Education Technology: Examining Student Privacy and Security

Eric Coopey

Abstract

Education is undergoing a revolution, with technology increasingly becoming an integral part of the learning process. Computers, tablets, and cell phones are quickly infiltrating the classroom, for better or worse. Digitally native students and technology savvy teachers are looking for resources to support their learning activities. Research has shown that technology can be effectively leveraged to support a range of pedagogical approaches in the classroom.

This paper will examine recent incidents of privacy breaches in education technology, and the lessons learned. A set of security best practices that education technology companies should follow will be developed.

Government regulations are often playing catch up when dealing with technology, but several states, most notably California, have taken an initiative on student data privacy. What rules do these emerging laws require ed-tech companies to comply with?

What risks, in the form of privacy and security, are students and teachers exposing each other to by leveraging technology? What steps are education technology companies taking to mitigate these risks?

Technology in the classroom is here to stay, but understanding and mitigating the privacy and security implications will be an ongoing process.

Supporting Video

https://vimeo.com/114375607

Introduction

The increased use of electronic systems in education has led to a focus on how that data is stored, accessed, processed, and applied. The type of data collected is wide ranging, and the existence of personally identifiable information (PII) of children increases the sensitivity of the data stored. Issues over where the data is stored, locally at the school or in the "cloud," as well as over who owns the student data abound.

Increased Concern

Earlier this year a teacher using Coursera, a popular online learning platform, discovered that it was trivial to dump the entire database of 9 million names and email addresses (Mayer, 2014). Exploiting the vulnerability didn't even require privilege elevation or SQL injection, all you had to was ask. An autocomplete textbox used when searching for someone returned as many names and emails as you asked for, enabling you to get a complete user list. Additionally, Coursera suffered from a cross-site scripting request flaw allowed any site you visited to list your course enrollments.

EdModo, a popular "Facebook for the classroom," didn't use SSL encryption for its student sessions, leaving them vulnerable to session stealing via network sniffing. The publicity the flaw generated, but not the applicable privacy laws, forced them to move to encrypting all traffic.

InBloom, an education data warehouse platform was forced to shut down over parent protests concerning the amount and scope of data tracking, forcing districts to drop the service. An analysis of their software showed fields in the database to store extremely sensitive information, such as the nature of family relationships, learning disabilities, and even Social Security numbers. Parents were mainly concerned over who owned such sensitive data and how it could be sold to third parties.

All of these incidents came to light because of concerned parents and educators. None of these breaches or insecure practices were against the law. The next section will look at some of the applicable legislation in this space, and how it has failed to keep up with the modern school.

Legal Landscape

First it is important to understand what is considered PII. FERPA, covered shortly, considers PII "information that can be used to distinguish or trace an individual's identity either directly or indirectly through linkages with other information."

The National Center for Education Statistics considers the following information to fall under consideration (Seastrom, 2010).

1. The student's name;
2. The name of the student's parent or other family members;
3. The address of the student or student's family;
4. A personal identifier, such as the student's Social Security Number, student number or biometric record;
5. Other indirect identifiers, such as the student's date of birth, place of birth, and mother's maiden name;
6. Other information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person to identify the student with reasonable certainty.

Now that you know what information is considered sensitive, both schools and technology providers need to take inventory of where this data exists and who has access to it.

There are two main pieces of legislation on a national level that affect the retention and use of student data. Sadly, neither contains any concrete direction or mandates concerning industry best practices detailing the secure storage or encrypted transmission of this data.

The Family Educational Rights and Privacy Act (FERPA) is a law intended to protect student's educational records. FERPA gives students and their parents' rights to data including:

- The right to access educational records kept by the school;
- The right to demand educational records be disclosed only with student consent;
- The right to amend educational records;
- The right to file complaints against the school for disclosing educational records in violation of FERPA.

Essentially it intends for students and their parents to be able to access, verify, and potentially amend records held by the school. FERPA was approved in 1974 so the idea that external companies would hold vast quantities of student data was not a possibility they considered, meaning that the law only applies to data held by schools. The law fails to consider the new ways, and entities involved in, the capture and transmission of student data in the modern world.

In addition to these problems, there are parts of the law that seem downright contradictory. While FERPA has provisions to protect students' right to privacy and a requirement for consent to disclose information to unauthorized entities, it simultaneously allows the release of student directory information (name, grade

level, contact information). Schools will often sell lists of students to universities for marketing purposes, which is completely legal.

The second major piece of legislation is the Children's Online Privacy Protection Act (COPPA), which applies to websites and applications aimed at and used by children under the age of 13. The main goals of the legislation are to:

- Provide parents notice of their information practices;
- Obtain prior verifiable parental consent for the collection, use, and/or disclosure of personal information from children
- Provide a parent, upon request, with the means to review the personal information collected from his/her child;
- Provide a parent with the opportunity to prevent the further use of personal information that has already been collected, or the future collection of personal information from that child;
- Establish and maintain reasonable procedures to protect the confidentiality, security, and integrity of the personal information collected.

Essentially COPPA wants sites that target children to have and post clear privacy and data retention policies, as well as provide parents with a way to review the data held about their children.

While the legislation has good intentions, in reality it does little to effectively protect student privacy. Websites continue to collect data on students under 13, especially

in cases where the student lies about their age when signing up. Methods for better age detection, such as entering a credit card number, have proven both onerous and ineffective. Parents who are engaged in their child's online activities often aid them in bypassing restrictions to gain access to online platforms.

Boyd et all (2010) suggest five goals to target in amending COPPA:

1. Limit how data about minors can be shared with third parties
2. Limit how commercial interests can target minors;
3. Require that minors opt in to changes to privacy policies and sharing policies;
4. Consider requirements to allow minors to prompt deletion of information about them; and,
5. Provide mechanisms to allow users (and their parents, in the case of children) to know when their data is being shared and with whom, if they care to know.

The final piece of legislation worth mentioning is a newly enacted law that applies only in California, the Student Online Personal Information Protection Act (SOPIPA). SOPIPA prohibits operators of online educational services from selling student data or using such information to target advertising to students. There are also provisions to prevent companies attempting to "amass a profile" on students for a non-educational purpose. It also requires providers to "implement and maintain reasonable security procedures and practices appropriate to the nature of the

Covered Information and protect that information from unauthorized access, destruction, use, modification, or disclosure." However, since these best practices security measures have not been specified in any way, there are no clear indications as to what security measures will satisfy these obligations.

To The Community

As we have seen in reviewing the pertinent laws concerning student data, they all make vague references to "best practices" to keep student data secure. I chose this topic to look at how the legal landscape has failed to keep up with the fast moving technological landscape. As an education technology developer, I have seen too little focus on security and privacy at the expense of delivering updates and features more quickly. This problem has been exacerbated by the type of people drawn to the education technology space. Former teachers with little programming experience learning on their own time aren't going to even consider security until it is too late.

I would like to start the conversation on what is considered best practices for the education technology providers. Such a discussion would be beneficial for both new entrants and existing companies by providing a basic roadmap towards privacy and compliance with the law.

Action Items – Best Practices

This is not meant to be an exhaustive, or overly technical, list of best practices technology providers can take to ensure student data is secure, but rather a set of guiding principles every education technology provider can use. The first step is to take security seriously upfront, not after the fact. Integrate security and data access policies into your workflow early. We will frame the discussion around the CIA principles of security: **C**onfidentiality, **I**ntegrity, and **A**vailability.

Confidentiality

This is roughly the same as privacy and is meant to prevent sensitive information from reaching the wrong people, while providing appropriate access to those who are allowed to access the data. It is important to have the appropriate access levels and ownership attributes on all sensitive information.

The most important step is to develop and publicly post a privacy policy and user agreement. This is one of the few specific things required by the applicable laws and will protect your, and your users, rights.

Take a step back from developing the next great feature and perform a review of your application. It might be helpful to start the conversation by performing an automated code review. Static analysis tools will scan your entire codebase and identify any potential security flaws. Dynamic analysis tools actually run the code or website and can test for real world scenarios like user inputs.

Think about any public API's and the data they return. The Coursera flaw let users see everyone's information, not just the students in their own school. Verify the access level and role of every request and ensure that unauthorized access of data does not occur.

Anonymize, or even destroy, PII data after a period of time has expired or the data is not longer needed.

Integrity

Integrity involves maintaining the consistency and accuracy of the data, both in storage in in transit. You must ensure that unauthorized people cannot alter or delete data.

For web applications, first and foremost encrypt all traffic. Many sites will only encrypt the login page, leaving the rest of the session in clear text, much like the Edmodo vulnerability. Every browser and operating system supports HTTPS and the performance hit is minimal in todays devices. Use a 2048-bit RSA or 256-bit ECDSA keys, which are secure and should stay secure for a considerable amount of time. If you have 1024-bit RSA keys in production, replace them with stronger keys.

Consider encrypting any personally identifiable information stored in a database, but especially passwords. If your database is compromised, the passwords can be

used to either access your service fraudulently, as well as other sites since people often use the same email and password combination in multiple places. I actually prefer to hash passwords, and simply compare the hashes on login. Hashing is sometimes called “one way” encryption because hashes cannot (generally) be “unencrypted.”

Finally use role based access to ensure that data cannot be accidentally or intentionally deleted or altered except by the appropriate people.

Availability

Availability ensures that people can access the data when they need it, meaning the servers have to be up and the data available. It also concerns physical availability, such as access to a disk or drive.

If you are using a cloud provider for hosting, ensure that you understand their security practices and uptime agreements. If your company is responsible for server management, have a regular schedule for installing updates. Keeping track of security vulnerabilities via central authorities such as the Common Vulnerabilities and Exposures (CVE) database.

Ensure that any platforms you depend on are regularly updated as well. Critical flaws in popular platforms like Wordpress are discovered all the time. Being on the appropriate mailing lists to know when and how to upgrade is critical.

Finally, know where your data is stored and who has access to it. Everyone knows to limit access to the production data, but backups locations like a file server and dev/test environments are often full of sensitive information and not closely monitored. Inventory your data and ensure access is limited to those who need it.

Conclusion

This paper has examined the evolving manner of technology and privacy in the classroom. This is an issue that will never be solved, the expectations parents and schools place on technology providers will continue to evolve. The legal landscape facing technology providers was examined, with a realization that the applicable laws lack little direction towards detailing industry best practices. The laws are outdated and only now becoming more a focus of discussion. It is important that the industry work harder to develop concrete best practices when it comes to securing student data. To that end, a series of best practices was detailed framed by the concepts of confidentiality, integrity, and availability.

References

Mayer, Johathan. A Funny Thing Happened on the Way to Coursera http://webpolicy.org/2014/09/04/a-funny-thing-happened-on-the-way-to-coursera. Accessed Dec 10, 2014

Boyd, Danah, Gasser, Urs, & Palfrey, John. How the COPPA, as Implemented, Is Misinterpreted by the Public: A Research Perspective. Statement to the United States Senate, Subcommittee on Consumer Protection, Product Safety, and Insurance of the Committee on Commerce, Science, and Transportation. April 28, 2010

COPPA, Children's Online Privacy Protection, http://www.coppa.org/

FERPA, Family Educational Rights and Privacy Act (20 U.S.C. § 1232g; 34 CFR Part 99). http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html

Seastrom, Marilyn. Data Stewardship: Managing Personally Identifiable Information in Student Education Records. (2010, Nov.). IES National Center for Education Statistics. SLDS Technical Brief. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011602