SCI StorInt™ Dispatch – Seagate071015
Announcing new options for securing disk data

This Silverton Consulting (SCI) Storage Intelligence (StorInt™) Dispatch provides a summary of Seagate, LSI, and IBM's collaboration in new options for securing disk data on LSI mega-RAID and control units based storage.

Seagate's strategy

Just about once a month we see news of another data security breach releasing 1000s to millions of credit card and/or social security numbers. Up til now these have been mostly due to laptop or tape cartridge loss. However, data security is not limited to laptops and tape. All that data residing on tapes originated on data center disk storage.

Consequently, one problem is all disk drives are ultimately removed from data center protection for re-sale, re-furbishing, or servicing. Today, once a drive leaves the protection of a data center, data on that disk can be read at will. Tomorrow, if the data is on an encrypted drive with protection enabled, data cannot be read at all.

Other options

There are many points where data being stored could be encrypted - at the host, in the network/appliance, at the controller, or at the drive. There are advantages and disadvantages to each of them:

- Host encryption burns up host based MIPs and encrypted data cannot be compressed, or de-duplicated. Indexing and searching only work if the data is indexed/searched from hosts who have access to the encryption keys/software. Host encryption protects not only data-at-rest but also data-in-flight. Also, to support DR requirements the host software and the encryption keys would need to be available at the secondary site. Having lots of cipher text available to anyone listening allows certain cryptographic attacks which can be used to crack the keys being used.
- Network/appliance encryption burns network/appliance cycles and data being stored cannot be compressed or de-duplicated past the network/appliance doing the encryption. Searching and indexing may be ok if the data is accessed over network/appliance hardware that has access to the encryption keys/software. Network/appliance encryption provides limited data-in-flight protection (from the point the data is encrypted) but it does provide full data-at-rest security. Also, for DR you would need a copy of the network/appliance hardware plus access to the encryption keys. These products also allow cryptographic attacks used to crack keys because both the plain-text and the cipher text are available
- Control unit encryption burns control unit cycle and depending on where encryption is done this may or may not impact data compression and de-duplication. Searching and indexing are not impacted. Control-unit encryption has no protection of data-in-flight but does protect data-at-rest. Also, for DR purposes you may or may not require copies of the control-unit and the encryption keys at the remote site. One issue is when drives are shared across control units the cryptographic keys would also

need to be shared. Also the cipher text is available once a drive is removed from the system and can once again be used in cryptographic attacks to guess keys

- Drive based encryption burns drive cycles, has no impact on controller data compression or de-duplication. Searching and indexing are not impacted. Similarly drive encryption has no protection for data-in-flight but does offer protection of data-at-rest. Finally for DR purposes there is no intrinsic requirement for keys or encryption at the remote site (but see below for LSI constraints). Finally as cipher text is not externally available this option is relatively immune from cryptographic attacks

A key question is what threats are you answering with encryption.

- One threat is having clear data outside the data center's protected arena. To answer this threat you must protect data-at-rest and any of the above ways of encrypting data will suffice
- Another threat is having clear data accessible within your data center to un-authorized users/applications. To answer this threat you must protect data-in-flight and you must use host or network/appliance based encryption.

How drive encryption works

Encryption hardware is available on the drive and encryption firmware and keys are added in a manufacturing feature personalization step. At the first power up the drive is able to read and write without authentication but it is always encrypting data. A control-unit can use a special authentication key to lock a drive down so that it requires an authentication key to provide clear data back to the subsystem. From that point on every power up - the drive does not respond to normal I/O commands until it receives proper authentication. If authentication fails enough times the drive destroys its encryption keys and renders all old encrypted data un-decipherable. This also happens when the drive is told to do a secure erasure. Drive encryption keys are kept on the drive in all zones and on all heads so any individual defect will not cause loss of encryption keys.

LSI controllers support cryptographic authentication in one of two ways:

- With IBM's enterprise key manager supplying authentication keys
- Internally with the LSI controller supplying authentication keys

Either approach is secure and provides appropriate mechanisms for the drive to insure its talking to the right control unit. If the drive needs to be moved to another control unit it's authentication key can be provided to the other control unit. Drive data recovery can also be accomplished with the proper authentication key. LSI and IBM EKM drive authentication keys can be backed up securely using yet another encryption key together with a pass phrase.

One advantage of drive encryption is that the drive supports a very quick and secure data erase by securely destroying its encryption keys. The encryption is AES-128 and can be changed in the future without impacting system architecture. Also, encrypted data is never available outside the drive, which makes many potential cryptographic attacks virtually impossible.

Problems overcome

Obviously just having encrypted drives in your environment does not protect data-at-rest. One must enable the drive authentication and whenever you replicate the data the data you must also use encrypted drives. LSI has a concept of a "security domain" and a "secure volume group". A secure volume group requires all drives in a RAID group to support drive encryption and to have encryption enabled. A security domain is having one or more secure volume groups for data to be secured. LSI constrains local and remote copies to be within a security domain, i.e., encrypted drive data cannot be snapped, cloned, or remotely replicated to non-encrypted drives.

What's the risk

Historically, data backup was the highest security risk. Getting a backup copy of data was relatively easy and once obtained all the data was in the clear. IBM, HP, Sun, and others have begun to encrypt tape data, which has quickly moved to close this exposure.

Consequently, disk drives are the next choice for a weak point. Security conscious government agencies have typically not allowed disk drives to leave their premises – voiding device warranties. Some security conscious commercial entities have shredded or crushed disk drives in the past after having done hour or daylong security erasure passes.

Accordingly, enterprise class drives have a MTBF around 1.3 million hours and low-end drives have a MTBF of 600,000 hours or more. With 8,766 hours/year this seems like a long time to wait for a drive to fail. But it's not unusual to have an enterprise data center with a PB of data or 1000s of disk drives. For 2000 disks you would lose one enterprise disk every 650 hours, ~1 drive every month and you would lose one low-end disk every couple of weeks. Most shops have a mixture of enterprise and low-end disks so it's not unusual to see a disk leave the data center every month.

However, what's actually on a drive leaving the shop and how easy it is to extract sensitive data off the drive is subject to some debate? Its not unusual for over 50% of the drives returned to a factory to have no defect found on them – meaning that one out of every two replaced drives read and write flawlessly. These (non-encrypting) drives could easily be plugged into any compatible interface and all the data could be read out, indexed and searched.

In contrast, reading drive data and knowing what that data represents is somewhat difficult. Most enterprise disk storage is configured in RAID groups, where one drive is but one of 4 or more drives in a group. Which RAID block addresses map to which drives in the RAID group is non-trivial, and knowing which RAID group blocks represent which LUN blocks is also non-trivial. Also, both of these are vendor specific and may depend on features enabled. Not easy to crack but not that hard for someone familiar with vendor internals.

Silverton Consulting
Strategy, Storage & Systems
P.O. Box 492|Broomfield, CO 80038-0492|
720-221-7270|info@silvertonconsulting.com

Fortunately for hackers (and unfortunately for data centers) all this may not be necessary as once one can access the data one could easily scan for likely text strings such as "Payroll", "SSN", "Credit Card Number", etc. This sort of "brute force" approach works well if you have the time and (un-)fortunately once a drive is outside the protection of the data center a hacker has the time.

If this is so easy why don't we see it on the news already? It's fairly easy for a data center to identify data on a lost backup tape or laptop but it's much more difficult to know what data was on a drive leaving their shop. Also legislation requires disclosure for lost data tapes and laptops but have yet to catch up to require disclosure for lost disk drives. Furthermore, re-sellers, re-furbishers, and/or service organizations are not currently required to report on missing disk drives. Someday soon legislation will catch-up and require such disclosure and tracking and then this all will be commonplace.

Announcement significance

It won't take long for drive and storage vendors to standardize on drive encryption to protect data-at-rest – for the cost of cryptographic authentication and encrypting drives this problem can be solved. Most large storage vendors already have key management and the rest can quickly coming up to speed. Look for other drive vendors and storage subsystem suppliers to quickly come onboard.

Security is a never-ending journey. Once you close one loophole another pops up quickly. Drive encryption is a straightforward, inexpensive and secure defense against drives falling into the wrong hands outside the data center. However, the next logical threat involves unauthorized access to clear data within the data center – to close this loophole protecting data-in-flight is needed. Infrastructure put in place to support drive encryption should be easily extensible to this as well – stay tuned.

Silverton Consulting, Inc. is a Storage, Strategy & Systems consulting services company, based in Colorado, USA offering products and services to the data storage community

Silverton Consulting
Strategy, Storage & Systems
P.O. Box 492|Broomfield, CO 80038-0492|
720-221-7270|info@silvertonconsulting.com