Explanation of Our Data

SecurityScorecard

SecurityScorecard collects a broad range of threat intelligence data across the public Internet and dark web. Our global security intelligence engine, called ThreatMarket™, continuously collects and analyzes a broad range of highly relevant, but non-intrusive, cybersecurity signals for millions of digital assets across the internet. This white paper provides detail on the active and passive collection methods and signal types that are in use by the SecurityScorecard solution. The breadth and depth of the security data discussed is the foundation from which SecurityScorecard can deliver the most comprehensive cybersecurity ratings in the industry.

Signals Collection is the function of collecting information about every Internet connected device, resource, organization and service provider in the world. This information is analyzed to track the historical and current state as well as predict the future state of the Internet. Our mission is to deeply understand the Internet so that we can improve the collective security of companies that use it.

We employ both passive and active collection means to obtain our data. Active collection involves initiating a connection towards some remote host and participating in some initial part of their protocol. Passive collection can be performed in two ways; either a remote host connects to us or we obtain copies or summaries of some protocol transaction from a network sensor or intermediary device. The quality of data collected is directly proportional to the diversity of collection locations around the Internet combined with the frequency of data collection. Below is a listing of the passive and active techniques we employ.

Active

• **Service discovery** - Using network service query techniques, SecurityScorecard collects information on active services that are running on publicly facing hosts. Services are part of the protocols that allow users to communicate with Internet-based applications including web servers, application servers, or any addressable Internet hosts). Service discovery occurs via a 2-step process: (1) find all hosts that are communicating on the public Internet, (2) for all active hosts, find all available services (e.g., web service, database service, application service). Service discovery is critical to understanding service and port-based vulnerabilities on a host.

• **Content capture –** This process performs additional, non-intrusive, network-based discovery to uncover potential vulnerabilities of active services. The content capture process uses publicly available network protocols that can discover cybersecurity exposure for active network services. SecurityScorecard maintains an entire team of cybersecurity experts that have a deep understanding of network services and that build broad content capture capabilities that uncover service based vulnerabilities across the broad range of networked services found on the public internet.

• **Fingerprinting –** As an extension of service discovery and content capture, fingerprinting performs deeper inspecting to understand and map the type and version of active services. For example, fingerprinting might detect and document that a web server is running on Apache instead of Microsoft IIS web server software. Furthermore, fingerprinting might detect and document that additional web services are in use (e.g., Wordpress, SSL, PHP, and much more). Furthermore, fingerprinting might detect what version a specific web service is running (e.g., PHP/7.1.14). Fingerprinting is an important process to collect information that helps narrow application-based vulnerabilities that may reside on a host.

• **Configuration enumeration –** As an extension of service fingerprinting, configuration enumeration looks to understand additional service configuration attributes that are obtainable using non-intrusive means. For example, SecurityScorecard may use a configuration enumeration to uncover attributes of a networked service that may have vulnerability implications.

• **Botnet interrogation –** Using network interrogation techniques, SecurityScorecard collects important data on botnet infections and networks. Botnets are a network of devices (e.g., servers, hosts, IoT devices) that have been infected by malicious software and operate together without the owner's knowledge. In most cases, botnets exist for nefarious purposes. Key data collected via botnet interrogation include botnet peer lists (i.e., lists of devices that are participating in a botnet) or that are acting as a botnet command and control channel (i.e., devices that issue commands to compromised peers).

• **Certificate discovery –** Using active network discovery techniques, SecurityScorecard collects information about issued or in use encryption certificates. Networked based data encryption is a fundamental security control that helps protect the confidentiality and integrity of data traversing the Internet. The exchange of cryptographic keys is a foundation of networkedbased data encryption. X.509 is an industry standard that defines how public key certificates are formatted. X.509 keys are used to encrypt traffic for multiple network protocols, including HTTPS (i.e., the protocol used for browser-based web data encryption). Understanding the information available in public X.509 certificates is important to uncover issues with the encryption of Internet-based applications (e.g., encryption is not in use or a certificate has been revoked).

• **Name resolution –** Using the Domain Name System (i.e., DNS), SecurityScorecard collects information about the naming and addressing of Internet-based hosts (i.e., computers, servers, IoT). DNS is a foundation of the Internet and provides a map between easy to remember host names and associated numerical IP addresses. DNS provides a capability to query the DNS directory for publicly available naming information about a host (e.g., hostname and IP address). SecurityScorecard's ability to collect and analyze DNS name resolution mappings help uncover DNS misconfiguration issues or DNS misuse.

• **Names and numbers –** As an extension of DNS name resolution collection, SecurityScorecard collects additional host information that is available publicly via the domain name system. Using DNS queries, SecurityScorecard collects public information that DNS has available about Internet-based hosts (e.g., registrant contact info, administrative contact info, and technical contact info). DNS provided information acts as a foundation of SecurityScorecard IP attribution where the platform maps Internet-based hosts to the companies that own and manage them.

Passive

• **Honeypots –** SecurityScorecard maintains a network of nonintrusive “honeypots” that detect Internet-based malware. Honeypots often appear as legitimate networked hosts but are deployed as a decoy to attract illicit activities. Honeypots often introduce a security trap for multiple emulated network-based services. Once a hacker compromises a honeypot, security researchers can understand, in great depth, how malware use by bad Internet actors. Information collected from honeypots is used by SecurityScorecard to issue malware security advisories and to detect and report specific issues of active malware on hosts.

• **Sinkholes –** SecurityScorecard maintains a network of nonintrusive “sinkholes” that detect Internet-based malware. Our sinkhole network ingests millions of malware signals from commandeered Command and Control (C2) infrastructures from all over the world. The system processes incoming data sinkhole data, and attributes detected malware to corporate entities. Information collected from sinkholes is used by SecurityScorecard to issue malware security advisors and to detect and report specific issues of active malware on hosts.

• **Passive DNS –** SecurityScorecard utilizes passive DNS monitoring techniques to understand both legitimate and illegitimate use of the domain name system. Manipulation of DNS has become a frequent attack vector for bad actors. An example attack vector in this area is called DNS cache poisoning where domain records are corrupted and result in a legitimate DNS request being re-directed to an illegitimate host. Information collected from passive DNS sensors help detect and report DNS related security issues within a company.

• **Advertising exchange –** SecurityScorecard uses passive data collection techniques to monitor advertising exchange networks to understand and detect specific browser-based and operating system-based security concerns. The prevalence of online advertising networks has introduced numerous advertisement related attack vectors on the Internet. SecurityScorecard’s sensors use non-intrusive monitoring techniques to help detect and report the nefarious use of advertising exchange networks including digital fraud.

• **SPAM senders –** SecurityScorecard uses passive data collection techniques to uncover hosts that are generating SPAM (i.e., unsolicited network messages). In many cases, hosts that are generating SPAM are compromised by botnet-related malware and receiving Spamming instructions from a command and control channel, without the knowledge of the host owner. The most common form of SPAM is email; however, many other Internet-based applications fall prey to SPAM attacks. Information collected about SPAM senders helps SecurityScorecard detect and report issues of compromised hosts that are generating SPAM

• **Credential dumps –** SecurityScorecard uses specialized data collection techniques that uncover unauthorized publication of credentials that would unlock corporate systems or specific user accounts. Each year, hundreds of millions of credentials are stolen, leaked, and shared freely within the hacker community. Credential dump monitoring identifies exposed passwords from data leaks, keylogger dumps, database dumps, and a variety of other types of information leak. SecurityScorecard uses this data report issues that indicate exposed corporate credentials on the Internet.

• **Registered emails –** SecurityScorecard uses specialized data collection techniques that uncover the unscrupulous use of corporate email addresses on fraudulent sites. For example, data collection techniques can uncover a legitimate customer service email address published on an illegitimate site\ disguised as part of a phishing scheme to collect sensitive information like personal identifying information, usernames, and passwords. SecurityScorecard uses this data to report issues that indicate misuse of legitimate corporate email addresses.
Analysis of this data is complex due to the nature and scale of the Internet. Remote hosts can present different information depending on where you reside within the world. The network itself can take youto different hosts depending on your location, the current load of the network or the details of your request. Some portion of the Internet is not going to be visible to your sensors for various reasons: system failure, route fluctuations, bad configurations or state security and political restrictions. All of these challenges must be accounted for and overcome in performing analysis of this data. Analysis must also take into account errors and omissions present in the data, below is a listing of some of the sources of these errors and omissions.

Network Partitioning

• **Segmentation –** The ability to segment networks is a core pillar of network security. By design, many organizations segment networks to restrict traffic to systems that maintain sensitive corporate information. A common network partitioning technique is to build a DMZ (i.e., demilitarized zone or perimeter network) that provides an air gap between the Internet and internal networks and systems. Although network partitioning schemes hide many corporate digital assets, SecurityScorecard passive sensors can often detect important security concerns not surfaced by other means.

• **Failure –** A wide variety of operational failures can impact the availability of networked systems. It is common for networked systems to become unavailable because of hardware failure, poor system performance, or system misconfiguration. Failure of a networked system can cause other systems to adjust their configurations in such a way to dynamically route traffic around the failed system. A similar change in routing tables will occur after a system returns to an available state. The nature of dynamic routing protocols allows routing table updates to occur without human interaction. SecurityScorecard's data collection has been designed to monitor and map a wealth of important security data across millions of assets on the internet, but to also be resilient to the dynamic nature of networks when system failures do occur.

• **Reconvergence –** Convergence (and re-convergence) are the process where dynamic routing tables are updated to reflect system state changes (e.g., available and not available). Convergence requires that all relevant routers contain the same routing information for all systems on the network. Reconvergence requires that relevant routers get updated with identical routing information after a state change. The time it takes for re-convergence can vary greatly depending on the routing protocol in use. The timing of re-convergence can span from near-instantaneous (for a faster converging routing protocol like OSPF) to many minutes (for a slower converging routing protocol like RIP). In some cases, re-convergence will not occur without human interaction because of some underlying configuration issue. SecurityScorecard's data collection has been built to adapt to the dynamic nature of routing protocols and reconvergence.

• **Congestion –** Congestion (e.g., heavy traffic or system performance loads) can appear to other networked systems as an outage. The concepts discussed above in the areas of failure and re-convergence are relevant topics when a routing protocol treats a congested system as a failed system. SecurityScorecard's data collection has been built to adapt to the dynamic nature of routing protocols that interprets a system as unavailable because of congestion.

Stale Records

• **Misconfiguration –** Misconfiguration by system operators is a frequent source of network outage and errors. System operators frequently introduce network configuration errors by accident (e.g., inadvertently blocking a system, protocol, or application) or because of a lack of knowledge of how to properly configure a device (e.g., writing dynamic routing configuration policy). Misconfiguration, by system operators, can result in unavailable or poorly performing systems. SecurityScorecard's data collection has been built to adapt to potential system misconfigurations by an operator.

• **Lag –** The time lag of system configuration changes can affect the recognized state of a networked system. For example, a system may appear to be offline or congested while going through a software configuration change or other system maintenance activity. Many organizations require changes to occur during specific change control windows to minimize the impact of configuration changes. SecurityScorecard's data collection has been built to adapt to the time lag, and potential state changes from time lag, of configuration changes.

• **Omission –** Omissions occur when critical networking data is left out of pertinent records either intentionally or by accident. Data omissions can have unintended consequences including the introduction of network errors, system failure, or introducing unnecessary confusion. SecurityScorecard's data collection has been built to detect and report omissions in data records that should be present.

The Internet is not homogeneous, it is a disparate connection of independently operated connected networks. Network operators run their portions of the Internet with a focus on their primary line of business and its objectives. For instance a mobile phone operator will use more of the IPv6 space and optimize their network for high rates of endpoint roaming and handoff. A transit network is more focused on network availability, throughput and cost-based routing. Put in another way the networks of that compose the Internet look and behave differently. When analyzing collected data from the Internet it is important to understand what type of network the data is coming from to help guide the analysis. Below is a list of some of the types of network operations.

Types of Networks

- Residential ISP - retail broadband Internet access to consumers
    - High fan-out, or oversubscription, of last mile access
    - Dynamic IPs (DHCP) are used for customer modems/routers
    - Provider edge routers might encode regional names into interfaces
    - IP assignments are almost always IPv4 with a VLSM of /32
    - IPs are never assigned to the customer
- Business ISP - Internet access to small and medium business
    - Medium to no fan-out, or oversubscription, of last mile access
    - IPs can be served dynamically (DHCP) or statically assigned
    - Provider edge routers might encode regional or company names into interfaces

    - IP assignments are almost always IPv4 with a VLSM range of /24 - /32
    - IPs are almost never assigned to the customer
- Premise networks - local networks run by enterprise branch and campus offices
    - High fan-out is possible in small offices, no fan-out in larger offices
    - IPs are statically assigned for large offices, small offices may get dynamic assignment
    - Provider edge routers might encode regional or company names into interfaces
    - Networks might be multi-homed, have an ASN number and therefore could receive IP assignments from a numbering authority
- Content Distribution Networks - networks focused on distributing Internet content
    - These networks are collocated with ISPs based on customer density
    - Networks will have an ASN, be multi-homed and run BGP announcements
    - IPs will be assigned, or could be allocated, from a numbering authority
    - Network router interfaces might encode their location and function
- Transit Providers - carry traffic between network operators

Networks could be regional or global

Networks will have multiple ASNs, be multi-homed and participate in BGP

    - Networks will receive assignments and allocations from a numbering authority

- Cloud Service Providers - offer network based compute, storage, etc. services
    - Networks could be regional or global
    - Networks will have multiple ASNs, be multi-homed and participate in BGP
    - Networks will receive assignments and allocations from a numbering authority
- Mobile Operators - provide voice and data access to mobile subscribers
    - Networks are generally national or global
    - Heavy usage of IPv6 in 3G and 4G+ environments
    - Networks will have multiple ASNs, be multi-homed and participate in BGP
    - Networks will receive assignments and allocations from a numbering authority

Networks by their nature are dynamic and can morph at any time due to system failure, poor performance, and misconfiguration. SecurityScorecard has built a robust data collection capability that uses an adaptive network of active and passive collection sensors that can learn potentially hundreds of security-related signals that have relevance to assessing the security posture of networked systems and companies. SecurityScorecard's network of sensor technology is a pillar of the solution and enables the following key capabilities:

(1) discovering networked systems, protocols, and applications on the public Internet, (2) attributing systems to the companies that own them, and (3) detecting cybersecurity concerns on networked systems.
Data collected by the processes and techniques presented in this white paper are fundamental the SecurityScorecard rating system to collect and analyze a broad range of relevant cybersecurity data. SecurityScorecard's data collection and security rating capabilities are helping thousands of companies counter advanced threats using a broad range of security telemetry, intelligence, and collaboration tools that actively identify vulnerabilities, quickly remediate exploits, and continuously monitor the cyberhealth of an ecosystem of companies.