Dissecting Server-Discovery Traffic Patterns Generated By Multiplayer First Person Shooter Games

Sebastian Zander
Centre for Advanced Internet Architectures
Swinburne University of Technology
Melbourne, Australia
+61 3 9214 4835
szander@swin.edu.au

David Kennedy
Centre for Advanced Internet Architectures
Swinburne University of Technology
Melbourne, Australia
+61 3 9214 4846
dkennedy@swin.edu.au

Grenville Armitage
Centre for Advanced Internet Architectures
Swinburne University of Technology
Melbourne, Australia
+61 3 9214 8373
garmitage@swin.edu.au

ABSTRACT

We study the 'background traffic' resulting from tens of thousands of networked first person shooter (FPS) clients searching for servers on which to play. Networked, multiplayer games utilise the network in two distinct ways. Game play is typically built around a client-server communication model, and the resulting traffic patterns have been well studied to date. However, the discovery of available game servers is itself a client-server process. Operational game servers register themselves with well-known 'master servers', which are then queried by game clients looking for available servers. Game clients then probe the servers and retrieve information such as game type, number of other players, currently active map, and latency (ping time). We instrumented two active and public "Wolfenstein: Enemy Territory" servers over 20 weeks, developed a simple method to differentiate client probes from game-play traffic, and then characterized and contrasted the time-of-day, geographical distributions and traffic characteristics of both traffic types. We find that a significant amount of a server's traffic is probe traffic and the geographical origins are very different for both types of traffic. We propose techniques to improve server location and to decrease the amount of probe traffic.

Categories and Subject Descriptors

C.2.3 [**Computer-Communication Networks**]: Network Operations - *Network monitoring*;
C.4 [**Performance of Systems**]: Measurement Techniques

NetGames'05, October 10–11, 2005, Hawthorne, New York, USA.

General Terms

Measurement.

Keywords

Game Traffic Analysis, Geographic Distribution, Server Location.

1. INTRODUCTION

The growth in popularity of multiplayer, networked first person shooter (FPS) games [1] is in part due to the ease with which anyone can set up a game server on their own Internet connection. Tens of thousands of game servers are active on the Internet at any given time. Initially game servers would be located at university sites, but in recent years there has been a significant rise of Internet Service Providers (ISPs) running their own dedicated game servers, and personal game servers being hosted over consumer (home) broadband connections.

Networked FPS games utilise the network in two distinct ways. Game play is typically built around a client-server communication model [2], and the resulting traffic patterns have been well studied to date. However, the discovery of available game servers is itself a client-server process. Operational game servers register themselves with well-known 'master servers', which are then queried by game clients looking for available servers. (Master-server identities are pre-configured into each game client.) The master-server returns a list of currently available servers. Game clients then probe each of the servers retrieving information such as game type, number of other players, currently active map, and latency (ping time) from the client to each game server. The players use this information to select the server on which to play.

A game client can end up probing hundreds or even thousands of remote servers before the player makes a choice. In addition, 3rd-party stand-alone tools (such as qstat [3] and the original gamespy [4]) or websites (e.g. gamespy.com [4])) also perform server probing on behalf of potential players or to create regularly updated server statistics for public consumption.

We have chosen to measure the level of traffic that impacts on a public game server as a consequence of tens of thousands of game clients probing it from around the Internet (probe traffic), and contrast this with the traffic resulting from people actually playing on the server (non-probe traffic). Although non-probe traffic has been well studied, we believe there exists no other work comprehensively characterising the probe traffic experienced by a typical game server.

This analysis has two outcomes. The insight will help server operators estimate the network capacity that will be consumed by hosting a game server, regardless of whether the server itself is popular. In addition, since probe traffic reveals the IP address of clients an operator can analyse probe traffic trends to discover client usage patterns across regions of the Internet.

First we develop a simple method to differentiate probe and non-probe traffic, and then gather raw data over 20 weeks from two separate (but identically configured) "Wolfenstein: Enemy Territory" [5] servers. We characterise the total traffic of both types, and then describe and contrast the traffic patterns as a function of time (over various periods) and geographical distribution.

We make two observations about the aggregate traffic – probe traffic is a significant percentage of the overall traffic volume (easily over 6%-7% for an active server and in the region of 35% for a poorly utilised server), and the actual probe traffic load is relatively independent of a server's popularity. (We surmise this latter fact is due to probes being generated regardless of the server's potential relevance to a client, even though people tend to only play on servers they 'like'.) Although the volume of probe flows is fairly small for active servers the number of probe flows is very large. This can have a significant impact on routers or software that keeps per-flow information. For both our servers 99.9% of the observed flows were probe flows.

Finally, we propose that master servers optimise the list of game server addresses passed back to clients. On the theory that a player will be more likely to choose servers having geographical or topological locality to the client, Internet-wide probe traffic could be reduced if master servers returned the list of game-servers in order of likely locality to the client.

The rest of the paper is organized as follows. Section 2 presents an overview about related work. Section 3 describes how we collected our data set and Section 4 describes how we identify probe and non-probe traffic. Section 5 contains the results of the data analysis. Section 6 proposes improvements to the server location algorithm to decrease the amount of probe traffic and the time users need to find a suitable server. Section 7 concludes and outlines future work.

2. RELATED WORK

A number of papers exist on the modelling of game traffic, almost entirely focussing on what we term non-probe traffic. Early work in [6] presented a traffic model for Quake 1 and Quake 2. A traffic model for the newer Quake 3 is proposed in [7]. The network traffic and server workload of the game Half-Life is characterised in [8], [9] and [10]. The authors of [11] present a traffic model for the Xbox game Halo 1. We have developed a traffic model for the Xbox game Halo 2 [12]. The authors of [13] have developed models for the games Quake 2, Grand Prix 3, Ages of Empires II and Panzer General 3D.

The authors of [14] explore the current geographic distribution of a global set of servers for several popular on-line games as well as the geographic distribution of a set of players for a Half-Life: Counter-Strike server. The results quantify the break-up of current game servers across continents and show that players do not necessarily play on servers that are geographically close. Related work in [15] presents a redirection service for on-line games based on the geographic location of players relative to servers.

3. DATA COLLECTION

Our study is based on traffic statistics collected over a period of 140 days (20 weeks) on two public servers connected to the Internet at two different locations. Both servers were running identical versions of the team-based FPS game "Wolfenstein: Enemy Territory" [5] (which we refer to as "ET").

One of these servers was located on a university campus in Melbourne, Australia ("CAIA" server). The other server was located in Canberra, Australia, on the backbone of GrangeNet's high performance research network [16] ("GrangeNet" server). All times and dates in this paper are centred on the Australian east coast time zone where these servers reside.

Both of the servers had a 100Mbit link to the Internet, the bottleneck being the Ethernet links at the servers. Both of these servers were also configured identically, (except for their names and the messages of the day). No more than 20 players could join at any one time, and the map rotation included all 6 standard maps. Both servers ran a popular 'mod' (game software modification) called ETPro 3.1.0 [17] (so popular that there are more ETPro servers online than those without). This ensured our server was as up-to-date as possible to attract player interest.

We ran NetMate [18] on each server to capture all of the traffic on the server, and then analyse and log the characteristics of each observed traffic flow. NetMate classifies packets into bidirectional flows using source IP address, destination IP address, source port and destination port. Statistics that NetMate captures for each flow include end time, duration, number of packets and bytes, packet length, and inter-arrival times. All statistics aside from duration are computed separately for both directions of the flow. NetMate includes the link layer headers and trailers in all of its measurements. This means all packet length and byte volume statistics presented include Ethernet headers and trailers.

NetMate terminates an existing flow if no further packets are observed during a preconfigured timeout. 60 seconds was chosen as the flow timeout, so a flow would remain 'open' if no more than 60 seconds passed between packets. This is because it was discovered in a preliminary study that some longer-term probe flows have approximately 30-second packet inter-arrival times (see Section 5.4). Consequently, if a source probed the server once every 30 seconds for 24 hours NetMate would report a single 24 hour-long flow for this source.

We configured NetMate to stop measuring and rollover the current log file each morning at 4am. This means that any flows still active at that time would be chopped into two parts (although, as we note in section 5.2, we did not experience any long-term flows active at this time).

4. PROBE TRAFFIC IDENTIFICATION

We used two simple thresholds to differentiate between probe and non-probe traffic. First, a flow is classified as a probe if less than 20 packets were sent from server to client. Second, a flow is also classified as a probe if the average inter-packet interval of server to client traffic is larger than 500ms, which is an order of magnitude higher than the inter-packet interval during regular game play and smaller than probe flows that have inter-arrival times of only a few seconds (the probe and non-probe traffic characteristics are discussed in section 5.4).

To verify our classification of the flows, we analysed the log files produced by the ET server. These files contain a record of all the clients that actually played on the server. From the log file we extracted a list of game sessions with the following information for each: client IP address and port number, connect time, disconnect time and duration.

When analysing the log files we combined sessions of the same IP address and port if there was less than 60s time between them because these sessions would end up in a single NetMate flow due to the 60s timeout.

We compared the game sessions reported by the server log files to the flows reported by NetMate. The start and end timestamps in the two data sets are always different due the NetMate data including extra traffic outside the actual game play time. For each entry in the game session list we searched the flow list for a flow with the same IP address and port and the 'closest' start and end times. To find the closest timestamps we compute the sum of the magnitudes of both start and end timestamp differences and map the flow that has the minimum difference to the game session entry. Finally, we compared the flows selected via the server log with the flows selected by our simple threshold approach.

We found that ~7% of flows appearing in the server log were incorrectly classified as probe flows but their volume is only ~0.8% of the total volume of non-probe flows from the server log. On the other hand we found that ~7% of flows our simple heuristic classifies as non-probe flows do not appear in the server log. These have a volume of ~0.3% of the total volume of non-probe flows given by the server log.

There are a number of reasons for these errors. Most of the flows in the server log that are 'misclassified' as probes by our approach are very short. The median is ~1 minute and 90% of them are shorter than 5 minutes and some of them have very high inter-arrival times. We believe that these flows are connection attempts (players connecting and disconnecting shortly afterwards) and therefore could be seen as kind of probes. Most of the flows that have no corresponding entry in the server log file are also short (95% are less shorter than 5 minutes) and we can only speculate about their nature.

Strangely enough, we also discovered some 'errors' in the server log. Some client-disconnect entries were missing which lead to very long session times. Because we found only very few of these problems we manually corrected them in the dataset.

5. ANALYSIS

Our analysis excludes probe and non-probe traffic generated locally by students playing on the university (CAIA) server and test probes initiated during development of this paper. In-house players accounted for up to 0.25% of all flows and 20% of all bytes per day on the CAIA server (but the 90 percentiles are only 0.08% and 12% respectively).

5.1 Overall Statistics

Table 1 shows the overall volume of our dataset. Both servers saw almost equal numbers of flows, but differed significantly in the number of packets and bytes transferred because the CAIA server was far more popular with players.

Table 1: Overall statistics for both servers

	CAIA	GrangeNet
MFlows	16.18	16.93
MPackets	791.59	147.68
GBytes	124.76	22.66

Table 2 breaks down the traffic into probe and non-probe categories. Here we can see that the vast majority of flows for both servers are probe flows (over 16 million in each case) whereas the vast majority of packets and bytes come from non-probe flows. Non-probe traffic originates from actual player activity, and is thus proportional to server popularity. Probe traffic is independent of server popularity, which explains why both servers saw very similar probe traffic statistics despite differing greatly in their non-probe statistics.

Table 2: Total amount and percentages of probe traffic vs. non-probe traffic

	CAIA		GrangeNet	
	Probe	Non-Probe	Probe	Non-probe
Flows	16.18e6 (99.95%)	7993 (0.05%)	16.93e6 (99.99%)	1757 (0.01%)
Mpackets	36.46 (4.61%)	755.13 (95.39%)	36.94 (25.01%)	110.74 (74.99%)
GBytes	8.18 (6.56%)	116.58 (93.44%)	8.10 (35.75%)	14.56 (64.25%)

Figure 1 shows the number of probe and Figure 2 the number of non-probe flows measured per day for each server (note that the y-axis scales differ by several magnitudes). The number of probe flows is fairly constant and very similar for both servers over the measurement period. The GrangeNet server shows two noticeable drops caused by power outages around day 40 and day 100.

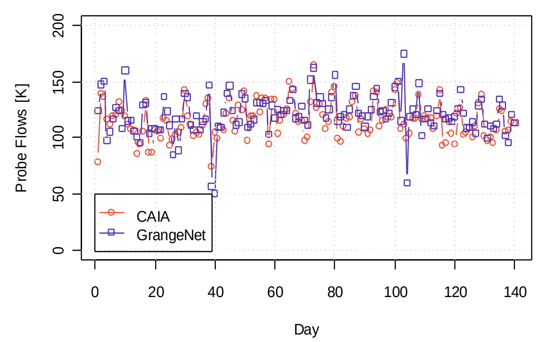

Figure 1: Number of probe flows per day

The number of non-probe flows per day changes dramatically from day to day for both servers, and it is clear that the CAIA server attracted many more players than the GrangeNet server.

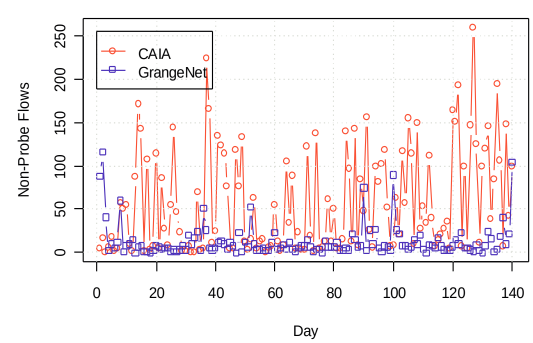

Figure 2: Number of non-probe flows per day

The volume per-day graphs are very similar to the previous two figures, with probe traffic volume fairly consistent over time (around 0.06 Gbyte a day) and non-probe traffic volume fluctuating in response to player activity (between 0 and 5 Gbyte a day). Because the trends are identical the graphs are not included in the paper for space reason.

As noted earlier, there is a substantially higher number of probe flows than non-probe flows, but the probe flows count for a much lower level of traffic volume. The percentage of probe volume per day changes between 0% and close to 100% depending on the utilisation of the server.

5.2 Weekly and Daily Traffic Patterns

We also analysed the weekly traffic patterns for both servers. Figure 3 and Figure 4 show the number of probe and non-probe flows averaged for each day of the week for both servers. The probe traffic curves show identical behaviour (the value is slightly higher for the GrangeNet server). The amount of probe traffic seems to be almost constant during the working days and is increasing to its maximum on the weekend. We believe this is because more people run their game clients, and thus trigger more probe traffic, on the weekends.

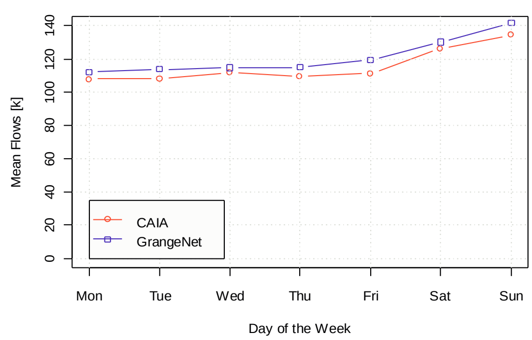

Figure 3: Weekly pattern of number of probe flows

The non-probe traffic on the CAIA server is increasing from Monday to Friday and then drops a bit on Saturday and increases again on Sunday. We think local players playing on workdays after work cause this behaviour. Although the local traffic was excluded from our analysis, these players on the server attracted other players. The GrangeNet server shows a similar behaviour on a much lower level with the number of flows increasing from Monday to Friday. But the maximum for the GrangeNet server is clearly on the weekend. This also shows that without regular in-house players, the server is less attractive for other people to join.

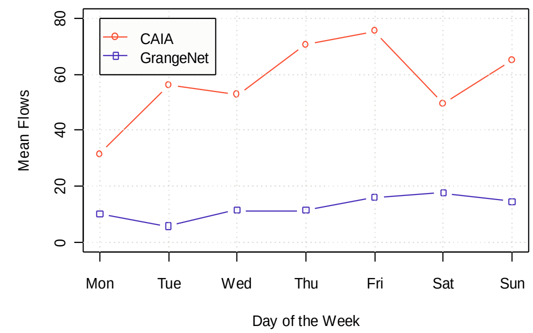

Figure 4: Weekly pattern of number of non-probe flows

The curves for weekly traffic volume are very similar to Figure 3 and Figure 4 and therefore not included. The only notably difference is that the CAIA server's probe volume equals or slightly exceeds that of the GrangeNet server (despite having slightly less probes per day). We believe this is caused by CAIA server's popularity – when the server is busy there's more information returned in (longer) probe response packets, and more information is requested by potential players (by default, the ET client only requests brief information from the server, but players can trigger a request for more detailed information).

Next, we analyse the number of flows and the volume depending on the hour of the day. A problem with this analysis is that flows – especially non-probe flows – can easily span multiple hours. To provide an accurate estimation of the number of flows and the volume per hour, we instrumented NetMate to output interim flow records each hour. When the number of flows per hour is used in the following analysis, it means the number of active flows in that hour. Each flow can be counted in more than one hour. This is slightly different from the previous analysis where flows were only counted once. This affects mainly the non-probe flows because most probe flows are usually very short.

Figure 5 and Figure 6 show the average number of probe and non-probe flows. The probe traffic pattern is identical for both servers, and provides the first evidence of uneven client distribution around the planet. As we see in the next section, the peak around 5am is largely due to probes originating from Europe (locations approximately 8 to 10 hours behind our local time zone).

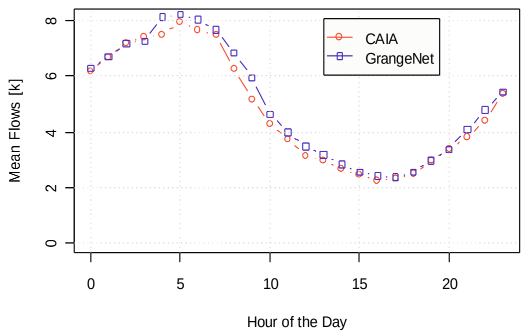

Figure 5: Number of probe flows per hour of day

Players in our local time zone dominate the non-probe traffic, with playing activity starting in the early afternoon and continuing until 2am to 3am.

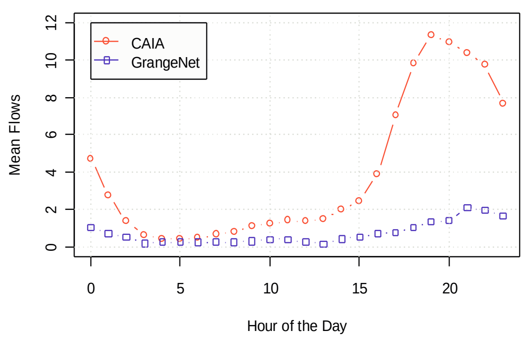

Figure 6: Hourly number of non-probe flows

The volume patterns are very similar to the previous graphs and not shown. The probe volume of the CAIA server is slightly higher in the afternoon/evening. We believe there are two reasons for this: some failed client connection attempts are misclassified as probes by our heuristic, and probe response packets are larger when the server is populated.

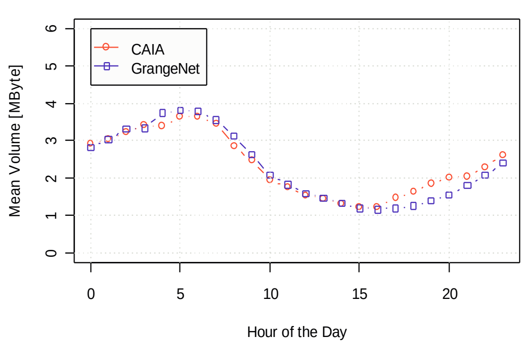

Figure 7: Mean hourly probe volume during workdays

Figure 7 and Figure 8 compare the mean probe volume during the working days (Monday to Friday) and on the weekend. The shape of the pattern looks very similar. The maximum is roughly the same in both graphs but the number of probe flows is slightly higher on weekends.

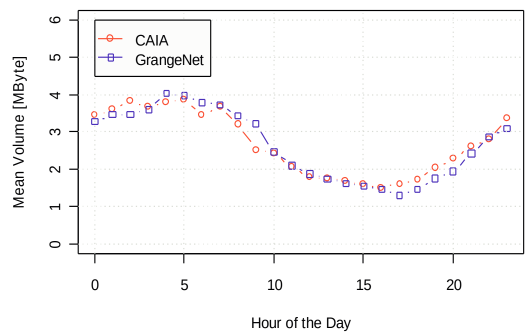

Figure 8: Mean hourly probe volume on the weekend

Figure 9 and Figure 10 and show the mean volume of non-probe flows on workdays and weekends. As one would expect there is a distinct difference between both patterns. During the week playing is clearly limited to the time between 5pm and 1am.

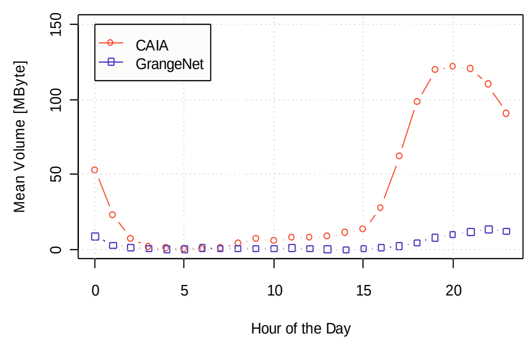

Figure 9: Mean hourly non-probe volume during workdays

On the weekend playing activity is more spread across the whole day starting at lunchtime and ending around 2am. Still the most activity is limited to the evening.

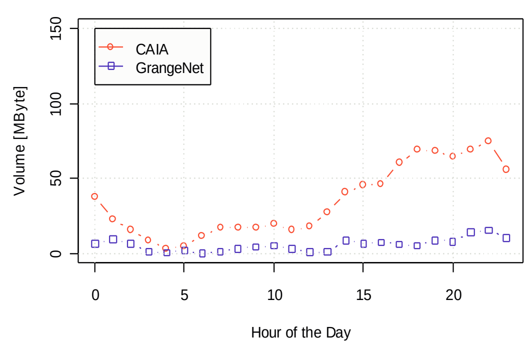

Figure 10: Average non-probe volume per hour on the weekend

5.3 Geographical Analysis

We also analyse the geographical origins of the traffic. We use a freely available geo-location database to identify which country is associated with a particular IP address (we use the database of May 2005, which has a claimed 97% coverage) [19]. It was not possible to identify the country for each IP address. However, ultimately only for 0.05% of flows and 0.04% of bytes we could not identify the country of origin.

Because flows come from a number of countries (up to 80-90 per day) Table 3 groups countries into a smaller number of regions (similar to the regions used in [14]). We subsequently ignore Africa as it contributes only 0.03% of flows and 0.002% of bytes.

Table 3: Regions and corresponding countries

Region	Countries
Europe (EU)	European countries
Australia (AU)	Australia and New Zealand
North America (NA)	USA and Canada
South America (SA)	All middle and south American countries
Asia (AS)	All countries from Asia and the Pacific region
Africa (AF)	African countries

Figure 11 and Figure 12 show what percentage of flows and bytes for probe and non-probe traffic at both servers comes from which region. Interestingly, although the utilisation was quite different, the difference between the traffic origins of the two servers is fairly small.

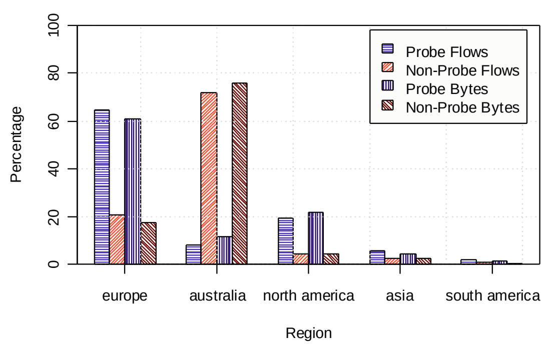

Figure 11: Percentage of flows and bytes for probe and non-probe traffic for CAIA

The majority of probe flows originate in Europe (about 60%) followed by North America (20%) and Australia (10%). Not surprisingly, most non-probe flows come from Australia (70-80%) because most players tend to play on 'local' servers (having lowest latency). Further demonstrating this is the fact that although there are less non-probe flows from Asia, we can see that players from this continent are more likely to join our server once they have probed it than players from Europe and North America.

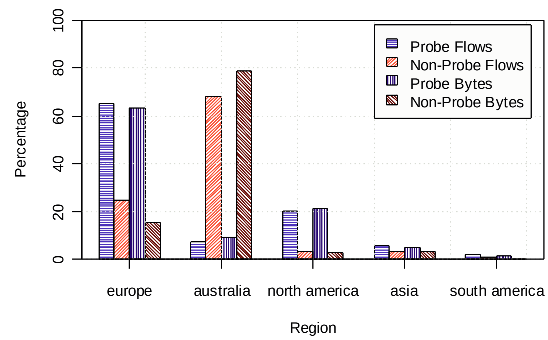

Figure 12: Percentage of flows and bytes for probe and non-probe traffic for GrangeNet

Figure 13 shows the amount of probe flows over time for both servers. It shows that the average amount of probe flows was almost constant for each region and roughly equal at both servers for the duration of the data collection. A similar figure for the non-probe traffic shows that the amount of non-probe traffic heavily fluctuates over the measurement period and we cannot find any region-specific trends. Because this figure is hard to read we have omitted it.

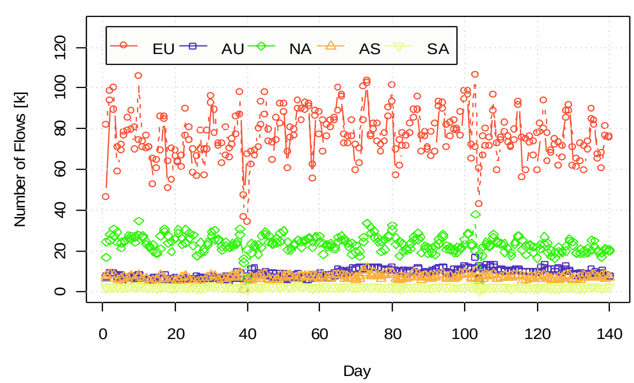

Figure 13: Probe flows from the different regions for CAIA (solid lines) and GrangeNet (dashed lines)

Figure 15 shows the amount of probe and non-probe flows per weekday, broken down by region and server (CAIA server in solid lines, GrangeNet server in dashed lines). For probe traffic, the trends are similar for all regions (fairly stable during workdays and increasing on the weekend).

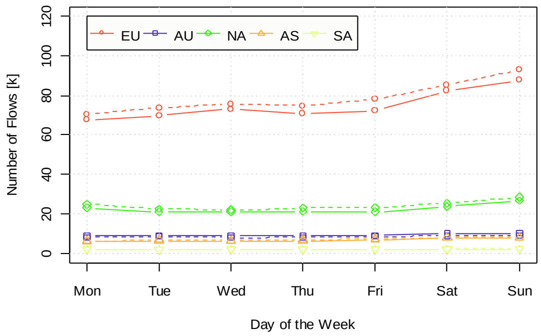

Figure 14: Number of probe flows for different regions over weekdays

The non-probe flows graph shows that the patterns for both servers are dominated by traffic from Australia.

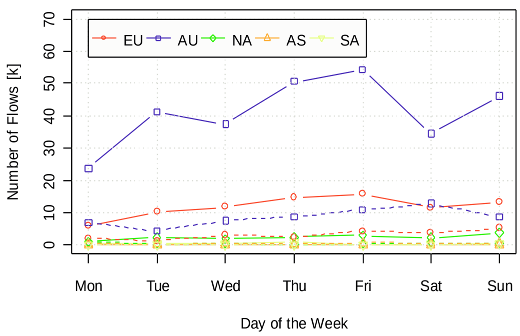

Figure 15: Number of non-probe flows for different regions over weekdays

The volume of the probe and non-probe traffic for the different regions follows the same patterns (figures excluded for space reasons). One difference that was observed before is that the size of the probe flows is slightly larger at the CAIA server.

Although the daily and intra-week plots show broad regional differences, even more interesting insights come from intra-day breakdown of probe traffic by region. Figure 16 and Figure 17 show the number of probe and non-probe flows for each region per hour of the day (solid line for CAIA and dashed line for GrangeNet).

It becomes clear that the overall sinewave-like shape of probe traffic in Figure 5 is mainly due to European probes, which peak around 5am. North America provides a noticeable (but much smaller) contribution a few hours later.

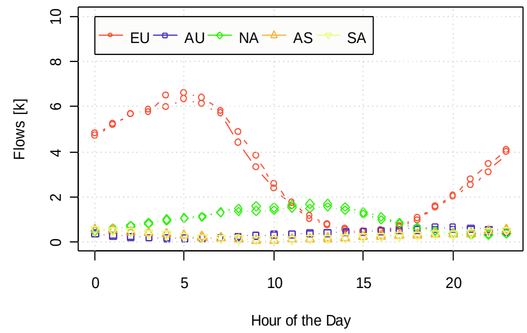

Figure 16: Number of flows for probe traffic depending on hour of the day and region

The non-probe traffic is dominated by traffic from clients in Australia. Later in the evening (after 7pm) there is a noticeable bump in non-probe traffic from Europe (which we suspect is due to European game clients beginning to come online for their day and noticing the CAIA server active with local players).

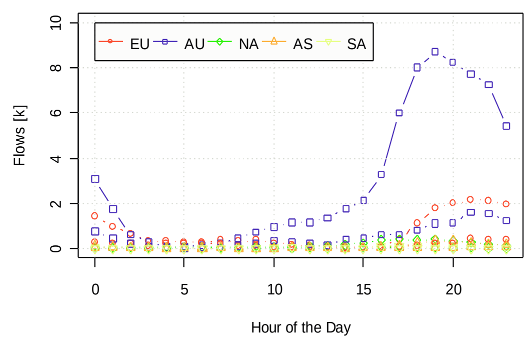

Figure 17: Number of flows for non-probe traffic depending on hour of the day and region

The same graphs for the volume of probe and non-probe flows are very similar and therefore not shown. One notable difference is that the larger probe flows observed at the CAIA server all came from Australian clients.

Figure 16 suggests we can extract interesting demographic information about the distribution of any particular FPS client around the planet by analysing the hourly trends of probes versus their source IP addresses. This technique does not require players to actually play on the server, and is thus independent of the server's attractiveness to players (whether due to topological/geographical or other considerations). In this case we are able to infer that ET is quite popular in Europe even though our servers did not see many actual players from outside Australia.

5.4 Probe and Non-probe Flows Traffic Characteristics

In this section we analyse probe and non-probe flows according to their duration, size in bytes, mean packet inter-arrival times and mean packet lengths. Aside from duration, we compare both directions of the flows (client-to-server (c->s) and server-to-client (s->c)). Please note that non-probe flows have been studied in far greater detail before (see Section 2) but we provide their statistics here to allow a comparison between the characteristics of probe and non-probe flows.

Because the number of probe flows is very large (more than 32 million) and many of the very short probe flows have the same characteristics, we used stratified random sampling to reduce the number of flows in the analysis. We select all probe flows with >4 packets and sample 5% of the flows which have ≤4 packets. When computing the empirical density functions, we weight probe flows with ≤4 packets 20 times as much as those have more than 4 packets. Our analysis is based on a total of 9,750 non-probe flows and 2,151,509 probe flows for both servers.

Figure 18 and Figure 19 show the cumulative density functions (CDFs) for the duration of probe flows and non-probe flows. We use a different scaling for the y-axis for probe flows to better show the distribution of longer probe flows.

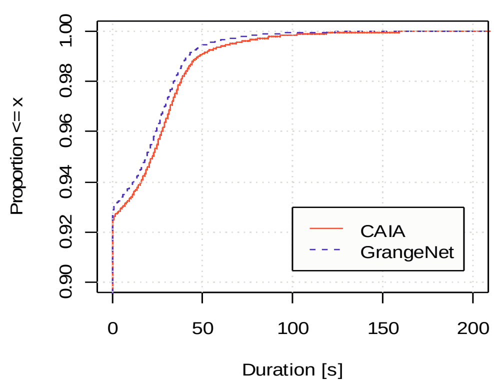

Figure 18: CDFs for the duration of probe flows

Over 92% of probe flows are shorter than one second. There are some longer probe flows but most of them are still very short (<90s) compared to non-probe flows. Very few probe flows have long durations (1-2 hours) probing the server in very regular intervals (30s). These flows come from hosts whose purpose seems to be the scanning of game servers e.g. a number of them came from the following two hosts: scan1.ase.games.re2.yahoo.com and scan2.ase.games.re2.yahoo.com.

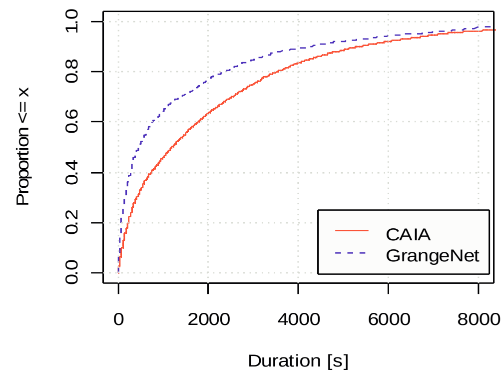

Figure 19: CDFs for the duration of non-probe flows

Non-probe flows are longer on the CAIA server because it is more popular and people stay longer. Around 40% of the CAIA flows are longer than 30 minutes (which is confirmed by the server log data).

Figure 20 and Figure 21 show the CDFs for the mean packet size in both directions for probe and non-probe traffic. Probe traffic mostly uses few characteristic mean packet sizes. All packets sent by the clients are 60 bytes long. Packets sent by the server are 60 bytes, 320-330 bytes (the majority) or 800-810 bytes long. The 320-330 byte packets are responses to standard polls, whereas the 800-810 byte packets contain more detailed server information that the client has requested. The difference between the two servers is very small.

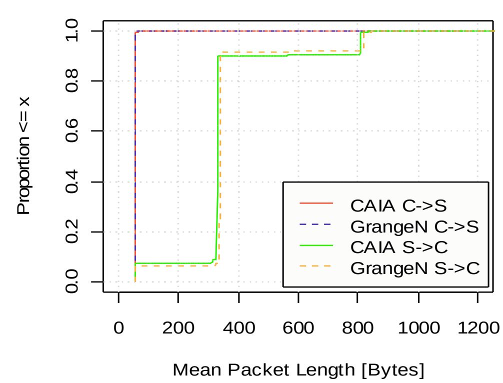

Figure 20: CDFs for the mean packet size of probe flows

Non-probe traffic has many different packet sizes. The client-to-server distributions are equal for both servers (99% of mean packet sizes between 60 and 160 bytes but 1% is larger). The server-to-client mean packet size distributions are different for both servers because it is proportional to the total number of players. Because the CAIA server usually has more players, it has larger packet sizes. This effect only occurs for in-game traffic – mean packet size smaller than 400 bytes – but we it seems we include other traffic in the non-probe traffic (e.g. mod downloads), which uses large packet sizes.

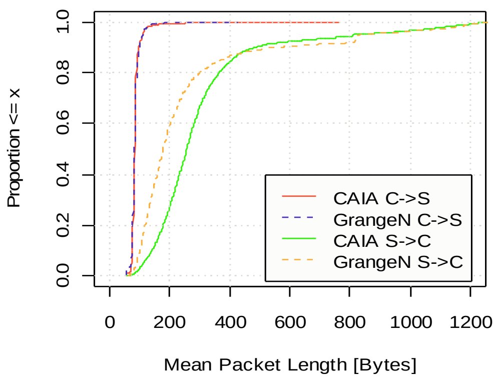

Figure 21: CDFs for the mean packet size of non-probe flows

Figure 22 and Figure 23 show the CDFs for the mean inter-arrival times in both directions (client to server and server to client). Over 92% of the probe traffic has no mean inter-arrival time because these flows consist of only two packets (one in each direction). Less than 0.5% of the flows have inter-arrival times in the sub second range and the rest of the inter-arrival times range between 1 and 50 seconds. For both servers the distributions in each direction are exactly the same because the servers almost immediately answer client requests. The high inter-arrival times are slightly larger for the CAIA server.

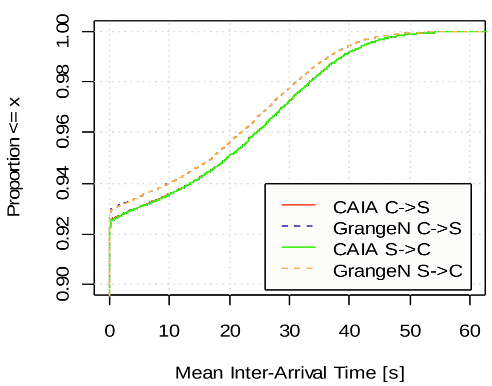

Figure 22: CDFs for the mean inter-arrival time of probe flows

The distributions for non-probe traffic are all similar. One notable difference between traffic sent by the client and traffic sent by the server is that the server will never send more than 20 packets/s (50ms inter-arrival time) whereas the client send rate depends on the frame rate, which in turn depends on the CPU and graphics card performance of the client. The majority (80%) of the inter-arrival times are less than 100ms. Surprisingly there are some high values. These could have been caused e.g. by congestion, flows that were not in-game traffic or slow clients.

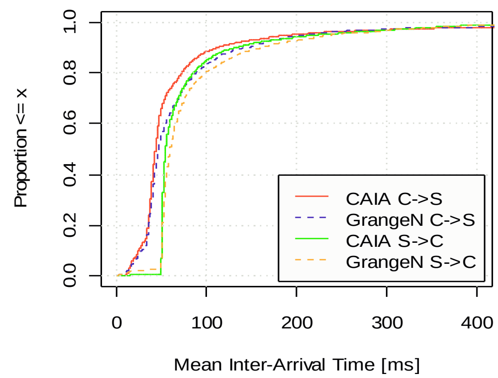

Figure 23: CDFs for the mean inter-arrival time of non-probe flows

Figure 24 and Figure 25 show the CDFs of the byte volume transmitted in both directions. The volume of the probe flows is small. The client to server probe traffic is mostly one 60 byte packet. The volume of the server responses reflects the same few characteristic packet lengths as shown earlier. The difference between the two servers is very small.

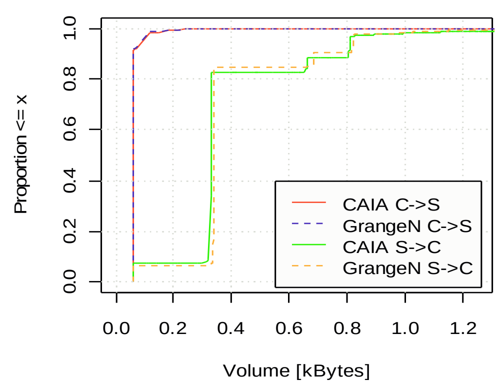

Figure 24: CDFs for the probe flow volume

The volume of the non-probe traffic depends on the server. Server-to-client flows are larger than client-to-server flows for both servers. Flows observed at the

CAIA server are larger because more people were playing longer due to its higher popularity.

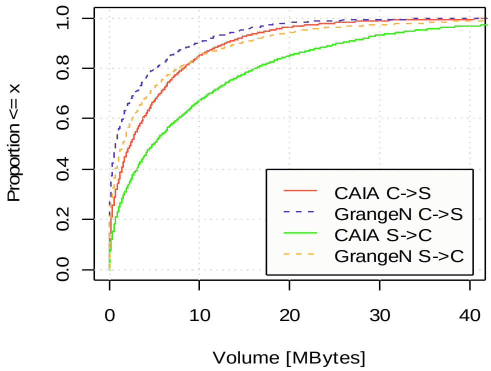

Figure 25: CDFs for the non-probe flow volume

6. IMPROVED SERVER LOCATION ALGORITHM

Probe traffic can clearly constitute a substantial fraction of the actual bandwidth consumption for an infrequently played server, and depends on the size of the potential player population around the planet rather than the number of people who actually decide to play on the server. In our case European clients, who generally do not end up playing on our servers, dominate as source of probe traffic.

The distribution of probe traffic is influenced by the order in which the master server sends the game server IP addresses to clients. Players will typically watch as their client probes all game servers in the order returned by the master server, and then stop the search process when enough 'close' or 'interesting' servers have been located. (Although we do not know the exact algorithm used by the ET master server, our experience suggests it is an un-ordered list, or perhaps ordered by time of game server registration.)

We propose a server-side enhancement to the address distribution algorithm: the master server should sort the advertised list of game servers in order of likely locality (or closeness) to the querying client. This modification would not require any client-side changes.

It is not essential that locality/closeness is calculated with great precision. The closeness of a client to particular game servers can be estimated using IP address-based geographic location tools (e.g. [19]), route table databases (to infer topological locality), or even certain latency estimation techniques (e.g [20]).

If the database used for the IP address-based location provides latitude/longitude or other coordinates, a distance can be easily computed. If only country information is available, coordinates can be defined for each country making a rough estimation about its epicentre. (Or alternatively, a table with all distances between countries pre-calculated using whatever metric seems appropriate.) Latency estimation techniques would theoretically create a more useful notion of closeness (since latency plays a huge part in game satisfaction). However, schemes such as proposed in [20] would increase network traffic load in return for only questionable accuracy.

Our suggested enhancement would have two positive consequences – it would decrease the overall probe traffic load on servers from clients who will never play anyway, and increase the chances that a player will see an interesting or playable server early. We believe a solution based on country location and a distance lookup table can be easily implemented and would provide sufficient accuracy.

A related problem occurs when game servers use redirect mechanisms. Most current game servers are able to redirect players to a different server when the server is full. Players should be redirected to a server that not only has similar game characteristics but also is close to the player. This problem was discussed in [15] and has nothing to do with optimizing the master server algorithm, but it is important to not 'compromise' an improved master server algorithm by local redirects that lead players to distant servers.

7. CONCLUSIONS AND FUTURE WORK

We have studied the 'background traffic' resulting from ten thousands of clients searching for servers on which to play, based on data collected over 20 weeks at two "Wolfenstein: Enemy Territory" servers. We developed a simple method to differentiate client probes from game-play traffic, and then characterized and contrasted the time-of-day and geographical distributions of both traffic types. We find that:

- The amount of probe traffic is independent of server popularity and can represent a significant fraction of total traffic in and out of a game server (in our case, 6-7% of the volume for a fairly popular server and up to 35% of the volume for an unpopular server).
- The number of probe flows is very high. Even for a fairly popular server 99.9% of the flows were

probe flows, which can have a significant impact on routers/software that keeps per-flow state.

- The geographic origins of probe and non-probe traffic differ greatly – probes come from all over the Internet whereas players originate from locations topologically (and usually geographically) closer to the server.
- The density of probe traffic as a function of time from around the Internet can reveal where communities of interest lie for particular game clients. It is not required to run a popular or well-located server to attract enough probe traffic and extract this information.

We propose a simple server-side method to improve the server location discovery process. When returning a list of available game servers to a querying client, master servers should sort the returned list of game servers in order of probable closeness between each server and the client. As a result, players will find that their client discovers and reports usable servers more quickly than is currently the case, and servers will see a reduction in probe traffic from clients who are too far away to choose the server for play anyway.

In the future we plan to perform a similar study for a more recent game (e.g. Half-Life 2) and estimate the benefits of the server-side sorting algorithm proposed above (based on simulations using collected real-world data). Finally, we plan to evaluate the feasibility of implementing a proxy for master-servers that transparently sorts the IP addresses returned to a querying game client.

8. ACKNOWLEDGMENTS

We thank Chris Meyers for letting us to run a game server in GrangeNet's [16] high-speed network and his continuous support. We also thank the anonymous reviewers for their helpful comments.

9. REFERENCES

[1] S. McCreary and k. claffy, "Trends in wide area IP traffic patterns - A view from Ames Internet Exchange", in ITC Specialist Seminar, Monterey, CA, 18-20 Sep 2000.

[2] E. Cronin, B. Filstrup, A. Kurc, S. Jamin, "An Efficient Synchronization Mechanism for Mirrored Game Architectures", Proceedings of NetGames2002 Workshop, Germany, April 2002.

[3] qstat, http://www.qstat.org/ (as of June 2005)

[4] gamespy, http://www.gamespy.com/ (as of June 2005)

[5] Enemy Territory, http://www.enemy-territory.com (as of June 2005)

[6] M.S. Borella. Source models of network game traffic. Proceedings of networld+interop '99, Las Vegas, NV, May 1999.

[7] T. Lang, P. Branch, G. Armitage. A Synthetic Traffic Model for Quake 3. ACM SIGCHI ACE2004 conference, Singapore, June 2004.

[8] J. Faerber. Network game traffic modelling. ACM NetGames Workshop 2002, Germany, April 2002.

[9] W. Feng, F. Chang, W. Feng, J. Walpole. Provisioning On-line Games: A Traffic Analysis of a Busy Counter-Strike Server. SIGCOMM Internet Measurement Workshop, November 2002.

[10] T. Lang, G. Armitage, P. Branch, H-Y. Choo. A Synthetic Traffic Model for Half Life. Australian Telecommunications Networks & Applications Conference (ATNAC) 2003, Melbourne, Australia, December 2003.

[11] T. Lang, G. Armitage, "A ns-2 model for the System Link game Halo", Australian Telecommunications Networks & Applications Conference (ATNAC) 2003, Melbourne, Australia, December 2003.

[12] S. Zander, G. Armitage. A Traffic Model for the Xbox Game Halo 2. 15th ACM International Workshop on Network and Operating System Support for DIgital Audio and Video (NOSSDAV 2005), Washington (USA), June 2005.

[13] J. Lakkakorpia, A. Heinerb, J. Ruutuc. Measurement and characterization of Internet gaming traffic. Research Seminar on Networking, Helsinki University of Technology, Networking Laboratory, Espoo, Finland, February 2002.

[14] Wu-chang Feng, Wu-chi Feng, "On the Geographic Distribution of On-line Game Servers and Players", in Proc. of NetGames 2003, May 2003.

[15] Chris Chambers, Wu-chang Feng, Wu-chi Feng, Debanjan Saha, "A Geographic Redirection Service for On-line Games", in Proc. ACM Multimedia 2003 (short paper), November 2003.

[16] GrangeNet, http://www.grangenet.net/ (as of June 2005)

[17] ETPro, http://bani.anime.net/etpro/ (as of June 2005)

[18] NetMate, http://sourceforge.net/projects/netmate-meter/ (as of June 2005).

[19] MaxMind GeoIP Free, http://www.maxmind.com/app/geoip_country (as of June 2005)

[20] K. Gummadi and S. Saroiu and S. Gribble, "King: Estimating Latency Between Arbitrary Internet End Hosts", Internet Measurement Workshop, Marseille, France, November 2002.