Sautter Award Submission: Secure Box

Project title: Secure Box

Submitted by: Erik Wieland, Manager – Document Management Systems, UCSF IT, 415-502-7822, erik.wieland@ucsf.edu

Project team: Jill Cozen-Harel (Box product owner and support), Jann Fong (first Box product manager, project manager), Mattice Harris (IT architecture and solutions management), Lalit Lakshmanan (desktop engineer), Phil Lunde (IdAM application analyst), Jennifer Pham (Box intern), Susil Rayamajhi (IdAM application programmer), Lakshman Reddy (IdAM application contractor), Christian Sisenstein (security analyst and architect), Erik Wieland (second Box product manager), Wesley Yip (desktop engineer)

Synopsis

We launched UCSF Box in 2013, and our users immediately asked "can we store PHI on Box?" Box signed a BAA in 2014, but we still had to tell our users "no PHI" until we had a technical solution for protecting it. Enter CipherCloud, which monitors changes in Box and passes files to our DLP system to be scanned for UCSF PHI. PHI is encrypted, and stays encrypted when downloaded. Users get a Secure folder where all content is encrypted. We launched Secure Box in 2016, and use exploded. We can report on PHI, stop people from sharing it outside of UCSF, and be certain that no one from outside of UCSF can read it. Say "yes" to storing PHI in the cloud!

Description

Our users want their data everywhere they are, but "everywhere" is a pretty scary place for restricted data like PHI. But if we don't offer a solution, our users will create their own, which is even scarier! So what do we do? Put our data in the cloud, and watch that cloud!

UCSF is UC's only campus working exclusively on health science, and 40% of our workforce is employed by the Medical Center. With clinicians, educators and researchers all over the globe "advancing health worldwide," we need to provide access to data everywhere our users are. Cloud services provide the features and scale that we need at a price we can (usually) afford, but the cloud is...cloudy on security. All that easy access means more ways for people to make poor decisions.

Assuming we would have to use two Box instances to secure our restricted data, UCSF sent out an RFP in Spring 2015. We asked for proposals which would leverage our existing data loss prevention (DLP) solution, already used to report on and secure restricted data at UCSF. Two vendors reached the proof of concept stage, and only CipherCloud passed all critical use cases and earned a rating of *excellent*. Critically, CipherCloud was the only solution providing persistent file-level encryption with central key management. This means the file stays encrypted even if a user shares it with someone via other methods (like email or removable storage). The file cannot be accessed/decrypted without authorized UCSF credentials (via Box SSO) and the CipherCloud decryption agent that is supplied with their offering. This extends to the mobile client too. Encryption is built-in, and the key management comes with the CipherCloud product. Because of these features, we were able to implement CipherCloud on top of our existing UCSF Box instance, which minimized user confusion, and greatly increased user adoption.

Architecture

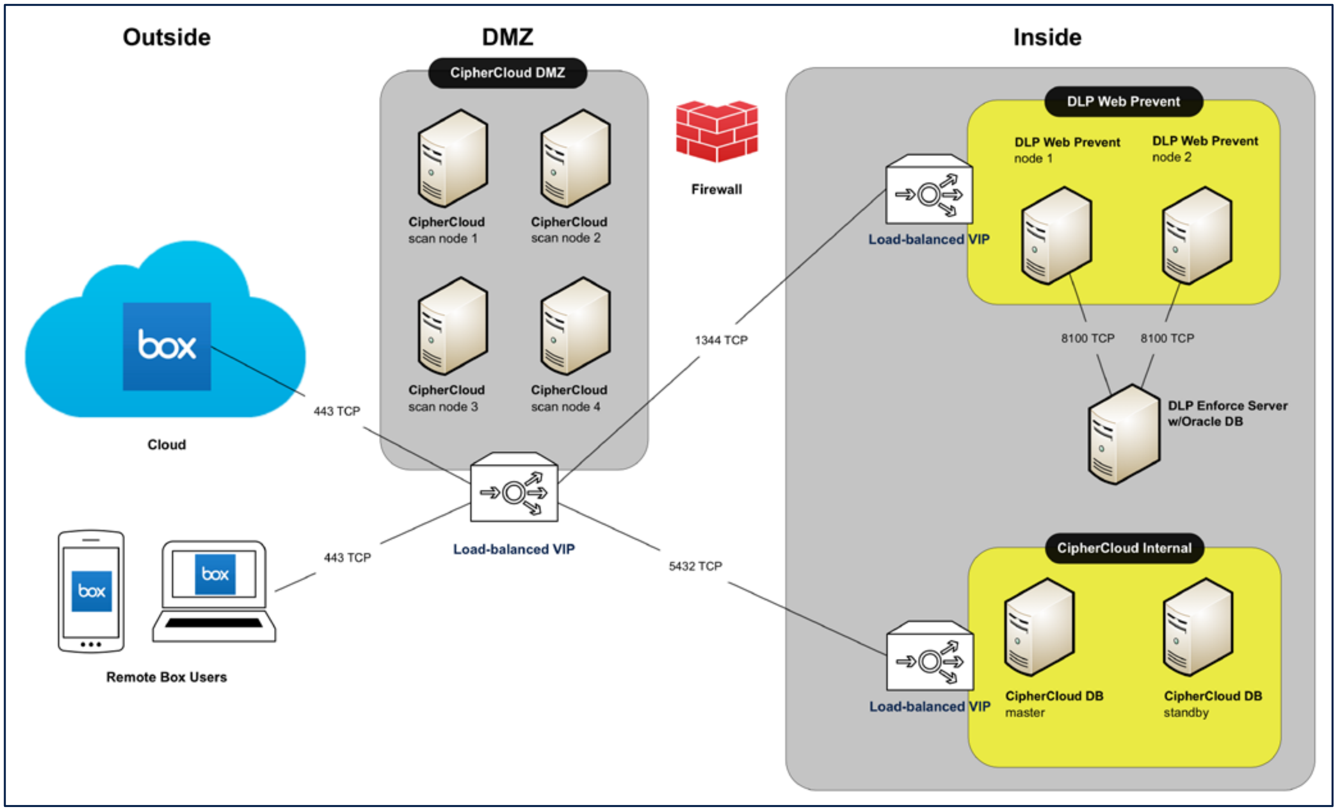


Figure 1: CipherCloud system architecture

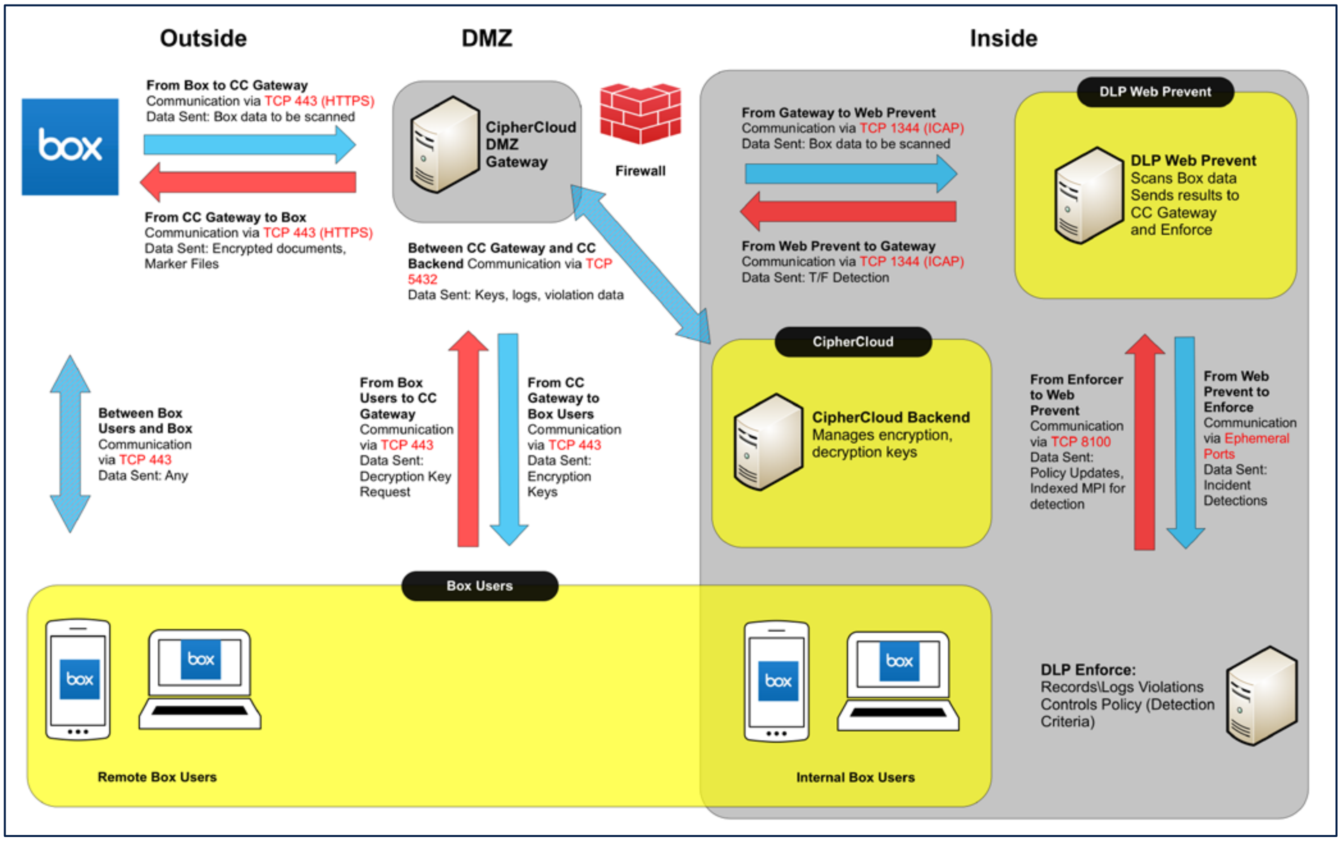
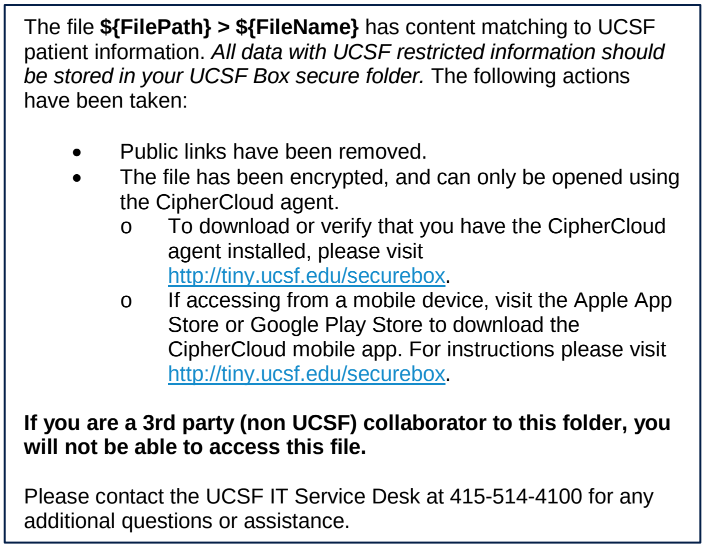

Outside
DMZ
Inside
box
From Box to CC Gateway
Communication via TCP 443 (HTTPS)
Data Sent: Box data to be scanned
From CC Gateway to Box
Communication via TCP 443 (HTTPS)
Data Sent: Encrypted documents, Marker Files
CipherCloud DMZ Gateway
Firewall
Between CC Gateway and CC Backend Communication via TCP 5432
Data Sent: Keys, logs, violation data
From Gateway to Web Prevent
Communication via TCP 1344 (ICAP)
Data Sent: Box data to be scanned
From Web Prevent to Gateway
Communication via TCP 1344 (ICAP)
Data Sent: T/F Detection
DLP Web Prevent
DLP Web Prevent
Scans Box data
Sends results to CC Gateway and Enforce
Between Box Users and Box
Communication via TCP 443
Data Sent: Any
From Box Users to CC Gateway
Communication via TCP 443
Data Sent: Decryption Key Request
From CC Gateway to Box Users
Communication via TCP 443
Data Sent: Encryption Keys
CipherCloud
CipherCloud Backend
Manages encryption, decryption keys
From Enforcer to Web Prevent
Communication via TCP 8100
Data Sent: Policy Updates, Indexed MPI for detection
From Web Prevent to Enforce
Communication via Ephemeral Ports
Data Sent: Incident Detections
Box Users
box
box
Remote Box Users
box
box
Internal Box Users
DLP Enforce:
Records\Logs Violations
Controls Policy (Detection Criteria)

Figure 2: CipherCloud data flow

UCSF IT found a solution that leverages our existing data loss prevention (DLP) solution, so we can apply the same detection criteria we already use to protect restricted data. CipherCloud monitors the Box API for changes, and passes them to DLP to detect policy violations in real time. Violating files are encrypted and written back to Box, and they stay encrypted when downloaded. We also provide a secure folder, where everything is encrypted without DLP scanning, and nothing can be shared outside of UCSF. For violations outside of the secure folder, we drop a marker file along with the encrypted file. This marker file informs the user exactly what happened and why, and helps us reinforce the proper handling of restricted data.



The file **${FilePath} > ${FileName}** has content matching to UCSF patient information. *All data with UCSF restricted information should be stored in your UCSF Box secure folder.* The following actions have been taken:

- Public links have been removed.
- The file has been encrypted, and can only be opened using the CipherCloud agent.
  - To download or verify that you have the CipherCloud agent installed, please visit http://tiny.ucsf.edu/securebox.
  - If accessing from a mobile device, visit the Apple App Store or Google Play Store to download the CipherCloud mobile app. For instructions please visit http://tiny.ucsf.edu/securebox.

If you are a 3rd party (non UCSF) collaborator to this folder, you will not be able to access this file.

Please contact the UCSF IT Service Desk at 415-514-4100 for any additional questions or assistance.

Figure 3: Marker file text

We provision accounts just in time, inactivate separated users nightly (with notifications to supervisors and collaborators), and delete inactive accounts after 90 day. We are stopping people from creating public links to restricted data. We can even scan for other types of restricted data (e.g., PCI, PII), and decide based on real data whether we want to create new DLP rules.

The UCSF Box team worked with Stanford, who had already implemented a Box secure folder structure for another solution, to devise our secure folder structure. The method we chose keeps users from renaming or removing their secure folder, and all secure data is owned by the CipherCloud service account. This method makes it easier to centrally manage all secure folders. For example, we have restricted all folders owned by the CipherCloud service account to disallow sharing outside of UCSF. This folder structure obviated the need for a separate Box instance, as the secure folder became the location where users could store all restricted and sensitive data, regardless of DLP rules. DLP can't scan multimedia files for PHI, for example, but clinicians can store their patient interviews and case notes in their secure folders without worrying about their safety and security.

The UCSF Box team worked with the UCSF Privacy office to develop audit procedures. When we launched Secure Box we immediately started encrypting files containing PHI in real time, as they were uploaded or edited. Since all UCSF faculty, staff, and students are trained and authorized to handle PHI, we focus on files violating policy which may have been shared with people outside of UCSF. We came up with an audit procedure where we reconstructed the sharing and access history of any file violating policy, so we could determine if it might have been accessed by someone outside of UCSF. We also began running on-demand scans of historical data (data on UCSF Box which hadn't changed since we launched Secure Box), and reconstructing the access history for these files. We then determine whether further investigation is necessary. In this manner, we are securing all of UCSF Box's files, with confidence about how it has been handled and by whom since we launched our Box instance.

Timeline

- UCSF's Box instance turned on for testing: September 26, 2012
- UCSF Box pilot: December 2012 through February 2013
- UCSF Box launched: February 28, 2013
- Box signs UC BAA: June 20, 2014
- RFP issued for Box DLP solution: March 9, 2015
- Finalists present proofs of concept: July 2015
- CipherCloud implementation project kick-off: February 2016
- CipherCloud architecture finalized: April 2016
- Build: May 2016
- Internal testing, configuration tweaking: June-July 2016 – This took longer than anticipated, due to much higher loads than CipherCloud had ever seen. They produced multiple system patches and enhancements over this period.
- Enterprise communications and pilot: August-September 2016
- Launch: October 3, 2016
- Stabilization: October 2016-February 2017 – With the announcement that we were finally allowing PHI on Box, usage jumped. Our storage went from 56TB to over 130TB in this period. CipherCloud continued to enhance the system to accommodate this scale.
- Automated deprovisioning begins: February 2017

Limitations

With any security initiative, we had to balance risk with usability. As we had seen many times in the past, if we tilted too far toward security we would drive customers to unsanctioned solutions. If we provided only a veneer of security, we would be wasting resources on a solution that did little to secure our most sensitive data. With that in mind, and in concert with Security, Privacy, and UCSF leadership, we made the following compromises.

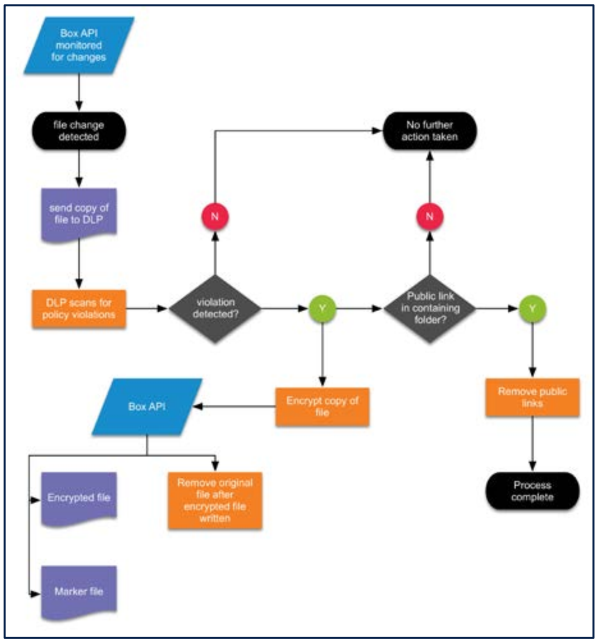


Figure 4: CipherCloud real-time actions

We do not scan files before they are written to Box. A delay of a few seconds was seen as unacceptable by users, and would have pushed them to other platforms. In testing we saw a violation rate of 0.3%, including false positives, so we chose to scan the Box API for changes after they occurred. Under normal load, the round trip from Box to DLP and back is under 80 seconds (depending on the size of the file being encrypted). In subsequent audits, we were able to show Privacy that no violating files were accessed within this window. It is still a risk, but as the alternative is having no visibility or control of our restricted data, it was felt to be a minor one.

Because of the huge volume of files on UCSF Box, we could not perform a complete scan of the data before launch. This means we are retrospectively scanning archival files for restricted data using on-demand scans, focusing on files within folders with public links. However, as real-time scanning captures all subsequent changes to these files, we are catching all file activity, and will eventually scan all of the data on UCSF Box.

DLP scans for restricted data matches using text, so we are unable to scan multimedia files. We can still scan the names of these files for restricted data matches, though, so we are doing all we can to reactively protect this content without implementing image and audio recognition.

Scanning and encryption takes time, and in order to balance system performance with security, files over 400MB are not encrypted (either inside or outside the secure folder). A review of the UCSF Box data showed that the vast majority of files over 400MB would not be scanned anyway, as they were multimedia files. With recent improvements in the CipherCloud product, we are considering raising this limit to scan larger files. Since decryption takes place on the client, we may yet reach the limits of older clients' ability to decrypt very large files, so we will proceed with caution.

Impact

UCSF implemented Box in 2013, with a 25000-user license. Our active user base was about one third that before launching Secure Box, with little use by Medical Center staff. CipherCloud sized their solution with this in mind, and incorporating room for moderate growth. During build-out we were storing 47TB of data, and at launch we had 56TB of data in Box. Four months after launch we were at 130TB of data, and our active user population more than doubled, including a significant number of users from the Medical Center and trainee populations. At six months, when this was written, we were at 150TB, three times the scale we had built around. We have over 20TB of data in secure folders alone. Most of this data is new, but all of it would have been stored elsewhere: on computers, in email, on flash drives, or in unsanctioned cloud services. All of this data is now on UCSF Box, and all of it is secure.

Customer reaction was surprising. There were a few people, notably in the research community, who were against encryption of any kind. The vast majority of customers said nothing, but those who did thanked us. We did something unexpected: we gave them a solution to a security problem, rather than just saying no. The response to automated deprovisioning was also surprising, as most users thanked us for letting them know what they were losing access to, and departments added ownership transfer to their offboarding checklists. The continued customer education, which we thought would be the most difficult aspect of this project, has created a positive ripple effect on security, privacy, and business continuity.

The road to Secure Box was not smooth, and we learned a lot about working with small vendors (and they learned what "a lot of data" means). But ultimately, through a combination of good architecture and good communication, we provided a solution that users thank us for!

Reference

- CipherCloud
- UCSF Box service page
- Secure Box service page
- UCSF Box Automated Deprovisioning information page