Terabyte Threat Analysis
Project Plan

Team Lima
For Paul Reid on behalf of BT

12th February, 2013

Contents

1 Scope

This document is intended to provide a high-level plan of the key milestones, assumptions and decisions made in a software project to implement a high-performance, scalable network threat analysis system. The detailed specifics of the project, including its design and implementation, are the subject of another document, the *Requirements Specification*, which is published alongside this document.

2 Project Details

This project is a 6 week collaboration between six Computer Science students at the University of Cambridge Computer Laboratory, hereinafter referred to as *Team Lima* or simply *the group*, and an external client, Paul Reid, from BT. Due to the scale of the project and the nature of the design decisions involved, it has taken some time to capture the requirements, complete a design to a suitable standard and, subsequently, to prepare this document for publication.

2.1 Responsibilities

2.1.1 Client

The client's responsibilities include the provision of the original project brief, which defines the background to the project, and supplying any resources which will be necessary to make progress possible. In particular, the client is responsible for providing the group with sample data in an appropriate format to allow in-house testing and demonstration that the system works correctly. The client also oversees Team Lima and their progress by means of three client meetings, to be held at the Computer Laboratory periodically during the course of the project. It is intended for the group to operate, for the most part, on an autonomous basis between such meetings.

2.1.2 Team Lima

Using the requirements captured from the client's brief, the initial meeting and any subsequent communications which may be exchanged, the group is responsible for the definition and management of the project, its design and arranging for its implementation.

3 Project Phases

3.1 System Design (∼2 weeks)

The design of the system has been the subject of much deliberation, discussion and planning on the part of the group. As of the time of this writing, the overall design of the project, including the structure of the modules and their interactions, has been completed. The design requirements represent the overall structure of the system and its services to which the group will commit itself to developing with the client. As such, it is likely that any modifications to the design requirements at a later date without good reason will be seen unfavourably by all parties.

3.2 Implementation (2 weeks)

A correct design and project plan will lead naturally into an implementation phase, in which the group writes the programming code to implement the behaviour of each module. This code will be written in accordance to the design document. Particular attention must be paid to the dataflow diagrams, interface design and database schemas for each module. Correctly adhering to such interfaces will streamline the integration of each module at a later date.

It should be noted that the period of time predicted for the implementation is equal to that which elapsed during the design phase. While this may seem counter-intuitive, the complexity of the system has necessitated an extended design phase. During this period, the group located research papers to understand the algorithmic analysis of the data, and experimented with the various technologies to be used. This has contributed useful information to the design, which the group is confident is exhaustive in every regard. In addition, the modular structure divides the project cleanly into modules which, when stood alone, are relatively trivial to implement. Finally, it is not expected for the group to make further design decisions during the implementation, which is typically a contributing factor to slow progress in any ad-hoc development project.

3.3 Testing (3–4 days)

A critical phase of the project will be its testing. This confirms that the individual modules conform to their respective requirements and, in turn, the entire system implements the overall behaviour expected and conforms to its design requirements. This phase of the project may take place alongside the implementation such that particular modules are tested in parts as they are developed. If this is the case, an overall period of testing, during which all modules are integrated and expected to communicate successfully with each other, will still be required to prove system functionality, and to test those components which cannot be easily tested on a standalone basis.

3.4 Integration and Handover (3–4 days)

The final stage in the project, this completes the system through the collation of documentation suitable for further development work or system integration on the client's network. The codebase, along with such documentation, is passed to the client for future use. The client reports on their satisfaction with the work completed, and officially signs off on the project if it is appropriate to do so.

4 Milestones

In order to track project progress and to be accountable to the client, it will be necessary to identify several key points throughout the project. These milestones are derived from reasonable estimates as to the time required to design, implement and test certain modules, including the ancillary work required for a project of this scale, such as documentation.

- 12th February, 2013: The group are expected to have completed their discussions around how best to implement the project, and be preparing to move forwards towards an implementation. Deliverables: Requirements Specification, Project Plan.
- 14th February, 2013: Completion of the system scoping and design will move the group forwards into beginning to think about project implementation. A forthcoming client meeting requires the group to report on its implementation progress to the client. Deliverables: Implementation Progress Report, Integration Progress Report.
- 15th February, 2013, 2PM: Client Meeting. Deliverables: Prototype code suitable for private demonstration, demonstrating the functional components of the system thus far.
- 28th February, 2013: Project completion. All systems implemented to a suitable standard and in accordance with the agreed design documentation. Deliverables: Group Report, Personal Reports.
- 4th March, 2013: Completion of the codebase for submission to the Computer Laboratory. Deliverables: Final codebase, Full documentation.

5 Resources

A significant number of resources will be produced during the course of this project. It will be necessary to store those resources in locations suitable for access by the entire group. The appropriate location will be dictated by the nature of the resource and the number of changes which must be accommodated.

Code will be subject to version control using the *git* version control system. This is a distributed version control system, so a copy is stored on the machines of every group member who initiates a checkout of the repository. This provides a primitive level of backup since every group member holds a complete copy of the data (including the entire history of the codebase). It is unlikely that the machines of multiple group members will be damaged, lost or rendered otherwise inaccessible during the course of the project. To allow for central sharing of repositories, a private repository on the *GitHub* website will provide the authoritative central store for the code repository, another level of backup and additional functionality, such as a wiki, for central collaboration.

Documentation will be stored in two locations. Draft documentation and internal meeting documentation, such as design documents and meeting minutes, will be stored in a central folder on *Google Drive*. This facility provides convenient tools for multiple group members to collaborate simultaneously on documents and spreadsheets. The team is relying on Google's knowledge of distributed and clustered computing to protect the data, and working on the assumption that the company is of a sufficient size and level of technical knowhow to fend off attacks or

attempts to harm or remove our data from its infrastructure. Nevertheless, the data stored on Google Drive will be purely incidental data, particularly for formal reports, which will eventually be typeset and shared with all group members through the git repository for safekeeping.

Computing resources will be required to perform development work and to execute the code on the system. The use of the Hadoop framework and other system services eliminates the use of the University's Managed Cluster Service (MCS) for this purpose, as none of the group members are permitted to install or start system services on privileged ports on those machines. It is for this reason that group members must use their personal machines to run the Cloudera CDH4 virtual image, which contains the correct version of Hadoop upon which the project will be built. In addition, the size of the datasets (in excess of several gigabytes) will preclude the group's use of the 100 MB of provided storage on the MCS. The group noted that on the code handover day, the Computer Laboratory intends to freeze and clone the workspace on the MCS. The group will endeavour to investigate closer to the handover time how best to transfer both code and documentation, which may consist of a single git clone into the group project space on the MCS, or arrange for some other method of handover.

In addition to this, the services of a central virtual private server (VPS) have been made available to the group. Although this will not be capable of performing heavy-lifting of any data - a task which must be confined to individual machines - it will act as a central repository for data which has been aggregated and anonymised. This permits group collaboration and enforces a single instance of a database among all members of the group, without exposing any of the raw data beyond the group's personal machines and the University's data network.

6 Risk

The group are required to keep confidential any sample data provided to them by the client, and preparations have been made to ensure it is stored securely, in an encrypted form where appropriate. The risk of such data being acquired has been minimised.

A considerable period of time has been spent researching and developing the designs for the project. This is largely due to the very open brief which the group was presented with. This was much appreciated, as it allowed the scope of the work to be defined by the group, but it has taken a considerable period of time to narrow that scope to a level appropriate to the project and the intentions of the client. The group is aware that good planning is never a bad idea, and that past experience means time spent in planning and design will often lead to fewer issues during implementation and the subsequent phases. However, the group is also cognisant of the delay which its design has had on the progression of the project. It should be noted that a milestone relating to implementation progress at the next client meeting is a matter of days from the date of this report, and that as such there is a risk the implementation may not have progressed as far as the client would have hoped by that time. However, the internal management of the project, completion of the designs, and good skills with many of the third-party systems leads the group to believe this is not a significant cause for concern.