). MIT’s procedure is that static IP addresses should be requested through IS&T, so as to avoid conflicting address assignments, which would result in disruption of service.33

As noted above (footnote 8), Aaron Swartz registered five times in 2010: September 24, October 2, 8, and 22, and November 28. When the laptop was located in January 2011, it had a static IP address. At some point,34 Swartz had switched from using DHCP-provided IP addresses to using a static self-assigned IP address.35

I.E.2 JSTOR and eControl

MIT operates a very open network. Anyone can come onto campus and plug their computer into an MIT network port, or connect to the wireless network.36 Connecting to the wired network, and getting connected automatically, requires registering the computer the first time it is plugged in. Connecting to MIT’s wireless network does not require registration.

Prior to January 2011, any computer connected to the MIT network could access JSTOR. In the wake of the October 2010 downloading incident, and as a direct result of that incident, the Libraries and IS&T decided to deploy an authorization system for JSTOR called “eControl” that had been designed by the MIT Libraries to more narrowly restrict access by the MIT community to certain electronic databases. Under eControl, requests to access JSTOR would require a valid MIT certificate and be verified against MIT’s Human Resources directory, and only MIT faculty, students, or staff—not guests—would be granted access to JSTOR. Guests seeking access to JSTOR would now have to come to the MIT Libraries and use a library computer there.

MIT was prepared to implement eControl as early as October 2010. JSTOR and MIT were mindful that an abrupt change would diminish user convenience for the MIT community. JSTOR asked MIT to delay deployment of eControl to allow JSTOR to add an explanatory message to the JSTOR web page that would advise MIT users of the change, and redirect them through the eControl process. JSTOR informed MIT that this

33 See “Request an IP Address/Host Name,” <http://ist.mit.edu/network/ip-request>.
34 MIT is not sure when, as there is no MIT system of record that would have indicated this.
35 He used two different addresses, 18.55.6.240 and 18.55.7.240.
36 MIT’s framework for network security follows the general principle that, while access to individual resources on the network could be restricted, access to the underlying network should be as open as possible. This reflects a general computer system design methodology called the End-to-End Principle.
change would not be ready to implement until after December 18, effectively putting off the planned activation of eControl until after MIT’s winter holiday break.

On the morning of January 3, MIT and JSTOR agreed to expedite the implementation of eControl, and the system was activated on January 10, 2011. Since then, guests at MIT can access JSTOR only from certain workstations located in the libraries.
cutting-edge Internet technologies. By the age of 19 he had started and sold a successful web-publishing company.¹ He was also the cofounder of Demand Progress, an Internet blog and activist group self-described as focusing on civil liberties, civil rights, and government reform.²

### II.A.1 Aaron Swartz in Cambridge

Aaron Swartz was neither a member of the MIT staff, nor an enrolled student nor alumnus, nor a member of the faculty. He was a regular visitor to the MIT campus and interacted with MIT people and groups both on campus and off. His web-publishing startup was developed with the help of an entrepreneurship accelerator company “boot camp” that arranged for him to be housed on the MIT campus for the summer of 2005.³ After a short period in San Francisco, he returned to Cambridge in 2006 and lived in an apartment on Massachusetts Avenue in Central Square, between Harvard and MIT. He was a member of MIT’s Free Culture Group,⁴ a regular visitor at MIT’s Student Information Processing Board (SIPB), and an active participant in the annual MIT International Puzzle Mystery Hunt Competition.⁵ Aaron Swartz’s father, Robert Swartz, was (and is) a consultant at the MIT Media Lab. Aaron frequently visited his father there, and his two younger brothers had been Media Lab interns.

Aaron Swartz was a respected contributor to the World Wide Web Consortium’s Semantic Web, HTML, and TAG (Technical Architecture Group) activities. He attended gatherings of the Semantic Web working group that met at the MIT Computer Science and Artificial Intelligence Laboratory, and he was an invited speaker at one of the gatherings in 2008.

In 2010, Aaron Swartz became a research fellow at Harvard University’s Edmond J. Safra Center for Ethics,⁶ invited to conduct experimental and ethnographic studies of the political system and to prepare a monograph on the mechanisms of political corruption.

¹ The company was Infogami, which was used to support the Internet Archive’s Open Library Project. Infogami later merged with Reddit, which was subsequently acquired by Condé Nast.
³ The accelerator company was Y-Combinator (<http://ycombinator.com>). Aaron Swartz was housed in Simmons Hall. See Swartz’s web log of June 11, 2005, for his comments about arriving in Cambridge.
⁴ <http://freeculture.mit.edu/>, now inactive.
⁵ MIT Mystery Hunt, <http://www.mit.edu/~puzzle/>
⁶ <http://www.ethics.harvard.edu>
He had an office at the Center, and he was a regular contributor to discussions and activities there.

After Aaron Swartz’s arrest, Harvard suspended his fellowship and banned him from the Harvard campus, pending the outcome of an investigation into whether he had also used Harvard’s computers or network for similar activities. MIT took no action itself, but at Aaron Swartz’s arraignment on January 7, 2011, Cambridge District Court Judge Thomas ordered him to stay away from MIT property as part of the conditions of release. U.S. Magistrate Judge Dein imposed the same ban as a condition of Aaron Swartz’s release at his initial appearance and arraignment for the federal indictment on July 19, 2011. At the time of the federal arraignment, Aaron Swartz was residing at two locations: one in Brooklyn, New York, near his employment as an independent contractor in New York City; and the other in Cambridge, Massachusetts.

II.A.2 Possible motives for downloading

As far as the Review Panel knows, Aaron Swartz made no statement after his arrest regarding what he had planned to do with the downloaded documents. The Review Panel views the question of what he intended to do with the information that he was downloading from JSTOR as remaining open. Speculations about his motives reference a statement about free information to which he contributed, as well as two previous, large download experiences.

The federal indictment states that the downloading was “with the purpose of distributing a significant proportion of JSTOR’s archive through one or more file-sharing sites.” That is, the alleged motive is that Aaron Swartz intended to place the material on the Internet so that it could be freely distributed around the entire globe. In support of this

---

7 Although the investigation apparently did not find that he did this, his fellowship expired before this finding was made, and before the ban was lifted. Harvard’s OGC declined to comment for this review.
8 Recall from Part I that Aaron Swartz was arrested under Massachusetts Law.
9 The Federal District Court conditions of release are at <http://ia600504.us.archive.org/29/items/gov.uscourts.mad.137971/gov.uscourts.mad.137971.6.0.pdf>. As far as the Review Panel has been able to determine, the Cambridge District Court and Federal Court bans were imposed at the recommendations of the District Attorney and the federal prosecutor, respectively, and no one in the MIT administration asked for them or knew about them. The federal conditions of release also included the requirement that Aaron Swartz “undergo medical or psychiatric treatment as directed.”
10 Defendant’s Motion for Leave to Change Residential Address, Doc. 15 (filed September 8, 2011), United States v. Swartz, Case No. 1:11-cr-10260-NMG.
11 Also open is the question as to why Aaron Swartz used the MIT network for the downloading, as opposed to the Harvard network, to which he already had registered access. Lawrence Lessig, Director of the Safra Center and Professor of Law, Harvard Law School, suggests that Aaron Swartz did the downloading at MIT so as not to create trouble for Lessig and the Safra Center. (Lawrence Lessig on “Aaron’s Laws—Law and Justice in a Digital Age,” <http://www.youtube.com/watch?v=9HAw1i4gOU4&feature=player_embedded>.)
interpretation of his purpose, the government pointed\textsuperscript{12} to a “Guerilla Manifesto” he posted on the Internet in 2008. This “manifesto” included the following:

We need to take information, wherever it is stored, make our copies and share them with the world. We need to take stuff that’s out of copyright and add it to the archive. We need to buy secret databases and put them on the Web. We need to download scientific journals and upload them to file sharing networks. We need to fight for Guerilla Open Access.\textsuperscript{13}

Federal law enforcement apparently took the first sentence, “We need to take information, wherever it is stored, make our copies and share them with the world,” as the motive and purpose behind his extensive downloading—some 4.8 million articles, or 80\% of JSTOR’s database of journals. (“It is alleged that Swartz avoided MIT’s and JSTOR’s security efforts in order to distribute a significant proportion of JSTOR’s archive through one or more file-sharing sites.”)\textsuperscript{14}

Collecting the JSTOR articles through the MIT network was not the first time Aaron Swartz had engaged in large-scale downloading, although it was the first time he was charged with a crime. In 2008, he downloaded about 20 million pages of documents from the government-run PACER (Public Access to Court Electronic Records) system. Unlike the JSTOR documents, these were all in the public domain.\textsuperscript{15} “He donated the 19,856,160 pages to \url{http://public.resource.org}, an open government initiative spearheaded by Carl Malamud as part of a broader project to make public as many government databases as Malamud can find.”\textsuperscript{16} The FBI opened an investigation, but apparently concluded that no laws were violated, and thus no charges were filed.\textsuperscript{17}

\textsuperscript{12} Government’s consolidated Response to Defendant’s Motions to Suppress at 3, Doc. No. 81 (filed November 11, 2012), \textit{United States v. Swartz}, Case No. 1:11-cr-10260-NMG.

\textsuperscript{13} <\url{http://archive.org/stream/GuerillaOpenAccessManifesto/Goamjuly2008_djvu.txt}>. Aaron Swartz was not the sole author of the memo, and it is unknown whether he authored the sentences that were quoted. Quinn Norton told the Review Panel that she did the final editing of the piece, and that she does not know who the other authors were, or who contributed which part.

\textsuperscript{14} \textit{Supra} at 4 n.15; \textit{see also} USAO Press Release July 19, 2011, <\url{http://www.justice.gov/usao/ma/news/2011/July/SwartzAaronPR.html}>


\textsuperscript{16} <\url{http://www.wired.com/threatlevel/2009/10/swartz-fbi/>}

\textsuperscript{17} Aaron Swartz, together with public-domain advocate Carl Malamud, identified numerous instances of personal identifying information that was supposed to be redacted or hidden in these public documents that had been left available for viewing: “names of minor children, names of informants, medical records, mental health records, financial records, tens of thousands of social security numbers.” They then sent their results to 31 district courts. The federal Judicial Conference eventually changed its privacy rules. See the comments by Public.Resource.Org Director Carl Malamud, in his January 23, 2013, memorial “Aaron’s Army.”
Aaron Swartz also participated in a study of downloaded articles concerning the payment by interested organizations to experts, including law professors, to publish papers in academic journals. He wrote a script that downloaded articles from Westlaw, and a second script that extracted the relevant information about the funding sources from the footnotes of each article.\(^\text{18}\) This has been cited as support for a different possible motive for his actions: an intention to cross-reference the entire JSTOR database by author, publisher, and funding source, so as to demonstrate the extent to which JSTOR’s service, and thus the fees it charged, was enabled and funded by public money.\(^\text{19}\) In support of this interpretation is Aaron Swartz’s self-description on the first page of his blog: “He [Aaron Swartz] is a frequent television commentator and the author of numerous articles on a variety of topics, especially the corrupting influence of big money on institutions including nonprofits, the media, politics, and public opinion. From 2010–2011, he researched these topics as a Fellow at the Harvard Ethics Center Lab on Institutional Corruption.”\(^\text{20}\)

One can also speculate that Aaron Swartz had not decided what he would eventually do with the articles at the time of the downloading.\(^\text{21}\)

\(^{18}\) S. Barday, *Punitive Damages, Remunerated Research, and the Legal Profession*, 61 Stanford L.R. 711 (2008). Aaron Swartz’s name does not appear in this publication, but Professor Lawrence Lessig, now the Roy L. Furman Professor of Law and Leadership at Harvard Law School, and who at the relevant time was a Professor of Law at Stanford University, told the Review Panel that Ms. Barday did this study for a seminar he was teaching at Stanford Law School, and that he suggested that she and Swartz work together. Aaron Swartz and Ms. Barday collaborated on research for the piece, downloading the articles using Ms. Barday’s user ID. “The database was compiled using Python source code extracting all entries contained in the Westlaw “Journals and Law Reviews” database, including full-text articles. The first three footnotes and the Westlaw “cite as” field were then extracted from the articles. Articles receiving outside funding were identified using . . . search terms as they appear in one of the first three footnotes in each article.” We note that one of the articles cited in this study was authored by Review Panel member Peter A. Diamond.


\(^{20}\) On the other hand, the Review Panel notes that JSTOR provides a service, JSTOR Data for Research <http://dfr.jstor.org>, that researchers can access (including for downloading) to obtain some of the information for such a study.

\(^{21}\) Carl Malamud, in his memorial to Aaron Swartz, writes, “I’m convinced that Aaron had not made a decision to release those articles, and I am certain he would not have released them without a great deal of post-download analysis.” (“Aaron’s Army: On Crime and Access to Knowledge,” <https://public.resource.org/crime/pamphlet.pdf>). Or, as his friend Quinn Norton told the Review Panel, “He liked to collect data sets.” Norton also told the Review Panel that Swartz was shocked by the arrest: he didn’t regard what he’d done as a big deal and was surprised that people were making so much of it. His (third) attorney Elliot Peters also told the Review Panel that Swartz had been shocked by the arrest.
Report to the President
MIT and the Prosecution of Aaron Swartz

Review Panel

Harold Abelson
Peter A. Diamond
Andrew Grosso
Douglas W. Pfeiffer (support)

July 26, 2013
UNITED STATES DISTRICT COURT
DISTRICT OF MASSACHUSETTS

Crim. No. 11-CR-10260-NMG

VIOLATIONS:

18 U.S.C. § 1343 (Wire Fraud)

18 U.S.C. § 1030(a)(4),(b) (Computer Fraud)

18 U.S.C. § 1030(a)(2), (b), (c)(2)(B)(iii)
(Unlawfully Obtaining Information from a Protected Computer)

(Recklessly Damaging a Protected Computer)

18 U.S.C. § 2 (Aiding and Abetting)

1030(i) (Criminal Forfeiture)

UNITED STATES OF AMERICA

v.

AARON SWARTZ,

Defendant

SUPERSEDING INDICTMENT

The Grand Jury charges that at all relevant times:

PARTIES

JSTOR

1. JSTOR, founded in 1995, was and continued to be a United States-based, not-for-profit organization that provides an online system for archiving and providing access to academic journals and journal articles. It provides searchable digitized copies of articles from over 1,000 academic journals, dating back for lengthy periods of time.

2. JSTOR’s service is important to research institutions and universities because it can be extraordinarily expensive, in terms of both cost and space, for a research or university library to maintain a comprehensive collection of academic journals. By digitizing extensive, historical collections of journals, JSTOR enables libraries to outsource the journals’ storage, ensures their preservation, and enables authorized users to conduct full-text, cross-disciplinary
searches of them. JSTOR has invested millions of dollars in obtaining and digitizing the journal articles that it makes available as part of its service.

3. JSTOR generally charges libraries, universities, and publishers a subscription fee for access to JSTOR's digitized journals. For a large research university, this annual subscription fee for JSTOR's various collections of content can cost more than $50,000. Portions of the subscription fees are shared with the journal publishers who hold the original copyrights. In addition, JSTOR makes some articles available for individual purchase.

4. JSTOR authorizes users to download a limited number of journal articles at a time. Before being given access to JSTOR's digital archive, each user must agree and acknowledge that they cannot download or export content from JSTOR's computer servers with automated computer programs such as web robots, spiders, and scrapers. JSTOR also uses computerized measures to prevent users from downloading an unauthorized number of articles using automated techniques.

MIT

5. The Massachusetts Institute of Technology ("MIT") was and continued to be a leading research and teaching university located in Cambridge, Massachusetts.

6. JSTOR provided MIT with its services and content for a fee.

7. MIT made JSTOR's services and content available to its students, faculty, and employees. MIT also allowed guests of the Institute to have the same access to JSTOR, but required guests to register on the MIT network. MIT authorized guests to use its network for no more than fourteen days per year, and required all users to use the network to support MIT's research, education, and administrative activities, or at least to not interfere with these activities; to maintain the system's security and conform to applicable laws, including copyright laws; and to conform with rules imposed by any networks to which users connected through MIT's system. These rules explicitly notified users that violations could lead to state or federal prosecution. Guest users of the MIT network agreed to be bound by the same rules that applied to students,
faculty, and employees.

8. JSTOR’s computers were located outside the Commonwealth of Massachusetts, and thus any communications between JSTOR’s computers and MIT’s computers crossed state boundaries. JSTOR’s and MIT’s computers were also used in and affected interstate and foreign commerce.

Aaron Swartz

9. Aaron Swartz lived in the District of Massachusetts and was a fellow at Harvard University’s Safra Center for Ethics. Swartz was not affiliated with MIT as a student, faculty member, or employee or in any other manner. Although Harvard provided Swartz access to JSTOR’s services and archive as needed for his research, Swartz used MIT’s computer networks to steal millions of articles from JSTOR.

OVERVIEW OF THE OFFENSES

10. Between September 24, 2010, and January 6, 2011, Swartz contrived to:

a. break into a restricted-access computer wiring closet at MIT;

b. access MIT’s network without authorization from a switch within that closet;

c. access JSTOR’s archive of digitized journal articles through MIT’s computer network;

d. use this access to download a substantial portion of JSTOR’s total archive onto his computers and computer hard drives;

e. avoid MIT’s and JSTOR’s efforts to prevent this massive copying, efforts that were directed at users generally and at Swartz’s illicit conduct specifically; and

f. elude detection and identification.
MEANS OF COMMITTING THE OFFENSES

11. Swartz alone, or in knowing concert with others unknown to the Grand Jury, (hereafter simply “Swartz” in this section) committed these offenses through the means described below.

September 24 through 27, 2010


13. Later that day, Swartz connected the Acer laptop to MIT’s computer network from a location in Building 16 at MIT and registered with MIT’s computer network as a guest.

14. When Swartz registered on the network, he took measures to hide his identity as the computer’s owner and user:
   a. Swartz registered the computer under the fictitious guest name “Gary Host.”
   b. Swartz specified the computer’s client name as “ghost laptop.” (A computer’s client name helps to identify it on a network and can be chosen by its user.) In this case, the “ghost” client name abridged the pseudonym “Gary Host” by combining the first initial “g” with the last name “host.”
   c. Swartz identified the fictitious “Gary Host’s” e-mail address as “ghost@mailinator.com”, a temporary e-mail address. Mailinator advertised itself as a free e-mail service that allows a user to create a new temporary-mail address as needed. Mailinator advertised that it would accept mail for any e-mail address directed to the mailinator.com domain without need for a prior registration or account. Mailinator also advertised that all mail sent to mailinator.com would automatically be deleted after several hours, whether read or not, and that the company kept no logs of e-mail access.

15. On September 25, 2010, Swartz used the Acer laptop to systematically access and
rapidly download an extraordinary volume of articles from JSTOR by submitting download
requests faster than a human could type, and in a manner designed to sidestep or confuse
JSTOR’s computerized efforts to restrict the volume of individual users’ downloads.

16. The effect of these rapid and massive downloads and download requests was to
impair computers used by JSTOR to provide articles to client research institutions.

17. As JSTOR, and then MIT, became aware of these events, each took steps to block
communications to and from Swartz's computer. Swartz, in turn, altered the apparent source of
his automated demands to sidestep or circumvent JSTOR’s and MIT’s blocks against his
computer, as described below:

   a. On the evening of September 25, 2010, JSTOR terminated Swartz’s
      computer’s network access by refusing communications from the computer’s
      assigned IP address.
         i. An IP (short for “Internet Protocol”) address is a unique numeric
            address assigned to each computer connected to the Internet so that the
            computer’s incoming and outgoing Internet traffic is directed to the proper
            destination. Most Internet service providers control a range of IP
            addresses. MIT controls all IP addresses that begin with the number 18.
         ii. Swartz’s computer had been assigned an IP address of 18.55.6.215.
         iii. On September 25, 2010, JSTOR blocked communications from
             that IP address, thus preventing Swartz from requesting and receiving any
             more JSTOR articles.
   
   b. On September 26, 2010, Swartz established a new IP address for his
      computer on the MIT network – 18.55.6.216 – which sidestepped the IP address
      block and allowed the laptop to resume downloading an extraordinary volume of
      articles from JSTOR. Accesses from this address continued until the middle of
      the day, when JSTOR spotted the access and blocked communications from this
new IP address as well.
c. Because the downloads on September 25 and 26 originated from shifting MIT IP addresses beginning with 18.55.6, and because JSTOR’s computers used to provide articles to research institutions had been impaired and significant portions of its archive was at risk of misappropriation, on September 26, 2010, JSTOR began blocking a broader range of IP addresses. The block prevented a researcher assigned any one of over 250 other IP addresses available at MIT from being able to access JSTOR’s archive until September 29, 2010.
d. After JSTOR notified MIT what was happening, MIT sought to block Swartz in particular. It did so by prohibiting Swartz’s laptop from being assigned any IP address on MIT’s network. MIT did so by blocking communications with any computer bearing the laptop’s MAC address.
i. A MAC address is a unique identifier assigned to each computer’s network interface, in this case, Swartz’s Acer laptop’s network interface card.

ii. When a user plugs his computer into MIT’s wired network on campus, the network reads the computer’s MAC address to determine whether the computer is authorized to use the network. As part of the registration process, “Gary Host’s” computer, i.e., Swartz’s Acer laptop, had identified its network interface’s MAC address as 00:23:5a:73:5f:fb.

iii. Consequently, on September 27, 2010, MIT terminated the laptop’s guest registration and barred any network interface with that MAC address from obtaining a new IP address.

October 2 through 9, 2010

18. On October 2, 2010, just over a week after JSTOR and MIT had blocked Swartz’s Acer laptop from communicating with JSTOR’s and MIT’s networks, Swartz sought and
obtained another guest connection on MIT’s network for his Acer laptop.

19. Once again, Swartz registered the Acer laptop on the network using identifiers chosen to avoid identifying Swartz as the computer’s owner and user:

a. Swartz once again registered the computer under the fictitious name "Gary Host" and the client name "ghost laptop."

b. To evade the MAC address block, Swartz “spoofed” the Acer laptop’s computer’s MAC address. A MAC address is usually assigned to a network interface card by the card’s manufacturer, and therefore generally remains constant. But a user with the right knowledge can change the MAC address, an action referred to as “MAC address spoofing.” Swartz spoofed the Acer laptop’s MAC address by changing it from 00:23:5a:73:5f:fb to 00:23:5a:73:5f:fc (that is, the final ‘b’ became a ‘c’).

c. By re-registering the laptop, the laptop received a new IP address, which disassociated Swartz’s Acer laptop from the IP addresses that JSTOR had blocked when Swartz had used them in September.

20. On October 8, 2010, Swartz connected a second computer to MIT’s network and registered as a guest, using similar naming conventions: Swartz registered the computer under the name "Grace Host," the computer client name "ghost macbook," and the temporary e-mail address “ghost42@mailinator.com.”

21. On October 9, 2010, Swartz used both the “ghost laptop” and the “ghost macbook” to, again, systematically and rapidly access and download articles from JSTOR. The pace of Swartz’s automated downloads was so fast and voluminous that it significantly impaired the operation of some computers at JSTOR.

22. In response, beginning on or about October 9, 2010, JSTOR blocked MIT’s entire computer network from accessing JSTOR. The block lasted several days, again depriving legitimate users at MIT from accessing JSTOR’s services.
November and December, 2010

23. During November and December, 2010, Swartz again used the “ghost laptop” (i.e., the Acer laptop) at MIT to download over two million documents from JSTOR, more than one hundred times the number of downloads during the same period by all legitimate MIT JSTOR users combined.

24. During this period, when Swartz connected to MIT’s computer network, he circumvented MIT’s guest registration process altogether. Rather than let MIT assign his computer an IP address automatically, Swartz instead simply hard-wired into the network and assigned himself two IP addresses. He did so by entering a restricted network interface closet in the basement of MIT’s Building 16, plugging the computer directly into the network, and operating the computer to assign itself two IP addresses. To further cloak his activities, Swartz also hid the Acer laptop and a succession of external storage drives under a box in the closet, so that they would not arouse the suspicions of anyone who might enter the closet.

January 4 through 6, 2011

25. On January 4, 2011, Swartz entered the restricted basement network wiring closet and replaced an external hard drive attached to the laptop.

26. On January 6, 2011, Swartz returned to the wiring closet to remove his computer equipment. This time he attempted to evade identification at the entrance to the restricted area. Apparently aware of or suspicious of a video camera, as Swartz entered the wiring closet, he held his bicycle helmet like a mask to shield his face, looking through ventilation holes in the helmet. Swartz then removed his computer equipment from the closet, put it in his backpack, and left, again masking his face with the bicycle helmet before peering through a crack in the double doors and cautiously stepping out.

27. Later that day, Swartz connected his Acer laptop to MIT’s network in a different building — the student center — again registering on the network using identifiers chosen to avoid identifying Swartz as the computer’s owner and user:
a. Swartz registered the computer under the fictitious name "Grace Host" and the client name "ghost laptop."

b. By re-registering the laptop, the laptop again received a new IP address, which disassociated Swartz’s Acer laptop from the IP addresses Swartz had used up to that point.

c. To again evade the MAC address block, Swartz had spoofed the Acer laptop’s MAC address a second time, changing it from the blocked 00:23:5a:73:5f:fb (or from the later-spoofed 00:23:5a:73:5f:fc) to 00:4c:e5:a0:c7:56.

28. Swartz’s Acer laptop contained a software program named “keepgrabbing.py,” which was designed to download .pdf files (the format used by JSTOR) from JSTOR and sidestep or confuse JSTOR’s computerized efforts to prevent repeated and voluminous downloads.

29. When MIT Police spotted Swartz on the afternoon of January 6, 2011 and attempted to question him, Swartz fled with a USB drive that contained the program “keepgrabbing2.py,” which was similar to “keepgrabbing.py.”

30. In all, Swartz stole a major portion of the total archive in which JSTOR had invested.

31. Swartz intended to distribute these articles through one or more file-sharing sites.