

Contextual Network Navigation; Situational Awareness for Network Administrators

Cameron C. Gray*
Bangor University, Wales, UK

Panagiotis D. Ritsos†
University of Chester, England, UK

Jonathan C. Roberts‡
Bangor University, Wales, UK

ABSTRACT

One of the goals of network administrators is to identify and block sources of attacks from a network stream. Various tools have been developed to help the administrator identify the IP or subnet to be blocked, however these tend to be non-visual. Having a good perception of the wider network can aid the administrator identify their origin, but while network maps of the Internet can be useful for such endeavors, they are difficult to construct, comprehend and even utilize in an attack, and are often referred to as being “hairballs”. We present a visualization technique that displays pathways back to the attacker; we include all potential routing paths with a best-efforts identification of the commercial relationships involved. These two techniques can potentially highlight common pathways and/or networks to allow faster, more complete resolution to the incident, as well as fragile or incomplete routing pathways to/from a network. They can help administrators re-profile their choice of IP transit suppliers to better serve a target audience.

Index Terms: C.2.1 [Computer-Communication Networks]: Network Architecture and Design—Network Topology; C.2.3 [Computer-Communication Networks]: Network Operations—Network Management; I.6.8 [Computing Methodologies]: Simulation and Modelling—Visual

1 INTRODUCTION

System administrators face many challenges when dealing with network attacks. These challenges include identifying the attack is taking place, working out the best course of action, and finally enacting the solution. In addition, mitigating against future attacks is also important. If administrators can make their systems more robust then attacks could be handled differently.

In fact, if administrators can identify the initial position of the attack then it would be possible to block that traffic. Thinking holistically, we could imagine creating a visualization of the whole Internet, and map any attacks onto it. This visualization could subsequently aid the user to locate weak-spots in the network and make more informed decisions about such attacks. In the late 1990’s Burch et al. created a visualization of the whole Internet [8]; however, while Burch’s work is seminal and an iconic work-of-art, it is especially difficult to understand and would be difficult to create today with the complexity and intricacy of the modern Internet.

Indeed, attempting to map the Internet in terms of devices is simply not a useful task. The vast number of hosts shown will act as noise; masking the wider, more interesting patterns of connectivity between ISPs. Ultimately it is the backbone connectivity that defines the structure of the Internet and not the number of hosts connected at every node. Therefore, in order to map the backbone

we are able to draw on a convenient level of abstraction already present in Internet routing. An Autonomous System (AS) is a network that has enough suitable connections and infrastructure to make its own routing decisions. ASs, by definition, are required to have connections to at least two different ISPs. Consequently, including networks that do not qualify as ASs in the visualization, does not add any further structural information. These networks can be thought of as within their parent network’s sphere of control, just as a connected host would be.

Being able to conceptualize network pathways can be critical in mitigating certain security incidents, such as Distributed Denial of Service (DDoS) attacks. Traditional tools and representations work on the micro-scale, dealing with individual hosts, networks or sub-networks that are either the source or target. The paths in-between are not considered and, from a network planning point of view, this is a big loss of information. The design of the Internet dictates that traffic will attempt to be delivered even if that means that ‘useful’ or ‘desired’ traffic is delayed or dropped.

Quite often researchers have focused their efforts to improve the signal-to-noise ratio of logs or network streams. There are few tools to assist in network planning for multi-homed Internet connectivity. In fact, most tools assist in designing local, LAN, to metro area, MAN, sized networks. Even less that take into account both structure and commercial concerns of Internet bandwidth supply.

In our research, we have been focusing on these pathways. Our hypothesis is that if we can visualize these pathways better, then the user would have another way to locate attacks, make informed decisions about their network, and potentially gain insight on where future attacks may occur. Our main design strategy is to utilize the idea of ‘Contextual Navigation’, where users see the Internet (network pathways) from their viewpoint rather than as a whole system.

Within this paper we present five contributions;

1. three specific use-cases that explicate and motivate the need for Contextual Navigation,
2. development of our visualization design for Contextual Navigation,
3. an implementation of Contextual Navigation, which includes data gathering, cleaning and display interaction to allow visual analytics of the Internet.
4. four case studies utilizing that system over differing types of network.
5. discussion of related issues, including topological changes.

2 RELATED WORK

Quite often, efforts to visualize network pathways are tied to an implicit physical or human geography, resulting in tools and methods that are closer to traditional cartography than visualization. An excellent collection of these endeavors is by Dodge and Kitchin, in their Atlas of Cyberspaces [16]. For example, the topological network maps they list show graphs mapped on to a projection of countries or continents. Kitchin has also published works mapping cyberspace onto the physical and human spaces [23]. They investigate the Internet as a graph of networks, controlled by various national and trans-national (mostly) commercial entities.

*e-mail: c.gray@bangor.ac.uk

†e-mail: p.ritsos@chester.ac.uk

‡e-mail: j.c.roberts@bangor.ac.uk

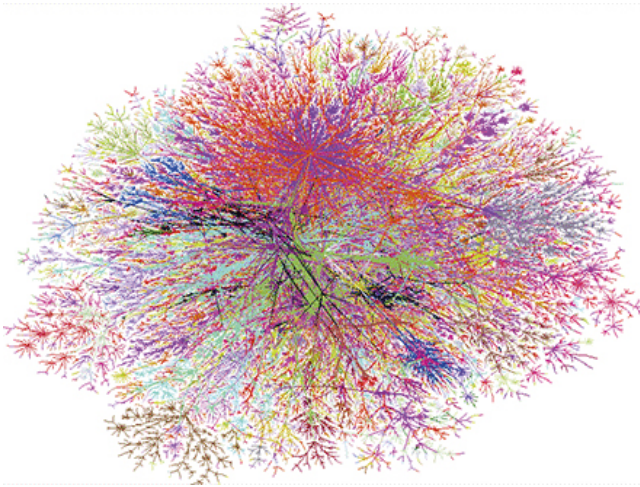


Figure 1: 'Map of the Internet' 1998, published in Wired Magazine. (Reproduced with kind permission of Bill Cheswick [8])

In this view, the political, socio-economic, industrialization or physical boundaries that define world geography are irrelevant.

The Réseaux IP Européens Network Coordination Centre (RIPE NCC) provides a set of visualizations under their RIPEstat product (stat.ripe.net). This suite of tools include some visualizations that describe a network's routing status or efficacy. These visualizations are taken from the point of view of RIPE's RIS Routing Beacons, strategically placed around the Internet. Whilst administrators are able to access details on the exact path through the various ASs, there is no flagship visualization showing this information. Tools, such as BGPlay [15], deployed in RIPEstat, can depict the shifting effect of network changes over time. However, while this information is useful to network administrators, it can only report the near real-time recorded dynamics of the actual system, rather than those of a simulation, answering 'what if' questions.

There have been several prior attempts to generate a static 'map' of the Internet [8, 14, 35]. Researchers have used traditional cartographic tools and developed GIS systems [33, 43]. Most of these, including one of the most iconic pictures (by Burch [8] republished in Wired, shown in Figure 1), were taken in a different era of the Internet. In particular, by the end of the dot-com bubble in 2001, the ISP market was swamped with small providers ripe for take-overs. During the following years there has been some consolidation in the Far East and US markets, but Europe has maintained its wide spread of Internet providers. This is evidenced by 28,851 [37] of the 50,754 active ASs (56.8%) [22] being issued by the RIPE NCC.

Other researchers have focused on visualizing attack scenarios in one of the main four aspects: (1) Traffic flows/volumes [25, 27, 44]; (2) Ports/services affected [1, 30, 34]; (3) Threat type/vectors [13, 29, 32] and (4) Source/destination IP(s) [3, 12, 24]. In some cases, the process of visualizing network attacks can be more difficult due to the vast amount of data. This requires utilising 'drill-down' and other filter and focus tools [31]. Whilst filtering may allow the noise or other uninteresting data to be removed, it can remove important contextual information, that can be useful for an analyst, in order to understand the displayed patterns [19].

Lakkaraju et al. [26] propose a method named 'closing the loop', where visual patterns are translated into symbolic rules for configuring network devices. In other work, Shneiderman and Aris [40] demonstrate how the visual analysis can provide the all important context to allow operational decisions to be made.

Exploring network configurations and topologies are another prime use-case of visual representations. There is a wealth of general visualization tools to display network data, such as Gephi [6], Cytoscape [28] and MAVisto [39] for biological network data, which would be feasible to use for situation awareness tasks. However, there are several disadvantages: first there are no layout algorithms to display the data that fits our requirements, second these are infrastructures that contain many parts, that are not required by a network administrator; third, it could be possible to create an algorithm in these tools, but we wish to have more control over the system and develop a specific tool for network administrators.

Often, an administrator would have to use several tools in order to discover and diagnose problems. Multiple views are therefore useful [38]. For instance, Padmanabgan and Subramanian [33] present different geographic mapping techniques; Noel et al. [32] use coordination, and Subramanian et al. [41] characterize the Internet hierarchy from different vantage points. In fact, hierarchy is a critical component to visualize. Hierarchical structures have been shown to be useful in Network applications [9] and other hierarchical charts such treemaps have been used in network visualization [42, 41].

3 INTENDED USE-CASES

We present three use-cases, to motivate our work, and place it in context with other network security visualization tools.

3.1 Upstream Coverage

Service providers will have a geographic customer profile, i.e., where their customers are located, but this is not necessarily the same as the network location. Connectivity providers build their network to serve a particular geographical area as well a network segment, which often matches the locations and topology of their physical cables. Content providers on the other hand will need to interconnect with several connectivity providers to match access customers to content customers.

In this use-case, network operators would need to use the visualization to show a balance between their upstreams, i.e., showing the same amount of destinations through each. Having an imbalance could introduce a fragility to the provider's network should that upstream become unavailable. The results of a single (or few) point(s) of failure are hard to predict as the dynamics of re-routing, when a connection is lost, is an entirely individual situation. The process may result in more spread with longer paths or on the other end of the spectrum, shifting to another single upstream. The impact of this is, likewise, individual and it may affect only a single customer or be entirely catastrophic with major customers inconvenienced. An abundance of significantly longer paths may mean that the provider's current mix of upstreams does not favor a particular geographic or network area. In most situations shorter paths equate to better, faster connections. This may or may not be an issue depending on the exact circumstances. For example, their customers may not be in that area, so good connections may not be a priority. Also given the dynamics of the system, the lengths will depend on the connections available throughout the Internet at any point in time.

A visualization for these goals becomes a matter of business intelligence, rather than network operations. However, without appropriate tools the network administrator would not be able to answer the questions posed without significant time or capital cost. As such, the visualization must be a repeatable process in near real-time.

3.2 Path Commonalities

Once a network administrator has decided, technology assisted or not, that their network is under attack steps must be taken

to mitigate the effects. Often this is achieved by reconfiguring a firewall or null-routing the incoming traffic. When it is a Distributed Denial of Service (DDoS) attack, these rules are often unfeasible to create due to the number of individual networks or hosts involved. Some providers may simply settle on contacting their upstream providers to have the destination of the attack blocked to at least limit the disruption to the attacked party.

It is uncertain how the network will react to this change. The Internet, as a whole, is built on the premise of always delivering packets via any means necessary. Whilst a large number of attacks do not use TCP [10], that includes guaranteed delivery, the self-healing nature of Internet routing ensures that delivery will occur until all possible pathways have been blocked/removed/filled.

The pathway into a network may differ from the necessary pathway to return traffic, producing a different view on the problem at every network. In this situation, an exploratory visualization can be useful to see common patterns in the pathways from/to attackers and the administrator's network. All possible paths must be considered when making any re-routing decision. Again, the visualization would need to provide near real-time results in order to capture the state of the Internet at the point in time. The information cannot be drawn directly from the routing information received by the administrator, which will only show one possible viewpoint. When detecting potential points to 'cut off' an attack, it is crucial to have the largest possible amount of information.

3.3 Multi-Exit Destinations / Disaster Planning

A pure exploration task, locating the diverse pathways leading to the same destination can be of most use in disaster recovery planning. A network administrator may wish to find a set of upstreams that provide diverse pathways to a critical market or network. In the event of a major outage, e.g., a transatlantic fiber-break or power outage at a large interconnect point, the likelihood that at least one pathway remains operational is relatively high.

In this use-case, the discovery is reversed. The operators want to observe the pathways out of their prospective providers, the objective being to locate their network in the resulting graph. The journey becomes more important than the destination. However, there may also be cause to examine what sorts of connections the path uses out of the prospective provider. This result is providers prioritizing the availability of chargeable, in whatever direction, connections. Therefore, peering connections are likely to be more brittle and less likely to be fixed rapidly. To satisfy this requirement, the commercial structure must be overlaid to provide the complete picture. As we will not have access to the commercial terms of every relationship between networks, some inference will need to be made.

4 CONTEXTUAL NAVIGATION

Traditional cartography would produce a single, uniform projection with all nodes and links plotted in the same geographic or relative location. Contextual Navigation, instead, presents a different viewpoint, and therefore a different projection for each location. The plot then presents the 'best' route from that location to every other. 'Best' is always context-specific and left to the implementer to define. For example, in public transport networks 'best' may be defined as shortest elapsed time with fewest changes of route, line, or type. Alternatively, for a road atlas the definition may be tied to highest average speed or best fuel economy.

This does not necessarily mean that a Contextual Navigation map does not represent relative distances or importance. The concept of distance will also be context-specific, and will not apply to every context. For example, when examining Internet pathways there are several metrics that could be substituted for distance; link capacity, round-trip (ping) time, physical distance or perceived speed. To a certain extent, the choice of distance metric will

be governed by the availability of suitable data. In the Internet example, two of the three presented options are either commercially sensitive or not available at all. The third, ping time, is entirely variable depending on the load of traffic at every point along the path. These reasons make including a distance dimension of little added value in the Internet scenario. However, transport networks have widely available and dependable metrics – such as the ones informing the routing algorithm (time, changes, speed, etc.).

As the map is presented from a certain viewpoint, it lends itself to exploration – answering the question 'where can I go from here?'. This makes Contextual Navigation an ideal candidate for electronic visitor guides to cities or large attractions, like theme parks. The concept, however, fails to transition to print or other more permanent media as the user can relocate making the projection no longer relevant. With this being said, a Contextual Navigation projection could be displayed in print at specific, fixed points as the map does not relocate with the user.

5 DESIGN REQUIREMENTS AND DEVELOPMENT

5.1 Data Collection

The source data for the Contextual Navigation projection of the Internet is taken from various Internet Routing Registries (including ARIN, RIPE and MERIT RADB). The coverage of this data-set represents four of five Internet regions. The Asia-Pacific region is no longer published by APNIC. This data-set includes objects, known as 'aut-num's, which represent an Autonomous System and their routing policies. A previous study finds that this data is suitable to use as an analogue for the Internet [20].

Routing policy is specified by the Routing Policy Specification Language (RPSL). This language allows administrators to detail which networks connect to each other, the priority and preference settings and which set of routes are exchanged. By categorising these statements of policy, into inbound/outbound, all/specific and whether the specific set contains the connecting network, an inference can be made on the type of commercial relationship in place. These commercial relationships fall into three main categories. Transit (also known as upstream) where the network pays for access to the wider Internet. Downstream where the network receives payment from another network for access to the wider Internet through it. Lastly, peering - which is most often settlement free - where two networks agree to pass traffic between themselves but not provide access to a wider area. The notable exception to that rule is the sharing downstream routes. This is a purely financial decision; the network get paid to pass the traffic, but then does not pay to pass it on making the fees almost pure profit.

Each morning, during a scheduled job, our process converts the RPSL for every network contained in the data-set into a list of typed adjacencies. The key to this process is utilising a second data-set, known as 'as-set's, also from the IRRs. An AS Set is a collection of related networks, for example the complete set of a network's customers. There are some 'universal' sets such as AS-ANY (any or all) and AS-NONE (nothing), otherwise networks are free to define their own proper subsets of AS-ANY.

The following rules were used in the classification. Let AS represent the network being examined, C is the connection, D is the direction (import or export) of the action and S is the set the policy acts on.

$$S \equiv \text{AS-ANY} \ \& \ D = \text{export} \Rightarrow \text{Classification}(C) = \text{Downstream} \quad (1)$$

The second set of rules are dependent on both the import and export policies. Therefore, let S_1 be the set imported, S_2 be the set exported, and AS_2 be the partner network.

$$AS \in S_1 \ \& \ AS_2 \in S_2 \Rightarrow \text{Classification}(C) = \text{Peer} \quad (2)$$

$$AS \in S_1 \ \& \ S_2 \equiv AS\text{-ANY} \Rightarrow \text{Classification}(C) = \text{Transit} \quad (3)$$

Once all of the collection and classification is completed, a real-time service constructs the pathways from the requested node. This uses a modified Breadth First Search algorithm to construct the shortest path to a node from the root. The modification prioritizes peering connections as they will always terminate a pathway — peering by its nature can only guarantee access to the peered network. Every network that has been visited before is then excluded from future evaluations to avoid an infinite loop and ever expanded pathways.

5.2 Visual Representations

Whilst all possibilities are in the realm of node-link and adjacency diagrams, the first choice we must make is whether or not to present the ‘full’ graph/map. Presenting the full map provides extra structural information, effectively all the possible paths, at the expense of readability due to cycles and local ‘cliques’ [5]. Techniques have been developed [11, 17] to explicitly handle these issues, including all the pathways will dilute the usefulness under the intended use-cases.

This choice means that the resulting structure will be some form of hierarchy. Using a hierarchy has other benefits, chief among which is the ability to mimic path selection algorithms inherent in the network routing (no matter what protocol is used). Within the hierarchical visualizations, there are still several options. Firstly, the distance between networks is not a real-world measurement or even relative. Therefore the length of the links hold no significance. Secondly, the branching factor of a diagram of this type can be significant. With almost 51,000 potential nodes, the visualization must be able to scale showing this data whilst remaining readable.

Initially, the visualization was implemented using a sunburst diagram, shown in Figure 2. To combat the biases that are inherent in relative space diagrams, such as the sunburst diagram and pie charts [18], the percentage of pathways prefixed with the selected sector is shown (figure centre), as well as a breadcrumb trail explicitly showing the selected path (figure top-left).

Heuristic testing, by visualization experts, proved initial suspicions that the resulting visualization was too hard to correctly interpret. The sunburst, whilst representing one measure of importance, distorted the relative importance at each level as the size of the sector is based on the number of nodes in the sub-tree.

The next experiment replaced the sunburst with a radial dendrogram (seen in Figure 3). The colors are kept consistent between these versions, however this makes the resulting image harder to interpret as there is not enough contrast between the brown and orange without the white divisions of the sunburst.

This implementation represents the branching nature of Internet pathways better, represented by the grey link lines. However the links now become the limiting factor, crossing and blending together. At the lower levels of the tree, it becomes almost impossible to visually trace their path back to the root. The dendrogram does have a coincidental, and unanticipated, feature of representing the amount of coverage. This would, for example, produce gaps in the outer ring if the network in question could not reach the entire Internet.

To address the failings of the second iteration, colors were selected from the ColorBrewer [21] service (4 data classes, print friendly, qualitative). This serves to make the classes more distinct. The plot was also changed to a Radial Tree, laid out using the Reingold-Tilford algorithm [36]. The principal aim of this algorithm is to separate sub-trees to limit the number of cross-overs

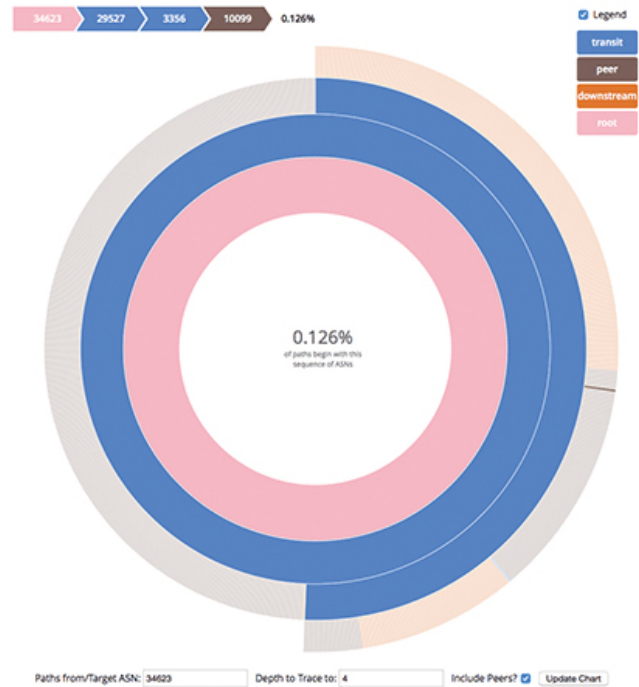


Figure 2: Initial Development; Sunburst Diagram. The major benefit of this type of diagram is relative size is readily identifiable. However, there is a finite number of leaves that can be shown.

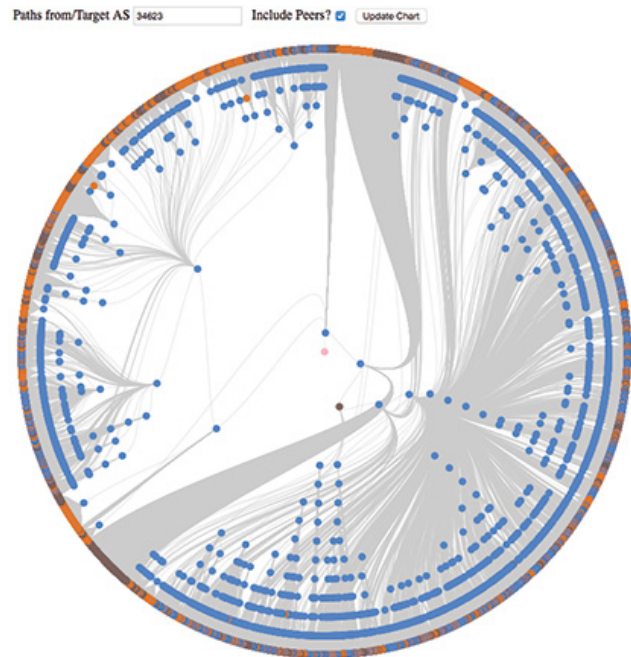


Figure 3: Second Iteration: Radial Dendrogram. Edges are bundled but colored identically making individual pathway tracing difficult. Also note the amount of crossing edges. Color meanings remain the same as the first iteration.

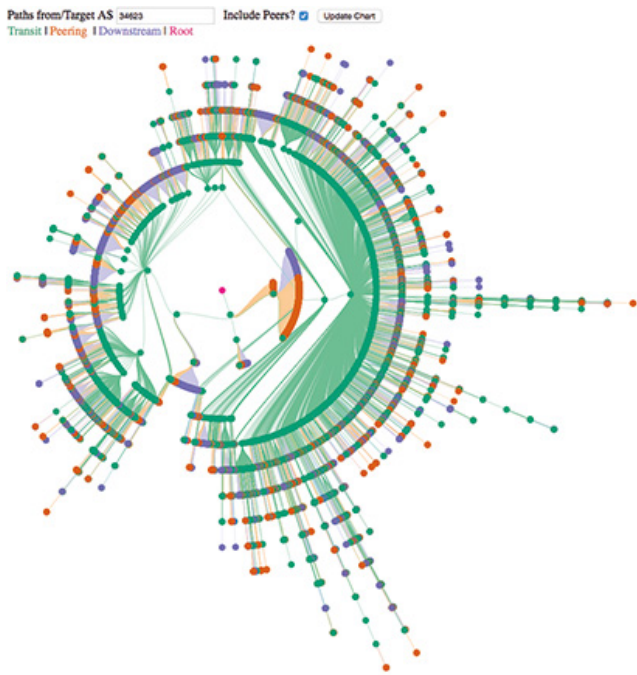


Figure 4: Third Iteration: Radial Reingold-Tilford Tree. The new palette and coloring links increase distinction between pathways. The new layout reduces but does not always eliminate crossing links.

in links and over-plotting. This iteration is shown in Figure 4. The resulting contextual projection has the following beneficial qualities:

- *Fidelity of Distance*
The relative distances are preserved throughout the projection, each concentric circle represents one level in the tree (or in our case one node in the pathway).
- *Fidelity of Importance*
Important, or significant common, nodes are clearly visible by a wide spread, dense set of colored links.
- *Fidelity of Position*
The relative positions of related nodes in a sub-tree are preserved. This produces a projection which is suitable for visual analytics as the paths can be visually traced to the root.
- *Fidelity of Separation*
The commercial classes can now be clearly separated even in bunched areas.
- *Fidelity of Coverage*
The Radial Tree retains the ability to demonstrate coverage. A full circle, when the level are flattened, demonstrates global reachability.

Depending on the pathways that are found for a target network, there is still some cross-over of the links (as can be seen in the figure). Cross-over in a radial layout is somewhat unavoidable. However, as the links are colored the same as the target node these can be more easily distinguished. When demonstrating the visualization to non-expert colleagues and students, they were able to pick out the visual patterns despite not being able to interpret the real-world meaning of them.

6 VISUAL ANALYTICS CASE STUDIES

The following case studies were selected through discussion with network administrators with substantial expertise and to demonstrate distinctive patterns and shapes from different network

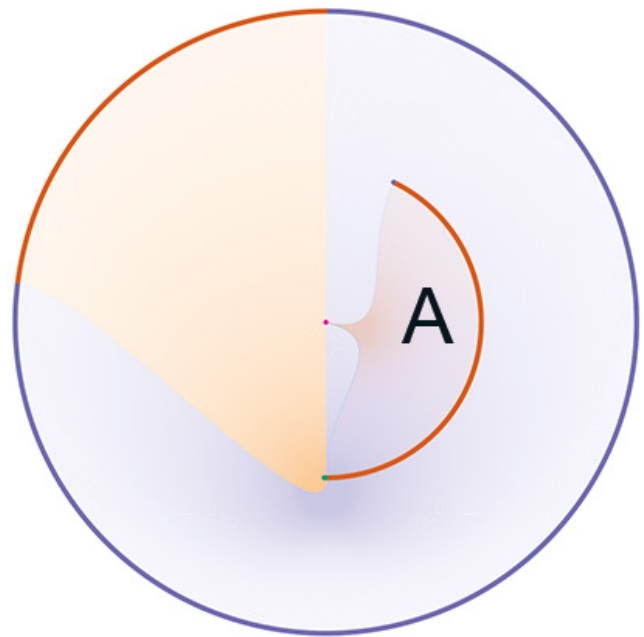


Figure 5: Contextual Navigation Projection for AS20773. The wide range of local peering is readily apparent. As is the dependence on a single major upstream network.

types. There are two fundamental commercial activity groups that take place on the Internet. Content supply networks house the web sites, videos, e-mail, etc. that get viewed by the masses. Access networks provide connectivity to the masses to be able view content. A US example of each would be Netflix (content) and Qwest (access), UK examples would be the BBC (content) and BT Internet (access).

6.1 AS20773: A Content Network with Global Customers

Figure 5 shows the projection for AS20773 (HostEurope GmbH), a major European hosting provider. As such, the aim of their network build-out is to connect to as many access providers in as short a distance as possible. This is evidenced by the (almost) 180° orange 'slice' at the right-center of the projection (labelled A). This represents the direct peering for the network and a significant number of 'local' routes. This network then has a single Tier-1 provider to access the rest of the Internet. We are able to conclude that the upstream is most likely a Tier-1 provider as there are no upstream connections (shown in green) radiating from that node to wider circles. In actuality (and from examining the raw data); there are two upstream networks. The major one is Deutsche Telekom, AS3320, which is one of the largest pan-European ISPs. This connection provides most of the outbound pathways. This is depicted by the single green link from the root and branching to the entire outer circle.

6.2 AS5511: A National 'Hub' Access Provider

In some European countries, there is still an incumbent major communications provider. Figure 6 shows the projection for one of these such companies, Orange S.A. (formerly France Télécom). These incumbent providers act as a national 'hub', essentially interconnecting the smaller local providers and the country with the outside world. The visualization suggests this status, with a dense purple section in the center (at one hop distant). We can also see a significant level of peering. These two observations would

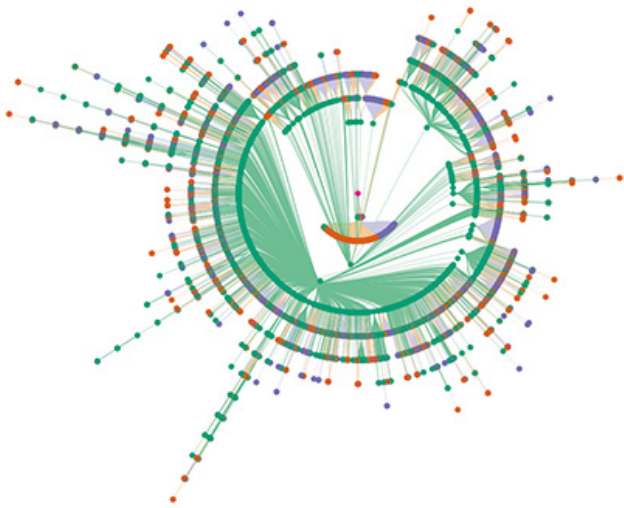


Figure 6: Contextual Navigation Projection for AS5511: Note; the hub portion can be seen by the significant purple wedge in the center of the projection. The large (bottom) green/transit wedge represent the 'foreign' connections to the wider Internet.

support a hub-like usage. However, even a dominant position in this market does not automatically raise this network to Tier 1 status. This AS requires additional upstream connections to ensure that the whole Internet is reachable - the green shaded nodes. AS5511 uses a balanced mix of several to attain this access.

6.3 AS6677: A 'Traditional' National Access Provider

Figure 7 shows the projection for AS6677, Iceland Telecom. This provider has a 'traditional' setup, where the ISP is responsible for the last mile, and uses larger backbone ISPs for connecting to the wider world. The projection demonstrates the lack of connectivity with sparse inner rings, just two nodes. In stark contrast to other networks, this ISP does not have any peers, just upstreams. The lack of local inter-connectivity may be a function of geography as much as network planning, as Iceland sits on the edge of the European Internet, with no 'local' competition and only physical links to the UK and Greenland. It highlights the importance of strong physical (as in cables) support behind any major interconnection point.

The network's choice of upstreams keeps the overall quality of the connectivity high, with a comparable set of long paths. This is at the expense of a local vulnerability should those links be lost. Using this information, the administrator may choose to invest in more diverse physical pathways, i.e., a redundant link to the same upstream but via a different cable route. Or, they may choose to invest in a different connectivity mix. A traditional last mile ISP, however, is more often concerned with the physical infrastructure connecting their customers within the geographical territory.

6.4 AS5089: National ISP with International Upstreams

Figure 8 shows an interesting projection for Virgin Media's UK network, AS5089. The projection highlights an issue with any kind of mapping technique, which are effectively only as good as their backing data-set. In this case the visualization is relying on the RIPE NCC database, which only includes details of resources issued in Europe. Virgin Media have incredibly large set of peers (the large orange circle), but only two major upstreams — Level(3) (AS3356) and Savvis (AS3561), both major US connectivity providers. As such, their onward data is not (currently) fully available to visualize.

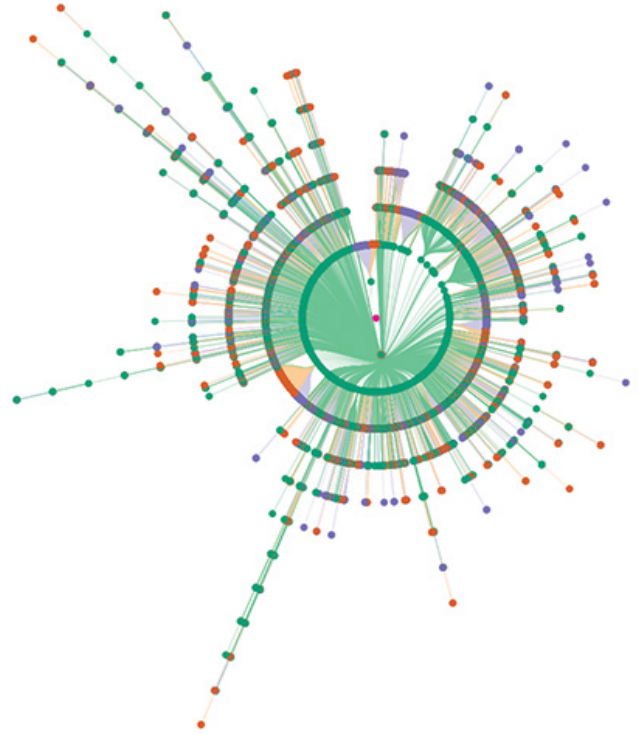


Figure 7: Contextual Navigation Projection for AS6677. The sparse inner rings demonstrate that this network relies solely on two major upstreams with no independent 'exits'.

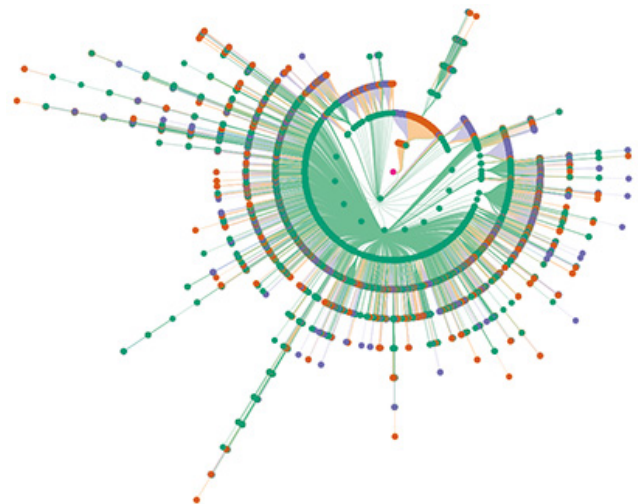


Figure 8: Contextual Navigation Projection for AS5089. The international Tier-1 providers are clearly visible as the 'dotted' inner green ring. There are six significantly longer paths, highlighting a 'bottleneck' to those networks.

However, this projection does show a potential weakness for this network. Apart from strong UK and western European peering connections, the network is reliant on these two US providers. Whilst those providers have ‘local’ infrastructure and may be resilient within their own networks, there are no other potential routes if those fail. However, the probability of both failing at the same time is incredibly small – unless a systemic failure occurs such as power outages or an act of terrorism.

7 FUTURE EXTENSIONS

The most useful future extension is to integrate more data. Whilst the current visualization is built from all publicly available Internet Routing Registry data, there are some territories (e.g., Asia-Pacific) that no longer publish their information in a usable form. This is not to say that all of Asia-Pacific is missing, but would appear as leaf nodes lacking any onward connections.

The next major element of improvement would be to introduce a full RPSL parser, this would allow the full technical meaning of the IRR data to be understood. The visualization would benefit as all connections would be correctly classified (transit, peer, downstream), rather than defaulted if the current tests fail. It may also allow more ‘scientific’ identification of the tiers of providers. However, ultimately this would still be in the hands of domain experts as the data will never contain any commercial considerations.

This technique can be built upon to create a more tailored visualization for attack scenarios. The simplest of these is merely highlighting a destination network, for example by increasing the size of the node or changing its color. Next, a more ‘lenient’ routing algorithm can be installed to evaluate every possible path to that destination. That algorithm would differ in that our present limited Breadth First Search is optimized for use in real-time by avoiding all of the possible loops of connections.

There may be scope for creating a collapsible version of this visualization. It would label entire sub-trees as, for example Level(3), and reduce that entire sub-tree to a single node. We envisage that this would be a highly specific enhancement, and require significant input of domain knowledge to keep current. Whilst the current technique could be simply enhanced to hide sub-trees behind a parent node, domain knowledge of what is sensible to reduce would improve the signal-to-noise ratio more.

Labelling remains an issue for dense hierarchies/graphs. Techniques have been developed, such as the Extended Excentric Labelling scheme [7], which create summaries to reduce the number of labels required. This approach is certainly valid in this case, however suffers from the same domain knowledge requirement as collapsing the tree. To effectively label a generic case dense graph would require a new scheme or to introduce a ‘magic lens’ or other filtering technique.

8 RELATED ISSUES

8.1 Changing Topology

The current implementation has a scheduled job to re-retrieve the source data every 24 hours, at which time the adjacencies are re-computed and re-classified. This is a fair compromise between constant recalculation (and a potentially unwelcome load on the data suppliers) and a purely point-in-time, static projection. Changes in the Internet topology are almost always intended to be local. Some of the most costly international or inter-continental outages have been caused by incorrectly controlled changes [4]. Therefore, whilst the change at the local level may be drastic; the overall ‘grand scheme’ view of the Internet does not really differ. This is because the Internet (as has been found previously by Albert et. al. [2]) exhibits ‘small world’ properties.

As the intended purpose of this visualization is to show the current status, there is no long term storage of the raw data. Neither

do we provide a temporal view of the visualization. There are use-cases that would benefit from being able to visualize these longer term patterns. However, this will only model one side of the dynamics of the Internet – the commercial growth. As we do not show the selected paths (by BGP), we cannot visualize the operational dynamics of the whole system.

8.2 Difference with Reality

There are two factors that introduce differences between reality and the tree produced by the visualization. The first is timing-based, the second is ‘private’ connections that are not registered in the IRRs. Updates to the IRRs are non-real time. Therefore, the visualization will naturally lag behind actual network conditions. We have accounted somewhat for this effect by utilising the breadth first search, showing more pathways than will most likely be used.

To totally overcome this issue would require privileged access to a network’s border routers. This would allow collection of the Routing Information Base (RIB) in near real-time to base the visualization on. The downside to this approach is that it is non-portable and that instance would necessarily be tied to that network.

Some network administrators may in fact prefer this approach, providing tailored information. Utilising RIB data would also solve the second source of differences, as the previously unregistered connections would show in the RIB information. Unless the technique/visualization were packaged as a ‘self-install’ product, it is unlikely that administrators would allow the requisite access to their routers. It is possible to run feasibility studies using public ‘route servers’, however these will only show one ‘best’ route to all destinations.

8.3 Including ‘Distance’

The current design for the Contextual Navigation projection, as tailored to Internet mapping, does not include a ‘distance’ metric. The concept/design can include it, changing the visual representation of the resulting graph. Reingold-Tilford Radial Trees can support links of varying lengths, however the layout does not guarantee placement of nodes at the same ‘distance’ from the root in the same radius due to other layout constraints. There are expansion options; an Elbow Dendrogram layout (using right angles on the connections) whether using the Reingold-Tilford algorithm or not, or a Radial Phylogram layout.

Due to the layout method, the Elbow Dendrogram will require more canvas space. As the label for the node is included in the horizontal axes, the projection will become wider with long labels - forcing the user to trace long lines. This is undesirable, as links can be confused especially the longer and more numerous they become. As the links are now guaranteed to be separated by, at least, the label font’s height the grouping at branching points is lost visually.

The preferred option to including distance is to use the Radial Phylogram. Nodes at a distance are presented on the same radius, as with the Radial Tree. The links from that radius can be varied in the vertical dimension (away from the root/center), and horizontal distance (around the circle) does not factor into the distance. Grouping is preserved as more dense ‘forks’ with each vertical stem being plotted perpendicular to the tangent at the point on the arc. This separation of the horizontal and vertical dimensions conveys more effectively the sense of cumulative distance from the root. However, the Phylogram will still suffer over-plotting or confused link routing if the number of nodes at a low distance is greater than the space available. This may be able to be counteracted by introducing a scaling factor to increase the distance at the root.

9 CONCLUSIONS

Our Contextual Navigation concept and method is able to produce valid, usable and informative visualizations for the current state of the network, from a given viewpoint. We have shown that

the current Internet can be mapped topologically and include the commercial relationships. The contextual approach can also be used in other mapping situations, as long as a topological representation makes reasonable sense. These projections can be used for both exploratory and search purposes by including highlighting for a destination node.

As discussed in Section 7 various improvements could be introduced. Some of these are simple customizations of the plotting, others require more in-depth changes and greater access. We have shown that the plotting method is not perfect (as seen in Figure 5) when the branching factor near the root of the tree is significantly higher. During our experiments we have only found a handful of cases where this occurs, as this is encountered in specialist situations. We have also considered alternative layouts for when a ‘distance’ metric needs to be included – increasing the visual utility of the projection. Labelling has been considered and would require either a new high density mechanism or more interactive filtering tools.

REFERENCES

- [1] K. Abdullah, C. Lee, G. Conti, and J. A. Copeland. Visualizing network data for intrusion detection. In *Proc. IEEE SMC IAW*, pages 100–108. IEEE, 2005.
- [2] R. Albert, H. Jeong, and A.-L. Barabási. Internet: Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999.
- [3] R. Ball, G. A. Fink, and C. North. Home-centric visualization of network traffic for security administration. In *Proc. ACM Workshop Vis. and Data Mining for Comput. Sec.*, pages 55–64. ACM, 2004.
- [4] H. Ballani, P. Francis, and X. Zhang. A study of prefix hijacking and interception in the internet. In *ACM SIGCOMM Comput. Commun. Rev.*, volume 37, pages 265–276. ACM, 2007.
- [5] H.-J. Bandelt and E. Prisner. Clique graphs and helly graphs. *Combinatorial Theory, Series B*, 51(1):34–45, 1991.
- [6] M. Bastian, S. Heymann, M. Jacomy, et al. Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8:361–362, 2009.
- [7] E. Bertini, M. Rigamonti, and D. Lalanne. Extended excentric labeling. In *Proc. Eurographics*. John Wiley & Sons, 2009.
- [8] H. Burch and B. Cheswick. Mapping the internet. *Computer*, 32(4):97–98, 1999.
- [9] K. L. Calvert, M. B. Doar, and E. W. Zegura. Modeling internet topology. *IEEE Commun. Mag.*, 35(6):160–163, 1997.
- [10] V. Cerf, Y. Dalal, and C. Sunshine. Specification of Internet Transmission Control Program. RFC 675, Dec. 1974.
- [11] T. T. Chen and L. C. Hsieh. The visualization of relatedness. In *Proc. IEEE SMC IAW*, pages 415–420. IEEE, 2008.
- [12] H. Choi, H. Lee, and H. Kim. Fast detection and visualization of network attacks on parallel coordinates. *Computers & Security*, 28(5):276–288, 2009.
- [13] G. Conti and K. Abdullah. Passive visual fingerprinting of network attack tools. In *Proc. ACM Workshop Vis. and Data Mining for Comput. Sec.*, pages 45–54. ACM, 2004.
- [14] A. Danesh, L. Trajkovic, S. H. Rubin, and M. H. Smith. Mapping the internet. In *Proc. IFSA World Congr./NAFIPS Int. Conf.*, volume 2, pages 687–692. IEEE, 2001.
- [15] G. Di Battista, F. Mariani, M. Patrignani, and M. Pizzonia. Bgplay: A system for visualizing the interdomain routing evolution. In *Graph Drawing*, pages 295–306. Springer, 2004.
- [16] M. Dodge and R. Kitchin. *Atlas of cyberspace*, volume 158. Addison-Wesley New York, 2001.
- [17] J. Edachery, A. Sen, and F. J. Brandenburg. Graph clustering using distance-k cliques. In *Graph drawing*, pages 98–106. Springer, 1999.
- [18] S. Few. Save the pies for dessert. *Visual Business Intelligence Newsletter*, pages 1–14, 2007.
- [19] J. R. Goodall, W. G. Lutters, P. Rheingans, and A. Komlodi. Focusing on context in network traffic analysis. *IEEE Trans. Vis. Comput. Graphics*, 26(2):72–80, 2006.
- [20] C. C. Gray and S. P. Mansoor. Finding the invisible: A comparison of irr data and routing pathways. *Submitted for Publication*, 2015.
- [21] M. Harrower and C. A. Brewer. Colorbrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.
- [22] G. Huston. CIDR Report - AS2.0. <http://www.cidr-report.org/as2.0/> collected 2015-06-03, May 2015.
- [23] R. M. Kitchin. Towards geographies of cyberspace. *Progress in human geography*, 22(3):385–406, 1998.
- [24] H. Koike, K. Ohno, and K. Koizumi. Visualizing cyber attacks using ip matrix. In *Proc. VizSEC*, pages 91–98. IEEE, 2005.
- [25] S. Krasser, G. Conti, J. Grizzard, J. Gribshaw, and H. Owen. Real-time and forensic network data analysis using animated and coordinated visualization. In *Proc. IEEE SMC IAW*, pages 42–49. IEEE, 2005.
- [26] K. Lakkaraju, R. Bearavolu, A. Slagell, W. Yurcik, and S. North. Closing-the-loop in nvisionip: Integrating discovery and search in security visualizations. In *Proc. VizSEC*, pages 75–82. IEEE, 2005.
- [27] K. Lakkaraju, W. Yurcik, and A. J. Lee. Nvisionip: netflow visualizations of system state for security situational awareness. In *Proc. ACM Workshop Vis. and Data Mining for Comput. Sec.*, pages 65–72. ACM, 2004.
- [28] C. T. Lopes, M. Franz, F. Kazi, S. L. Donaldson, Q. Morris, and G. D. Bader. Cytoscape web: an interactive web-based network browser. *Bioinformatics*, 26(18):2347–2348, 2010.
- [29] S. Mathew, R. Giomundo, S. Upadhyaya, M. Sudit, and A. Stotz. Understanding multistage attacks by attack-track based visualization of heterogeneous event streams. In *Proc. Vis. Comput. Sec.*, pages 1–6. ACM, 2006.
- [30] J. McPherson, K.-L. Ma, P. Krystosk, T. Bartoletti, and M. Christensen. Portvis: a tool for port-based detection of security events. In *Proc. ACM Workshop Vis. and Data Mining for Comput. Sec.*, pages 73–81. ACM, 2004.
- [31] S. Musa and D. J. Parish. Visualising communication network security attacks. In *Proc. IEEE INFOVIS*, pages 726–733. IEEE, 2007.
- [32] S. Noel, M. Jacobs, P. Kalapa, and S. Jajodia. Multiple coordinated views for network attack graphs. In *Proc. VizSEC*, pages 99–106. IEEE, 2005.
- [33] V. N. Padmanabhan and L. Subramanian. An investigation of geographic mapping techniques for internet hosts. In *ACM SIGCOMM Comput. Commun. Rev.*, volume 31, pages 173–185. ACM, 2001.
- [34] J. Pearlman and P. Rheingans. Visualizing network security events using compound glyphs from a service-oriented perspective. In *Proc. VizSEC*, pages 131–146. Springer, 2007.
- [35] M. P. Peterson. *Maps and the Internet*. Elsevier, 2003.
- [36] E. M. Reingold and J. S. Tilford. Tidier drawings of trees. *IEEE Trans. Softw. Eng.*, (2):223–228, 1981.
- [37] RIPE NCC. RIPE Database - Split by Type - Autonomous System Numbers. <ftp://ftp.ripe.net/ripe/dbase/split/ripe.db.aut-num.gz> collected 2015-06-03, June 2015.
- [38] J. Roberts. State of the art: Coordinated multiple views in exploratory vis. In *Int Conf. CMV*, pages 61–71, July 2007.
- [39] F. Schreiber and H. Schwöbbermeyer. Mavisto: a tool for the exploration of network motifs. *Bioinformatics*, 21(17):3572–3574, 2005.
- [40] B. Shneiderman and A. Aris. Network visualization by semantic substrates. *IEEE Trans. Vis. Comput. Graphics*, 12(5):733–740, 2006.
- [41] L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz. Characterizing the internet hierarchy from multiple vantage points. In *IEEE Proc. INFOCOM*, volume 2, pages 618–627. IEEE, 2002.
- [42] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *IEEE Trans. Vis. Comput. Graphics*, 13(6):1121–1128, 2007.
- [43] Y. Wang, P. Lai, and D. Sui. Mapping the internet using gis: The death of distance hypothesis revisited. *Journal of Geographical Systems*, 5(4):381–405, 2003.
- [44] X. Yin, W. Yurcik, M. Treaster, Y. Li, and K. Lakkaraju. Visflowconnect: netflow visualizations of link relationships for security situational awareness. In *Proc. ACM Workshop Vis. and Data Mining for Comput. Sec.*, pages 26–34. ACM, 2004.