ABSTRAK

DESAIN DAN IMPLEMENTASI SISTEM CYBER THREAT INTELLIGENCE BERBASIS WEB DATA EXTRACTION PADA REDDIT

Oleh

FATHAN ANANTA NUR

NIM : 18219008

Ancaman siber telah merambat menjadi masalah sosial. Perusahaan telah banyak berinvestasi dalam pertahanan siber dalam upaya memerangi hal ini. *Cyber Threat Intelligence* (CTI) adalah salah satu mekanisme pertahanan yang akan dibahas. Organisasi baru-baru ini mulai melakukan investasi yang signifikan dalam pengembangan CTI untuk memerangi meningkatnya ancaman serangan siber. CTI dicirikan sebagai sekelompok data faktual yang mencakup konteks, metode ancaman, indikator, dan penanggulangan potensial. Komunitas dan forum *hacker* atau *bad actor* dapat memberikan CTI proaktif yang substansial. Forum, jika dibandingkan dengan platform lain, menawarkan metadata terlengkap, data permanen, dan puluhan *tools*, *technique*, dan *procedure* (TTP) yang dapat diakses secara terbuka. Salah satu platform yang mewadahi komunitas dan forum adalah Reddit. Perlu dibuatkan model atau *prototype* yang menyimulasikan bagaimana informasi dari Reddit diambil. Reddit menyediakan izin untuk mengambil data-data esensial yang dapat dikonversikan menjadi CTI. Praktisi profesional dapat lebih mudah memutuskan tindakan masa depan mereka dengan menggunakan kesimpulan analisis CTI sebagai bantuan visual. Oleh karena itu, data-data yang dikumpulkan dari Reddit dapat diolah dan diberikan visualisasi sesuai konteks, sehingga dapat membantu perusahaan maupun organisasi menyiapkan langkah preventif terkait ancaman siber.

Kata kunci: CTI, Reddit, *web scraping*, *data visualization*

ABSTRACT

DESIGN AND IMPLEMENTATION OF WEB EXTRACTION-BASED CYBER THREAT INTELLIGENCE SYSTEM ON REDDIT

By

FATHAN ANANTA NUR

NIM : 18219008

Cyberthreats have crept into social problems. The company has invested heavily in cyber defense in an effort to combat this. Cyber Threat Intelligence (CTI) is one of the defense mechanisms that will be discussed. Organizations have recently started making significant investments in CTI development to combat the increasing threat of cyberattacks. The CTI is characterized as a group of factual data that includes context, threat methods, indicators and potential countermeasures. Hacker or bad actor communities and forums can provide substantial proactive CTI. Forums, when compared to other platforms, offer the most complete metadata, permanent data, and dozens of tools, techniques, and procedures (TTP) that can be accessed openly. One platform that accommodates communities and forums is Reddit. It is necessary to make a model or prototype that simulates how information from Reddit is retrieved. Reddit provides permission to extract essential data that can be converted into CTI. Professional practitioners can more easily decide on their future actions by using the CTI analysis conclusions as a visual aid. Therefore, the data collected from Reddit can be processed and visualized according to context, so that it can help companies and organizations prepare preventive measures related to cyber threats.

Keywords: CTI, Reddit, web scraping, data visualization