Report on inadvertent data leakage
School of Informatics
5th March 2019
Authors: Stephen Quinney and Alastair Scobie

Summary
On Wednesday 20th February 2019, we discovered that for a period of 3 weeks, authorisation checks for login access to a number of Informatics' servers had been unintentionally bypassed. The result of this was that certain medium to high risk data may have been available for access by all members of the School. Whilst we have evidence that some users made use of this login access, we have no evidence that they accessed the medium to high risk data - but nor do we have any evidence that they did not. The correct authorisation checks were reinstated within an hour of learning of the problem.

Details
On Wednesday 20th February 2019 we identified an issue affecting all Informatics managed (DICE) machines which meant that authorisation checks for Secure Shell (SSH) login access had been unintentionally bypassed for the previous 3 weeks. Throughout this period, authentication still functioned as normal which means that access remained strictly limited to valid DICE accounts. DICE accounts are only available to the staff, students and visitors of the School and to a small number of other University of Edinburgh staff. The impact of this mistake was that users had access to machines which would not normally be granted (for example, undergraduates being able to access the staff SSH server).

This problem was first introduced onto DICE development machines on Friday 2nd November 2018 when a change was made to the authorisation section of the Linux PAM configuration for SSH. The change subsequently reached all other DICE machines as part of the weekly stable (“production”) release on the afternoon of Wednesday 30th January 2019. The primary reason for this incorrect configuration change being deployed to all stable DICE machines was insufficient testing of potential outcomes.

For auditing and security investigations, and to ensure that the School's resources are accessible only to authorised users, we retain all login information for a period of 120 days. Using this data, all successful SSH login events for the period have been compared against the expected access rules for the relevant machines and all potential unauthorised access has been identified.

We have identified potential unauthorised access to 41 DICE machines involving 80 user accounts. The authorisation checks were done using current access rights for users and current rules for the machines. We do not retain a complete historical record for all changes to access rules, so we expect a certain level of "false positives". To minimise the impact of this issue we have filtered out login events associated with accounts which have since expired (and thus no longer have any access rights).

It appears that most of the unauthorised access events were unintentional. For example, there is a service named student.inf.ed.ac.uk which would normally only be accessible via SSH for administrators, this name
Lessons learnt and resulting actions

The authentication and authorization of user login access for Linux is typically configured using the Pluggable Authentication Modules (PAM) system. This uses a powerful and flexible but rather complex configuration syntax. The direct result of increased complexity is an elevated risk of a configuration change containing an error. It is clear that when coupled with the likely high impact of any PAM misconfiguration there is a need to approach all such changes with a good degree of caution.

**ACTION:** All future PAM configuration changes will be peer reviewed and thoroughly tested using a variety of test user accounts with different access permissions.

This incident showed that we are wholly reliant on our PAM configuration for authorization of user access.

**ACTION:** We will investigate whether there are ways we can enhance our "strength in depth" with additional requirements that are applied at other levels in the login process. For example, we will investigate whether we can configure the SSH daemon to also check users are in the correct groups.

Our review of data which is accessible by user accounts without special privilege (e.g. root) revealed that on many servers we are reliant upon the login-time authorization rules for protection of sensitive data. This is clearly less than ideal. Our entire computing team has login access to all servers using their personal accounts: if any of those were to become compromised, an attacker would have read access to a wide range of sensitive data.

**ACTION:** We will carry out a review of sensitive data held on our servers and enhance the filesystem access controls to ensure that access is strictly limited to relevant users. We will also review the default settings for filesystem access controls with the aim of assisting our users in storing their data securely.

During the investigation of this incident it became clear that we do not have an adequate register of what data is held on individual servers. We currently do not hold sufficient information on location, purpose or sensitivity of important data. This makes the process of reporting on potential data loss time-consuming and likely to be error-prone. Related to this is that we also do not hold adequate information on the roles of all servers or a complete list of who is responsible for each server. There is a current development project to develop a data asset register and a mechanism to populate, and keep up-to-date, this register.

**ACTION:** We will raise the priority of the development project to deliver a data asset register.