ISSN 1940-5758

Issue 12,

Electronic Resources Security: A look at Unauthorized Users

Much of the literature written on electronic resources security focuses on systematic downloading. However, when the unauthorized use from two cases of stolen identities at the University of Saskatchewan was studied in more depth, a different pattern emerged. By analyzing proxy server data, we found that the unauthorized use was coming from all over the world, was focused on science, technology and medical resources, and included both small-scale and excessive downloading. This article outlines some steps that libraries can take to detect and prevent small-scale unauthorized use and implications as libraries move towards Shibboleth authentication.

Introduction

Quietly and regularly, international electronic journal thieves are downloading your electronic resources with stolen logins and passwords. Sound far-fetched? Not really. In late 2008, the University of Saskatchewan was refused access to a major publisher because a user tried to download more than 150 articles in a session. When the incident was investigated, the library did not find an overzealous patron in need of a stern warning. Instead, the library found a stolen login ID and password that had been used undetected 15 other times that day from various international locations. While incidents like this one are not uncommon, the topic of electronic journal theft is not often considered by librarians and administrators. Additionally, the topic is not often delved into more deeply to determine the breadth of electronic journal misuse. So, how are electronic resources being misused? How do you stop it?

Types of E-Resource Misuse

The license agreements for electronic resources typically limit access to authorized users, as defined in the agreement, and also often explicitly prohibit downloading in a systematic way. When either of these conditions is not met, the library is in violation of its legal contracts and can face serious

consequences including partial or full revocation of access. Baker and Tenopir identify three examples of misuse[1]:

1. excessive downloading with a stolen login and password;
2. excessive downloading from an authorized user; and
3. downloading by an unauthorized person from an open proxy server that did not require authorization.

This case at the University of Saskatchewan appeared, at first, to be similar to Baker and Tenopir's second example. The first sign of a problem was the e-resource vendor revoking access to the library's proxy server (EZProxy), which allows remote access to resources. Instead of access to journal articles, library users saw a webpage stating that access had been denied because of excessive downloading. Examination of the EZProxy files identified the library user that had triggered this situation and the user was contacted immediately. However, when contacted, the user stated that they had no part in this incident. At this point, we dug a little further into the log and audit files and found that the user's ID was being used from 16 different locations around the world that day. The client's ID and password were stolen. At this point, the example looked like Baker and Tenopir's first example, excessive downloading with a stolen login and password

But while taking a closer look at the log and audit files, another pattern of misuse emerged. The other 15 times that day the stolen password was used was for limited downloading. These other incidents went undetected because the usage was not excessive, and the login appeared to be authentic. This type of misuse may be more widespread than libraries realize, because it appears to be legitimate use. In fact, after a detection script was implemented to prevent this type of misuse, a second stolen login was immediately identified. The second stolen ID had been used undetected for some time because it was not being used for excessive downloading. This type of misuse fits into a fourth pattern:

1. limited downloading with a stolen login and password.

How the logins were stolen in the first place is not known. The library user may have simply shared their password with a friend overseas. Once contacted, the owners of the library accounts were not able to,

or perhaps not willing to, shed any light on how it might have happened. There are several methods by which passwords can be stolen, but with IT security in place at most organizations, one of the simplest ways is to take advantage of human gullibility through phishing, "the impersonation of reputable companies in order to induce individuals to reveal personal information, such as passwords and credit card numbers, online".[2] For example, simply by posing as an employee of the University IT department in an email, the original attacker could have convinced the user to share personal information. The IDs are then sold or shared on the internet through online forums or web pages. It is unknown if phishing was employed in these cases of stolen identity, but it does illustrate the point that no system is bulletproof, at least not yet. Stolen identities are not going away anytime soon.

In both cases, the library clients immediately changed their passwords when notified, which stopped the unauthorized access and the vendor also quickly reinstated our access. It was a fairly standard resolution. However, curiosity remained about the identity of the unauthorized users and what they were doing. The EZProxy log and audit files from these two stolen identities provide an interesting case study.

Who are our Unauthorized Users?

Once it was determined that there were many unauthorized users, questions arose about who they were and what they were doing. One could romantically envision these people as the Robin Hoods of the e-journal world, stealing from the information rich to give to the information poor. Or as thieves driven only by profit, repackaging and reselling resources. Perhaps they are simply genuine scholars and medical professionals looking for the access that they need. Are they Open Access vigilantes, taking access into their own hands? While we cannot tell their motives from the log files, this case study does shed some light on the question of identity and download behavior in the areas of location, subject matter and quantity.

First, I created a database and imported the EZProxy audit files and log files into separate tables. I was able to join the audit data, which contains user IDs, and the log data, which contains download information, by joining the two tables on the session field. I then limited the data to only the

compromised IDs. From this data I could identify IP addresses, download URLs and download sizes. Another table was created with location information for each IP gathered from http://ip-lookup.net/ . For privacy and storage reasons, we delete files after a set period of time. This meant that I could only obtain four days worth of data, in other words, only a snapshot in time.

Location

During those four days, the two stolen IDs were accessed from a total of 17 different countries from all over the world (Figure 1). Some of the IP addresses could not be identified, but most could be traced to a suspected country of origin. It is possible that some of the IPs were intermediate hosts or proxies, and the user actually resides elsewhere, but that is not possible to detect from the data available. Interesting to note is the wide variety of countries from many continents.

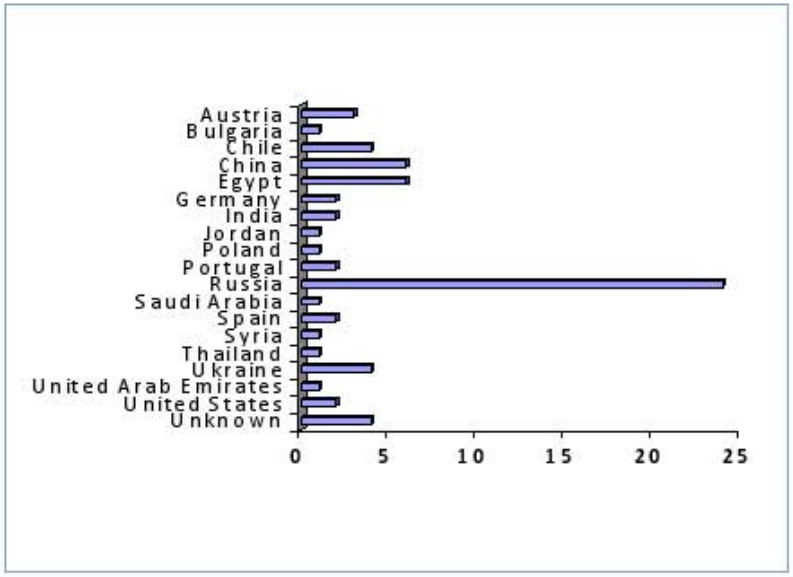

Figure 1.Number of IP Addresses by Country.

Host information was also looked up and investigated further with Google searching. Hosts fell into seven categories (Figure 2). The biggest category was legitimate commercial and/or residential

Internet Service Providers (ISPs). The second biggest category was unknown. If the host of an IP is unknown, this may indicate that it is not reputable, but not necessarily. Perhaps the most interesting finding, is that some use was coming from known university or healthcare organization IP ranges.

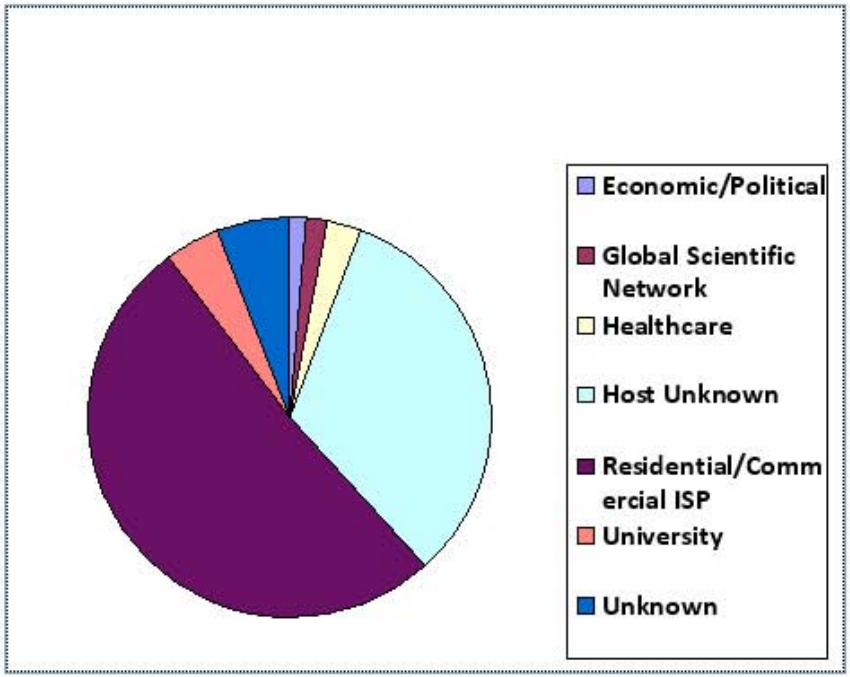

Figure 2.Organization or ISP Types of IP Addresses.

Subject Matter

Although one could investigate the subject matter of each article downloaded, for the sake of speed and simplicity, I instead looked at the domain names and determined the vendors or publishers. Not surprisingly, the majority of the visited domains were vendors that specialize in science, technology, medicine (STM) or have strong STM collections within their large multidisciplinary holdings. The complete list of vendors/publishers accessed is in Appendix A.

Quantity

There is no question that some excessive use was occurring that was likely systematic downloading. After all, that was the behavior that triggered this investigation. However, if you look at all the sessions over that four day period, there are only a few spikes in download quantity, with many sessions showing moderate quantities. Please note that the bytes represented in Figure 3 are total bytes downloaded in the session including not just articles, but also information like search results and webpage components.

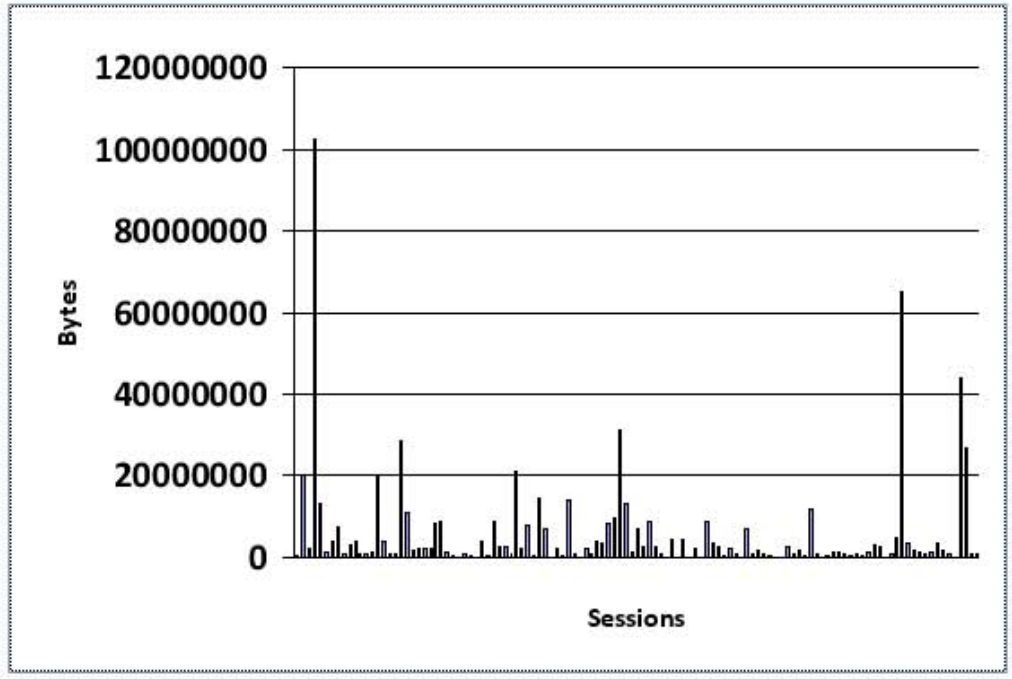

Figure 3.Download Quantity for Each Session.

Limitations

This data is only from two IDs over a short period of time and cannot be extrapolated to represent the nature of all unauthorized use. This case study is not meant to prove that there exists any particular percentages, rather it attempts to shed light on the fact that these types of unauthorized use do exist. For example, not all unauthorized use is coming from China. Not all unauthorized use is excessive downloading or spidering activity. Unauthorized use from hospitals and universities does exist.

Prevention

Regardless of the type of unauthorized use, it is still important to try reasonable prevention measures to uphold our licenses, and also to prevent downtime when vendors shut off access. The systems in place for security must be reasonable for libraries to implement. For example, it is not reasonable to think that libraries are going to eliminate stolen passwords completely. In addition to general good practices for IT security, there are a couple of proxy-specific prevention methods that are also reasonable to implement.

One relatively simple method to detect stolen passwords, even if the usage is not abnormal, is to have a script run nightly on the EZProxy audit files and scan for logins to a single ID that are coming from multiple locations. When multiple locations are detected, the script outputs a warning message that is put into an email message to alert staff so they can ensure the library user changes their password. If local IP ranges are excluded, and a minimum threshold set, any usernames caught will likely be stolen. So far in our experience, this script has not falsely fingered a library user on a legitimate whirlwind international tour. Also, in our experience, simply having the library user change their password is enough to stop the misuse.

```
#!/usr/bin/perl

use strict;
use Socket;

my $today = `date +"%Y%m%d"`;
$today =~ s/\s*$//;
$today = $ARGV[0] if ($ARGV[0]);
my $file = "/usr/local/ezproxy/audit/" . $today . ".txt";
my $min_threshold = 3;
my %Logins = ();

&DoFile($file);

# Multiple (min_threshold) successful login locations outside local city for
a user
# should be an indication of a stolen password
#
foreach my $user (keys %Logins) {
  my @IPs = split /::/, $Logins{$user};
  if ($#IPs >= $min_threshold) {
```

```
    print "\nAlert: Multiple login locations for $user:\n";
    foreach my $ip (@IPs) {
      print "\t",$ip,"\n";
    }
  }
}

exit;

sub DoFile {
  my ($filename) = @_;

  open(IN,"<$filename") || die;
  while() {
    chop;
    my ($xime,$Event,$IP,$Username,$Session,$Other) = split /\t/;
    if ($Event eq "Login.Success") {
      $Logins{$Username} = ListBuild($Logins{$Username},$IP);
    }
  }
  close IN;
}

sub ListBuild
{
   my ($list,$add) = @_;

   # Our own IPs
   return $list if ($add =~ m/^10\./);
   return $list if ($add =~ m/^128\.233\./);

   # Other local ISPs
   return $list if ($add =~ m/^71\.17\./);
   return $list if ($add =~ m/^216\.197\./);

   if ($list eq "") { return $add; }

   my @items = split /::/, $list;
   foreach my $item (@items) {
       if ($item eq $add) { return $list; }
   }
   return $list . "::" . $add;
}
```

Another relatively simple method of finding stolen passwords is to occasionally Google your proxy server domain name. The logins and passwords are sometimes shared or sold on websites and discussion boards in a fairly open way. This method would only catch a small number, but it is worth a look.

Future of Access and Unauthorized Use

The move from IP and proxy based authentication towards Shibboleth authentication has begun and is growing in popularity. There is a hope that Shibboleth may reduce the workload of libraries in maintaining proxy configurations and submitting IP ranges. At the same time, it protects patron privacy by continuing to authenticate users at their home institution and is based on trust in the local authentication.[3] A stolen password will appear to be legitimate and will be authenticated against the local organization in the same way it is now. Shibboleth will not protect against stolen passwords any more than our current system does. The difference is that when a vendor detects excessive use, they will only need to deny access to that single user, rather than blocking an entire proxy server and effectively blocking all remote access. Limited unauthorized use may go undetected unless the vendors themselves look for multiple locations or other markers of suspicious activity. With Shibboleth, libraries will often rely on their parent organizations to be the local authentication (identity provider). If libraries want to learn more about who is accessing resources, they will need to communicate with the identity provider and/or their vendors (service providers), rather than simply look at their own proxy server activity.

In the print-based world, security and access were not so diametrically opposed. Stopping theft from the library was necessary to ensure continued access for the rest of the patrons. Limiting use of rare or fragile books was necessary to preserve them for use by future generations. Of course, access was also restricted due to the physical nature of the resources and costs/budgets. Those things were not under the libraries control and so we could focus on providing access. In the electronic world, things changed. People far away seem a lot closer and walk-in use, although important, now seems paltry. Resources were freed from their more cumbersome paper bodies, but not from copyright. E-Resource license agreements and authentication systems have become as much about keeping people out as they are about letting people in and libraries now have an active role in keeping people out. It is a policing role that many libraries will be happy to give up to their parent organization and the Access Federations under Shibboleth.

But if libraries do less of the policing and gate keeping, will the awareness of the issues

surrounding unauthorized access also be reduced? The notification email that is generated with the data from the detection script in this article has become not just a practical tool for managing access, it is also a regular and uncomfortable reminder of the issues of unauthorized use: the difficulty in restricting information in a digital age, pricing barriers and access inequalities. When I no longer receive them, will unauthorized use be out of sight, out of mind?

Conclusion

Many resources, such as time, hardware, and software are spent on protecting our electronic resources and, indeed, this is the agreement that we have made with many of our content providers and we must uphold our end of the bargain. But when we see some of the people that we are keeping out, they start to look an awful lot like the people we are usually trying to serve. It is not a coincidence that some of the resources being accessed by the unauthorized users are among the most prohibitively expensive in the world. We can't let the unauthorized users in now, but we can promote and support Open Access policies, negotiate for more consistent interlibrary loan rights in our licenses and continue to be vocal about excessive pricing. If libraries do realize efficiencies in moving to Shibboleth, perhaps this time and energy needs to be spent on opening up access. Out of sight, but not out of mind.

References

1. Baker G and Tenopir C. 2006. Managing the unmanageable: Systematic downloading of electronic resources by library users. Journal of Library Administration 44(3/4):11-24. (COinS)

2. Phishing [Internet]. 2008. In: Oxford English Dictionary Online. Oxford University Press; [cited 2010 10/13]. Available from: http://www.oed.com

3. Shibboleth [Internet]. [copyright 2010]. Internet2; [cited 2010 10/13]. Available from: http://shibboleth.internet2.edu/

About the Author

Heather Tones White is an Information Technology Librarian at the University of Saskatchewan in Saskatoon, Saskatchewan. She has been with the U of S since 2005 after completing her M.L.I.S. at the University of Western Ontario. Heather can be reached at heather.white@usask.ca

Acknowledgements

Special thanks to Doug Macdonald for the script and also for his patience in explaining the ins and outs of library authentication.

Appendix A: Vendors/Publishers Accessed by Unauthorized Users

Vendor or Publisher	Subject Category
ACLS Humanities E-Book	Humanities
American Association for Cancer Research	Medical
American Association for Clinical Chemistry	Science
American Chemical Society	Science
American Physiology Society	Medical
American Society for Biochemistry and Molecular Biology	Science
American Society for Microbiology	Science
American Society for Pharmacology and Experimental Therapeutics	Medical
American Society of Hematology	Medical
Association for Computing Machinery	Technology/Engineering
Blackwell Synergy	Multidisciplinary
Book 24×7	Multidisciplinary
British Medical Journal	Medical
Canadian Foreign Relations Index	Social Science
Ebsco	Multidisciplinary
Elsevier	Multidisciplinary

Elsevier MD Consult	Medical
Elsevier Science Direct	Multidisciplinary
Elsevier Scopus	Multidisciplinary
Gale	Multidisciplinary
HeinOnline	Law
IEEE	Technology/Engineering
Informaworld	Multidisciplinary
Informs	Social Science
Journal of the American Society of Nephrology	Medical
Karger	Medical
Lexi-Comp	Medical
MetaPress	Multidisciplinary
MIT Press Journals	Multidisciplinary
Myilibrary	Multidisciplinary
National Academy of Sciences	Science
National Center for Biotechnology Information	Science/Medical
Nature	Science
New Science Press	Science
Ovid	Multidisciplinary
Ovid SilverPlatter	Multidisciplinary
Oxford Journals	Multidisciplinary
Proquest	Multidisciplinary
Proquest UMI	Multidisciplinary
Pub Med Central	Medical
Refworks	Other
Royal College of Psychiatrists	Medical
Sage Publications	Multidisciplinary
Shattauer Publishers	Medical
Society for Endocrinology	Medical

Society for General Microbiology	Science
Springer Link	Multidisciplinary
Springer Protocols	Science
STAT!Ref	Medical
Statesman Yearbook	Reference
SwetsWise	Multidisciplinary
Thieme Publishing Group	Medical
University of Michigan Digital Collections	Multidisciplinary
Wiley	Multidisciplinary
World Book	Reference

Subscribe to comments: For this article | For all articles

Leave a Reply

Name (required)

Mail (will not be published) (required)

Website

Submit Comment

Log in