Charting Censorship Resilience & Global Internet Reachability: A Quantitative Approach

Marina Ivanović
ETH Zurich
mivanovic@ethz.ch

François Wirz
ETH Zurich
wirzf@inf.ethz.ch

Jordi Subirà Nieto
ETH Zurich
jordi.subiranieto@inf.ethz.ch

Adrian Perrig
ETH Zurich
aperrig@inf.ethz.ch

***Abstract*—Internet censorship and global Internet reachability are prevalent topics of today's Internet. Nonetheless, the impact of network topology and Internet architecture to these aspects of the Internet is under-explored. With the goal of informing policy discussions with an objective basis, we present an approach for evaluating both censorship resilience and global Internet reachability using quantitative network metrics, which are applicable to current BGP/IP networks and also to alternative Internet network architectures. We devise and instantiate the metric on the network topology of multiple countries, comparing the BGP/IP network, an overlay network using a waypoint mechanism for circumventing undesired nodes, and the path-aware Internet architecture SCION. The novelty of the approach resides in providing a metric enabling the analysis of these aspects of the Internet at the routing level, taking into account the innate properties of the routing protocol and architecture. We demonstrate that the Internet topology matters, and strongly influences both censorship resilience and reachability to the global Internet. Finally, we argue that access to multiple paths accompanied with path-awareness could enable a higher level of censorship resilience compared to the current Internet, and reduce the centralization of Internet routing.**

***Index Terms*—quantitative metrics, censorship, networking, routing, reachability, next-generation Internet architectures**

I. Introduction

The issue of Internet censorship—the deliberate restriction or suppression of information [1], [2]—has emerged as a pervasive concern in the digital era. There have been long standing records of censored network communication practices employed by various entities, and most prominently governments [3]–[7]. Furthermore, the issue of dependency on certain countries has also grown in the context of the global Internet [8], [9], with various analyses that western countries have gained significant influence on the global Internet routing, as routing paths predominantly traverse them [10].

The innate properties of Internet topologies and architectures play a crucial role in determining how traffic flows through the network, and therefore, they are likely to influence the effectiveness of censorship efforts, and in general Internet reachability. In the context of the global Internet, we refer to the network topology as the interconnectedness of Autonomous Systems (ASes), while the Internet architecture encompasses the underlying structure and protocols that facilitate operation of the Internet. For instance, the core Internet routing protocol is the Border Gateway Protocol (BGP), which provides Internet inter-domain routing [11]. In traditional networks using BGP, ASes only consider the next hop when making routing decisions. Unlike traditional routing, SCION—a next-generation Internet architecture designed to provide secure inter-domain routing [12]—ensures that packets traverse predetermined paths and making end-nodes in the network path-aware [12]. Finally, the usage of Virtual Private Network (VPNs) has been a popular technique for Internet censorship evasion, given that it could not only provide an additional layer of secrecy using encryption, but also circumvent censoring devices altogether [13], [14].

Previous research underlines the evidence that the topology of the network could be an indicator of deployed censorship capabilities [15]–[17], and reachability to the global Internet [9], [10], [18]. Nonetheless, to the best of our knowledge, it is an open research challenge how traditional BGP routing, the use of waypoint network with VPN nodes, and in general fundamentally different approaches such as BGP and SCION could be quantitatively compared in this context.

Research Question. In this context, the following research questions arise: *Do the topology and the architecture of the Internet have an influence on Internet censorship, and in general global Internet reachability?* And if so, *can we quantify this influence?* Answering these questions does not only provide insights into the interplay between Internet topology and architecture, and censorship and reachability, but can also quantitatively inform policy-makers. To achieve this, we propose a concrete approach for evaluating censorship and reachability aspects using a quantitative network metric.

Key Contributions.

1) We design a quantitative metric instantiable to *censorship resilience* and *global Internet reachability*. The metric is agnostic to network topology, and applicable to the current Internet and captures path-awareness. (Section III).
2) We instantiate the metric on the current Internet topology of several countries, analyzing their network topologies with regards to Internet censorship. In the context of the influence to Internet reachability, we instantiate our metric using diverse groups of potentially influential countries (Section IV).
3) We perform extensive experiments using the contemporary Internet topology on both BGP, a waypoint network with intermediate nodes, and SCION, a path-aware Internet architecture. Ultimately, we provide a comparative analysis of the three analyzed architectures for Internet censorship and global reachability (Section V).

II. Background

In this section we provide background relevant for our work.

Internet Censorship. Internet censorship can be observed when an entity in power—a government, company, or an individual—restricts its citizens or users from certain online communication or content, if it is deemed harmful, politically inappropriate, sensitive or legally noncompliant [1], [2]. Censorship could happen at various communication points: at the end point devices, or on the link, by nodes that the traffic passes through [14]. The scope and the focus of this paper will be on the latter, where we will consider exclusively *on-link* censorship as relevant in this work.

Previous research shows that different countries employ different hierarchical structures when it comes to enforcing censorship policies. For instance, Iran had been characterized as having centralized operational deployment of censorship [3], [17], and Russia as de-centralized [4]. In the case of China, Xu et al. observe that most of the censoring activity indeed happens at the border ASes [19]. Furthermore, Ensafi et al. observe that Tor is not being censored when traffic enters the country via CERNET, the Chinese Educational and Research Network, ultimately reaching its final destination [20]. Ultimately, prior work indicates that network topology, the routing protocol, nodes that are being traversed, and in general also the Internet architecture play a role for censorship [16], [18], [21], stressing the need for a quantitative evaluation of this influence.

Censorship Circumvention. In the face of widely deployed censorship techniques, a plethora of censorship circumvention methods have arisen with the aim of providing connectivity [13]. In the case of circumventing the censor altogether, one could also resort to using intermediate nodes in the network as a waypoint, to circumvent the censors' influence [14], for instance by utilizing a Virtual Private Network (VPN) connection.

Global Internet Reachability. Prior research shows that certain countries exhibit hegemonic influence on global reachability [8], while many depend on Western countries to access common Internet destinations [10]. These instances not only jeopardize global reachability due to direct dependence on other countries and their Internet infrastructure [10], but also raise concerns about potential surveillance [22]–[24] and collateral damage [18].

SCION Next-generation Internet Architecture. SCION is a *next-generation Internet architecture*, which provides a high level of security and efficiency, but also availability, scalability, and transparency [12], [25]. It is already in use for production use cases and deployed in production networks [26]. SCION groups ASes into Isolation Domains (ISDs), with shared governance institutions. A grouping of ASes into ISDs provides a trust environment between the ASes in it, and potentially also a common jurisdiction. To that end, one possible model is to form ISDs along national borders [12]. ISDs are interconnected, thus providing global connectivity to their nodes. An ISD is administered by a certain number of ASes which are in SCION called *core* ASes, whereas all others are *non-core* ASes. Two core ASes can have a *core link* between each other, whereas a connection between any two other ASes can either be a peering or parent-child link—with the typical provider-customer business relationship.

SCION is a multi-path and path-aware architecture [27]. In the SCION network each node obtains path segments through the process of *beaconing*, which would later be used for the construction of a full path to the desired end-node. Hence, SCION has a clear separation between *control plane* and *data plane*. In general, each node can have a large number of end-to-end paths at its disposal, from which it can select any one of them. As a part of the data plane, each SCION border router performs inter-AS packet forwarding based on the packet-carried forwarding state representing the path selected by the end host [12].

III. Quantifying Avoidability in a Network

In this section we define *Avoidability Potential* and show its applicability to Internet censorship and Internet reachability.

A. Avoidability Potential

Avoidability Potential quantifies the potential of avoiding undesirable or potentially malicious nodes in the network.

Scope of the Analysis. In this work we focus on the topological organization of the Internet, providing an analysis at the inter-AS routing level. We note that Internet censorship could be deployed at any communication point: either at the source/destination devices, or between the two [14]. The scope of this paper is the latter, where we exclusively consider *on-link* censorship and in general inter-AS communication.

Network Model. We model the network as a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$. Each node $v \in \mathcal{V}$ represents an Autonomous System (AS), whereas each edge $e \in \mathcal{E}$ represents a link between nodes. Edges between nodes are labeled with the standard business relationships on the Internet: *customer-provider* and *peer-peer* [28]. In the case of the SCION model, the edges between core nodes can also be labeled as *core* [12]. Furthermore, the nodes in the SCION network are grouped into ISDs. This does not affect the model of the network as a graph, but rather only provides grouping of nodes composing it.

Threat Model. Given the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, communication between any two nodes $s, d \in \mathcal{V}$ could be censored, intercepted, blocked, or in any way tampered with, ultimately harming communication between nodes s and d. This can be done anywhere on the path between s and d, and caused by nodes on this path which are unreliable, potentially malicious, with interests of deteriorating communication between s and d, or simply not trusted. Given that our assessment is done at the level of ASes, we assume that the ASes as a whole pose this threat. The number of Byzantine ASes can vary, and they can collude. Finally, in the context of Internet censorship, the adversary could have various motives for engaging in censoring activity, although we do not consider them explicitly in our model.

The Metric. We define set $\mathcal{S} \subset \mathcal{V}$ as the set of all *source* nodes from which paths of interest originate, and the set $\mathcal{D} \subset \mathcal{V}$ as the set of all *destinations*, where the paths terminate. Finally, we define $\mathcal{X} \subset \mathcal{V}$ as a set of ASes that should be avoided when communicating between nodes of interest.

Given the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, the set $\mathcal{X}$ of nodes whose avoidability is analyzed, a source node $s \in \mathcal{S}$ and a destination node $d \in \mathcal{D}$, we define $e_{\mathcal{X}}(s \to d)$ as a binary flag of whether a path between s and d exists, which completely circumvents nodes in $\mathcal{X}$. If it exists, we say that these nodes have the full potential of establishing a connection.

$$e_{\mathcal{X}}(s \to d) = \begin{cases} 1, & \exists r, \text{ a path } s \to d, \text{ s.t. } \forall x \in \mathcal{X}, x \notin r \\ 0, & \text{otherwise} \end{cases}$$

From there, we define the *Avoidability Potential* by allowing for all possible sources $s \in \mathcal{S}$ and destinations $d \in \mathcal{D}$. This yields the final metric, presented in the Equation (1).

$$AP_{\mathcal{X}}(\mathcal{S}, \mathcal{D}) = \frac{\sum\limits_{\substack{s \in \mathcal{S} \\ d \in \mathcal{D}}} e_X(s \to d)}{\|\mathcal{S}\| \cdot \|\mathcal{D}\|} \tag{1}$$

The value $\|S\| \cdot \|D\|$ in the Equation (1) is the number of all pairs of sources and destinations, which leads to a normalized value $AP_{\mathcal{X}}(\mathcal{S}, \mathcal{D}) \in [0, 1]$. Here, 1 means that the nodes from $\mathcal{D}$ can always receive traffic from the nodes in $\mathcal{S}$, without traversing any node in $\mathcal{X}$, whereas 0 would mean that this traffic would *always* traverse some of these nodes.

The above introduced metric is general, applicable to any graph and sets of nodes in the graph $\mathcal{S}$, $\mathcal{D}$ and $\mathcal{X}$. It is also independent of the network model, routing protocol, and captures architectures which allow for multiple paths. Thereupon we lay out two important applications of this metric, which are largely relevant for today's Internet: censorship resilience, and global Internet reachability.

B. Censorship Resilience Potential

We apply the *Avoidability Potential* to the case of censorship resilience, deriving the *Censorship Resilience Potential* metric.

The Metric. Given the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, we define $\mathcal{C} \subset \mathcal{V}$ as a set of ASes that could pose a threat of censoring paths that traverse them. From here, we define *Censorship Resilience Potential* as *Avoidability Potential*, where the set $\mathcal{X}$ is the set of censoring nodes $\mathcal{C}$.

$$CRP_{\mathcal{C}}(\mathcal{S}, \mathcal{D}) = AP_{\mathcal{C}}(\mathcal{S}, \mathcal{D}) \tag{2}$$

Example: National Outflow Traffic. One might focus on a certain country, and define the set $\mathcal{C}$ as a set of ASes that have interests or capabilities to censor national outflow traffic. In the case of the outflow traffic—namely, the traffic that originates within the country and is intended for a foreign AS—all national ASes can be considered as sources, and thus present in the set $\mathcal{S}$. By analogy, all ASes outside the country would be in the set of destinations $\mathcal{D}$.

This metric requires the set of censoring ASes, $\mathcal{C}$ to be pre-defined. However, it is often difficult to *a priori* attribute censorship interests to particular ASes. Therefore we present in addition a method which defines $\mathcal{C}$, purely based on their *potential* to censor outflow traffic.

Towards a Metric Agnostic to Normative Claims. As we already briefly discussed in Section II, censorship has been performed around the world by various entities, often for political reasons. Therefore, one might consider that quantifying censorship must require embedding these interests into the metric itself, while defining the aims of censorship and providing normative arguments of such an activity. Nonetheless, we develop a metric that treats all nodes in the network as potential censors, without arguing whether any of them perform censorship[1].

Defining Censoring ASes. In the case where censoring ASes are not known *a priori*, we define them based on their potential to choke the highest number of paths from $\mathcal{S}$ to $\mathcal{D}$. Following the intuition of the outflow traffic —an example relevant for the current Internet [15], [19]— we define the set of censors $\mathcal{C}$ as a subset of $\mathcal{S}$, which have the *highest potential* of choking outflow paths that go from $\mathcal{S}$ to $\mathcal{D}$. In other words, we focus on the situation where $\mathcal{S} \cap \mathcal{D} = \emptyset$ and $\mathcal{C} \subset \mathcal{S}$.

A border AS [15] is an AS in $\mathcal{S}$, with at least one direct link to an AS outside of $\mathcal{S}$. Set $\mathcal{B}$ is the set of all border ASes.

$$\mathcal{B} = \{b \in \mathcal{S} | \exists e = (b, x) \in E \text{ s.t. } x \notin \mathcal{S}\} \tag{3}$$

Following the work by Leyba et al. [15], we adapt the concept of choke potential to capture the concept of path-awareness. For that, consider a subset of border ASes, $\mathcal{B}' \subset \mathcal{B}$. Their *Cumulative Choke Potential* (CPP) is the fraction of outflow paths that they could choke together. The rigorous definition of $CCP_{\mathcal{S},\mathcal{D}}(\mathcal{B}')$ is shown in Equation 4, and due to normalization yields to $CCP_{\mathcal{S},\mathcal{D}}(\mathcal{B}') \in [0, 1]$.

$$CCP_{\mathcal{S},\mathcal{D}}(\mathcal{B}') = \frac{\sum\limits_{\substack{s \in \mathcal{S} \\ d \in \mathcal{D}}} f_{\mathcal{B}'}(s \to d)}{\|\mathcal{S}\| \cdot \|\mathcal{D}\|} \tag{4}$$

$$f_{\mathcal{B}'}(s \to d) = \begin{cases} 1, & \forall r, \text{ a path } s \to d, \exists b' \in \mathcal{B}', \textit{ s.t. } b' \in r \\ 0, & \text{otherwise} \end{cases}$$

From there, we define the set of censoring ASes $\mathcal{C}$ as a subset of $\mathcal{B}$ with cardinality $\|\mathcal{C}\| = N$, which cumulatively have the potential of choking the highest number of outflow paths. As an intuition, if all of these border ASes were to censor, ($\mathcal{C} = \mathcal{B}' = \mathcal{B}$) their cumulative choke potential would equal 1, which in turn results in $CRP_{\mathcal{C}}(\mathcal{S}, \mathcal{D}) = 0$. However, in reality if is often challenging for a country to ensure strict enforcement of censorship to all border ASes. Therefore, it is useful to select only a subset of these border ASes, thus defining the set $\mathcal{C}$. In addition, thus defined set $\mathcal{C}$ can provide

[1]We do however note that our metric can also be applied to a set of censoring ASes that are *a priori* labeled as such.

us with insights into centrality of a small number of border ASes to the connectivity to the global Internet. We perform experiments for various countries and various values N, which we comment on in Section IV.

Algorithm for *CRP* Metric. Given a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, it is necessary to define the set of source and destination nodes, $\mathcal{S} \subset \mathcal{V}$ and $\mathcal{D} \subset \mathcal{V}$, respectively. These sets can be chosen arbitrarily, or based on certain properties of the nodes, such as their country of origin. Finally, the set of censoring ASes $\mathcal{C} \subset \mathcal{V}$ can be defined in two ways.

The first method requires the set of censoring ASes $\mathcal{C}$ to be known *a priori*. Thus, the value of the metric could directly be determined by calculating the portion of paths that start in $\mathcal{S}$ and end up in $\mathcal{D}$, not passing through censoring ASes. The full algorithm pipeline of this method is laid out on Figure 1. The inputs to the algorithm—here the graph and sets $\mathcal{S}$, $\mathcal{D}$ and $\mathcal{C}$—are marked as gray, intermediate steps as purple, and the output and the final value of the metric as yellow.

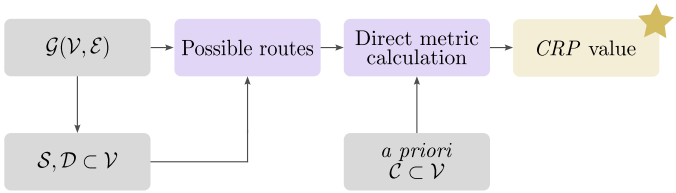

Fig. 1: *Censorship Resilience Potential* (CRP): the algorithm pipeline, with censoring ASes known *a priori*.

The second method determines the value of CRP using *Cumulative Choke Potential* (CCP). In this case, the underlying assumption is that censoring ASes would be border ASes from the set $\mathcal{S}$. The CCP value of the subset of them provides us with both the set of censoring ASes $\mathcal{C}$ with high potential of cumulatively choking outflow traffic, and the value of the final metric. The full algorithm pipeline of this method is laid out on Figure 2, with the same color-coding from Figure 1.

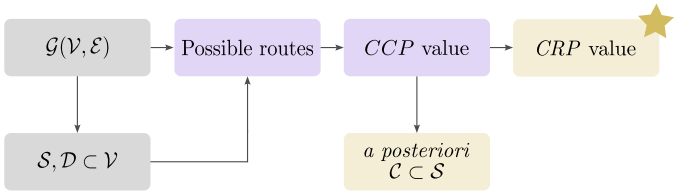

Fig. 2: *Censorship Resilience Potential* (CRP): the algorithm pipeline, where the *Cumulative Choke Potential* (CCP) is used for defining the set of censoring ASes $\mathcal{C}$ *a posteriori*.

***CRP* as Means for Comparative Analysis.** One of the desired properties of our metric is to be suitable for comparative analysis of different network models and in general Internet architectures. Therefore, if the set of censoring ASes is known in advance, the method depicted on Figure 1 should be applied to all Internet architectures. However, if the set $\mathcal{C}$ is not known *a priori*, the the set $\mathcal{C}$ should be defined independently. For a comprehensive analysis, the set $\mathcal{C}$ should be defined as per the pipeline on Figure 2 for all architectures independently. We employ this approach in our simulation, which we comment on more in Section IV, ultimately gaining results suitable for comparative analysis of BGP, a model of waypoints in the network, and SCION.

C. Global Reachability Potential

Our second goal is to gauge the potential of reaching the global Internet, while avoiding undesirable nodes in the network. We achieve this independently of the specific network topology, or the Internet architecture, by adopting the *Avoidability Potential* metric to this use case.

The Metric. Given the graph representing the global Internet, $\mathcal{G}(\mathcal{V}, \mathcal{E})$, it is possible that certain nodes are more central to for global connectivity than others. To measure how much influence a group of nodes $\mathcal{X} \subset \mathcal{V}$ have to nodes $\mathcal{S} = \mathcal{V} \setminus \mathcal{X}$ to establish paths with each other, we employ the *Avoidability Potential* metric, for convenience calling it *Global Reachability Potential*.

$$GRP_{\mathcal{X}}(\mathcal{S}, \mathcal{S}) = AP_{\mathcal{X}}(\mathcal{S}, \mathcal{S}), \ \mathcal{S} = \mathcal{V} \setminus \mathcal{X} \tag{5}$$

We note that the set of nodes desirable to be avoided, $\mathcal{X}$, can be predefined according to any property (e.g, their country of origin). However, we note that the metric only considers their capabilities in terms of global connectivity and routing influence, but does not take into account any of their underlying interests for tampering with communication on the global Internet, thus being agnostic to this aspect.

Example: Collateral Damage of Internet Censorship. Although censorship techniques are most prominently used for control of certain groups of nodes in the network [7], spillover effects to other nodes are possible, ultimately causing collateral damage outside of the desired area of influence [29]. For instance, in their work Acharya et al. conclude that countries that are known for their censoring activities might have influence on global reachability [18]. To that end, our metric can be applied to this case, providing an analysis of collateral damage of censorship at the AS level.

Example: Influence of Hegemonic Groups. As various authors observed, a small number of ASes are commonly used as global transit networks [30]. This leads to the problem of hegemonic influence of certain ASes—and not rarely countries—when it comes to global Internet reachability [8]. A *Global Reachability Potential* metric can directly be instantiated for this example, as it can analyze the potential of circumventing certain groups of nodes in the network, which might have hegemonic influence to the global Internet. Ultimately this provides comparative analysis of different network models, and quantitative evidence of how centralized or democratized Internet routing is.

IV. Experimental Evaluation

In this section we explain the evaluation setup and approach of the extensive experiments we perform, whose results we further comment on in the following Section V.

Overview of Analyzed Network Models. In our study, we apply our metrics to three different network models. First, we looked at the current BGP/IP Internet design, using the BGP routing protocol as a basis, and without considering any routing attacks. Second, we consider a scenario where a waypoint mechanisms are widely used to bypass undesirable nodes in the network. Lastly, we examined SCION, a path-aware next-generation Internet architecture, where each end-host is able to choose the whole path to its destination.

A. Datasets

For all our experiments, we use the datasets tht represents real and contemporary relationships between ASes.

AS Relationships. We model the network graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ using the CAIDA AS Relationships dataset [31]. This dataset provides us with the topology of the Internet, which we directly use for all our BGP simulations, and the waypoint model. For our SCION simulation we use the same topology, in order to have comparable and applicable results.

AS Country Origin. To obtain an accurate country origin of each AS, we utilize two detasets. First, we use the CAIDA AS to Organizations Mappings, which utilizes data from the Regional and National Internet Registries, ultimately providing us with the information of each AS's legal entity country origin [32]. Second, we use complementary RIPEStat Geo Map dataset [33], which includes the physical locations where an AS is announcing BGP prefixes. To provide realistic results, we assume that Tier-1 ASes are present in multiple countries, and that these branches are interconnected.

Waypoints in the Network. For the model of a network with waypoint nodes used for circumvention of undesirable nodes, we employ the anonymous dataset provided by MaxMind, which includes information of ASes that have previously been characterized as potential hosts of such—most notably VPN—services [34].

B. Country Network

We instantiate the *Censorship Resilience Potential* metric on nodes in the network based on their country of origin. The countries we select can be considered diverse based on a number of indicators: geographical location, size in terms of population and national network, and the *Internet Freedom Score* (IFS) [35].

Nodes Forming a Country Network. Let $\mathfrak{X}$ be a country of interest, and $\mathcal{K}$ a set of ASes with $\mathfrak{X}$ as their country of origin. We define $\mathcal{K}$ as following:

$$\mathcal{K} = \{k \in \mathcal{V} |\ \text{country}(k) = \mathfrak{X}\} \tag{6}$$

We use thus defined set $\mathcal{K}$ to define the country network for a given country, and exclude potential outliers. Namely, Guillermo et al. analyzed the global Internet, observing the presence of so-called *islands*—groups of nodes partitioned from the main Internet [36]. Extending their work from the global Internet to local national networks, we find that for all assessed countries the nodes from the set $\mathcal{K}$ do not form a connected component. We label the biggest connected component the *country network*, whereas all others we label as outliers for the given country network.

C. Network Model with Waypoints

Given that waypoint mechanisms are commonly used for bypassing censoring systems, we model a network where a set of ASes are hosting such services. With this model, we assume that any of these ASes could be used as an intermediate point for circumvention of undesirable nodes.

D. SCION Topology

Although SCION is already deployed, its current production network footprint does not yet reach the scale of the BGP infrastructure. Therefore, we construct the SCION topology based on the graph from the CAIDA AS Relationship dataset [31]. Thus constructed, it corresponds to the topology rooted in the real-world deployment, hence being both plausible and comparable to the BGP/IP network.

Core ASes. As Krähenbühl et al. already discussed, a global SCION deployment would likely have no more than 2000 core ASes [26]. As they already addressed, ASes with higher customer cone size would likely be regarded as most influential, thus being susceptible to being core ASes in the SCION network [26]. To that end, we use the customer cone size as heuristic for defining core ASes, while keeping links between them as per the initial graph, labeling them as *core links*. We use the network of core ASes when determining the value of the *Global Reachability Potential*, which we will elaborate in the remainder of this section.

Grouping into ISDs. Grouping nodes into an ISD is crucial for the application of the *Censorship Resilience Potential* metric, given that we instantiate this metric on a per country basis. We assume that all ASes connected to the network infrastructure of a given country $\mathfrak{X}$—or more precisely its *country network*, thus forming a connected component—would naturally group together, thus forming "national" ISD—or a group of ISDs in the general case. Additionally, an AS could belong to multiple ISDs, which is by design allowed in the SCION Internet Architecture [12]. However, we do not anticipate how ASes would split into groups globally, that way forming ISDs. Rather, given the country $\mathfrak{X}$, we assign the set of nodes in its country network to its ISD. Finally, we retain links between all ASes in an ISD defined in this way, and if a non-core AS in the ISD previously had a link to another AS that is now not in the ISD, we disregard this link.

E. Simulation on Diverse Network Models

In this section we lay down further implementation details of our extensive simulations.

1) Simulation of BGP. We perform inter-AS BGP simulation, based on the *routing tree* algorithm proposed by Gill et al. [37]. Using this algorithm, we determine preferred paths between *any* two ASes in the given Internet topology. These results are used for all further simulations of *Censorship Resilience Potential* and *Global Internet Reachability*, including

the definition of the set of censoring ASes $\mathcal{C}$ in the context of censorship resilience.

Defining Censoring ASes. For the BGP network model, we define the set $\mathcal{C} = \mathcal{C}_{\mathcal{BGP}}$ as a subset of border ASes in a country network, which cumulatively have the highest potential of choking the outflow traffic. As a viable heuristic for defining such most capable ASes in a network, we build upon the prior work done by Leyba et al., who note that the sum of choke potentials of all border ASes must sum into 1 [15]. Therefore, we define censoring ASes by attributing a potentially choked path only to the last border ASes on the path, in which case Leyba's assumption and analysis hold. Finally, given that the waypoint model relies on the exact same topology and routing algorithm as the BGP model, we use the same set of censoring ASes for the waypoint model.

2) Simulation of the Waypoint Network Model. A path between a source s and a destination d would be composed of two segments: from the source to a waypoint host, and further from the host all the way to destination. These segments are created based on the algorithm previously explained in Section IV-E1, and put together constitute a whole path between s and d. For a fixed source s and destination d, a multitude of available paths could exist.

3) Simulation of SCION. Once defined, the SCION topology defines the paths that would be discovered and is used to perform the inter-AS SCION simulations. In this section we further elaborate on how they are performed.

Defining Censoring ASes. Given a SCION network topology, we select a subset of border ASes with a predefined cardinality N, that have the highest customer cone size, thus defining the set of censoring nodes $\mathcal{C} = \mathcal{C}_{\mathcal{SCION}}$. This method represents a viable heuristic for selecting nodes that have the highest potential to cumulatively choke outflow paths, as it encompasses the number of customers each border ASes has in the country network.

Censorship Resilience Potential. When it comes to the CRP metric, our goal is to determine the number of *outflow* paths originating in a country network—i.e. national ISD in the SCION simulation—that could circumvent censoring ASes. Given that a border ASes in a SCION topology defined in such a way could only be one of the core ASes, it is enough to determine whether a path—that originates in the national ISD, leaves it through its core ASes, and along the way circumvents the censoring ones—exists, since each source has full freedom to select the whole end-to-end path [12]. These paths still have to follow the business relationships [28], since SCION does enable network operators to preserve this property. Finally, with non-core ASes between different ISDs, CRP metric for SCION would be higher, as this would increase the number of outflow paths that are at the disposal to end hosts [12]. Nonetheless, as already discussed in Section IV-D, we do not consider such links in our analysis.

Global Reachability Potential. In a SCION network, each country network (or ISD) forms a connected component. To assess interconnectedness between ISDs, we focus on the graph of core ASes, derived from the current Internet topology, as explained in Section IV-D. After selecting countries for whole influence to global reachability we want to analyze, we determine how many of other core ASes could still reach other, while circumventing ASes from the given countries.

V. Experimental Results

We apply our metric to two distinct cases: censorship resilience of various countries, and global Internet reachability.

A. National Censorship Resilience Potential

We evaluate the *Censorship Resilience Potential* metric on BGP, waypoint model, and SCION, using multiple diverse countries. Ultimately, we show that the network topology plays a crucial role when it comes to Internet censorship. We report the results on Figure 3, and in the remainder of this section we provide their substantive analysis.

Border ASes as Central Choking Points. First, our results indicate that a surprisingly small number of a country's border ASes could choke a surprisingly high percentage of outflow paths in the current Internet. The number of border ASes as low as 20 can choke as high as 50% outflow paths in cases we analyzed, regardless of the country network size. For instance, in BGP with only one AS, the United States have the potential of choking 26% of the outflow paths ($CRP = 0.74$).

To further complement our analysis, we present network statistics in Table I, showing the size of the network, and the number of border ASes in all analyzed models. For instance, comparing the Iranian BGP-network topology with the Swiss BGP-network topology in Table I, one can see that Iran has a rather centralized model, i.e., fewer of them are border ASs in the network—which further confirms its centralized model of censorship [3], [17]—whereas for the Swiss case, there is a high density of border ASes. This has an impact on the CRP value for both countries: a single AS from Iran has the potential of choking as high as 55% of the outflow paths ($CRP = 0.45$), whereas the same number of censoring ASes in Switzerland would have the potential of choking more than two times less, or 21% of the outflow paths ($CRP = 0.79$). Similar reasoning can be applied to other countries and the number of ASes that could collude in their censoring effort.

Network Topology Matters. The number of border ASes is not the only influential factor, but also the connectedness of ASes within the network, and in general how prominent each border AS is as a transit network. The network topology of a country immensely influences its censorship resilience, and in most of the cases regardless of the Internet architecture. In other words, although the number of exit points from a country network might be high, the routing among those exit points might not be democratized, and the network could still depend on a low number of ASes. As an example, the United States—globally the biggest network, with almost 18 thousand nodes—has relatively low value of CRP of 0.32 with only 5 censoring ASes, given their influence on routing.

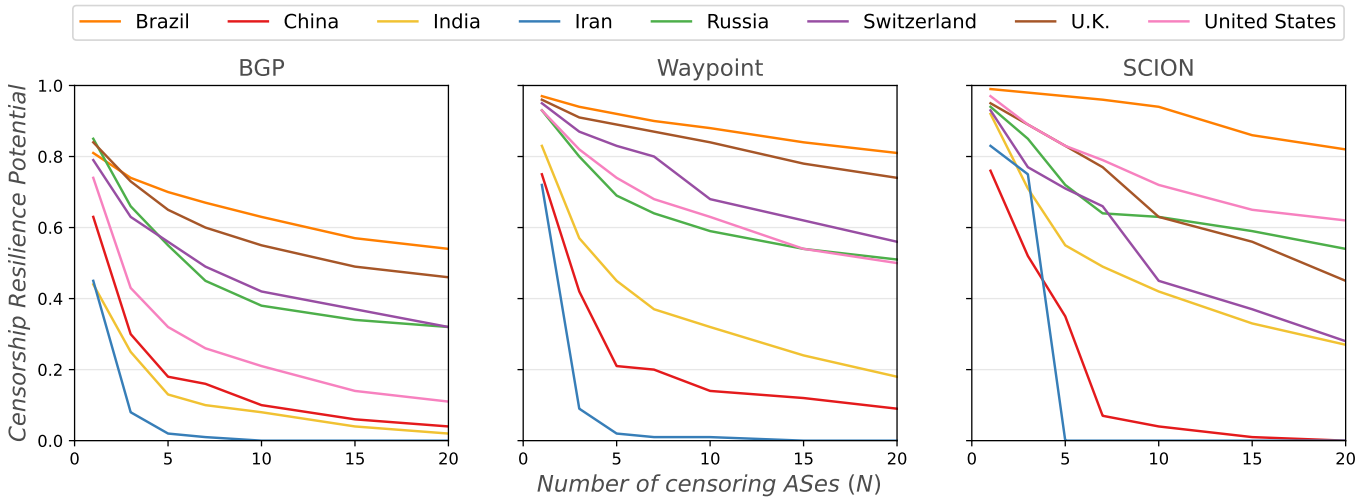

Fig. 3: *Censorship Resilience Potential* for BGP, waypoint model, and SCION, presented for various countries and varying number of censoring ASes, N.

	Number of ASes	Border ASes (BGP/Waypoint)	Border ASes (SCION)
Brazil	8174	2285 (28%)	243 (3%)
China	534	94 (18%)	21 (4%)
India	2537	209 (8%)	36 (1%)
Iran	481	25 (5%)	5 (1%)
Russia	4957	1139 (23%)	82 (2%)
Switzerland	654	308 (47%)	28 (4%)
U.K.	1562	861 (55%)	62 (4%)
United States	17934	2173 (12%)	236 (1%)

TABLE I: Country network statistics: absolute number of ASes in each country network, including how many of them are at the border of the network for all analyzed architectures.

Path-selection and Censorship Resilience. Various paths may be available when leaving a country, which could be leveraged to enhance censorship resilience. However, they often go unused. In path-aware technologies like SCION, the likelihood of a single autonomous system (AS) controlling all outflow paths is significantly lower compared to BGP. A similar pattern can be observed in the waypoint model, where multiple intermediate nodes can be accessed for the final circumvention of undesirable nodes.

It should be noted that the number of Border ASes varies substantially between the BGP network and the SCION network, due to the technical details of the underlay protocol and governance models (see Table I). However, although the absolute number of border ASes is lower on SCION relative to BGP, routing is more democratized in SCION, and due to end-to-end multiple path-selection the network does not depend on any of the nodes. In summary, our quantitative analysis shows that path-selection has the potential to drastically mitigate the influence of censoring border ASes as choking points, by providing the choice of circumventing undesirable ASes.

B. Global Internet Reachability

We evaluate the *Global Reachability Potential* metric presented in section III-C on BGP, waypoint model, and SCION. The groups of influential countries we analyze here follow the prior work, which concludes that certain countries either have detrimental potential for Internet reachability [8], [30], or the potential to impose collateral damage due to their censorship activities. Note that once again we use the word *potential*, as we do not comment on the underlying interests of either these AS or countries to be central for global Internet reachability, but rather comment on their capabilities based on the current Internet topology. The results are presented in Table II.

Nodes Centrality. The obtained results show that the ASes originating from the United States and the Five Eyes countries[2] are transit nodes for more than 40% of all paths originating and terminating in a node from another country. Similarly, our findings show that ASes from the European Union are also partially central for the global reachability. We also report the value of *Global Reachability Potential* of three countries most commonly mentioned in the context of censorship: Iran [3], China [20] and Russia [4], as they are commonly analyzed in the context of collateral damage of their censoring activities [18]. Our analysis concludes that they do not have a significant influence on the global reachability in any of the three analyzed Internet architectures, as they are not centrally positioned in the current Internet topology.

Path-selection and Global Reachability. As the results from the Table II indicate, there is a variance between the three analyzed Internet architecture. Specifically, the impact of the examined groups on global Internet reachability is lower in the waypoint model compared to BGP. In this model, various waypoint hosts can serve as intermediate nodes, suggesting

[2]The Five Eyes is an alliance of five countries: Australia, Canada, New Zealand, the United Kingdom, and the United States.

	BGP	Waypoint	SCION
United States	0.59	0.92	0.9951
Five Eyes	0.52	0.88	0.9941
European Union	0.87	0.98	0.9975
Iran, China, Russia	0.98	0.99	0.9995

TABLE II: Results of the *Global Reachability Potential*, with various groups of countries analyzed for global influence.

the existence of multiple paths, consequently increasing the potential to avoid undesirable ones. Nonetheless, given that a number of the waypoint nodes are originating from the United States, the *Global Reachability Potential* is not fully achieved even with intermediate nodes. Finally, our results indicate that SCION's end-to-end path-awareness could provide means to circumvent undesirable nodes altogether, thus achieving high value of the *Global Reachability Potential* for all analyzed groups of countries.

VI. Related Work

Country Network Analysis. In the context of Internet censorship, a range of authors have previously confirmed the importance of network topology of a country to Internet censorship success. For instance, Ensafi et al. observe that Tor traffic—commonly censored in China—is not being censored when entering the country via CERNET, the Chinese Educational and Research Network, providing insights that network topology and routing could play a crucial role in Internet censorship resilience [20]. Analyzing Iran, Gill et al. assess the country network as centralized [7], whereas Salamatian et al. reveal the limited number of the country network's direct links to foreign ASes, concluding that BGP could be strategically employed for censorship purposes [17]. On a similar note, Wählisch et al. investigated the network architecture of Germany, employing a sector-based approach to classify nodes within the network, and assessing their betweenness centrality [38].

Control of National Outflow Traffic. Expanding on the analysis of country networks, researchers proposed new ways to understand how networks are structured and controlled. Roberts et al. introduced a measure for the network complexity, ultimately revealing underlying network properties that could indicate censorship capabilities of a country [16]. In addition, they assess points of control in countries around the world—ASes that control at least 90% of the country IP address space [16]. Similarly, Leyba et al. took a similar approach by looking at the choke points in the networks of various countries, revealing that the number of nodes that could choke huge fraction of outflow paths are not only low, but in general also decrease over time [15].

Global Internet Reachability. Other researchers studied global Internet reachability, evaluating betweenness centrality on a global scale, identifying nodes and countries that hold pivotal positions in facilitating global Internet connectivity [8], [30]. Furthermore, these analyses explored the potential repercussions of censorship efforts, elucidating the collateral damage that could arise from censorship activities [18].

Broadened Prior Work and Contributions. Our proposed metric builds upon these prior contributions to devise a comprehensive tool for quantifying both censorship resilience and global Internet reachability. Notably, its adaptability spans diverse network models and topologies, and captures path-aware Internet architectures. Finally, in the context of censorship resilience it does not require *a priori* defined censoring ASes.

Next-generation Internet Architectures and Censorship. We underscore the significance of scrutinizing next-generation Internet architectures in the context of censorship and Internet reachability. While the studies by Kohler [39] and Wrana et al. [21] have delved into this subject using a qualitative approach, our main contribution is to provide a quantitative metric that can be used for comparative analysis.

VII. Discussion

Routing Attacks on BGP. When evaluating both *Censorship Resilience Potential* and *Global Reachability Potential* metrics on BGP and the waypoint model, we determine genuine paths between any two nodes, without incorporating any routing attacks by malicious actors in the network. A potentially malicious node in the network could launch a series of routing attacks [40], redirecting traffic, and thus compromising both censorship resilience and access to the global Internet.

Waypoint Model on the Internet. Our waypoint model provides an insight into the impact of systems—for instance VPN connections—used for circumventing undesirable nodes. However, this model may convey a simplistic idea that the censorship circumvention depends on the waypoint service providers, whereas in reality censoring entities simply block IP addresses from such known providers. Nonetheless, it provides quantitative evidence of the benefits of multiple paths for Internet routing.

Deployment of SCION. BGP is deployed as the sole inter-domain routing protocol, thus making our results on BGP directly applicable to the current Internet. On the other hand, SCION is deployed on a much smaller scale [26], and therefore the results of this work provide a future outlook. In addition, the waypoints—for instance using VPN connections—can be used in addition to SCION, as their deployments are orthogonal and compatible.

Policy Impacts. In our analysis we refrain from making normative statements about Internet censorship or global Internet reachability. Instead, we offer an objective metric for their evaluation, serving as a quantitative tool for assessing these aspects of the Internet. We believe that this work can offer a quantitative basis for policy analysis, informing decision-makers about the design, development and deployment of network technologies.

VIII. Conclusion

In this paper, we have shown that the network topology and Internet architecture can influence the potential of a country's

network to exhibit resilience to Internet censorship, as well as their impact on global Internet reachability and dependence on certain nodes in the network. We have proposed a novel approach that utilizes quantitative network metrics to evaluate these aspects of today's Internet, which we evaluated on contemporary Internet topologies of various countries, utilizing the Border Gateway Protocol (BGP) as a routing protocol, a model of waypoints in the network commonly used through Virtual Private Networks (VPNs), and the SCION path-aware Internet architecture. Our findings indicate that the network topology matters when it comes to those resilience aspects of the Internet, as well as that a path-aware Internet architecture has the potential to democratize routing on the global Internet by reducing centralization and to facilitate a higher level of censorship resilience.

REFERENCES

[1] J. L. Hall, M. D. Aaron, A. Andersdotter, B. Jones, N. Feamster, and M. Knodel, "A Survey of Worldwide Censorship Techniques," Internet Engineering Task Force, Internet-Draft draft-irtf-pearg-censorship-09, Jan. 2023, work in Progress.

[2] C. S. Leberknight, M. Chiang, H. V. Poor, and F. Wong, "A taxonomy of internet censorship and anticensorship," *Fifth International Conference on Fun with Algorithms*, 2010.

[3] S. Aryan, H. Aryan, and J. A. Halderman, "Internet censorship in iran: A first look," in *3rd USENIX Workshop on Free and Open Communications on the Internet (FOCI 13)*. Washington, D.C.: USENIX Association, Aug. 2013.

[4] R. Ramesh, R. S. Raman, M. Bernhard, V. Ongkowijaya, L. Evdokimov, A. Edmundson, S. Sprecher, M. Ikram, and R. Ensafi, "Decentralized control: A case study of Russia," in *Network and Distributed System Security*. The Internet Society, 2020.

[5] K. Singh, G. Grover, and V. Bansal, "How India censors the web," in *Web Science*. ACM, 2020.

[6] B. Marczak, N. Weaver, J. Dalek, R. Ensafi, D. Fifield, S. McKune, A. Rey, J. Scott-Railton, R. Deibert, and V. Paxson, "An analysis of China's "Great Cannon"," in *Free and Open Communications on the Internet*. USENIX, 2015.

[7] P. Gill, M. Crete-Nishihata, J. Dalek, S. Goldberg, A. Senft, and G. Wiseman, "Characterizing web censorship worldwide: Another look at the OpenNet Initiative data," *Transactions on the Web*, vol. 9, no. 1, 2015.

[8] J. Karlin, S. Forrest, and J. Rexford, "Nation-state routing: Censorship, wiretapping, and BGP," *arXiv*, 2009.

[9] A. Shah, R. Fontugne, and C. Papadopoulos, "Towards characterizing international routing detours," in *Proceedings of the 12th Asian Internet Engineering Conference*, ser. AINTEC '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 17–24.

[10] A. Edmundson, R. Ensafi, N. Feamster, and J. Rexford, "Nation-state hegemony in internet routing," in *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, ser. COMPASS '18. New York, NY, USA: Association for Computing Machinery, 2018.

[11] Y. Rekhter, S. Hares, and T. Li, "A Border Gateway Protocol 4 (BGP-4)," RFC 4271, Jan. 2006.

[12] L. Chuat, M. Legner, D. A. Basin, D. Hausheer, S. Hitz, P. Müller, and A. Perrig, *The Complete Guide to SCION - From Design Principles to Formal Verification*, ser. Information Security and Cryptography. Springer, 2022. [Online]. Available: https://doi.org/10.1007/978-3-031-05288-0

[13] M. C. Tschantz, S. Afroz, Anonymous, and V. Paxson, "SoK: Towards grounding censorship circumvention in empiricism," in *Symposium on Security & Privacy*. IEEE, 2016.

[14] S. Khattak, T. Elahi, L. Simon, C. M. Swanson, S. J. Murdoch, and I. Goldberg, "SoK: Making sense of censorship resistance systems," *Privacy Enhancing Technologies*, vol. 2016, no. 4, pp. 37–61, 2016.

[15] K. G. Leyba, B. Edwards, C. Freeman, J. R. Crandall, and S. Forrest, "Borders and gateways: Measuring and analyzing national as chokepoints," in *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*, ser. COMPASS '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 184–194.

[16] H. Roberts, D. Larochelle, R. Faris, and J. Palfrey, "Mapping local Internet control," in *Computer Communications Workshop*. IEEE, 2011.

[17] L. Salamatian, F. Douzet, K. Salamatian, and K. Limonier, "The geopolitics behind the routes data travel: a case study of Iran," *Journal of Cybersecurity*, vol. 7, no. 1, p. tyab018, 08 2021.

[18] H. B. Acharya, S. Chakravarty, and D. Gosain, "Few throats to choke: On the current structure of the internet," in *2017 IEEE 42nd Conference on Local Computer Networks (LCN)*, 2017, pp. 339–346.

[19] X. Xu, Z. M. Mao, and J. A. Halderman, "Internet censorship in China: Where does the filtering occur?" in *Passive and Active Measurement Conference*. Springer, 2011, pp. 133–142.

[20] R. Ensafi, P. Winter, A. Mueen, and J. R. Crandall, "Analyzing the Great Firewall of China over space and time," *Privacy Enhancing Technologies*, vol. 2015, no. 1, 2015.

[21] M. Wrana, D. Barradas, and N. Asokan, "The spectre of surveillance and censorship in future internet architectures," *arXiv*, 2024.

[22] A. Edmundson, R. Ensafi, N. Feamster, and J. Rexford, "A first look into transnational routing detours," in *Proceedings of the 2016 ACM SIGCOMM Conference*, ser. SIGCOMM '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 567–568.

[23] ——, "Characterizing and avoiding routing detours through surveillance states," *CoRR*, vol. abs/1605.07685, 2016. [Online]. Available: http://arxiv.org/abs/1605.07685

[24] J. Obar and A. Clement, "Internet Surveillance and Boomerang Routing: A Call for Canadian Network Sovereignty," *SSRN Electronic Journal*, 2013.

[25] S. Bechtold and A. Perrig, "Accountability in future internet architectures," *Commun. ACM*, vol. 57, no. 9, p. 21–23, sep 2014.

[26] C. Krähenbühl, S. Tabaeiaghdaei, C. Gloor, J. Kwon, A. Perrig, D. Hausheer, and D. Roos, "Deployment and scalability of an inter-domain multi-path routing infrastructure," in *Proceedings of the 17th International Conference on Emerging Networking EXperiments and Technologies*, ser. CoNEXT '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 126–140.

[27] B. Trammell, J.-P. Smith, and A. Perrig, "Adding path awareness to the internet architecture," *IEEE Internet Computing*, vol. 22, no. 2, pp. 96–102, 2018.

[28] L. Gao and J. Rexford, "Stable internet routing without global coordination," *IEEE/ACM Transactions on Networking*, vol. 9, no. 6, pp. 681–692, 2001.

[29] Sparks, Neo, Tank, Smith, and Dozer, "The collateral damage of Internet censorship by DNS injection," *SIGCOMM Computer Communication Review*, vol. 42, no. 3, pp. 21–27, 2012.

[30] R. Fontugne, A. Shah, and E. Aben, "The (thin) bridges of AS connectivity: Measuring dependency using AS hegemony," *CoRR*, vol. abs/1711.02805, 2017. [Online]. Available: http://arxiv.org/abs/1711.02805

[31] Center for Applied Internet Data Analysis, "AS Relationships (serial-2)."

[32] ——, "Inferred AS to Organization Mapping Dataset."

[33] RIPEstat, "RIPE Stat."

[34] MaxMind, "GeoIP2 Anonymous IP Database."

[35] Freedom House, "Internet Freedom Scores."

[36] G. Baltra and J. Heidemann, "What is the internet? (considering partial connectivity)," University of Southern California, Tech. Rep., 2022.

[37] P. Gill, M. Schapira, and S. Goldberg, "Modeling on quicksand: Dealing with the scarcity of ground truth in interdomain routing data," *SIGCOMM Comput. Commun. Rev.*, vol. 42, no. 1, p. 40–46, 2012.

[38] M. Wählisch, T. Schmidt, M. de Brün, and T. Häberlen, "Exposing a nation-centric view on the german internet – a change in perspective on the as level," in *International Conference on Passive and Active Network Measurement*, vol. 7192, 03 2012.

[39] K. Kohler, "One, Two, or Two Hundred Internets? The Politics of Future Internet Architectures," *CSS Cyberdefense Reports*, 2022.

[40] S. L. Murphy, "BGP Security Vulnerabilities Analysis," RFC 4272, Jan. 2006.