Where In The World is My Credit Card Data?

Greg Swedosh
Knightcraft Technology
Senior Security Consultant

With more than twenty-five years Tandem experience, Greg Swedosh is a specialist in security and PCI compliance for the HP NonStop platform. Greg has presented on HP NonStop server security and compliance in the USA, United Kingdom, Netherlands, India and Australia. He is the primary author of the technical white paper PCI Compliance for HP NonStop Servers and a contributing author of the book Securing HP NonStop Servers in an Open Systems World. Greg can be contacted at greg.swedosh@knightcraft.com. A copy of the PCI compliance white paper can be downloaded free from www.knightcraft.com.

Introduction: What is a PAN and what is PCI DSS

Many organizations running HP NonStop servers process credit or payment cards as a part of their business. This means that they are subject to the Payment Card Industry (PCI) Data Security Standard (DSS), which mandates a strict set of requirements for the protection of payment card cardholder data. At the core of the standard is the Primary Account Number, usually abbreviated to PAN. This is the card number embossed on credit or payment cards that is used in payment card transactions, along with the card expiry date and often the Card Verification Value (CVV). The PCI DSS mandates in requirement 3.4[1] that all PAN data that is stored on the system must be protected by means of encryption or tokenization. But how do you know where all of your PAN data exists?

The need to identify where PAN data is located

Organizations that are subject to PCI DSS must have an annual assessment of their system by a PCI Qualified Security Assessor (QSA) to determine whether or not they are compliant with the standard. This is a thorough audit of system and application settings, procedures and documentation as well as an analysis of the methods used to protect PAN data, in any system or network component on which cardholder data is processed, stored or transmitted.

PCI DSS v2.1 states: "The first step of a PCI DSS assessment is to accurately determine the scope of the review. At least annually and prior to the annual assessment, the assessed entity should confirm the accuracy of their PCI DSS scope by identifying all locations and flows of cardholder data and ensuring they are included in the PCI DSS scope."[2]

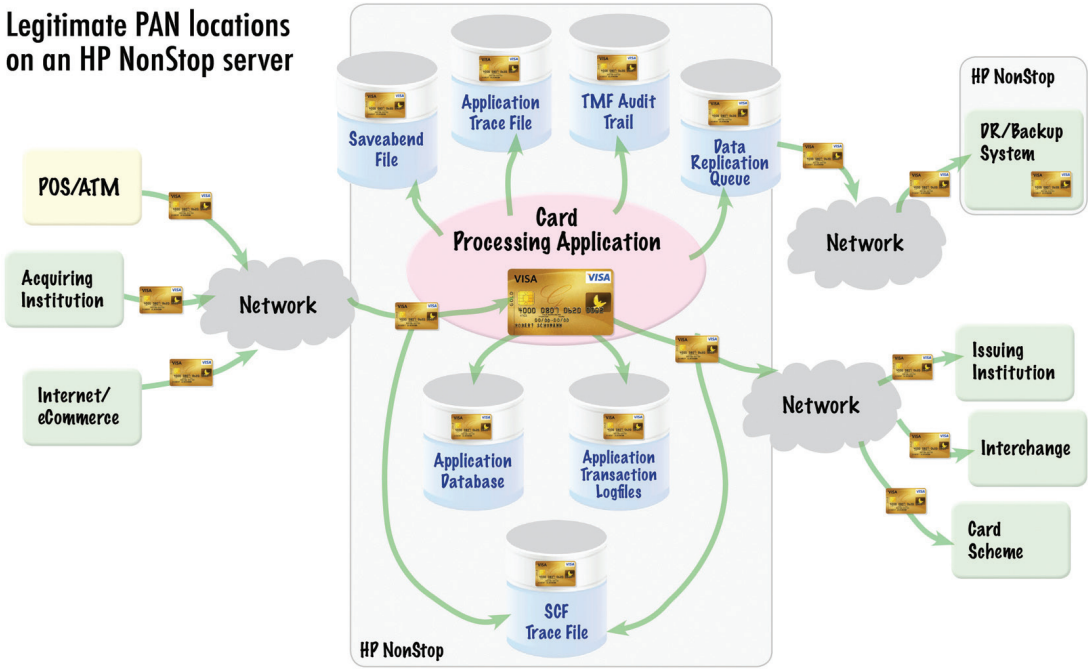
Organizations will typically have a good idea of where PAN data is stored as part of their core application. But is this the only place where PAN data resides on the system? Could copies of data containing unprotected PANs exist somewhere on the system, possibly relating back to the troubleshooting of some old application problem? Has live production PAN data been copied by some member of staff for their own testing purposes? Worst of all, has any employee copied PAN data for their own unauthorized purposes? It is essential as part of any PCI assessment that all of the places where PAN data is stored are identified and documented, not just for compliance reasons, but also to mitigate the risk of credit card fraud. But how can you be sure that you know all of these locations?

But my application is tokenized or "encryption enabled"…

A number of the widely deployed card payments applications are user customizable. In general the customization tends to be in regards to different message formats and interfaces. The storage locations of cardholder data will often remain as standard. There may however be non-standard legitimate locations where cardholder data is also stored, depending on specific organizational requirements. There certainly should be no assumptions made that all cardholder data on the system and in the network is protected, simply on the basis that a new version of a card processing application is "encryption enabled". A full analysis needs to be conducted of all possible locations where cardholder data may exist and a determination made as to whether all data elements are protected as required by PCI DSS.

It is mandated in PCI DSS that "*the assessed entity identifies and documents the existence of all cardholder data in their environment, to verify that no cardholder data exists outside of the currently defined cardholder data environment (CDE).*"[3]

[1] PCI DSS Requirements and Security Assessment Procedures, Version 2.0, October 2010, Requirement 3.4
[2] PCI DSS Requirements and Security Assessment Procedures, Version 2.0, October 2010, Page 10 (Scope of Assessment for Compliance with PCI DSS Requirements)
[3] PCI DSS Requirements and Security Assessment Procedures, Version 2.0, October 2010, Page 10 (Scope of Assessment for Compliance with PCI DSS Requirements)



There are a number of places that PAN data may legitimately exist on the system and all of these need to be specifically identified as part of the scope definition for a PCI DSS assessment. All such identified files must be appropriately protected in accordance with PCI DSS requirement 3.4. Typical locations where cardholder data may be stored are:

- Cardholder database. Often the PAN (cardholder number) is used as a primary key.
- Transaction log files
- TMF audit dumps and audit trails
- Backup media
- Data replication transaction queues
- Application trace files
- Communication line/process trace files (e.g. SCF traces)
- Saveabend (or other memory dump) files

And of course it is essential to also determine if any unauthorized copies of PAN data exist anywhere on the system.

What happens if the scope of my PCI assessment is incomplete?

Not accurately defining the scope of a PCI assessment can ultimately lead to non-compliance with PCI DSS. Failure by an organization in this regard can mean that appropriate consideration may not have been given to certain files containing PAN data, which means they may not be appropriately protected in accordance with PCI DSS.

If a QSA finds cardholder data outside of the defined scope, they will often take it that the organization does not have an appropriate understanding of where all of their cardholder exists or have not taken suitable measures to determine this. If "live" PAN data is found to be unprotected, it will be considered a serious issue of non-compliance. In recent times, a number of organizations around the world have been subject to very large financial penalties as a result of being non-compliant with PCI DSS. For tier 1 organizations, these sums will often be in the millions of dollars. However, the cost of this could pale into insignificance compared to reputational damage, or the cost to an organization, if a cardholder data breach occurs resulting in fraudulent transactions.

QSA testing for compliance with PCI DSS requirement 3.4

A number of QSAs will run the following test on an organization's system during a PCI assessment. Firstly, perform a number of different end-to-end transactions through the customer application with a single, known payment card number. Then, once the transactions have completed, scan all files on the system searching for that specific known number. This is an effort by the QSA to find out where exactly the card number is stored in an unprotected manner. This is a particularly useful test for organizations to run themselves in the early days of moving towards PCI DSS compliance, before they have introduced encryption or tokenization into their application. This test will help identify all files where PAN data is stored by the application.

What about development and test systems?

PCI DSS requirement 6.4.3 mandates that "*Production data (live PANs) are not used for testing or development.*" So it is essential that an organization has some documented and provable mechanism in place to ensure that this is the case. In less regulated times, organizations may have used a copy of live production data for testing purposes on development systems or perhaps used to troubleshoot an application problem. This is not permitted under PCI DSS. Development and test systems often have security and auditing requirements far less stringent than ▶

production machines. So if production data for some reason were to exist on a development or test system, the possibility of data breach is likely to be significantly greater. Measures must be taken to ensure that no production data exists on these systems.

The difficulty of finding PAN data

As stated earlier in this article, an organization will usually know the key locations related to their application where cardholder information exists. The problem that an organization faces is proving to a QSA that these are the ONLY locations where cardholder data exists.

The PCI DSS introduction states that it is necessary that an organization "*retains documentation that shows how PCI DSS scope was confirmed and the results, for assessor review and/or for reference during the next annual PCI SCC scope confirmation activity.*"[4] That is, the means by which the locations of all cardholder data have been determined needs to be described in a document and provided to the QSA. So how do you prove that no PAN data exists outside of the areas that you have identified? It is not practical or reliable to use any kind of manual method to achieve this. The amount of data to be searched is invariably too vast. Some kind of automated search for PANs in all files on the system is the only viable and accurate way of being certain that you have identified all files that contain cardholder data.

What are False Positives?

The makeup of a PAN is essentially a numeric string that fits to certain conventions. A VISA card PAN, for example, is a 16 digit number beginning with 4. American Express PANs are 15 digit numbers beginning with either 34 or 37. To simply search large quantities of files for these basic numeric strings will likely yield a very large number of false positives. That is, number patterns that exist in files that may match these criteria and are detected as PANs (falsely), but do not actually represent PAN data. For example, there may be many occurrences of 16 digit numbers beginning with a 4 that reside on the system, for example as part of larger numeric strings that have nothing whatsoever to do with payment cards or their data. Any tool that returns a significant number of false positive results makes analysis an extremely onerous task. All of the identified files would then need to be examined manually to determine if the data detected truly does represent PAN data.

What are False Negatives?

As distinct from false positives, false negatives are files that are identified as "clean" but do in fact contain unprotected PAN data. This may be due to the file being stored in an unexpected format such as binary or EBCDIC, for example.

Challenges in creating a tool yourself

On first glance one could think "this is an easy task, I will just quickly whip up a program for that myself". Here are some points to consider which might make the job harder than you think:

- You want to minimize False Positives
- You want to be able to search for PANs stored in other formats (i.e. binary, EBCDIC)
- You want to be able to search all kinds of files (ENSCRIBE structured, ENSCRIBE unstructured, SQL, PAK files and so on)
- The tool has to create very clear PCI compliant reports, which ideally can be viewed both on the NonStop and on the PC
- A feature such as "partial scan" or "re-scan while retaining context from last scan" can be helpful
- You may want results of PANs detected to be written directly "off box" to a SIEM device such as HP ArcSight or RSA enVision
- You don't want the tool to adversely affect the performance of your core application
- You may want the tool to run in an agent mode, where it detects unprotected PAN data as soon as possible after it appears on the system

Automated intelligent PAN discovery with PANfinder

PANfinder is a tool that was developed by 4tech Software specifically to address the issue of PAN detection on HP NonStop servers. It has been designed with intelligent scanning functionality that minimizes false positives while finding all real PANs and identifies also full track data. Other features include a partial scan mode that scans only files that have been modified since the previous scan, restartable scans so that scanning can be performed automatically at "off peak" system hours only, minimization of system resource utilization and also the ability to send events to a centralized SIEM device on the corporate network. PANfinder also has the ability to run in "QSA test" mode where it searches for a single known PAN.

PANfinder is distributed internationally by comForte 21 GmbH. See the comForte website at http://www.comforte.com/products/protect/panfinder/ for details.

Summary

In our work with customers, it is a rather common scenario to find that a lot of time and money has been spent on protecting PANs known to be in specific locations. In a surprisingly high number of cases we have found out that the "data discovery phase", introduced earlier in this article, was basically skipped. Instead, a quick Excel spreadsheet was put together and labeled "this is where our PAN data is located".

We believe this is the wrong approach and the PCI DSS standard wording strongly supports our view. Without having a tool such as PANfinder available, data discovery can prove to be a long winding road. Using a proper tool makes the process painless for the company, as well as making it easy for the QSA to verify that all is as it should be.

[4] PCI DSS Requirements and Security Assessment Procedures, Version 2.0, October 2010, Page 10 (Scope of Assessment for Compliance with PCI DSS Requirements)